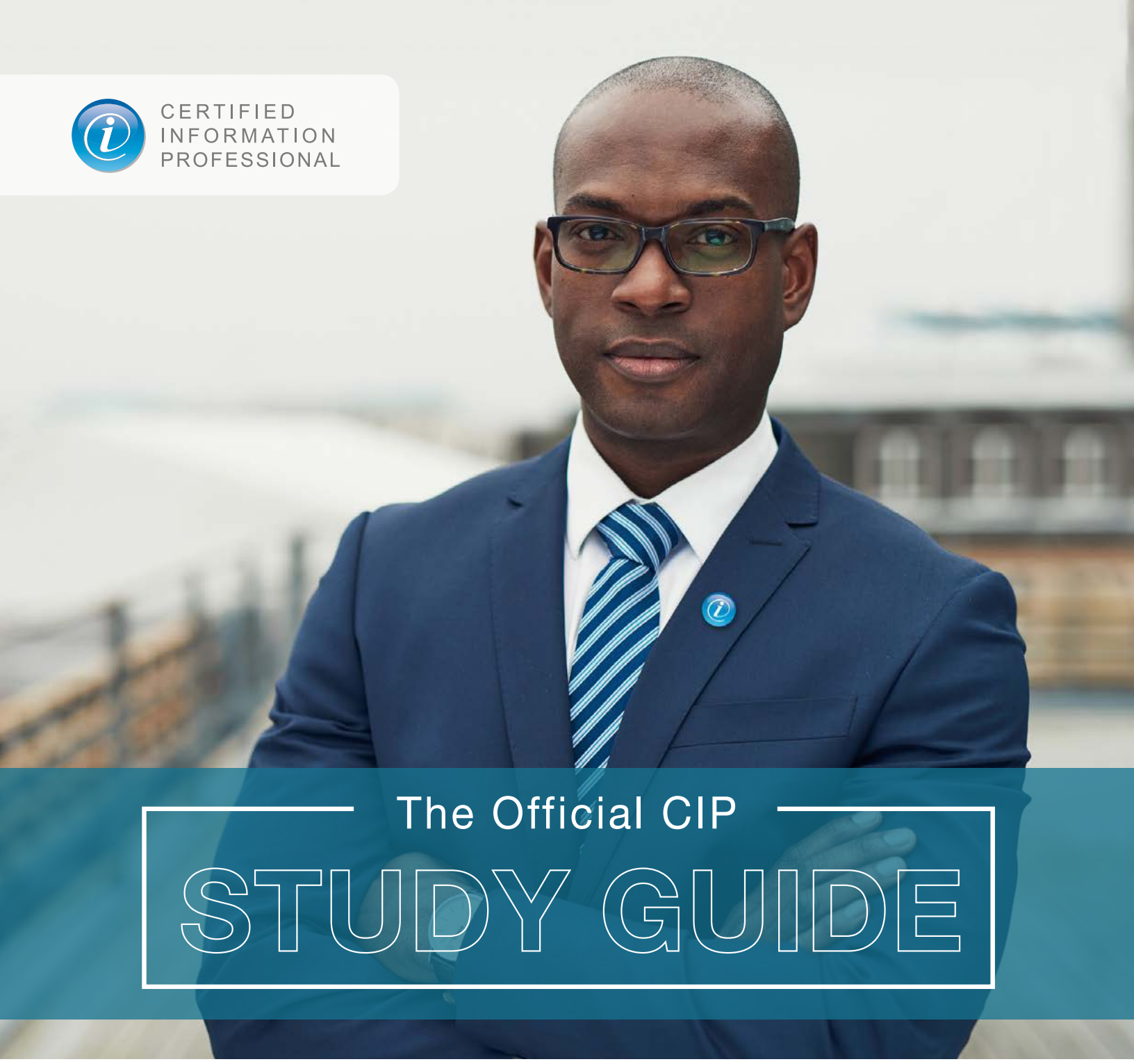


CERTIFIED
INFORMATION
PROFESSIONAL

The Official CIP

STUDY GUIDE

Are you the next Certified Information Professional?

This study guide contains the body of knowledge necessary for information professionals to be successful in the Intelligent Information Management era.

aiim



Welcome to the
Foundations of Intelligent Information Management

The Official CIP

STUDY GUIDE

2019, AIIM

aiim

© AIIM – 2019

Table of Contents

CIP 2019 Update Participants 8

CIP 2019 Exam Blueprint and Outline 9

Creating and Capturing Information 14
Introduction 14
Selecting the Appropriate File Format for Business Applications 14
Common Business File Formats 14
Selecting the Right File Format 16
Proprietary File Formats 16

Introduction to Capture 19
Determining the Best Points of Capture for Different Kinds of Information 20
Introduction to Document Management 22
Systems of Record 23

The Capture Process 27
Requirements for Multichannel Capture 29
The Digitization Strategy for Paper-based Information 30
Information Management Repositories 31

Virtual Teams and Information Management 38
Collaboration Across Organizational Boundaries 38

Collaboration and Information Management 42
Legacy Collaboration Issues 42
Document-centric Collaboration 43
Collaboration and Governance 45

Introduction to Knowledge Management 49
Knowledge Management 49
Knowledge Management and Expertise Location 50

The Migration Plan 54
Migration Issues 56

Extracting Intelligence from Information 61
Introduction 61
The Metadata Strategy 61
The Benefits of Metadata 61

The Metadata Strategy 63
Guidelines to Determine Metadata 63
Tasks to Determine Metadata 63
Mandatory vs. Optional Metadata 64
The Metadata Model 64
Metadata Standards 64
Metadata Automation 64

Capturing and Managing Metadata 67
How to Capture and Apply Metadata 67
The Challenges of Sharing Metadata Across Systems 69
How to Improve Metadata Quality 71

Classification Schemes 74
Findability 74
Classification Approaches 77
Stakeholders for a Classification Scheme 80

Extracting Intelligence from Information 83
Extracting Information from Structured Applications 83
Extracting Intelligence from Scanned Images 85
Automating Information Extraction 86

Analytics and Artificial Intelligence 89
Use Cases for Analytics and Artificial Intelligence 89
Issues Associated with Analytics and Artificial Intelligence 94

Enterprise Search 98

© AIIM – 2019

Digitalizing Information-intensive Processes ***102***
Introduction *102*
Introduction to Business Process Management *102*
Reasons for Process Change *102*
Selecting a Process to Change *104*
Process Scenarios *105*
Introduction to Business Analysis *110*
The Benefits of Process Automation *112*
Information Gathering Approaches *113*

Flowcharts ***116***
Introduction to Flowcharts *116*
Flowchart Limitations *119*

Troubleshooting and Improving Existing Business Process ***122***
Troubleshooting an Existing Business Process *122*
Planning Workflow Routes *124*

Selecting the Right Process Automation Solution ***128***
Process Metrics *129*

Robotic Process Automation ***134***
Introduction to Robotic Process Automation *134*
Use Cases for Robotic Process Automation *135*

Case Management ***140***
Introduction to Case Management *140*
Use Cases for Case Management *141*

Automating Governance and Compliance ***144***
Introduction *144*
Introduction to Information Governance *145*
Data and Information Stewardship *145*
Information, System, and Process Inventories *146*
The Business Context for Information Management *149*

The Information Governance Program ***153***
Information Governance Roles and Responsibilities *153*
Evaluating the Existing Information Governance Strategy *158*
Reviewing the Information Governance Program *159*

Personal Data ***163***
Introduction to Personal Data *163*
Strategies for Protecting Personal Data *164*

The Privacy and Data Protection Strategy and Assessment ***167***
The Privacy Assessment *167*
Privacy by Design *169*
Information Management Security Tools *170*

Privacy and the Information Management Program ***175***
Data Sovereignty *177*
Responding to a Data Breach *178*
Introduction to Records Management ***181***
Records and Non-records *183*
The Benefits of Automating Records Management Tasks *185*

Retention and Disposition ***189***
Determining Retention Requirements *189*
The Purpose of the Retention Schedule *191*
The Elements of the Retention Schedule *192*
Disposition *193*

eDiscovery ***198***
Legal Holds *198*
Collecting Information fromThird Parties *199*
The eDiscovery Process *200*

Digital Preservation ***206***
Digital Preservation Risk Factors *206*
The Digital Preservation Strategy *209*

© AIIM – 2019

Implementing an Information Management Solution ***214***
Introduction *214*
The Benefits of Intelligent Information Management *214*
The Impact Areas of an Information Management Initiative *217*

The Information Management Strategy ***221***
Information Management Roles and Responsibilities *222*

The Information Management Assessment ***227***
The Organizational Assessment *227*
The Technical Assessment *228*
Evaluating your Existing Information Management Systems *230*
Developing an Information Management Program Roadmap *232*

The Business Case for Information Management ***237***
The Costs of an Information Management Initiative *239*
Information Management Metrics *240*

The Role of Requirements in an Information Management Initiative ***246***
Gathering and Analyzing Requirements *247*

The Implementation Process ***252***
The Implementation Process *252*
Information Management in the Cloud *255*
Change Management *258*

Thank You! ***264***

About AIIM ***264***

CIP 2019 Update Participants

Updating a certification requires the expertise, passion, and time of many individual professionals. AIIM wishes to thank the following individuals for their contributions to the 2019 update to the Certified Information Professional (CIP) certification.

Certification Architect

Jesse Wilkins, CIP

Job Task Analysis

Alexander Cherkavsky, CIP
Alfred Mitchell, CIP
Amitabh Srivastav, CIP
Ayodeji Onipede, CIP
Bennie Kotze, CIP
Carlos Bassi, CIP
Dan Elam
George Goodall
Jill Stelter, CIP
John Daly, CIP
Jonathan Buford, CIP
Julie Harvey, CIP
Oladejo Fabolude, CIP
Mike Prentice, CIP
Ronald Hackbarth, CIP
Tondria Canty, CIP

Item Writing

Alexander Cherkavsky, CIP
Amitabh Srivastav, CIP
Ayodeji Onipede, CIP
Carlos Bassi, CIP
Ed Steenhoek, CIP
Jesse Wilkins, CIP
Jessica Marlette, CIP
Jonathan Buford, CIP
Leigh Isaacs, CIP
Mimi Dionne, CIP
Steve Smith, CIP

Item Review

Bob Larrivee, CIP
Dan Elam
Guy Mason, CIP
Jana Gallatin
Jessica Harman, CIP
John Daly, CIP
Julie Harvey, CIP
Paul Lanois, CIP

Score Setting

Alexander Cherkavsky, CIP
Amitabh Srivastav, CIP
Ayodeji Onipede, CIP
Bennie Kotze, CIP
Carlos Bassi, CIP
Chiny Onwuka, CIP
Edward Frankovic, CIP
Igor Shkamratov, CIP
Jessica Marlette, CIP
Jonathan Buford, CIP
Julie Harvey, CIP
Mike Prentice, CIP
Paul Lanois, CIP
Peter Rahalski, CIP
Steve Smith, CIP
Tom Motzel, CIP

© AIIM – 2019

CIP 2019 Exam Blueprint and Outline

The CIP exam is based on the following domains and topics. Individual questions all carry the same weight; the number of questions in each domain reflects the relative weight of that domain.

Domain	Exam Weight
Creating and Capturing Information	20%
Extracting Intelligence from Information	20%
Digitalizing Information-Intensive Processes	15%
Automating Governance and Compliance	30%
Implementing an Information Management Solution	15%

Domain 1: Creating and Capturing Information	Topics
Includes the following topics: • Multi-channel capture • Document management • Collaboration • Content migration & integration • Knowledge management	a. Select the appropriate file format for creating and capturing content based on business requirements, e.g. target audiences, access to content over time, regulatory requirements b. Determine the impact of using proprietary file formats on information creation, capture, and access over time c. Identify the process entry point for different information types d. Determine the best point of capture for different information types e. Describe the benefits of using document management capabilities, e.g. check-in/check-out, version control f. Identify the system of record/system of ownership for a given type of information g. Develop a process for capturing content, e.g. what to capture, approvals, audits h. Determine the requirements for multi-channel capture, e.g. platforms, routing and workflow, classification, and security i. Determine strategy for digitizing paper documents, e.g. day-forward, backfile conversion, on-demand, and the factors that contribute to each j. Compare and contrast the information management capabilities of enterprise content management solutions, point solutions, and enterprise file sync and share solutions and select the appropriate solution based on business requirements k. Determine information management needs and issues associated with virtual teams (e.g. synchronous vs. asynchronous collaboration, geographic issues) l. Identify issues associated with sharing content across internal and external organizational boundaries, i.e. between departments, with customers m. Identify issues associated with legacy collaboration approaches, e.g. email n. Identify key features required for effective document-centric collaboration, e.g. version control, workflow, access controls o. Determine whether and how to apply governance to collaboration environments/artifacts p. Differentiate between tacit and explicit knowledge and their impact on an information management program q. Define and compare approaches to expertise location, e.g. social graphing, analytics r. Identify the steps to include in a migration plan s. Determine issues associated with migrating legacy content, e.g. from one location or system to another

Domain 2: Extracting Intelligence from Information	Topics
Includes the following topics: • Metadata • Taxonomies • Data recognition/extraction/ standardization • Analytics/machine learning/AI • Search	a. Identify specific business benefits associated with effective metadata usage, e.g. lifecycle management, security management, improved findability b. Define a metadata strategy and the elements to include, e.g. consistency of metadata model & vocabulary, metadata maintenance, mandatory vs. optional metadata, metadata automation c. Describe and compare different methods for applying metadata to information objects, e.g. manual data entry, recognition technologies, inheritance, workflow, analytics d. Identify the challenges of sharing/propagating metadata across tools and systems e. Describe methods to improve the quality of metadata values, e.g. data validation, data masking, controlled vocabularies f. Compare and contrast the use of formal classification schemes, search, and navigation and their impact on findability g. Compare & contrast different approaches to developing classification schemes, e.g. thesaurus-based vs. hierarchical, organizational vs. matter/ topical vs. functional h. Identify the stakeholders for a formal classification scheme i. Determine methods for extracting and capturing information from structured applications j. Identify techniques for extracting information from scanned images, e.g. character recognition, barcodes k. Identify the business benefits associated with automating information extraction, e.g. consistency, accuracy, automation l. Identify common use cases for analytics and artificial intelligence, e.g. document categorization, topic recognition, named entity recognition and extraction, data loss prevention m. Identify common risks associated with the use of analytics and artificial intelligence, e.g. resource availability, training data, model management, black box AI n. Compare & contrast application and enterprise search capabilities

© AIIM – 2019

Domain 3: Digitalizing Information-Intensive Processes	Topics
Includes the following topics: • Business analysis • Business process management • Robotic process automation • Case management • Decision management	a. Articulate typical reasons for business process change b. Distinguish among different business processes and determine which are most suited for change c. Compare and contrast different process scenarios, e.g. routing, workflow, BPM d. Determine whether a process exists and can be automated e. Identify the expected benefits from automating a business process, e.g. financial, non-financial, consistency, reporting f. Compare different approaches to information gathering, e.g. interviewing, process mapping g. Develop a process map using best practices and standard methodologies h. Identify the limitations of process maps i. Troubleshoot an existing business process j. Determine how to plan routing of tasks or information using a workflow/BPM system, e.g. deadlines/time stamp, parallel processing, sequential processing k. Select the right business process management-related technologies for a given scenario, e.g. routing, workflow, BPM, case management l. Identify different metrics to capture about a process m. Identify the key capabilities associated with robotic process automation n. Identify key use cases that would benefit from robotic process automation o. Describe the key capabilities associated with case management p. Identify key use cases that would benefit from case management

Domain 4: Automating Governance and Compliance	Topics
Includes the following topics: • Information governance • Records management • Information security • Privacy and data protection • eDiscovery • Digital preservation	a. Define the concept of data and information stewardship b. Identify and compare different types of inventories, e.g. information, system, process c. Gather information about the business context of the organization, e.g. jurisdiction(s), nature of organization d. Identify and compare IG roles & responsibilities, e.g. stakeholders, champions, center of excellence, community of practice, IG-specific roles, IG support roles e. Evaluate existing IG strategy, processes, documents, and tools f. Describe the importance of reviewing IG program with senior management g. Describe key considerations for using security technologies effectively, e.g. redaction, encryption h. Define personal data including sensitive personal data, e.g. personally identifiable information, financial information i. Describe the elements of Privacy by Design and how they apply to the organization's information assets j. Describe strategies for ensuring protection of personal data k. Develop a process for conducting a privacy assessment l. Define data sovereignty and describe its impact on the organization, e.g. data sharing, policies, architectural considerations m. Describe how effective information management supports a privacy program n. Develop a process for responding to a data breach (e.g. discovery, security, notification to regulators/those affected) o. Explain the purpose of capturing and managing records p. Distinguish between records and non-records based on legal, historical, administrative, and operational requirements q. Explain the benefits of automating common records management and compliance-related tasks (e.g. capture, classification, disposition) r. Determine how long to retain different types of content based on legal, regulatory, and operational requirements s. Describe the purpose of a retention schedule t. Identify the elements of a retention schedule, e.g. records identifiers, retention periods, disposition instructions u. Compare & contrast different approaches to disposition of information based on the type and sensitivity of information and the type of media v. Define legal holds and the importance of legal holds in the information lifecycle w. Define the issues associated with collecting information from sources not owned/controlled by the organization, e.g. personal devices, commercial social media platforms x. Provide information from a variety of sources in response to requests, e.g. litigation, audit, regulatory inquiry y. Identify preservation risk factors, e.g. format obsolescence, media/hardware obsolescence, media degradation z. Identify the elements to include in a digital preservation strategy

© AIIM – 2019

Domain 5: Implementing an Information Management Solution	Topics
Includes the following topics: • Information management strategy • The business case for information management • Business requirements for information management • System design and implementation • Change management	a. Identify the strategic benefits of intelligent information management, e.g. improved engagement, process automation b. Determine the impact of an information management initiative, e.g. on ways of working, on business processes, on training and change management requirements c. Develop an information management strategy, e.g. vision, key performance indicators, critical success factors, success measures d. Identify the roles & responsibilities required for an information management implementation program, e.g. sponsor, champion, management, specialists, business users, others e. Conduct a baseline organizational assessment, e.g. business and regulatory environment, organizational culture f. Conduct a baseline technical assessment, e.g. existing enterprise architecture, system lifecycle stage g. Identify existing information management-related systems and determine whether they can be used/expanded/improved for a particular information management initiative h. Develop an information management program roadmap i. Compare & contrast metrics for determining the success of an information management initiative, e.g. financial, non-financial, non-quantifiable j. Develop a business case for improving information management k. Determine all costs associated with an information management initiative, e.g. acquisition costs, operating costs l. Determine the role of business and system requirements in an information management initiative m. Gather and analyze business and system requirements for an information management solution n. Identify the implications of cloud vs. on-premises deployment, e.g. costs, security, uptime, management/maintenance, lock-in o. Identify the key steps required to implement an information management solution p. Develop a change management plan (e.g. roadmap, communications plan, training plan)

Creating and Capturing Information

Introduction

We begin this course by reviewing how organizations create and capture information, from a variety of sources, for a variety of purposes. We start with a look at multi-channel capture – all the different ways in which organizations receive information and the impacts they make on capturing and managing it.

We look at document management and collaboration, and how they allow organizations to create information more efficiently and effectively.

Once information is created, it needs to be captured somewhere. We review different types of information solutions and their capabilities and limitations.

We'll take a quick look at knowledge management as a process and discipline.

Finally, some information needs to be kept for extended periods – often longer than the lifecycle of the systems that store it. We'll take a look at content migration as well as how to integrate different solutions.

Collectively, creating and capturing information form the first step in the intelligent information management lifecycle. Indeed, it sets the stage for everything that follows: extracting intelligence from information, digitalizing information-intensive processes, and even automating governance and compliance. And, of course, we implement information management systems in support of creating and capturing information and ensuring its usefulness to the organization.

What this means is that if the organization does not put effective processes in place to create and capture information, the results will often get very messy very quickly. We've described most organizations' information stores as "digital landfills" for some time because that lack of control and governance leads to information being created and stored haphazardly, with multiple copies and versions all over the organization.

Selecting the Appropriate File Format for Business Applications

Common Business File Formats

There are literally thousands of file formats available – it often seems like every application uses its own file format. Different file formats work better to meet certain business requirements, and selecting the wrong format can cause issues for organizations, their customers, their legal team, etc.

Here are some very common file formats used in almost every organization. We'll look at each of these in a bit more detail in the next few sections.

Adobe Portable Document Format (PDF)

PDF has been around since 1993 as a way to share rich documents, including formatting, links, and images. Over the years the format has been updated a number of times; its proprietary nature has resulted in significant issues with backwards compatibility.

In 2005 Adobe produced a subset of PDF, the PDF/Archive (PDF/A) which was standardized as ISO 19005. The purpose of PDF/A is to provide a stable format suitable for archiving, in part by prohibiting features that change the look of the document such as active code or font linking. The intent is that this should make PDF/A suitable for long-term preservation where such efforts are directed to faithfully reproducing a digital document in the future. This standard is regularly updated.

In2008 Adobe went further and made the PDF Specification a standard as ISO 32000. Adobe has also produced other specialized PDF formats including PDF/Engineering, PDF/X for prepress digital exchange, and PDF/UA for universal accessibility.

Today PDF is a widely used format for a variety of applications because of its support for multiple pages and a variety of content types within a single PDF and because it is natively supported in almost all web browsers through the ubiquitous PDF Reader. Many scanning and content creation applications can output directly to PDF.

© AIIM – 2019

Tagged Image File Format (TIFF)

Tagged Image File Format, or TIFF, is a graphical format presenting the document as a digital copy of the original using raster images. It is an ISO standard. It supports many different compression formats, particularly lossless ones – that is, all of the original data remains present in the file. This can result in significantly larger file sizes compared to other approaches.

TIFF was the most common format found in digital imaging applications because it was the first to be based on industry wide standards. It was the default file format for many scanners and digital imaging applications for a number of years. It supports black and white (bitonal), grayscale, and color scanning and can create multi-page files as well.

TIFF is not as popular today compared to PDF for scanning office documents for two main reasons. First, TIFF files are not searchable without taking additional steps to perform character recognition on them. This is often built into the PDF capture process. Second, the ubiquity of the PDF Reader means that PDFs are viewable on almost any device including mobile devices. TIFF readers and plugins are much less common.

Joint Photographic Experts Group (JPEG)

The Joint Photographic Experts Group, or JPEG, is both a standard compression algorithm and the file format that uses that algorithm. JPEG works by discarding up to 99% of color information that can't be discerned by the eye. This works best for continuous tone images such as digital photographs; it does not work as well for black and white images such as scanned business documents. JPEG is an ISO standard format.

JPEG is considered a "lossy" algorithm since data is actually discarded during the compression process. This is generally not an issue when creating a digital image but can become a problem if the image is repeatedly converted. The photo on the left is an original JPEG; the photo on the right is much more pixelated because it's been repeatedly converted, each time losing another 80-90+% of its data.

While JPEG can support some methods for displaying multiple pages in a single file, these are not very well supported in the marketplace.

JPEG has become much more common as mobile scanning and capture applications have matured – it is often the default file format for those devices and applications.

Portable Network Graphics (PNG)

Portable Network Graphics format, or PNG, is a more recent graphics format that supports very efficient, lossless compression and up to 32-bit color, making it very desirable for web graphics. Many graphics programs including those in digital photography applications support the creation of PNG formats; as with JPEG, PNG is natively supported in almost all current web browsers.

PNG was originally published as a W3C standard in 1996 and as ISO standard in 2004.

Microsoft Office (Word, Excel, PowerPoint)

Finally, we take a look at Microsoft Office. While there are other office productivity suites out there, including very good open source offerings, none of them has achieved the market share that Microsoft Office has – in fact, in some ways Office defines that market space.

Office includes a number of tools; the composition of these tools changes over time as do the individual tool capabilities. For this course we will limit our review to the three most common formats: Word, Excel, and PowerPoint.

Word is commonly used for creating and collaborating on business documents such as reports and contracts. Excel is a spreadsheet that can be used for financial calculations as well as presenting information in tabular format. PowerPoint is used to summarize and present information succinctly. Each tool offers a broad spectrum of capabilities that can make office workers more productive – but these broad capabilities also result in significant complexity in terms of what can be included in a given file or document.

Office file formats are all considered proprietary; there are standards-based XML versions of each file format, but they are much less commonly used than the default formats. Microsoft's sheer domination of the market ensures some compatibility between versions and among other tools, but complex authoring can result in potential incompatibilities in the future and in accessing significantly older versions of files.

Selecting the Right File Format

So how do you know which file format to use? It really depends on the business needs of the department creating the information, and, ultimately, the needs of the organization. If you want to make information available over the World Wide Web, you should select a format that can be displayed in a variety of browsers, isn't too large or cumbersome, etc. If you're looking for an archival format, that is likely a different format that will depend on the original. If your customers use Office, it will be difficult to engage them using an incompatible office productivity suite or format.

Standard file formats are preferred where possible, especially where you need to exchange information with others or when you need to retain information for long periods of time. As we'll see in another module, the more complex a file format is, the harder it is to use over time; proprietary formats start out with more complexity.

Proprietary File Formats

Introduction

In the market today there are many different file formats available – in fact it seems like almost every application has its own file formats.

One of the ways to look at these is based on whether, or how much, they are based on standards. Standards are generally good to follow because they represent consensus on a particular set of capabilities, behaviors, etc. On the other hand, proprietary file formats are specific to a particular application and sometimes even to a particular version of an application – for example Microsoft Word vs. Corel WordPerfect, or even Microsoft Word 95 vs. Microsoft Word 2016.

Proprietary File Formats

Most organizations don't give much thought to the file formats used to store their information – and this can cause problems in the short and long term when those formats are very complex and very proprietary.

There are two ways to look at proprietary file formats depending on their market share. For formats with significant market share, such as Adobe PDF, there are often other applications that claim to create compatible files. Whether the files are in fact compatible often depends on what is meant by "compatible" – for a simple document this may be easily done, but for a highly complex document it is likely that there will be issues.

On the other hand, for formats that are more niched, there may be no other way to access those documents than through the provider's application or service. If those shut down or change significantly, there is a real possibility of losing access to that information, and the longer it's retained, the more likely this will be the case.

At the same time, however, it's important to consider that converting a file from one format to another, especially from a highly complex proprietary format to a more open, but perhaps less functional one, presents some issues as well, especially with regards to authenticity and trustworthiness over time. The CIP needs to balance these two issues carefully, both in the short term and over time. We discuss long-term access to information in another module.

Whenever possible, CIPs should recommend standards-based file formats as they will tend to be more accessible over time. If this is not possible, selecting a proprietary file format with a larger market share will help to preserve access to that information in the short term.

© AIIM – 2019



Test Your Knowledge

Question 1:

An organization wants to make customer invoices available through a self-service web portal. Invoices can be single or multiple pages. Which file format would best meet their needs?

a. PDF
b. TIFF
c. JPEG
d. Office

Question 2:

Identify the following file formats as either proprietary or standard:

- Office (.docx)
- Wordperfect (.wpd)
- AutoCAD (.dwg)
- PDF (.pdf)
- TIFF (.tif)
- HTML (.html)

Question 3:

For organizations creating new documents today, which file format is most likely to be readily accessible over time?

a. Microsoft Excel
b. Adobe PDF
c. AutoDesk AutoCAD
d. CorelDraw Drawing



Test Your Knowledge - Answers:

Answer to Question 1:

An organization wants to make customer invoices available through a self-service web portal. Invoices can be single or multiple pages. Which file format would best meet their needs?

a. PDF
b. TIFF
c. JPEG
d. Office

The correct answer is A, PDF. While TIFF supports multiple pages, TIFF readers are not as common, and organizations should not expect their customers to be able to find and install TIFF readers. JPG is not well-suited for multiple page images. Office also requires readers, and while those are more common than TIFF, they are not as ubiquitous as PDF.

Answer to Question 2:

Identify the following file formats as either proprietary or standard:

- Office (.docx)
- Wordperfect (.wpd)
- AutoCAD (.dwg)
- PDF (.pdf)
- TIFF (.tif)
- HTML (.html)

The answer: Office, Wordperfect, and AutoCAD are proprietary file formats while

PDF, Tiff and HTML are standard file formats (based on universal standards).

- Office (.docx) Proprietary
- Wordperfect (.wpd) Proprietary
- AutoCAD (.dwg) Proprietary
- PDF (.pdf) Standard
- TIFF (.tif) Standard
- HTML (.html) Standard

Answer to Question 3:

For organizations creating new documents today, which file format is most likely to be readily accessible over time?

a. Microsoft Excel
b. Adobe PDF
c. AutoDesk AutoCAD
d. CorelDraw Drawing

The best answer here is B, Adobe PDF, which has been published as a standard, ISO 32000. Excel, AutoCad, and CorelDraw drawings are all complex and proprietary file formats that will likely require significant effort to access over time.

© AIIM – 2019

Introduction to Capture

Identifying Process Entry Points for Information

For any information-centric process, that information has to enter the process from somewhere. Where it comes from can make a difference because it can lead to certain assumptions about that information: its format, its quality, its state of approval, and so on. And for some types of processes, the fact that a piece of information has entered the organization may start certain workflows or responsiveness requirements.

Process entry points include:

- Email
- Paper mail/documents
- Fax – yes, in some industries fax is still a key entry point!
- Voice / voicemail
- Websites and web-based forms
- Internal and external workflows
- Smartphones, tablets, and mobile apps
- Uploads to a portal or file sharing solution
- And every other channel used to create or transmit information

That said, we can start with the basics: internal and external.

Internal. Internal information can be created specifically for/as part of a process, or it can enter a process as the output of another process.

> ***Example 1:*** *An employee creates an expense report, scans and uploads receipts to that report, and submits it to the expense report approval process.*
>
> ***Example 2:*** *The finance department aggregates all of the approved expense reports for the month, separates the expenses by charge code, and charges them to the appropriate departments on the monthly financial statement.*

External. External information has to be submitted to the organization, and then to a particular process, by someone. This could be an automated process – for example, a customer fills out a loan or insurance application online. There may be supporting documentation required of the customer, depending on the particular process and transaction. Once the initial application requirements are complete, the application is routed automatically to the applicable process for further processing.

Point of Service

One of the key considerations here and for multichannel capture more broadly is that, to paraphrase AIIM research, we want to "digitize everything that moves" and do so as early as possible. Every step that involves paper is less efficient than it should be. So ideally this capture and digitization happens at the point of service or the point of the transaction. This is not a production imaging process, but rather a very decentralized one that leverages standalone desktop scanners, multifunction devices, or, increasingly, mobile devices, and applications.

Here's a very common example: It used to be that in order to deposit a check, a customer would have to go to a bank and talk to a human being. Even ATM deposits required the creation of a deposit slip, placing the check to be deposited in an envelope with that slip, and waiting several days for it to be processed. Today many ATMs will scan the front and back of the check as it's fed into the machine and that money can be made available much more quickly. And there are mobile banking apps that will allow end users to take their own pictures of checks and use those to make a deposit directly through the app.

Key Considerations

Regardless of the specific process entry point, but especially as applies to point of service capture processes, there are some things to keep in mind to ensure the information received can be processed effectively: completeness; format and quality; and chain of custody.

There needs to be some sort of process step to ensure the completeness of the submission or transmission. That step also needs to account for the expected formats to be received and the overall quality of those documents, especially scanned images.

And these lead to a need for some sort of documentation of how the information was received and what happened to it – that is, a chain of custody. Depending on the process and the organization this may be more or less formalized, but there should still be some way to track where a particular piece of information is and that it *has* been received.

Determining the Best Points of Capture for Different Kinds of Information

Capture is the process of getting information from its source into some type of more formal information management environment or system, and then recording its existence in the system. This includes scanning or otherwise converting a paper document into a digital form.

To record the existence of the information in the information management system, you need to include relevant data about the paper document or other physical object that has been captured. This 'data about data' is known as metadata; we will discuss metadata in greater detail elsewhere in this course.

Sources of Content

For the CIP, the focus is on enterprise information. Enterprise information can take any number of forms. We may think of typical office documents such as spreadsheets or contracts, but any content with business value can be worth managing in an information management system. Think beyond the office suite to email, instant messaging or chat systems, or engineering drawings.

The popularity of collaboration platforms and social media tools means more business communication is occurring in those platforms. Rich media types such as video, audio, or digital photographs are also becoming increasingly common. They are useful ways for organizations to share information and communicate with employees, partners, citizens, and customers.

And we can't forget the legacy world of paper. Many organizations still use paper-based forms, hand-written signatures and other physical media to represent business transactions. We must still recognize and manage this content or plan for its digitization where appropriate.

Information can be in any digital file format – data, text, images, audio, video, as well as others, and even non-electronic content can be managed. It is the content of that information, not the format, that is significant when thinking about information management.

We know that enterprise content can come from a wide range of authoring tools. Modern organizations are also realizing that the ways in which content is created, retrieved and read also has changed. The work world is becoming increasingly mobile, and information management professionals need to look at the kinds of devices and applications their information workers use.

As desktop PCs have given way to laptops and notebooks over recent years, so may these devices give way to tablets and increasingly sophisticated smart phones. As new mobile and portable devices improve usability and gain enterprise acceptance, information management professionals need to plan for the short-term future and understand how and where enterprise content is originating.

Expensive high-volume scanning hardware is also giving way to smaller departmental tools, and better quality, inexpensive multi-function devices. Sophisticated smart phones have camera and snapshot capabilities that are beginning to see adoption by financial services companies for use cases such as check deposit.

Nor is paper disappearing from the mail rooms and file rooms of large companies in transaction-based businesses such as insurance, banking, or public sector.

And content continues to be created in line of business applications such as Customer Relationship Management (CRM) or Human Resources tools. This is ripe potential for integration, import and interoperability as your information management system deployment progresses.

© AIIM – 2019

The Point of Capturing Information

The entire point of capturing information is to establish some sort of control and context around it. The most effective way to do this is through formally capturing and storing information in some sort of repository. We address repositories in more detail elsewhere in this course, but in effect it's a place, often a database, where information can be stored, retrieved, accessed, and ultimately managed through its entire lifecycle.

Capturing information into a repository allows the organization to establish control over that information. Security and access control lists can be set as to who can retrieve, manipulate, or even delete it.

It also allows the organization to establish context around that information. What process is it part of? Who processed it and in what fashion? What other documents or records or processes does it relate to? Who owns and who uses it?

Capture at the Point of Service

We mentioned in another module that the best time to capture information is often at the initial point of service or transaction, because every step that takes place at the speed of paper is less efficient than it should be. This will help to reduce transaction times and the overall costs of the process, making it more efficient for the organization as well as for any customers or partners involved.

If it can't be captured directly at the point of service, it should still be digitized (if applicable) and captured as early in the process as possible. This should be the general rule regardless of the type or source of information. That said, there are some more specific approaches to consider.

Point of Capture – Business Documents

For business documents such as Microsoft Office or Adobe Acrobat PDFs, there are a couple of options for when to capture them:

- At the point of service or transaction as we just described. This would especially apply to documents received from elsewhere.
- Automatically as part of a workflow such as a review and approval workflow. In this case once the business rules have been satisfied, the final deliverable would be captured into a repository according to those rules. This would also apply to rich media types of business documents such as videos, photographs, or other deliverables with complex authoring and approval processes.
- As part of the authoring processes. Instead of creating a document on the desktop, consider changing the process so that documents are created within a document or enterprise content management solution. As soon as the document is created it's saved, versioning can be applied, etc., and it can be managed more effectively throughout the authoring process as well.

Point of Capture – Scanned Images

For business documents that come into the capture process in paper format, they need to be digitized as soon as possible. This means that in a transactional environment, paper documents should be digitized as they are received and then periodically uploaded to the repository of record.

In a production imaging environment where hundreds or thousands of pages are being scanned every day, the imaging process should include a step for releasing the images – to another process, to a repository, or even to a file share location.

Point of Capture – Email

Email has been a challenging information type for decades for many reasons:

- Very high volume for most users
- Very high volume of junk to sift through (solicitations, spam, spyware, Reply-All responses, etc.)
- Granularity and terseness of some messages
- Verbosity – multiple topics, multiple pages, all the replies in the thread – of other messages
- Informality – so many email messages contain information you would never say to someone's face;
- And of course, attachments and attachment spam

For all these reasons, but especially the first ones, the best approach to capture here is often to archive all email sent or received, or at least all of them sent to certain roles or functions, at the email server as they are being sent and received. This is a very broad-based approach but as we'll see in more in-depth discussion of email management later, manual approaches to email simply do not work.

Point of Capture – Structured Data

In most organizations, structured data is "captured" into its own application. Every line of business application has its own data tables, and perhaps its own database, and information is captured as it is entered into the system. In some cases, this structured data is used to generate reports, which may in turn need to be captured and managed. These may need to be captured manually or through a workflow of some kind to ensure all the data is present.

Structured data can also be captured more transactionally through the use of forms, whether paper-based or digital forms. In the case of paper-based forms, the most efficient way to capture this data accurately is through scanning the forms and then using recognition technologies to automatically extract and capture that data. Digital forms may have their data captured directly into a structured line of business application.

Introduction to Document Management

So what is document management? It is the use of a software application to track digital documents from creation through approval and publication. It serves in many ways to apply a formal governance framework to the document creation and collaborative editing processes. Today document management is generally incorporated as a set of capabilities in a broader enterprise content management (ECM) or information management solution. We will address those broader solutions in another module in this course. Document management capabilities are also key parts of the document control discipline, which is beyond the scope of this course.

Traditional document management includes the following capabilities:

- Check-in/check-out
- Version control
- Roll back
- Security and access controls
- Audit trails

Next we'll look at each of these in a bit more detail.

Check-in/Check-out

Check-in and check-out are very similar to how a library works – when a book is checked out, nobody else has access to it until it is checked back in. In document management, a user can check out a document in order to make changes to it. While the document is checked out, nobody else can edit it, and, depending on the solution, it may not even be accessible in a read-only mode.

Once the user has made any desired changes, the user checks the document back in. At this point a new version of the document is created – we'll discuss that shortly – and the document is unlocked and available for other users to review and/or check out.

The point of this capability is to ensure that multiple users aren't editing the same document simultaneously and overwriting each other's changes.

Version Control

As the name suggests, version control is used to manage or control different versions of a document as it goes through the authoring and approval process. New versions are automatically created through auto-save, by saving the document manually, or, in this specific context, by checking the document back in.

Some systems support major and minor versions of documents, while other simply consider any changes to result in a new version of the document. In either case, the system will also allow authorized users to compare different versions of the document to see what changes were made between any two versions of the document.

© AIIM – 2019

This feature also reduces the need to store multiple copies and versions, and their associated naming conventions, in order to retain a document's history. This manual approach – changing file names to e.g., "mydocument_final.doc" – is often overlooked or not of value, with the result that even with those naming conventions nobody is certain as to which version is the current, final, or approved one. Version control results in storing one document, with all of its versions, in one location so there is no confusion.

Roll Back

Many systems that offer document control capabilities offer the ability to roll back, or revert to the previous version. This is often done when a version is released prematurely or with some sort of error. This is most commonly seen in web content management and software development repositories as opposed to documents, but it is seen in some case management and contract management solutions as well.

Security and Access Controls

Security and access controls help to ensure that any changes made to a document are done only by authorized users. Some users might be able to make changes directly to a document, while others might be limited to only commenting on the document and still others to read-only access. They also help to provide accountability, as any changes made are also linked to the individual(s) who made them.

Audit Trails

Audit trails show what has happened in a system. In the context of document management, audit trails can track every change to a document throughout its lifecycle, including who made what changes, when, and in what sequence. As with security and access controls, this helps to ensure accountability and transparency in the authoring and approval process.

Systems of Record

Let's start by defining system of record. Wikipedia defines it as "an information management system that is the authoritative source for a given data element or piece of information." This is not a recordkeeping system per se, although it can be. Rather, this is the place where a particular type of information is stored. Ultimately, the organization should have a system of record identified for every type of business information it creates, receives, and manages – i.e., "a place for everything and everything in its place."

Where to Store Information?

So which system is right for you or your organization? There is no right answer because each organization, and each department or process within the organization, has different business needs. Records need to be stored more formally than drafts. Rich media has unique requirements for storage and retrieval compared to scanned documents. Personnel files are more sensitive compared to other types of information.

However, there is a sort-of right answer, which is that for any specific type of information, there should be a place designated for its storage. It doesn't matter *what* that system is necessarily. Rather, it matters that users understand that there is an answer to their question, "Where do I store my files/documents/stuff?" These answers in turn should be part of the governance framework, most likely in process and procedural documentation, and users should be trained and checked on regularly to ensure they are storing information in the appropriate location.

And that location should be a repository of some sort. Any repository is likely to be more fit for purpose than networked file shares, which in turn will always be better than storing information on users' individual computers, mobile devices, flash drives, and so on.

Similarly, legal, risk management, compliance, records management, IT, and so forth want to understand where and how information is stored so they can perform the tasks they need to do in support of the organization as well.



Test Your Knowledge

Question 1:

Why is it important to understand when a piece of information has been received by the organization?

a. Because the organization needs to make someone available to process that information
b. Because it could trigger workflows or responsiveness requirements
c. Because it can't be processed until it's entered in the tracking system
d. Because different formats result in different process steps

Question 2:

Why is it important to capture business information into a repository? (select 2)

a. To establish control
b. To determine ownership
c. To determine retention requirements
d. To establish context

Question 3:

What is the main benefit to using document management?

a. It ensures that important documents cannot be changed.
b. It provides a listing of all documents in the organization.
c. It outlines the flow of documents through the organization.
d. It tracks documents and versions from creation through approval.

Question 4:

An organization is trying to get its HR files digitized and organized. Where should the resulting files be stored?

a. On a secured network file share assigned to HR
b. On the HR manager's workstation
c. In the repository designated for HR files
d. In a folder on an encrypted flash drive

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Why is it important to understand when a piece of information has been received by the organization?

a. Because the organization needs to make someone available to process that information
b. Because it could trigger workflows or responsiveness requirements
c. Because it can't be processed until it's entered in the tracking system
d. Because different formats result in different process steps

The best answer here is B, because it could trigger workflows or requirements to respond within a certain time frame. The other three answers are much less important compared to a hard requirement to process something within a certain timeframe.

Answer to Question 2:

Why is it important to capture business information into a repository? (select 2)

a. To establish control
b. To determine ownership
c. To determine retention requirements
d. To establish context

The best answers are A, to establish control, and D, to establish context, in both cases so the information can be managed effectively throughout its lifecycle. For B, ownership may determine which repository a particular type of information is captured into, not vice versa. Similarly, for C, retention is determined by the content and value of information, not which repository it's stored in.

Answer to Question 3:

What is the main benefit to using document management?

a. It ensures that important documents cannot be changed.
b. It provides a listing of all documents in the organization.
c. It outlines the flow of documents through the organization.
d. It tracks documents and versions from creation through approval.

The best answer here is D, document management tracks documents and versions from creation through approval. Document management does allow important documents to be changed – that's really the point of that functionality. A document management solution would certainly be able to provide a listing of the documents it contains, but it wouldn't know about things in other systems or locations like file shares. And while an individual document does have a flow – creation, editing, review, approval – there are as many flows of documents in the organization as there are types and the document management system wouldn't track anything not created within that specific system.

Answer to Question 4:

An organization is trying to get its HR files digitized and organized. Where should the resulting files be stored?

a. On a secured network file share assigned to HR
b. On the HR manager's workstation
c. In the repository designated for HR files
d. In a folder on an encrypted flash drive

The best answer here is C, in the repository designated for HR files. A, the secured network file share, is a better answer than the other two, but it's not as good as storing the files in a repository designed for them. Business files should never be stored and managed on individual workstations or removable media.



© AIIM – 2019

The Capture Process

There are two distinct capture processes we will look at: paper and born-digital. There are some similarities, and once paper has been scanned it would generally follow the born-digital capture process, but there are some unique tasks to review in the paper capture process.

The Paper Capture Process

So let's begin with the process for capturing, or digitizing, paper documents. The digitization process includes several sub-processes:

Documenting these processes – and the organization's adherence to them – is the key to ensuring that the scanned images meet the organization's legal, evidentiary, and business requirements.

The Born-Digital Capture Process

Now let's look at the process for capturing born-digital information. You will need to make a reasonably quick assessment of exactly how digital documents and records are created, received, and used and where the key risks are in terms of information that needs to be captured into the information management repository. This will probably result in a prioritization of work.

Once you have identified your potential risk areas, you will need to examine the organization and its work processes to determine exactly what types of information need to be captured. By doing this you will also gain a much better idea of the business transactions that take place in the organization and the processes for creating and using information.

Next, you need to determine when in the process to capture a particular type of information. After all, capturing drafts and capturing final documents or records would occur at different points in the creation and publishing process.

Next, you will need to determine how to capture information effectively. Where possible this should be an automated process.

Once decisions have been made about what to capture and how to do it, those decisions should be defined in procedures that can be used with appropriate tools to capture specific digital records.

You will also need to address longer term sustainability issues covering the capture of digital records to ensure access to them over time.

The Capture Format

The decision to capture a particular type of information is only a part of the process. The organization will need to determine in what format that information should be captured. There are several options:

- **Original native format.** This is the preferred approach because it is the most faithful representation of the transaction. In case of legal challenge this is also the most likely to be requested by counsel or auditors.
- **Commonly-used format.** For example, an organization may have a number of different versions of Microsoft Office or Adobe Acrobat in place. Some documents might be created in the most current version but then saved to a final format that is earlier, such as Office 2003 or Acrobat 5, to ensure maximum compatibility across the organization. Similarly, an organization whose users use not just Word but also WordPerfect, StarOffice, and Google Docs might also select a common file format for final documents.
- **Standards-based format.** This approach requires users to save the record into a non-proprietary format such as plain text, HTML, or PDF/A. This is better for ensuring long-term access to the information but may come at the expense of losing certain functionality. This can be a significant issue for complex or proprietary file formats.

Why Manual Capture Doesn't Work

Many organizations operate in an environment in which users are encouraged, expected, or required to identify and capture their own information. There is some value to this approach – users are most knowledgeable about their business processes and activities and should be the best-positioned to determine what is important and where to store it.

But the reality is something different. In the vast majority of organizations, the vast majority of users do not identify and capture and manage their information properly. They simply don't.

There are a number of reasons for this. First, every organization is doing more with less. Where users already feel overwhelmed by their workloads, it is very difficult to expect them to periodically stop what they are doing to do this "information management" thing. And this is probably more true the more senior the user in question – which is even more of an issue because they are more and more likely to be creating important documents and records that document decisions, set strategy, etc. No matter how much they are trained, most users simply don't see information management as a priority – or certainly not higher than the actual work they are doing.

And many of them don't get training. Or they get a 30-min training when they onboard and maybe a refresher 30-min session every year or two.

More importantly, users are very likely to classify things inconsistently over time, and the more complex the classification structure, the more likely this is to be the case. If you don't think this is an issue, take a look at your own computer, your own folder structures on your desktop or where you store your documents, and especially your own email inbox. Have you been consistent in the way you save and classify messages over time?

And there is always the possibility of an error: the user drags the record into the wrong folder, or makes a typographical error in a key metadata field, etc.

So for all these reasons, it is vital that organizations streamline and automate the capture process to the maximum extent possible. Now let's look at some ways we can do this for different types and sources of information.

Automating Capture

There are a number of ways that the capture of born-digital information can be automated.

- **Bulk import.** This most often occurs when migrating information from one location or system to another.
- **Workflow.** One of the steps in a workflow could be to take action on a particular document. For example, once the final document has been approved, it can be converted to PDF, the PDF posted to the website, and the original finalized document declared a record and moved to the records repository.
- **Content types**, or document types, or record types, or whatever term a given repository uses. This approach provides for the definition of characteristics that define a certain information type such as invoices, contracts, etc. Every document that meets that definition is automatically processed in a certain way.
- **Analytics.** This approach leverages machine learning to understand the contents of documents or records and process them according to that understanding.

Approvals

The capture process may also need to address approvals. When paper documents are scanned, the final step in the quality control process is approval that the images were processed correctly. This is generally a less formal, more tacit approval process.

When more complex documents are being captured, such as contracts with references, or engineering documents with external reference files, it's often valuable to have a more formalized approval process. This could be done through a review and approval-type workflow, or through a more manual process. Either way, it's important to ensure that all parts of the document are complete and accurate, and that any supporting documentation that is required has been captured as well.

Auditing the Capture Process

No matter what approach or approaches you take to capturing information, it is important to audit the capture process periodically to make sure that information is being captured accurately and consistently. This audit needs to include the documents or records themselves, but it should also include metadata. After all, while capturing information can help safeguard it, it's not very useful if it cannot be located or retrieved.

© AIIM – 2019

Capture Process Metrics

Here are some capture process metrics that can be captured as part of the audit process. These metrics may help make the business case for effective information management.

- The percentage of information captured into a particular repository – documents, records, and metadata.
- Growth rates in terms of volume and in the rate of growth. This will help the organization plan for future storage needs.
- Access and retrieval rates – how much information is actually accessed and used.
- The percentage of information that has been captured but no longer has business value – or its value is unknown.

Requirements for Multichannel Capture

Multichannel Capture – Definition

Simply put, multichannel capture is capture from a variety of sources. These range from traditional production or desktop scanners, to multifunction devices, to mobile devices and applications, and everything in between.

And it's not just the capture of paper through a scanning process: multichannel capture may also include information received via web forms or website uploads; email attachments; and even fax and structured print streams. Ideally multichannel capture can work in every channel through which the organization receives information.

Ultimately the value of a document is in its content, not whether it was received as an email attachment, captured via a smart phone or tablet, or scanned using a multifunction device.

Types of Multichannel Capture

Multichannel capture also describes the point and approach to capture. It certainly includes traditional production capture but extends beyond that to include ad hoc capture at the point of service or of a transaction. And it includes on demand capture, often using a multifunction device or mobile device.

Classification and Routing

Once the content enters the organization, it will go through the multichannel hub. Here, the content is identified, classified, and routed to the appropriate work process. Different platforms take different approaches to this. Some solutions apply specific templates, rules and analysis per channel; others attempt to apply a uniform set of rules and analysis regardless of the format or source of the content.

Classification and routing can be done through static rules and template-based recognition; today there are also solutions available that can automate some or all of this process based on patterns. These patterns and relevant rules are derived through the application of machine learning to the incoming content.

Once the content is identified and classified, it can be sent automatically to the appropriate workflow if one exists.

And a key piece of the capture process is assigning metadata to the document. This can be automated as well using many of the same techniques.

Security

Security is even more important when digital information can come in from so many sources. As things come in, the organization needs to implement effective processes and protocols to ensure for example that personal medical information doesn't sit on an open fax or in a shared inbox.

Incoming digital documents need to be scanned for viruses and malware, too, to make sure they don't cause information security issues.

And access controls need to be set up to ensure that only authorized users have access to the incoming information, and only for the purposes they need it for. For example, someone assembling a mortgage application based on the customer's submissions probably doesn't need to be able to edit the documents submitted.

Quality Control

Quality control is always important, but it becomes even more important as organizations incorporate more of these ad hoc capture processes, perhaps using untrained "operators" (aka customers, line of business managers, or anyone whose main job doesn't involve scanning) and a variety of capture hardware including personal smartphones.

So the first thing to look at is the quality of the actual captured content. There needs to be a step in the process, manual or automated, that ensures that the source content is captured effectively. Images can't be blurry or out of focus. The entire document needs to be captured. And so forth.

Metadata also needs to be checked. While character recognition technologies are mature and better every year, they are still not perfect and can vary widely depending on the source and operator. Something as simple as a 1-off mismatch can result in every single document's metadata being incorrect.

Security needs to be checked periodically as well to make sure only authorized users are accessing and interacting with the captured content. Security of the various inputs and channels should be checked periodically as well.

The Digitization Strategy for Paper-based Information

Digitization Strategies

There are four primary strategies for digitization: full back file conversion, partial back file conversion, day forward, and scan on demand. Next, we will define and compare each of these.

Full back file conversion. The "back file" is a term generally used to describe inactive paper documents which might be stored in a file room, records center, or even an offsite warehouse. In this approach every paper document in that back file is digitized and, often, the paper originals are destroyed. We discuss the legality of this in another module.

This is the most comprehensive, but also the most expensive. If one of the drivers of the initiative is to reclaim office space, this may be required. In addition, some documents, particularly older ones, may be poor quality, offsized, etc. which will drive up the cost and resources required.

Partial back file conversion. In this approach only some paper documents are digitized. Which ones are and aren't could be a function of age – for example, only scanning one year of the back file and leaving the rest as paper. It could also be related to a business function – for example, scanning invoices and contracts but not other types of documents.

This approach is not as expensive, but could lead to users needing to search both physical and digital storage locations in order to find everything. In addition, if the primary benefit or business case is from reclaiming storage space, this approach is less beneficial.

Day forward. In this approach the organization sets a date and scans everything after that date. Paper documents older than that date are left in their original format. This approach can be very cost-effective because only newer, probably better quality, probably current documents are scanned. The biggest drawback is that, depending on the documents required, users may still need to search in more than one location, but this is offset somewhat by the fact that users will know where things are before and after the day forward date. In addition, it is vital that the organization be able to keep up with digitizing the volume of new or incoming documents. And this doesn't really save any current storage space – though it could help to reduce the need for additional storage space in the future.

Scan on demand. There are two ways to do this. One is to leave everything paper and only scan documents that are requested or accessed. This is the absolute cheapest approach, but it adds steps to each individual business transaction. The other is to combine day forward and scan on demand. The organization would scan everything after, say, January 1 of the current year, but would also scan older documents that are requested or accessed. This ensures that active documents do get scanned, while inactive ones do not.

This is a very cost-effective approach and focuses on active documents while not worrying about inactive documents. The downside is that this provides minimal storage space savings, and as noted may increase the time to process certain transactions.

© AIIM – 2019

Which to Choose?

How do you choose which approach will work best? It depends first on the purpose of the digitization initiative – is it to make content available online? Is it to reclaim storage space? Is it to make it easier to use the documentation in question to support customer service, or analysis, or something else?

One differentiator is whether the documents in question are active or inactive and what the relative volumes are – it makes more sense to digitize active documents as opposed to documents that have not been touched in 10 years. It's also important to consider the value of the information contained in those documents – which also goes to inactivity. You don't want to spend the resources to digitize records that are within a few months of their disposition date. We can summarize these two points as determining the overall value of the documents in question, and then prioritizing those with ongoing business value over those with less or no value.

A production scanning process is a lot more involved than a couple of people and a multi-function device or desktop scanner. Access to internal resources, including staff and space, needs to be considered, as does the potential cost to outsource to a scanning service bureau.

Finally, documents that are in good shape are better candidates for digitization than those that are ripped, old, fragile, or otherwise in poor shape. Similarly, if you have odd sized or weighted documents, such as cards, multi-part carbon or carbonless forms, oversized engineering drawings, onionskin or card stock, or anything other than standard paper documents, the digitization process will be more difficult and potentially more expensive – specialized equipment and even specially trained staff may be required.

Information Management Repositories

There are a variety of different solutions that organizations can use to manage information. These include enterprise content management (ECM) solutions, point solutions, and enterprise file sync and share solutions. It's important for organizations to identify their particular business requirements for information management, and then select the appropriate solution based on those requirements.

ECM Solutions

Enterprise content management (ECM) solutions provide significant information management capabilities across a broad variety of business and process needs. Instead of using a variety of different point solutions at every stage in the lifecycle, ECM solutions attempt to provide a single application to do it all. One challenge with many ECM solutions is that they've been built through acquisitions, meaning that they may have different code, different look and feel, and different structures between one set of capabilities and another.

Another challenge is that because they do so much, they tend to be very complex, and certainly much more complex than either point solutions or the enterprise file sync and share solutions we'll look at shortly.

Finally, comprehensive ECM solutions have a lot of capabilities to bring to bear – but they are not all best of breed and in fact may be only adequate or minimally sufficient to the task at hand.

Core ECM capabilities include:

- **Document management.** This includes the ability to check-out documents to ensure they cannot be altered or deleted while being edited by someone else; the ability to check-in documents; and the ability to track versions, either at the time the document is checked in or manually.
- **Records management.** This provides a mechanism to designate a subset of information as records so they can be managed more formally throughout the information lifecycle. Typical capabilities include formal retention according to content-based rules and specific disposition including destruction based on the same rules. We discuss records management in much more detail in another training course module.
- **Capture/scanning.** Almost all ECM solutions have a mechanism to capture scanned images, apply metadata, and manage them throughout the information lifecycle. Where they don't, the ECM solutions have the capability to ingest images and metadata from standalone imaging solutions.
- **Workflow.** This provides a capability to take specific actions on documents based on business rules according to the type of information, metadata, and other considerations. This is different from the more comprehensive business process management system (BPMS) approach, which we will discuss elsewhere in this course.

- **Search.** Here we refer to application search within the ECM repository as well as enterprise search with some ECM solutions.
- **Collaboration.** This refers primarily to document-centric collaboration, perhaps with a very lightweight review and approval workflow. Some vendors also offer process-centric collaboration tools.

Point Solutions

Point solutions are tailored to a very specific type of content and business process. These would include, but not be limited to:

- Document imaging
- Records management
- Digital asset management
- Email management
- Engineering drawing management

We'll look at each of these solutions in a bit more detail.

Where there are standalone records management, document management, or document imaging solutions these would certainly apply as well. The issue with these solutions is that very few of them exist because of the propensity of the ECM vendors to acquire them as a means of adding to their solution offerings.

Document Imaging

Document imaging is used to digitize physical documents – mostly paper, but there are some applications that can scan from microfilm as well. Paper documents are scanned in a variety of ways depending on their size, weight, condition, and other business and process considerations. As part of the digitization process, most imaging applications also allow images to be enhanced, such as by removing lines, de-skewing or straightening the images, removing holes or extraneous marks, etc.

Many imaging applications allow for the provision of metadata for each image. This metadata can be entered manually by the scanner operator or indexer, or it can be recognized and extracted automatically using recognition technologies. Once this step is accomplished, the image is released to a repository, folder, or other storage location for storage or further processing.

Multi-channel capture expands the scope of imaging from the use of scanners to include other sources and devices, including multi-function devices, web-based portals, fax, and even smart phones and other mobile devices. We address multi-channel capture elsewhere in this course.

Records Management

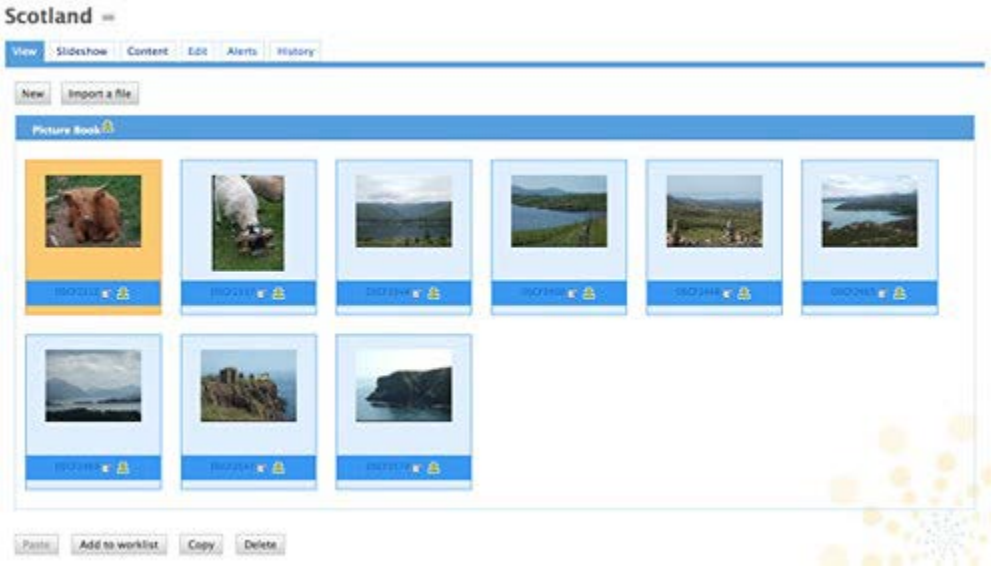
Records management solutions are used to capture and manage particular types of information more formally. By capturing and declaring a particular piece of information as a record, the organization is committing to manage it throughout the lifecycle in a way that preserves its evidentiary weight should it be needed for a legal case, a regulatory inquiry, etc. This means records have to be stored in such a way that they cannot be edited, altered, or deleted for their entire lifecycle. We address records management in much more detail elsewhere in this course.

Records management solutions can be used to manage physical records, digital records, or both – though most solutions are better at one or the other. In either case records are to be managed according to their content and their value to the organization, not by format. In other words, whether a record is a piece of paper, a Word document, a PDF, an engineering drawing, or any other type of format, it gets managed the same way, retained for the same period of time, and treated the same way at the end of the information lifecycle.

© AIIM – 2019

Digital Asset Management

Rich media is an increasingly important type of enterprise content, no longer restricted to only marketing departments. Audio (such as podcasts), video, and digital photographs are common types of rich media. Design documents, marketing assets, logos, architectural and engineering documents are all possible document types held in rich media formats.



Advanced users may use a dedicated digital asset management (DAM) system. Many ECM solution providers have additional modules or extended packages for the digital asset management power-user.

Rich media also often has extended metadata to indicate camera types, geographical data, or resolution. Metadata can also reflect any license or copyright restrictions. For example, a digital photo might be licensed from the owner for a one year campaign, and the rights to use it online expire after that period. An ECM or DAM system can help monitor these deadlines and secure content appropriately.

Email Management

Email management solutions in the information management context are probably more properly called email archiving solutions. Email archiving is an approach and a solution that attempts to capture, preserve, and often index and manage some or all emails sent or received by and within an organization. This is done for a number of reasons including performance and migration; our focus here is on the information management and governance aspects.

The key benefit here is that it happens automatically according to established business rules. That is, users don't have to select messages to save, and cannot select messages to delete or forget to save: the archive captures all of it. This means that this approach to saving emails is generally less burdensome to users and also more consistent with respect to capturing important messages.

Archiving can also include, and be used for, other types of communications objects including text messages, instant messages, social media, and others.

Once a message is archived, it is stored and maintained in the email archive until it is dispositioned. Users may or may not have access to archived messages depending on the solution and how it is configured.

Engineering Drawing Management

The last point solution we'll look at is engineering drawing management. As the name suggests, this solution is used to manage engineering drawings, which are quite common in architecture, engineering, and construction-focused processes. Engineering drawings today are almost always computer-aided designs, or CAD files, which consist of one or more actual files. These are generally structured as a series of layers; for example, for a building, there might be a layer for ventilation, a layer for plumbing, a layer for electrical, and so forth. Drawings might also include 3-dimensional models as well as external reference files that are reused across various drawings.

All of this leads to significant complexity in terms of managing those relationships. In addition, engineering drawings often go through a drafting and review process such that there are versions for the as-designed, intermediary, and as-built entities. As you might imagine, version control becomes a critical issue in the event something happens, and we need to know whether a particular valve was built a particular way – or indeed at all! This much more rigorous version control is more generally known as document control; the specifics of this discipline are outside the scope of this course.

Enterprise File Sync and Share

Enterprise file sync and share (EFSS) solutions are not new to the marketplace, but they are relatively new to business and business processes. These are solutions, mostly cloud-based, that allow users to share and synchronize documents easily across multiple devices. They are intended to be lightweight solutions that are easy to use for all employees because they are pretty simple in terms of the interface and in what capabilities they offer – generally some combination of document storage and simple collaboration. Frequently these tools also make sharing across organizational boundaries easier in terms of providing access to users outside the organization.

The challenge with many of these tools today is that they come from the consumer space, meaning they lack many of the features enterprises require and take for granted: centralized access control, robust security, metadata, lifecycle management, etc. They can also be seen as a way to evade existing IT and information governance requirements. Many of the providers in this space are taking both of these concerns to heart and starting to add enterprise-friendly capabilities; some have gone as far as to create connectors with ECM or other repositories such that the EFSS solution is the content creation/sharing/collaboration front-end and the ECM repository is used to store the final version of the document/record.

Selecting the Right Solution

So what's the right solution? There is no single answer – it depends on your business requirements. More specifically, comprehensive ECM solutions tend to work better for organizations that have more complex requirements and that have the staff to support them: system administrators, technical support, maybe database administrators, etc.

On the other hand, for an organization with limited IT capabilities, an enterprise file sync and share solution might be sufficient to meet its needs. If there is a particular requirement not satisfied by either, a point solution can fill that gap.

The other consideration is whether the organization wants best of breed capabilities – which suggests point solutions – or the consistency and comprehensiveness of a single solution.

Regardless of which approach you take, any repository is better than none at all, such as using file shares or individual PCs to store information. Storing documents in any repository, and applying appropriate access controls, is simply the best way to ensure documents can be retained appropriately and retain their potential evidentiary value.

© AIIM – 2019



Test Your Knowledge

Question 1:

Why is manual capture generally less effective than more automated approaches? (select 2)

a. Because users aren't trained properly
b. Because users don't understand their business processes
c. Because users have limited time to spend on entering metadata
d. Because users want to focus on their primary job responsibilities

Question 2:

What platforms should be included in the multichannel capture strategy? (select all that apply)

a. Mobile devices
b. Digital copiers
c. Book scanners
d. Email

Question 3:

A law firm has run out of space and is looking to lease a nearby building. The records manager notes that there are hundreds of boxes and file cabinets filled with internal and client files and suggests scanning them all to free up that space for offices. Which approach would best apply in this situation?

a. Partial back file conversion
b. Full back file conversion
c. Scan on demand
d. Day forward

Question 4:

Which solutions would be most applicable to use for collaborating on a document? (select 2)

a. Enterprise file sync and share
b. Records management
c. Digital asset management
d. Document management



Test Your Knowledge - Answers:

Answer to Question 1:

Why is manual capture generally less effective than more automated approaches? (select 2)

a. Because users aren't trained properly
b. Because users don't understand their business processes
c. Because users have limited time to spend on entering metadata
d. Because users want to focus on their primary job responsibilities

The best answers here are A, users aren't trained properly, and D, users want to focus on their primary job responsibilities. For B, users DO understand their business process, maybe better than anyone else. For C, how much time users spend on entering metadata is not directly relevant to the overall efficacy of the capture process.

Answer to Question 2:

What platforms should be included in the multichannel capture strategy? (select all that apply)

a. Mobile devices
b. Digital copiers
c. Book scanners
d. Email

A, mobile devices; B, digital copiers; and D, email are all among the platforms that should be considered in the multichannel capture strategy. C, book scanners, are very specialized and unique to the book digitization process and generally would not be included.

Answer to Question 3:

A law firm has run out of space and is looking to lease a nearby building. The records manager notes that there are hundreds of boxes and file cabinets filled with internal and client files and suggests scanning them all to free up that space for offices. Which approach would best apply in this situation?

a. Partial back file conversion
b. Full back file conversion
c. Scan on demand
d. Day forward

B, full back file conversion, is the best answer here because the purpose of the initiative is to reclaim space taken up by paper records so it can be used for offices. In each of the other approaches, there will be less storage space reclaimed.

© AIIM – 2019

Answer to Question 4:

Which solutions would be most applicable to use for collaborating on a document? (select 2)

a. Enterprise file sync and share
b. Records management
c. Digital asset management
d. Document management

The best answers are A, enterprise file sync and share, and D, document management. Records management solutions are specifically designed to keep information static and unchanging to preserve its potential evidentiary value. Digital asset management solutions are used to store and make information, especially rich media formats, available for use but they are not generally used to author or collaborate on content creation.



Virtual Teams and Information Management

Virtual Teams – Definition

Today, many organizations rely on virtual workers, and virtual teams, to get work done. Virtual workers can be every bit as productive as their in-office counterparts, if not more, if they are set up properly to succeed.

Virtual teams may be in the same geographic area and time zone, or may span the globe. That raises a number of issues, some of which fall under the guise of information management. Having workers, or customers for that matter, in different time zones means that scheduling calls and collaboration meetings is even more difficult than usual – who wants to be on a call at 11 pm or 2 am? What if there is a bank or federal or religious holiday the team isn't aware of?

Collaboration Tools

Virtual teams are successful when they have the necessary tools available to do the work that is required. This means access to collaboration tools; different tools have different use cases and approach collaboration differently.

Some tools are synchronous, meaning that everyone has to be available to use them at the same time. These would include tools like instant messaging, text messaging, web conferencing, and the like. These are great for co-authoring, reviewing, holding meetings, etc. but they do require everyone to be on – consider the meeting that includes team members in Australia, Germany, and the U.S. – there's just no time that works well for everyone.

Other tools are asynchronous, meaning they are designed to be used by different people on different schedules. Email is probably the most well-known of these; despite what many staff (and managers!) might think, emails do not generally need to be answered within 10 seconds of receiving them. And as we've seen, email has its own challenges for collaboration. Other tools here would include document management, enterprise file sync and share, and any collaborative tool that includes some sort of workspace and storage.

And some tools offer both synchronous and asynchronous capabilities that are used depending on what's needed.

Collaboration Across Organizational Boundaries

This module will help you to Identify the issues associated with sharing content across internal and external organizational boundaries, i.e., between departments or with customers.

Collaboration Issues

In addition to the usual challenges posed in collaborating effectively, there are a number of issues that are either unique to, or exacerbated by, collaborating across departmental and organizational boundaries, including:

- Culture
- Security
- Accessibility
- Time and geography
- Language barriers

Now let's look at each of these in more detail.

Cultural Issues

Cultural issues are always present in a collaborative environment, because of the clash between sharing and hoarding. This can come down to the individual contributor level – if you're a knowledge hoarder, it may be difficult to get you to share and participate in the process.

One level up, this often manifests in an unwillingness of specific departments to share and participate. It is not uncommon for departments to jealously guard "their" information and keep it from prying eyes elsewhere in the organization. With very few exceptions for very sensitive information like HR files, this is frankly ridiculous. It's the organization's information and cross-functional collaboration is always going to end up being more effective than the siloed approach.

© AIIM – 2019

And at the interorganizational level, culture clashes can prevent collaboration too. For example, in many jurisdictions digital signatures are every bit as legal as wet-ink ones, and there are significant arguments to be made that they are even more robust and defensible than wet-ink signatures. But that doesn't matter if your company or agency or country doesn't recognize digital signatures as valid. This could be a legal issue as well, of course. But this is just one example of how these issues could prevent effective collaboration between organizations.

Access Controls

The next challenge has to do with access control levels. This is a common issue inside organizations and to some extent may relate to the information hoarding we just discussed. This can be addressed through cross-functional teams and steering groups that can determine appropriate levels of security.

When outside users or organizations are involved this becomes significantly more challenging. In most cases your IT or security team will not want to simply grant network access to outsiders. The preferred approach will generally involve using some sort of collaboration tool that is designed for cross-organizational collaboration.

One quick note about email. Email makes it very easy to collaborate with external users. We discuss elsewhere in this course the significant challenges email presents to effective collaboration, but here's one more consideration: once you email something to an external party, you no longer have control over what they do with it. They can save it indefinitely, forward it to anyone they want, edit it however they want, etc. This may be OK depending on the nature of the document and the expected outcome of the collaborative process, but information professionals should be aware of this.

Accessibility and Findability

There certainly could be accessibility and findability issues between different organizations and even different departments, especially for larger, more complex organizations. While most organizations have standardized on market leaders like Microsoft Office and Adobe Acrobat, not all have, and different versions can present compatibility issues as well.

Metadata structures are often different between different systems in the same organization, much less between multiple organizations. Similarly, different departments will approach classification structures and folders differently, as will different organizations. This is why many more mature industries have developed standards and standard taxonomies – though this approach has issues as well. We discuss these issues in more detail elsewhere in this course.

Time and Geographic Issues

Any time geographically dispersed groups need to collaborate, synchronous collaboration will be an issue. Consider the challenges of setting up a web conference between offices in Los Angeles, CA; London, UK; and Sydney, Australia. Someone is getting up in the middle of the night for that call.

But there are other geographic issues to consider as well.

- What's the work culture like in each location – in terms of scheduling, the length of the work day, vacation days, even holidays celebrated here but not there.
- In the age of the General Data Protection Regulation (GDPR) and other privacy and data protection-related regulations, safeguarding information and not retaining personal data of participants can be a challenge.
- There could be specific legal requirements to consider around intellectual property, appropriate contracting vehicles and terms, etc. We mentioned digital signatures earlier as another specific example of this.

Language Barriers

It may seem obvious that there could be language barriers to take into account, but it's more than just actual languages (and their potential impact on legal terms and conditions, etc.). It's also things like departmental- or organization-specific acronyms, abbreviations, slang terms, or simply different terms and/or spellings – for example center (American English) vs. centre (most of the rest of the English-speaking world).

And it's not unusual for different organizations and even different departments within an organization to see a particular concept differently. Engineering and sales and HR will look at a particular resource or process quite differently.



Test Your Knowledge

Question 1:

Which tool would be most useful to engage the team in live review and markup of a draft document?

a. Email
b. Web conferencing
c. Text messaging
d. Digital asset management

Question 2:

How should access to internal resources be granted to external parties for collaboration?

a. External collaborators should be added to the organization's network.
b. Internal resources should be emailed to external collaborators.
c. The organization should use a collaborative solution that is designed for external collaboration.
d. The organization should set up a collaboration process using a commercial social network.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Which tool would be most useful to engage the team in live review and markup of a draft document?

a. Email
b. Web conferencing
c. Text messaging
d. Digital asset management

The correct answer is B, web conferencing. Email does not work very well at all for collaboration due to version control issues, and it's not a synchronous tool. Similarly, text messaging is not a good fit for collaborating on a document. Digital asset management is a repository for rich media, not a collaborative tool for document authoring.

Answer to Question 2:

How should access to internal resources be granted to external parties for collaboration?

a. External collaborators should be added to the organization's network.
b. Internal resources should be emailed to external collaborators.
c. The organization should use a collaborative solution that is designed for external collaboration.
d. The organization should set up a collaboration process using a commercial social network.

The best answer here is C, use a collaborative solution designed for external collaboration. As we discussed in this module, adding external collaborators to the organization's network is a security risk; so is emailing documents to them. For D, setting up a collaboration process using a commercial social network would present similar collaboration but different security risks.



Collaboration and Information Management

Legacy Collaboration Issues

The Problem(s) with Email

Like it or not, email remains the glue that ties an organization together. We use it to communicate with our bosses, colleagues, partners, and customers. We use it for storing important messages, and a lot of important collaboration happens in email.

That said, email presents significant problems to the organization. Below are six reasons for replacing email with new and better collaboration tools:

1. **Email turns collaboration into information chaos.** Email is a really good communication and notification tool, but we quickly end up with chaos when multiple people try to use email for discussing a topic or for developing something. We get reply one vs all, and reply first message vs later messages. And email volume only makes this worse. Consider: according to research, the average employee sends or receives more than 100 email messages per day. This means that that all-important message can quickly get lost in the deluge.
2. **Email locks down information and knowledge**. Gallup has found a correlation between level of employee engagement and customer service by 10%, productivity by 21%, and profitability by 22%. Email hinders this because it creates knowledge silos – one silo per mailbox! That means important information and knowledge gets locked down or lost in personal and corporate email boxes.
3. **Email distracts knowledge workers.** New incoming emails have a tendency to distract us. We end up in a responsive mode instead of spending our time being strategic and creative. We used to have a "You got mail" audio alert in the old days, - it's now best to turn off all email notifications to be productive.
4. **Email lacks information filters.** "I don't have an information overload problem, I have filter failure," says Clay Shirky, futurist and author. Most of us have only spam filters for our emails, - we don't know if the rest of the emails are important or not until we look at them. Google has tried to fix this by adding a filter for important emails, but this feels quite basic.
5. **Email makes it difficult to share large files.** One of the reasons so many enterprise file sharing solutions like Box and Dropbox have become so popular is that organizations still have a need to share large files both inside and outside the firewall, yet in many organizations, attachment sizes are arbitrarily limited to, say, 5 megabytes. This means that the tool is getting in the way of doing the business of the organization, and whenever that happens, users WILL figure out a way to work around it.
6. **Email leads to attachment spam.** Finally, even if you can share files as attachments, this creates its own problem as people send different versions back and forth, some with changes in the document, some with changes listed in the email, etc. etc. Version control is a huge issue in using email to collaborate.

The Problem with Paper

Similarly, paper files present significant issues to collaborative processes. First, paper is time-consuming to access. Even if it is perfectly filed in logical filing systems, someone has to get up, go to the paper files, and retrieve the file in question. And if the unthinkable happens and a paper file is misfiled, it might as well have been shredded because it will be extremely difficult to find again later.



The desk belongs to Steve Hubka, City of Lincoln, Nebraska Budget Director. He won the AIIM Messiest Desk competition a few years ago.

© AIIM – 2019

It's also difficult to share. If I have the file you can't really look at it at the same time. And if you're in another office or a remote worker, it becomes even more difficult (and time-consuming!).

Paper can be difficult to track – most schemes rely on users or document control staff or records managers to sign documents out manually.

Paper documents can take up a lot of space. Whether onsite at your facilities, or at an offsite storage provider, there is a real cost associated with maintaining paper documents.

Likewise, paper cannot provide as much security as electronic documents. You can't encrypt paper documents or apply digital rights management to them; if you have physical access to the document storage location, you can do whatever you want with or to them.

And paper offers very poor disaster recovery. Every incident and natural disaster results at some point in pictures of a sea of paper swirling or floating or otherwise being lost or destroyed. Because of the bulk and cost associated with managing paper documents operationally, including the cost and time required to make copies, most organizations identify their most critical stuff as "vital records," which are then stored in very safe, but very expensive storage: fireproof vaults, other offices, or offsite storage.

The Problem with Digital Landfills

This brings us to digital documents, which arguably address many of the issues we identified in the previous two sections. However, digital documents present their own issues, most notably:

- The volume is significantly higher compared to paper documents, and
- In the absence of a formal information management system, these documents are stored on network shared drives or perhaps an ungoverned SharePoint implementation, with the result that those systems come to resemble "digital landfills" overflowing with versions, outdated content, personal content, and all manner of other stuff.

This means that even though we have digital documents, we revert to information being difficult to find and not being able to find the right version or confirm what happened to a particular document. The net result is that it is increasingly difficult to trust that a particular document or set of documents is accurate, trustworthy, and reliable.

The other issue with these solutions is that they tend to focus on internal collaborative processes rather than external. If you have a business need to collaborate with suppliers, customers, partners, or outside parties, they can be difficult or even impossible to do from a practical information management and information security perspective.

What's the answer? As we'll see shortly, the answer is using an effective collaboration tool that enables more efficient collaborative processes while still ensuring appropriate levels of governance and security.

Document-centric Collaboration

This module will help you to identify the key features required for effective document-centric collaboration, such as version control, workflow, and access controls.

Why Collaborate?

Many of the core activities related to the lifecycle of business information are collaborative by nature. Creating, revising and reviewing content are tasks that often involve two or more people. Content that has been drafted, saved into an information management system, and organized with metadata often will still require revision cycles, proofreading and approvals before it is considered complete. An information management system provides an efficient, solid foundation for such content sharing, co-creation, or review activities.

Document-centric Collaboration

Document-centric collaboration is a combination of technologies, usually for asynchronous collaboration though some provide synchronous capabilities as well. This includes work spaces for collaboration on documents; mark-up and annotations; and version control. These tools can be quite robust and manage a complex draft and review process, or be very simple and get the tool out of the way in order to improve the collaborative process.

Important Features

Here is a more detailed listing of capabilities to look for in a document centric collaboration solution. These features would generally be found in document management solutions or in the document management capabilities of an enterprise content management solution.

How to collaborate on documents. Here we're looking for the ability to create a new document or open an existing one and begin the collaborative process. Necessary features here would include:

- Commenting – the ability for multiple users to add comments to a document in draft and to have those comments available to others for review.
- Version control – a mechanism to ensure that when changes are made or a new version is uploaded, there is a record of the change history including who made the changes.
- Workflow – business rules that help to ensure that documents are complete, all necessary reviews have been completed, and the document has been formally approved.
- Access controls so that only authorized users can make changes.
- An audit trail of all changes made by a particular user or to a particular document.
- And notifications and alerts for things like changes in status, requests for review, etc.

Finally, a few solutions offer **co-authoring capabilities**. This allows users to work in the same document simultaneously. Users can make edits and see others' edits as they are being made; some solutions also build in the ability to chat or send notifications in real time. Examples here include Google Docs and Microsoft SharePoint.

Use Case: Improving Meetings

Throughout this module we've reviewed the ways in which collaboration and content management can support better work practices and enable improved sharing and communication. Let's take a moment to apply some of these concepts, and step through how collaborative approaches can help streamline one of the most common business activities – the routine team meeting. We'll step through an example of how to introduce quicker, simpler collaboration methods and remove some of the email burden from routine processes.

Think about typical processes today: call participants dial in and designate someone to take notes on local laptop. Notes are emailed to the team hours or days after the call. Email threads make follow up tracking difficult, especially when the group grows over time and new roles are introduced. It can be hard to include new team members when background of project is stored in dozens of email threads. The convoluted thread of replies, forwards and cc's can be overwhelming for new team members. Content to support meetings are usually copied as email attachments, which begs the question: who has the right version weeks later?

Now let's consider how the same weekly team call could be more effective by incorporating more collaborative and content management tools.

The team could establish a simple online wiki page allowing it to be edited in real-time during the call, noting all decisions, action items and updates. The follow up activities can be recorded in one place to be updated over the course of the next work week.

Plans, schedule or other documents can be stored in an information management system with version control and access controls. This helps to reduce confusion, track progress, and show which users have contributed to ongoing updates. Links to these managed documents ensure up-to-date versions are always available from one central online location.

As the project progresses over weeks or months, new project members can quickly read the background and history in one place to get up to speed. The potential for productivity gains is high, saving time, reducing confusion, greatly reducing volume of email and copied documents, and minimizing time to productivity for new participants.

© AIIM – 2019

Collaboration and Governance

This module will help you to determine whether and how to apply governance to collaboration environments and artifacts.

Collaboration and Governance

So far we've been focused on the creative side of collaboration and how to make it useful. But we can't leave this topic without some discussion of governance. Effective governance supports effective collaboration by providing some guidance and boundaries. For example, in a complex authoring environment, governance would include the review and approval steps, as well as ensuring that all the parts that make up the final deliverable are present, the correct version, etc.

Governance also helps to ensure that the final output of the collaborative process is trustworthy and reliable. We know who contributed, who approved it, and that it hasn't been changed since that approval.

So governance does need to be applied to and throughout the collaborative process. The question then becomes how much, and how to balance the need for governance with the need for effective collaboration. If the process is locked down too tightly, the collaborative process will suffer.

Governance and the Platform

The first thing to consider is how the collaborative platform or environment supports governance. The most obvious element is access controls – that is, who can access the environment, who can access a particular project or document or site, and what can they do when they get there?

Next we can look for document management functionality: check-in/check-out and version control, so that changes can be tracked and rolled back if necessary.

We can implement business rules and workflows to guide the flow of the collaborative process, all the way through review, approval, and final publication.

And there's an element that is often overlooked: What do you do with the collaborative environment once the work is complete? All the drafts, all the supporting documentation, etc. all has to be dealt with in accordance with the broader information governance framework. This needs to include all the other artifacts of the process – email message threads, chats, recorded web conferencing sessions, and the like.

Roles and Responsibilities

We also want to consider the governance framework and how it can support the process. There are a couple of main areas to consider.

Roles and responsibilities. This goes hand in hand with the access controls we discussed earlier, but it includes roles like:

- The project manager, authoring lead, or whoever is in charge of a particular collaborative process
- IT – providing, administering, and supporting the system
- Legal – to address any legal considerations, as well as any potential requests for information about the process
- Records management – to ensure that any artifacts of the process are managed appropriately according to the retention schedule
- Privacy / data protection, where the output of the collaboration includes that type of data

Policies and Procedures

Policies and associated processes and procedures should be developed at the platform level, to address collaboration as a whole, and for individual collaborative processes. Here are some things to address:

- How collaborative environments are provisioned and made available: What does the request process look like? Are any approvals required? Or can users simply create a space, invite some collaborators, and start using it?
- What are the rules around external collaborators – are there certain processes or documents that external collaborators should not have access to for some reason?
- Are there any formal review and approval requirements? Even if not, a quality control and tacit approval process should be included.
- What happens at the end of the collaborative process? How does the team know when it's done, and what happens to the space and the artifacts within it?

Metadata and Findability

And here are some considerations for the metadata associated with a collaborative process.

- **Naming conventions.** This goes to how collaborative environments are named so they can be found by interested or appropriate collaborators. This might also apply to final deliverables. Naming conventions should be concise and describe the environment or deliverable accurately.
- **Metadata more broadly.** There is a lot of metadata generated during the collaborative process, ranging from the naming conventions above to date created, date last accessed, or date deleted, to many others. Good metadata improves the findability of the environment, any deliverables, and any supporting information. It also provides traceability to understand who participated, how they contributed, who approved the ultimate deliverable, etc. This means that there needs to be governance around who can change metadata in the collaborative environment.



© AIIM – 2019



Test Your Knowledge

Question 1:

What are the challenges associated with using shared drives to collaborate? (select 2)

a. Poor version control
b. Lack of effective audit trails
c. Limited support for some file formats
d. No security or access controls

Question 2:

Which solutions would effectively support document-centric collaboration? (select 2)

a. Records management
b. Document management
c. Enterprise content management
d. External file management

Question 3:

What is the relationship between governance and collaboration?

a. Governance makes collaboration less effective
b. Governance ensures that the collaborative process can be trusted
c. Governance actively hinders the collaborative process
d. Governance has very little bearing on collaborative processes



Test Your Knowledge - Answers:

Answer to Question 1:

What are the challenges associated with using shared drives to collaborate? (select 2)

a. Poor version control
b. Lack of effective audit trails
c. Limited support for some file formats
d. No security or access controls

The best answers here are A, poor version control, and B, lack of effective audit trails. Shared drives have no impact on whether file formats are supported or not. Shared drives do have the ability to set up security and access controls, although not as effectively as other tools can.

Answer to Question 2:

Which solutions would effectively support document-centric collaboration? (select 2)

a. Records management
b. Document management
c. Enterprise content management
d. External file management

The best answers are B, document management, and C, enterprise content management. Records management solutions explicitly prevent the changes required of a collaborative process. External file management doesn't really have anything to do with collaboration.

Answer to Question 3:

What is the relationship between governance and collaboration?

a. Governance makes collaboration less effective
b. Governance ensures that the collaborative process can be trusted
c. Governance actively hinders the collaborative process
d. Governance has very little bearing on collaborative processes

The best answer is B, governance ensures that the collaborative process can be trusted. For A, governance actually makes collaboration more effective. Similarly for C, governance supports the collaborative process. This also means that D is simply incorrect.

© AIIM – 2019

Introduction to Knowledge Management

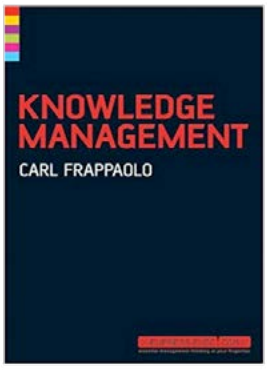
Knowledge Management

According to Carl Frappaolo, one of the leading practitioners and analysts in the knowledge management space, "Knowledge Management is the leveraging of collective wisdom to increase responsiveness and innovation."

In a similar vein, Lew Platt, the former CEO of HP, once noted that "If only HP knew what HP knows, we would be three times more productive."

Both of these show why knowledge management is important – it drives innovation and directly contributes to the bottom line. And not knowing what your organization knows is definitely a recipe for rework, stagnation, and inefficiency.

Encouraging the sharing of employee knowledge – which includes documents, records and other forms of enterprise content - to serve enterprise objectives remains an important goal for information professionals. Organizations with successful knowledge sharing cultures focus more on removing barriers to information flow, and less on capture and categorization, more dynamic, less static.



Source: Carl Frappaolo, "Knowledge Management", Capstone Publishing, 2006.

Tacit vs. Explicit Knowledge

One of the key concepts in knowledge management is the differentiation between *tacit* and *explicit* knowledge. Here you see a table that describes some of the differences; they can be summarized by saying that tacit knowledge is in someone's head, and the challenge is to make that knowledge explicit, or codified in recorded form so it can be shared. Wikipedia notes that facial recognition is a great example of tacit knowledge – you can recognize the face of someone you know among thousands or millions of faces, but you can't explain all the nuances of how you do that.

Companies that introduce on-the-job training (OJT) or mentoring programs do so in order to pass along the tacit knowledge of their more senior or experienced staff to junior, less-experienced staff.

Making Tacit Knowledge Explicit

There are some strategies for making tacit knowledge more explicit. One of these is to incentivize sharing. There are any number of ways to do this, from recognition to monetary or other tangible rewards. The right incentive will depend on the organizational culture and the individuals involved; some people are more prone to sharing, while others believe that hoarding knowledge will somehow make them indispensable.

It's also important to get the tool out of the way. One of the major issues with knowledge management has been that the tools have been quite cumbersome and "out of the flow" of work. Consider for example a tech support representative who figures out an innovative solution to a regular customer issue. In order to codify that knowledge in a knowledge base, that rep has to stop what he or she is doing, switch systems or at least screens within the system, laboriously navigate a too-complex taxonomy to find the right place to report the findings, and actually enter them into the right fields, select the applicable product, and on and on and on.

Instead, the rep could use a much simpler tool, perhaps a blog or wiki with good search capabilities, that is easier to use and therefore more likely to get used.

Finally, it's important to make that process part of the way of working. It's a best practice to have a debrief, or lessons learned, or something similar at key milestones of a project including at the end. At the same time, simple tools like blogs make it easier to keep an ongoing narrative of key developments and decisions.

Capturing Institutional Memory

The concept of institutional memory is important for organizations with specialized job roles. When expertise takes years to develop, suddenly losing skilled, experienced people can hurt the company from fulfilling core missions.

Easy to use and simple to access collaborative or content creation tools with hooks into an information management system can be a strong foundation for a knowledge management initiative. Documents and records are the tangible output of work activities, and companies that encourage consistent use of these systems will have smoother transition as individual experts come and go or change roles.

To further tap the knowledge of the specialized employee, however, lightweight collaboration and simple "social" sharing and communications tools should be considered. These applications can allow employees to contribute and share more easily. This is particularly important when trying to convey tacit knowledge across a distributed or virtual organization.

Knowledge Management and Expertise Location

Expertise Location – Definition

So what is expertise location? It helps organizations to manage knowledge by identifying experts on a particular topic and then leveraging that expertise in support of business goals and objectives. To some extent it also allows the organization to verify and qualify that expertise. And it allows individuals to claim, and support their claims, to verify expertise through demonstrated effort and work product.

Expertise Location – Profiles

There are a couple of different tools and approaches available for expertise location. The first one we'll look at is profiles. These are often user-generated and populated and their completeness and accuracy vary widely.

These profiles could be created and stored in specific tools for managing expertise. Many self-proclaimed experts will develop and manage social media profiles, such as personal blogs, LinkedIn profiles, and profiles on other public and private social media services.

Some organizations still use spreadsheets to track staff areas of expertise. Many smaller consulting firms in fact manage their expertise in this manner, with lots of columns for industry experience, more for horizontal/process experience, still more for types of information management projects they've worked on, etc.

The problem with all of these approaches is that they are heavily reliant on individuals to report their expertise completely, accurately, and without exaggeration. The other problem is that they are often out of date by the time they are completed by all the relevant staff.

Expertise Location – Interaction

The next approach is through interaction – that is, how do other people perceive your expertise? We just mentioned LinkedIn; in addition to the profile you create, it can also display skills and endorsements and recommendations, which are provided by other people.

Another interaction-based approach is through communities of practice. These are groups organized around a particular process, technology, or challenge. Individuals participate based on their interest and the value they perceive from participation; those who participate actively, speak, organize, etc. are often perceived as experts within that community. These could be in-person, like user groups, or online communities and forums.

Finally, there is a class of tools called question and answer (Q&A) platforms. As the name suggests, these are tools that let some users or, in some cases, anyone ask questions. Someone or anyone can answer questions, and just about anyone can rate the questions and answers. These make it fairly obvious as to who the experts are on a particular topic or issue. Interestingly, these can also surface gaps in existing documentation, training, etc. and point to potential opportunities to address them.

© AIIM – 2019

Expertise Location – Analytics

The last approaches we'll review here rely on the power of analytics to identify experts. One way to do this is to use content analytics to evaluate individuals' contributions to documents in the repository, with the idea being that the better and more frequent the contributions, the greater the level of expertise.

We can also leverage analytics to analyze email traffic or social media traffic. Who's being emailed about a particular topic, who's blogging on a particular topic, whose blog posts or articles or Tweets are reposted and forwarded? Social media in particular presents interesting opportunities to identify expertise through the so-called "social graph," or the interactions between individuals across a variety of platforms.

Issues with Expertise Location

There are a number of issues associated with locating experts. The first and perhaps most obvious is that of false expertise, generally because the "expert" claimed expertise he or she didn't actually possess. This could be for any number of reasons but should be straightforward to determine using the analytics – and interaction-based approaches.

Another issue relates to the frequent disconnect between an individual's official job title and their actual expertise. Sometimes this is because of the proliferation of job titles that are more creative than informative such as "Information Maven." But it's not uncommon for people to be hired for one thing and have a passion and skill at something else that would benefit the organization if only it knew.

The last issue we'll mention is that of "expert fatigue." This occurs when the organization relies on a particular expert so much that the expert starts feeling stress and overwork from what is generally an additional role or duty. In these cases, it may make sense to set up a center of excellence to let those experts serve in that capacity full-time.





Test Your Knowledge

Question 1:

Which processes would most effectively transmit tacit knowledge from senior to junior staff? (select 2)

a. On-the-job training
b. Providing reference manuals
c. Mentoring
d. Delivering online training

Question 2:

Which approach to expertise location is based on their contributions via email and social media?

a. Q&A platform
b. Recommendations
c. Profiles
d. Analytics

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Which processes would most effectively transmit tacit knowledge from senior to junior staff? (select 2)

a. On-the-job training
b. Providing reference manuals
c. Mentoring
d. Delivering online training

The best answers here are A, on-the-job training, and C, mentoring. Both of these allow the junior staff to learn from the senior staff and their experiences. B, reference manuals, and D, online training, are both examples of explicit knowledge.

Answer to Question 2:

Which approach to expertise location is based on their contributions via email and social media?

a. Q&A platform
b. Recommendations
c. Profiles
d. Analytics

The best answer here is D, Analytics. For A, Q&A platforms are specific tools. For B, recommendations are typically associated with social media but are often related to other factors such as job or volunteer experiences. For C, profiles are self-reported and have little to do with email or social media contributions.



The Migration Plan

The following definition for migration is taken from ISO 15489, the international standard about records management. It defines it as the "act of moving records from one system to another, while maintaining the records' authenticity, integrity, reliability, and usability."

Migration is much broader than just records, however. Any time a legacy system is decommissioned, it should be reviewed to determine what information is stored therein and whether that information still needs to be actively accessible.

The Migration Plan – 4 Key Sections

The migration plan and process consist of 4 key sections. Each of these includes one or more steps designed to ensure the migration is successful and is carried out as efficiently as possible. The four sections are:

- Strategy and objectives
- Planning
- Preparation
- Migration and post-migration



Strategy

The first section is the strategy. The strategy should first identify the purpose of the migration. How the migration will be conducted, what will be migrated, and where it will be migrated to will all depend on the purpose to some degree. These might include:

- A system is becoming obsolete. Data needs to be migrated from the obsolete system to a current system.
- The system is not usable, or perceived to be so, such that users do not want to use it.
- Standardization. Data is stored in one or more proprietary hardware or software systems or file formats. The data is migrated from the proprietary format to a standardized one.

The next step in the strategy is to identify the stakeholders for the migration and ensure their needs are taken into account. This will certainly include records management, IT, legal, and the business process owner that owns the information to be migrated.

Planning and Scoping

The next step is planning and scoping the migration. This step includes a number of activities:

- Determine scope for migration. We'll look at this shortly.
- Establish stakeholder expectations. Stakeholders need to understand the starting point, the desired end point, and how to get there. This includes what the expected outputs will be of the migration and what effort will be required of the organization and its employees.
- Determine the source and target systems and formats, for example from a network file share to a specific information management repository, or from a proprietary format to a standardized one.
- Determine the schedule. When will the migration start? When will it end? How will it be done – nights and weekends, or starting on a target date and continuing until it is complete?
- The team also needs to identify dependencies. For example, if the system to be migrated is in heavy use for a particular project or process, that needs to be identified and plans developed for when it can be switched over.

© AIIM – 2019

Scope of Migration

So what DOES get migrated? It will depend on the situation.

Some migrations will involve all of the information in the legacy system. This is particularly the case when the system is to be decommissioned or when media or format obsolescence threatens access to the information. Even in this instance, however, it is generally not necessary to migrate records whose retention period has expired. Legal should be consulted to verify this is the case prior to disposing of anything.

In other cases, it might make more sense to migrate only some of the information – for example, only records, or only active information.

Remember that any migration must include some or all of the metadata for the information being migrated. Without metadata it is difficult to search for information, for example, and it would be difficult to argue successfully that the records' integrity and authenticity has been preserved. When migrating to a new system, you need to reach agreement on what metadata will be transferred.

Lastly, don't forget the audit trails. Many information management systems, especially older ones, do not have the capability to export and import audit trails easily. Remember to consider how you will demonstrate the integrity of the migrated information if you do not migrate the audit trails.

Identify Migration Tools

Many applications have some type of administrative tool to allow for bulk transfer of records. Some applications focus on the ability to export from their repository to elsewhere, while others focus on ingesting information from outside sources. Regardless, these are often the best tools to use because they were created with that application's specific data structure in mind.

There may also be third-party tools available. This is especially true for applications and repositories built on top of standard relational databases like Oracle, DB/2, or SQL Server.

Unfortunately, for some applications there may be no tools available at all. This could be because the system is too old or was developed in a very proprietary fashion. In this case the organization may choose to develop its own tool to perform the migration. This can be expensive to do, and care must be taken to ensure that all of the information to be migrated is captured and migrated correctly.

Design the Target Solution

There are two sides to every migration – the export side and the import side. The information to be migrated has to end up somewhere, and the time to design the target solution is BEFORE the migration starts, not in the middle or at the end. This means setting up:

- Classification structures, such as folders
- Metadata structures, including controlled vocabularies
- Security structures, including access control lists, roles, and groups
- Templates for data entry
- And anything else required for users to access the system post-migration
- Clean Up the Data to be Migrated

The next step in preparation is to clean up the data prior to migrating it. The migration will be expensive and complex enough without migrating an average of 2.6 copies of everything.

The team should try to identify and remove old, inactive, or expired information that is not a good candidate for migration.

It's also important to identify and merge or remove duplicate records and documents. There are software tools that can assist with this, though they are generally better at identifying exact duplicates (identical bitstreams) than multiple versions or renditions.

At some point the existing metadata will need to be updated to match the target solution; this could be done before the migration or after, but it will be necessary.

And of course all of this must be done within the context of the records program. That means documenting any decisions to not migrate, documenting any deduplication, documenting metadata, etc. The more documentation captured as part of the process, the greater the likelihood of its being accepted later should a question arise as to provenance.

Before the Migration Process

The process for migration can only start when preparation is complete.

You need to make some basic checks before letting the specialists start using bulk export tools. Systems experts can be over-confident with data manipulation, so take the time to reassure yourself that the tools work for your types of record, the volume of data, and so on.

Are the users fully aware, and no longer accessing the system or that part of the system where the information is held? It has been known for users to access and 'lock' records in the database, thereby unexpectedly halting the migration process.

And is the destination system or holding medium ready for you? There are many checks to carry out before you begin the actual migration process, both on the old system and the new.

Once the team is confident everything is prepared, the migration can begin. As the migration progresses, it is important to monitor to make sure nothing bad happens. For example, an organization was migrating email messages from an older email archiving point solution to an ECM repository. Unbeknownst to the organisation, the ECM solution could not read ingest more than 32 GB of data at a time. The organization began the migration of 150 GB of email. The ECM solution wrote the first 32 GB to its repository and then began overwriting it with the next 32 GB. The system administrator was monitoring and saw that the repository size stopped at exactly 34,359,738,368 bytes (or 32 GB), less than a third of the way through the scheduled migration. He stopped the migration and a workaround was identified that allowed the migration to restart and continue. Had he not been monitoring, however, it's possible that the problem would not have been identified until the migration was complete – or later.

Complete the Migration

Once the migration appears to be complete, it is important to retain the original files and systems until the success of the migration can be verified. This is the most important step in the process because of the possibility of permanent loss of data if the migration was not successful.

One of the most basic checks is to do an item count. If 27,500 files were scheduled to be migrated, there should be 27,500 files in the migrated system. In the example from the previous section, 150 GB of data should have been migrated.

It is also recommended to do some more extensive checks to verify for example that the migration didn't result in 27,500 empty or corrupted files, that metadata was associated with the correct record, that any linked files retained their linkages, etc.

And the final check is to have users confirm that the migrated files made it and are readable, or that the data made it from the old system to the new one. Since the users are the ones using the system and the files, it stands to reason that they will be the most likely to find subtle errors in files or systems.

Migration Issues

This module will help you to determine the issues associated with migrating legacy content from one location or system to another.

Migration Issues to Consider

In planning and executing a migration, there are a number of issues to consider. We can group these into the following areas:

- File format issues
- System dependencies
- Data quality issues
- Process issues
- Decommissioning issues

Now let's look at each of these in a bit more detail.

© AIIM – 2019

File Format Issues

Any migration involving unstructured data, that is, individual files, is bound to run into issues migrating certain file formats. These issues include:

- Proprietary formats. The older these files are, the more likely they are to have issues. Special care should be taken to ensure that they need to be migrated and that they were migrated successfully.
- Complex formats. These are similar to proprietary file formats; in fact, most proprietary formats are also complex.
- Linked formats. Engineering drawings with linked external reference drawings, spreadsheets or PDFs that link to each other, or any other types of linked files often run into issues with the paths to the linked documents.
- Unknown formats. Most repositories can store any kind of digital data, but if you run into unknown formats there is a question as to whether to even bother migrating them.
- Duplicate files. Depending on which research you believe, organizations have anywhere from 3 to 10+ copies on average of every document they store. This is often because the legacy system in question is network file shares, and when users finally locate a long-sought resource, they download it to their own personal file stores. One of the outcomes of the migration, and frankly one of the first ones, should be to identify those duplicate files, determine which one should be the official copy, and mark the other copies for deletion.

The rest of this section deals with higher-level issues that affect everything in the legacy and target repositories.

System Dependencies

When migrating from one system to another, there are a number of specific issues to take into account.

- **Integrated systems.** Systems that have been tightly integrated may have unanticipated dependencies around data structures and output files.
- **Reports. I**ntegrated systems may also be generating reports that rely on data from both/all systems involved. And even single system repositories may be generating reports in a unique way that can't be done identically in the new system.
- **Process dependencies.** Work processes, both manual and automated, may rely on how a system works, what its reports contain, how its metadata is structured, and more. This is exacerbated in automated processes where workflow rules are very specific as to the conditions for a particular task or step.
- **Bandwidth and processing issues.** This doesn't seem like it would be a huge issue until you start transmitting 10 TB of data across the network – or from one data center to another thousands of miles away.

Data Quality Issues

Data quality issues are a huge concern for a migration project – what's the point of doing it if the end result is inaccurate, inconsistent, or ends up in data actually being lost? Some of the issues to consider here:

- Redundant, outdated (or obsolete), and trivial content (ROT). A migration takes long enough without including terabytes of outdated, personal, or irrelevant information. Where that information can be isolated (and it can be), it should not be migrated, and in fact it should be disposed of in accordance with the records management program.
- Lifecycle considerations. What we mean here is that if some of the data in the system to be migrated has met its retention requirements and there are no other legal, operational, or historical reasons to keep it, it doesn't really make sense to migrate that data just to turn around and delete it.
- Missing metadata. This is often the case because new fields were added; in many instances these new fields are also mandatory. As the target system and its data structures are being designed, attention should be paid to this to determine how best to fill in that missing metadata. This is also known as metadata enrichment.
- Inconsistent metadata. This is very common as different systems use different data structures. The way to approach this is generally to map the field in the legacy system to the one in the new system, either through a middleware application or by actually transforming the legacy metadata value into the new one during the migration process.

Process Issues

There are a lot of process-related issues to consider during a content migration.

Perhaps the one most thought about is accuracy and quality control. That is, was the migration accurate, and how can you verify it? The migration tool and process can provide some metrics around number of items, etc. but in all likelihood you won't *know* the migration was accurate until your users start to interact with their content using the new system.

The next question is timing and duration. How long will the migration take? Longer than you think, but it shouldn't be a never-ending effort. When does it start? Well, when can you freeze any additional changes to the legacy system, and what's the impact of that on your work processes? And of course when is the migration, and therefore the migration project, complete? And maybe as importantly, when do you cut off all access to the legacy system? Because if you don't users will continue to use it.

It should be clear that communication of all the previous points to those affected should be a high priority. In the absence of consistent, regular communication, rumors will fly and users may take counterproductive steps like saving all their content to a less governed location like their own computer or a flash drive.

Finally, where you have individual users participating in the migration process, getting them to do it is often a challenge, and when you do get them to do it, they often want to keep everything "just in case." You should ask the question whether they really need all that information, and why. One tool that can really be helpful here is a report that can show users that they haven't accessed a particular document or folder or repository in X number of years, that that document or folder hasn't been accessed by ANYBODY in X number of years, etc. These are easy to generate for many repositories.

Decommissioning Issues

You're done with the migration. You've addressed all of the issues in this section and your migration was a huge success. Now what?

More specifically, what do you do with the legacy system, and with the data on that system? In both cases, the answer needs to involve IT, legal, records management, data protection, etc.:

- IT will ultimately be executing whatever the decommissioning strategy is.
- Legal needs to determine whether there are any legal or other holds on the system or the data.
- Records management needs to determine whether there are any retention requirements that still need to be met.
- Data protection needs to determine whether there is any personal or sensitive data in the legacy system that needs to be handled differently.

Once all those issues have been considered and addressed, the system can be decommissioned according to the appropriate decommissioning process for that type of system.

© AIIM – 2019



Test Your Knowledge

Question 1:

When migrating from one system to another, what information should be migrated?

a. It depends on the information in the original system and the reason for migration.
b. Only the metadata from the original system should be migrated.
c. Everything in the original system should be migrated.
d. Whatever information legal determines is on legal hold should be migrated.

Question 2:

How should legacy systems be addressed once the migration is complete?

a. They should be reviewed for decommissioning.
b. They should be set to read-only access and left on the network just in case.
c. They should immediately be destroyed.
d. They should be left in place, with documentation referring users to the new system.



Test Your Knowledge - Answers:

Answer to Question 1:

When migrating from one system to another, what information should be migrated?

a. It depends on the information in the original system and the reason for migration.
b. Only the metadata from the original system should be migrated.
c. Everything in the original system should be migrated.
d. Whatever information legal determines is on legal hold should be migrated.

The best answer here is A, it depends on the information and the reason for the migration. There are very few migrations where B, only the metadata, would be a valid approach. For C, sometimes everything should be migrated, but many times some of the information has no ongoing value and should be deleted or simply left on the original system. For D, information that is on legal hold generally should not be migrated because it can cause issues around authenticity and chain of custody. Even if it were migrated, for example because the system is to be decommissioned, it would generally not be the only information that would be part of the migration.

Answer to Question 2:

How should legacy systems be addressed once the migration is complete?

a. They should be reviewed for decommissioning.
b. They should be set to read-only access and left on the network just in case.
c. They should immediately be destroyed.
d. They should be left in place, with documentation referring users to the new system.

The best answer is A, they should be reviewed for decommissioning. For B and D, leaving the legacy system in place is confusing to users and can result in liability because the information on them is still discoverable as long as it exists. For C, destroying them too soon could result in records or other important data being inadvertently lost, which could also increase organizational liability.



© AIIM – 2019

Extracting Intelligence from Information

Introduction

Once content has been created or captured by the organization, we need to extract intelligence from it to provide context to it. We start by looking at metadata – what it is, how to apply it, and how to make it meaningful and accurate.

Next, we review taxonomies – how to develop them and how to choose the best one for a particular set of circumstances.

We take a look at new way to automate how we create metadata and taxonomies, through the use of powerful data recognition and extraction technologies.

These in turn can be used to feed analytics and machine learning tools that can add additional insight and help to automate other aspects of information management.

Finally, we review approaches to search to ensure users can find the information they need to do their jobs.

Information in Context

This domain is really the gateway for leveraging and exploiting information in support of the organization's goals and objectives. Information needs context, and we need to provide that context in a way that doesn't burden users but instead supports them. This means we need to take full advantage of recognition and analytics technologies to streamline and automate how we develop that context.

This domain makes reference to a number of tools, but ultimately the tools have to serve the business needs and outcomes, not drive them. That said, these tools and processes can offer significant benefit in terms of understanding information in new ways and in being able to leverage that intelligence to drive innovation and the customer experience.

The Metadata Strategy

The Benefits of Metadata

Metadata defined. There is no one definition of "metadata" that is internationally and universally agreed – rather, there are many similar definitions or descriptions which mostly cover the same points and you should adopt the one most suitable and relevant to the context of your information management activities and the organization in which you work.

The ISO standard 15489, "Records Management," provides a simple definition of metadata, in its Terms and Definitions section. It defines it as "Data describing context, content and structure of records and their management through time."

The U.S. Department of Defense has a definition of metadata in its DoD 5015.2 standard, which is also similar to the ISO standards, namely "Data describing stored data: that is, data describing the structure, data elements, interrelationships, and other characteristics of electronic records." This illustrates several of the other purposes served by metadata in ERM systems.

Finally, NISO, the U.S. National Information Standards Organization, defines metadata as "Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource." Metadata is often called data about data or information about information.

Perspectives on Metadata

The act of entering metadata values is often called "Indexing", especially when a basic set of values about an item are being captured. It's also a form of shorthand – "Have you indexed those documents yet?" is easier to say than "Have you entered the metadata for those documents yet?"

ISO 23081 is careful to explain that metadata can support different needs, because different views and perspectives on metadata are possible and may coexist. These include:

- The business perspective, where at least some of the metadata supports business processes.
- The user view, particularly when seeking information, where metadata enable the retrieval and support understanding and interpretation of content.
- And the information governance perspective, which includes things like security, privacy considerations, and lifecycle management metadata.

Business Value of Metadata

The primary value of metadata comes with how it is aligned to and supports specific business goals and objectives. Metadata is a key to organizing content: the term for this is “classification”.

Metadata can be used to track things like the dates associated with a document's associated record schedule. Or, metadata can be used to flag a security setting, validating access and edit rights, and thus controlling distribution. Metadata can also be used as a way to capture users' rating of content; for example, indicating that content is "valuable or "useless" or even "dated”.

Metadata is an important part of the content capture, creation and organization phases of the content lifecycle. If associated metadata is not captured at the same time that the content is, you will quickly create a collection of content that is difficult to manage, find and retrieve. Metadata is extremely valuable as a search and retrieval enhancing mechanism.

Metadata also potentially provides much greater precision to an otherwise free-text query by allowing the user to target a query on a certain field, such as author, subject, date, etc. documents they create or use.

In short, metadata is one of the foundations for managing information efficiently and effectively.

Metadata can also provide value in the way content interfaces to business processes. For example, if loan applications have to be reviewed in the order that they were submitted, if the date of receipt was not captured as a metadata value, there would be no way to ensure the applications documents were addressed in the right order.

If documents need to be processed collectively, metadata provides the foundation for batch processing. For example, if resumes for all employees working on a new project are to be updated to include commentary on the project, tracking an employee's current project would allow all relevant resumes to be retrieved for a batch update. This also leads to enriched connections between people and content, helping build expert locator bodies of information.

Metadata also serves as a point of integration. Different content in different applications, even across different information management systems, becomes "linkable" through common metadata properties and values. A customer or project ID can be consistent across document repositories, as well as ERP or financial applications.

When large volumes of content need to be analyzed, metadata also provides input for business intelligence and analytic tools. For example: the ability to report on the number of invoices coming into the payables department, or the ability to dynamically determine the number of positive customer support calls received in a given time period, are valuable business intelligence insights, but are only possible if those characteristics and values are tracked in metadata. In these two examples, customer IDs and a document type of invoice, or the flags set on a closed customer support call are useful metadata for reporting activities.

Finally, metadata can be used to enrich knowledge management based on user profiles. When user profiles are associated with the people adding, using or retrieving content from the ECM system, then the metadata associated with those people can be used to find experts on various topics, captured as metadata associated to their profiles and the documents they create or use.

© AIIM – 2019

The Metadata Strategy

Many organizations use metadata. Every system uses metadata to store and retrieve data. But in too many organizations, every system uses similar but different metadata, with the result that different data structures and approaches make information harder to find and manage, not easier. Take a simple example of an employee name. In one system, it's "first name last name". In another, it's "last name (comma) first name". And in still another, it's two fields: "First name" and "Last name".

To avoid this, organizations should develop a metadata strategy. The goal of this strategy is to identify and define how metadata will be defined, captured, and managed across systems and processes, such that for any given concept, the metadata used to describe it is used consistently throughout the organization. This will improve findability and the ability to manage content over time; it also helps with emergent processes for discovery and automated management of information using analytics.

Guidelines to Determine Metadata

Ideally, the metadata design for an organization is based on a structured methodology and approach. However, to get a better idea of how to identify metadata, here are a few high-level guidelines.

Based on the content types already identified, determine what is, or should be, available to users to help them find that content. In other words, what do people know at each step of the work process when they retrieve documents? Retrieval of the same content type may be different at different points. For example, you may retrieve a client document based on a client identification number early on, but prefer to use the client's name when you become more familiar with the client.

Don't just recreate the current way information is retrieved. Create an "ideal" scenario. You may know how to retrieve information based on how the current folder system is set up, but that does not mean that is always the best way.

In general, when you create content, you will think about what you created rather than how someone else will want to retrieve it in the future. It's important to keep both needs in mind in order to avoid costly rework later.

Tasks to Determine Metadata

Identify documents and metadata that are shared among workgroups or departments. Make lists and content types consistent as much as possible. A record used in payroll – for example an expense report – may be shared with human resources or accounting. But each department may have unique needs or ways to retrieve that record. It can be tempting to make sure all possible scenarios and exceptions are covered when designing metadata. Some exceptions can occur so infrequently that spending time entering properties for the column may not add value.

Pay attention to the time it will take you to enter metadata values. If there are too many properties to fill in, your users may try to circumvent the data entry step, or store the documents elsewhere. If you don't want to fill in the data, no one else will either, so consider prioritizing and reducing what metadata is required to what is essential.

Analyze existing filing structures – paper documents in file cabinets, existing databases, etc. – since these likely have worked for you in the past.

Identify metadata types and formats, such as date, number or text, and align with an enterprise data dictionary. An enterprise data dictionary is a definition of data, formats and data relationships within the organization.

Include metadata requirements that may result from system needs as well human needs, such as workflow or integration with other line of business applications.

Mandatory vs. Optional Metadata

There is a trade-off between having too many mandatory elements (users may find entering their values a chore), and too few (users may enter almost nothing about an item, making its future use and management difficult).

Actually, some entries may be conditional, depending on the type of document or record of interest. Thus a "mandatory if applicable" entry is also a valid option described in some standards. This would apply, for example, to a letter or similar item of correspondence, where an entry for the "Addressee" element would be considered mandatory, but not relevant for a project schedule.

Similarly, the Location element may be considered mandatory when used for physical objects which can be loaned out or otherwise moved, but might be irrelevant for digital records stored in a recordkeeping system.

Note that records management staff often want as much mandatory metadata as possible, in order to achieve a metadata-rich repository. This is often at odds with the user community, who are often not prepared to spend the time and effort entering it! A sensible compromise is necessary, as well as the use of automation techniques, as will be explained later in this module.

The Metadata Model

There are many ways to develop a metadata model, but they should all include certain elements.

- **Field name.** This should be specific enough that it is unique and understandable. Date is almost always a bad field name – date of what? Rather, Date Paid, Date Published, etc. are clearer.
- **Data type.** A field that is typed as date, numeric, currency, etc. can then be searched for a range, e.g., "all invoices between Jan 1 and Feb 28" or "All expense reports over $1,000".
- **Mandatory/optional.** Many fields are less useful if they are not mandatory, but if this results in users having to enter many mandatory fields via manual data entry, there will be issues with uptake and completeness.
- **Source of the data.** Is it manual metadata entry, a system or default field, does it come from an external data source, etc.?
- What systems use that metadata field in exactly the way it's been outlined?
- The owner or steward of that metadata and what it's assigned to.

Given the recent and ongoing changes in many jurisdictions relating to privacy and data protection, it may also make sense to identify where fields contain personal or sensitive data.

Metadata Standards

As you consider the metadata model, you should be aware that there are a number of metadata standards available to consider. Dublin Core, ISO 11179, ISO 23081, and many others offer various lists and structures of metadata elements.

For the purposes of this discussion, we don't recommend any one particular standard, because in all likelihood none of them are an exact fit for a specific organization's needs. Standards are important, especially when exchanging information with third parties, but it's also important for the metadata model to reflect any unique needs or perspectives within the organization.

Metadata Automation

Because of the volume of information being created or captured in most organizations, manual metadata entry is a difficult sell. Users don't want to do it, and when they do it, they make mistakes. So one of the elements to include in the strategy is a bias towards automation. It is not necessary to include the specific approaches for this in the strategy, but it's important to address that metadata will be automated to the extent possible.

We address metadata automation in more detail elsewhere in this section.

© AIIM – 2019



Test Your Knowledge

Question 1:

How does metadata support more efficient information management? (select 2)

a. It makes content more findable and retrievable.
b. It determines retention requirements for digital records.
c. It supports access controls.
d. It ensures information is disposed of when no longer needed.

Question 2:

How should organizations use metadata standards?

a. They should synthesize a master list of metadata from the available standards.
b. They should use them as a starting point and tailor them to their needs.
c. They should pick the most applicable standard and implement it in its entirety.
d. They should avoid standards because they won't be relevant in most cases.



Test Your Knowledge - Answers:

Answer to Question 1:

How does metadata support more efficient information management? (select 2)

a. It makes content more findable and retrievable.
b. It determines retention requirements for digital records.
c. It supports access controls.
d. It ensures information is disposed of when no longer needed.

The correct answers are A and C. For B, metadata does not determine retention requirements, though it can reflect and document them. For D, metadata can support effective lifecycle management, but it cannot ensure that information is disposed of effectively.

Answer to Question 2:

How should organizations use metadata standards?

a. They should synthesize a master list of metadata from the available standards.
b. They should use them as a starting point and tailor them to their needs.
c. They should pick the most applicable standard and implement it in its entirety.
d. They should avoid standards because they won't be relevant in most cases.

The best answer here is B, use them as a starting point. A will be a complex endeavor that will likely result in many terms that are not used. C doesn't reflect an organization's unique requirements. D may be true for a particular standard, but they are standards for a reason – they do have value and relevancy for most organizations, at least to some extent.



© AIIM – 2019

Capturing and Managing Metadata

How to Capture and Apply Metadata

Systems and Default Fields

In this section we discuss how to actually capture and apply metadata to digital objects. There are a number of different approaches we can take depending on the business context, the systems being used, the presence of existing metadata, etc.

So let's start with fields where the metadata is already present. System fields are captured automatically by the system in response to some sort of triggering event. For example, when a document is captured into the repository, a unique ID is automatically generated and populated, as is file name or document title or something similar. System fields are generally not updateable.

A default field is similar in that it is filled in, but it *is* updateable if warranted. For example, a user captures a document by scanning it into the system. There is a field for date scanned, which is a default field populated with today's date. There's another field for the date the image was checked for quality control and released to the repository. That, too, defaults to today's date, but if it doesn't get checked until tomorrow, that field could be updated (manually or automatically) to reflect the correct date.

Manual Metadata Entry

The first and most obvious method for applying metadata is manual data entry. In this approach the user is required to enter metadata for the document into prescribed profile fields.

Manual metadata entry is expensive for a number of reasons. First, it requires someone to do the data entry, who has to be paid a salary and benefits. If these are expected to be authors or the persons capturing information, that takes time away from doing their regular job.

Manual data entry is also time consuming and can be expensive to find and correct errors.

And even if users are willing to do it, they may not have the background or training to do it correctly. Training and job aids can help, but they won't be able to prevent typographical errors, selecting the wrong value from a controlled vocabulary, etc.

The bottom line is that users don't want to do it, and if they do, they aren't generally very good at it. So one way to minimize these issues is to limit the amount of manual metadata capture that is required of end users.

We talk about how to improve the quality of metadata, especially manual metadata entry, in another section.

Metadata Extraction

Next, we can extract metadata directly from the document. In some cases, the repository or application can read the properties and extract that information. This is useful where organizations and users update the properties for new documents. But it can be misleading if they don't, because many users will use and reuse the same document or PowerPoint template, with the result that the title, subject, and author are the same for thousands of documents.

We can also use recognition technologies. The most common of these is optical character recognition and is used almost exclusively for images of scanned documents. In this approach the application renders the characters from the image and attempts to detect patterns of characters within the images. This can be upwards of 99% accurate but depends heavily on the nature of the documents, the quality of the images, and the abilities of both the software and the user.

Another type of recognition technology is barcoding. Barcodes are available in 1-d and 2-d flavors and are highly accurate but only hold a limited amount of information, particularly for 1-d barcodes. They are gaining traction in many electronic document applications to aid the recognition and extraction process.





And there are specialized recognition technologies available for recognizing and extracting data from audio, video, and other rich media types. These tend to be pricey and specific to the application and certain file formats, but may be quite useful for organizations that capture a lot of these types of documents.

Inherited Metadata

The principle of "inheritance" is a useful way of adding metadata to information objects.

When an item is captured into an information management system it will be placed into the appropriate file (or sub-file or volume) within the classification scheme, very probably joining other documents from the same business context and thus maintaining a useful collection of related and relevant documents for a given business area and its business activities.

The item will have had values allocated to at least its mandatory metadata elements, and several optional elements may also have entries, depending on the history and use of the item before its declaration as a record. The act of declaration will add additional metadata, for example - who made the declaration and when.

However, more metadata can now be applied to the item, based on metadata already present for classes and files in the classification scheme – very commonly this metadata provides particular security settings, and/or retention periods and disposal instructions for items stored in the many files. This metadata is said to be 'inherited' by the record, as a consequence of its declaration into a selected file (or volume) in the classification scheme.

Metadata from User Logins

Software can also be used to look up or capture the user's details. Many applications can use the Lightweight Directory Access Protocol, or LDAP, or other identity or directory-type applications to capture certain details, such as the current user's name, job title, department, etc. This could be done in the office productivity application, such that all documents created by the user have that information entered automatically, or in the information management application.

Metadata from Other Data Sources

Metadata can also be read or copied from existing data sources through a number of mechanisms. While the details vary, the idea is that the data is already stored in a database somewhere, probably de-duplicated and normalized, and that the organization would do well to reuse that data and the effort required to create it rather than entering data into many different applications. This process could be manual or automated; the more automatable, of course, the greater the benefits.

Metadata through Workflow

If the organization uses workflows to streamline and automate document-centric business processes, those same workflows can often be used to update metadata associated with a particular document. For example, a field called "Status" might have values of "In Draft", "In Review", "Approved", or "Revision Required". As the publication review and approval workflow progresses, the workflow can update the field to reflect its current status.

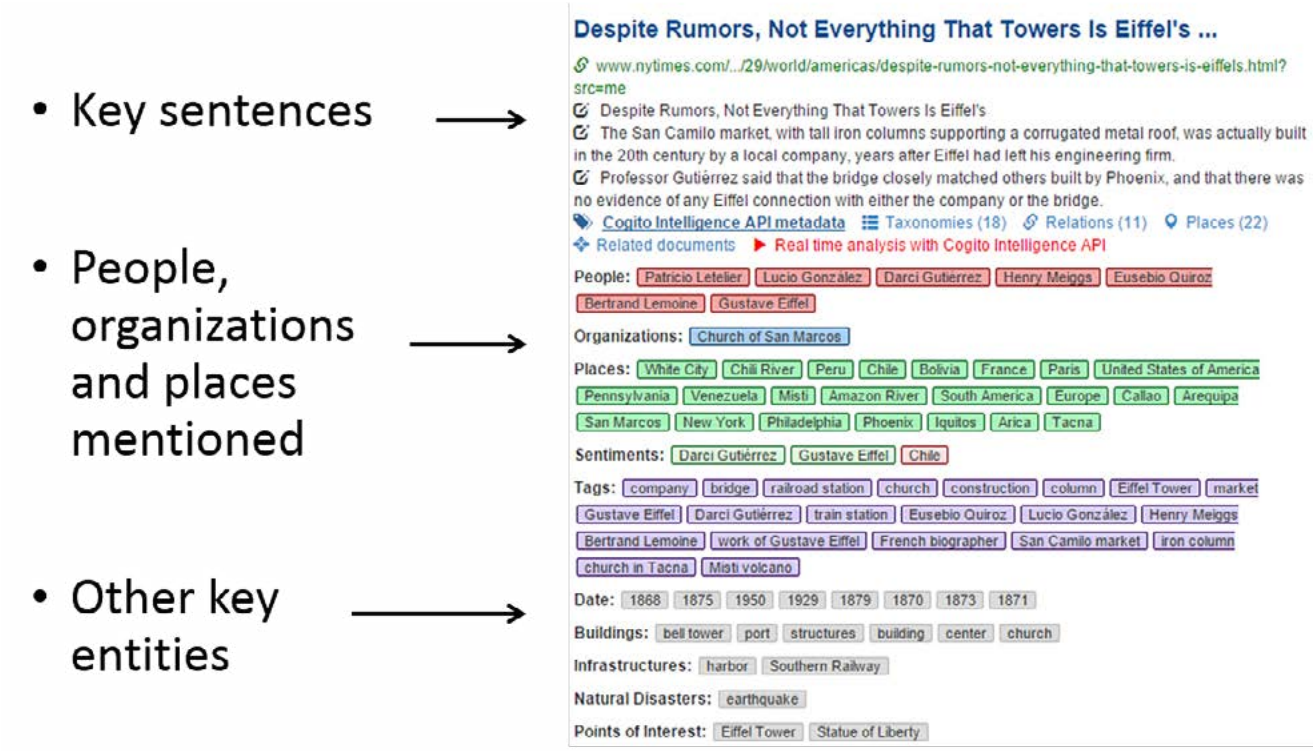
Similarly, an order processing workflow could use a similar field to display whether the order is "Received", "Processed", "Shipped", or "Delivered".

© AIIM – 2019

Metadata through Analytics

Analytics can also be used to extract and populate metadata – in this case directly from within the content of a document or record. Here's an example using an article from the New York Times. You can see that a number of different types of information have been extracted: key sentences; people, organizations, and places; and other key entities. Any of these entities could be used to populate a metadata field.

Analytics can also be used to "remediate", or fill in, missing metadata, for example as part of a system migration. This could be based on the contents of the documents in the old system, inference and mapping from the old system's metadata to the new one's; or converting deep hierarchical folder structures into metadata fields and values, among others.



Source: Expert Systems

The Challenges of Sharing Metadata Across Systems

This module will help you to identify the challenges of sharing/propagating metadata across tools and systems.

Sharing is Hard!

Sharing metadata across the myriad information siloes found in the typical organization is extremely important. Effective search and retrieval is based to significant extent upon complete and correct metadata. Yet in most organizations, finding, retrieving, and using information across systems is all but impossible, due to issues in how each system – or silo – is set up.

In most organizations, for any set of systems, these metadata-impacting issues are generally present:

- Different terms and values are present in each system. In one, we use the term "Customer name"; in another, "Customer"; and in a third, "Owner". Similarly, in one system we use a "first name last name" structure, and in another we use "last name (comma) first name". If controlled vocabularies are in use, it is not uncommon for different systems to have different values listed for the same field.

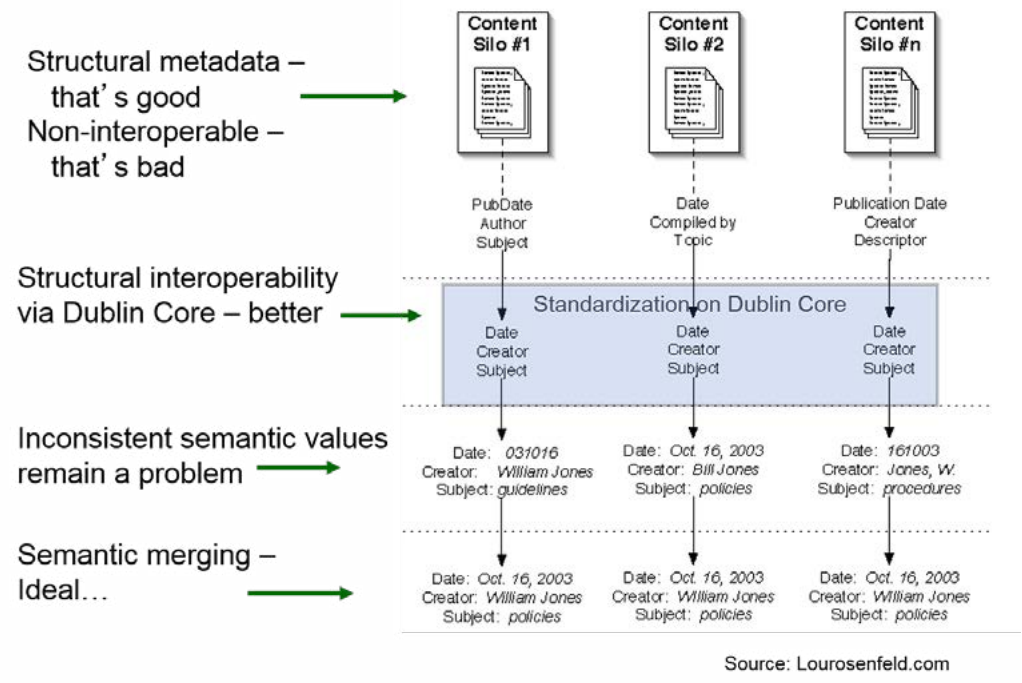
- Different structures. Continuing the example above, in a third system we use *two* fields: “first name” and “last name”. Maybe there’s a “middle name” or “middle initial” field as well. At least the different field names can be mapped; making this system interoperable with the first two requires a more complex solution.
- Different security/access controls. Access controls are metadata in a very real sense, and if you can’t access the document, you likely can’t access its metadata either. Again, resolving this becomes a significantly complex exercise.

At least as important as the technical constraints are strategic issues. Sharing can be difficult or impossible because of specific legal or regulatory requirements. We’ve discussed privacy and data protection requirements in a number of sections of this course because it touches so many things, and metadata is no exception.

Similarly, there may be other organizational constraints such as the need to maintain ethical walls between different parts of the organization due to conflict of interest rules. If you can’t legitimately access a document, you probably shouldn’t be accessing the metadata either.

Getting to Enterprise Metadata Interoperability

This is a key illustration, pointing out the path to metadata interoperability. At each stage, the content gets more and more interoperable.



At the top we have the typical enterprise with multiple content silos: ECM systems, shared drives, line of business systems, etc. We see three concepts: a date of publication, an author, and a subject/topic. Each of these has a different term and, as shown in the third line, a different structure (last name, first initial).

At the second level the organization leverages the Dublin Core metadata standard to ensure consistency and structural interoperability between the silos. All three systems now use the same field names: data, creator, and subject. However, there is still the inconsistency of the semantic values as shown in the third line: while the field names are the same, there are three different date structures; two different creator structures and inconsistent values in those that are similar; and three different values in the subject field for the same thing.

The ideal is then at the bottom, where both the fields and their values have been semantically merged to create consistency of both the names of the fields and their contents.

There are a number of technical approaches to achieving this; the details are beyond the scope of this course but would likely require a discussion with a number of parties including IT, information architecture, the various information management disciplines, and of course end users.

© AIIM – 2019

Information Interchange

It's not uncommon for organizations of any size or complexity to have multiple metadata models because of legacy systems, approaches, acquisitions, etc. So how do we deal with this today?

The fact is that many standards originated as a means to share or exchange information among enterprises or groups with common interests. So some large industries, for example, such as the auto and aerospace industries, have set up common content structures to support standardization. So companies that are within these particular areas are able to more easily exchange the information that they have because they follow the same format, the same standard. More industries are following auto and aerospace. Financial services and health care are also coming out with standards very similar to these other ones.

For a manufacturer, content standards become a mandatory format for anyone delivering to them. They provide consistently and predictability, and adopters can predict the structure of the information that will be provided to them. You can thus define content repositories around those particular standards. The predictability of structure and metadata opens up new opportunities for the automatic processing of content. The more these standards are followed, the more easily it's going to be for us to do things automatically based on how that content is structured or categorized.

How to Improve Metadata Quality

Improving Manual Metadata Entry

There are a couple of ways to capture metadata – manual and automated. Many organizations still rely on manual metadata entry – by scanner operators, by end users, etc. – in order to capture metadata. As we've previously discussed, this is not ideal for several reasons. Nevertheless, if this is the chosen approach, there are a couple of additional steps that can be taken to help improve metadata quality.

The first one is to make them mandatory. As noted earlier, mandatory fields can be problematic – but optional fields often don't get filled in at all.

Next, we can create "drop down lists" – aka controlled vocabularies. These are simply lists of terms from which users pick the appropriate entry. For example, most online ordering forms will show countries in such a list, and if you select certain countries, it will show states, provinces, or prefectures in another list. These work best with limited lists of closely related things, such as geographic locations, the departments in your organization, or project names. The benefit of this approach is that users are required to select from the list – no options, no misspellings or variations, etc.

We can also validate the data as it's being entered and captured. Data masking applies a pattern to a field such that any data entered has to match that pattern. This is often used with identification numbers, invoice numbers, dates, credit card numbers, or anything that always used the exact same pattern.

We can also significantly improve the quality of metadata through automation. Any time we can automate a task or process, that task is going to be completed more consistently and more efficiently. We address metadata automation in more detail elsewhere in this section.





Test Your Knowledge

Question 1:

Which approach to metadata capture is most likely to contain errors or omissions?

a. Workflow
b. System-generated
c. User-entered
d. Inherited

Question 2:

What are the strategic issues associated with sharing metadata across systems? (select 2)

a. Different metadata structures
b. Inconsistent semantic values
c. Differing privacy requirements
d. Need for ethical walls

Question 3:

What is the benefit to using controlled vocabularies for metadata entry?

a. It automates the metadata capture process.
b. It ensures that the correct value is selected.
c. It eliminates misspellings and variant terms.
d. It validates the entry with an external data source.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Which approach to metadata capture is most likely to contain errors or omissions?

a. Workflow
b. System-generated
c. User-entered
d. Inherited

The correct answer is C, user-entered or manual metadata entry. The other three approaches rely on varying degrees of automation and are much less likely to contain errors or missing metadata.

Answer to Question 2:

What are the strategic issues associated with sharing metadata across systems? (select 2)

a. Different metadata structures
b. Inconsistent semantic values
c. Differing privacy requirements
d. Need for ethical walls

The correct answers are C, differing privacy requirements, and D, need for ethical walls. Different metadata structures and semantic values are definitely issues, but we consider them more intrinsically metadata related rather than strategic.

Answer to Question 3:

What is the benefit to using controlled vocabularies for metadata entry?

a. It automates the metadata capture process.
b. It ensures that the correct value is selected.
c. It eliminates misspellings and variant terms.
d. It validates the entry with an external data source.

The correct answer is C, it eliminates misspellings and variations. For A, users still have to select the value manually in most cases. For B, there is no guarantee that the user will select the correct value. For D, generally controlled vocabularies are not validated with an external source.

Classification Schemes

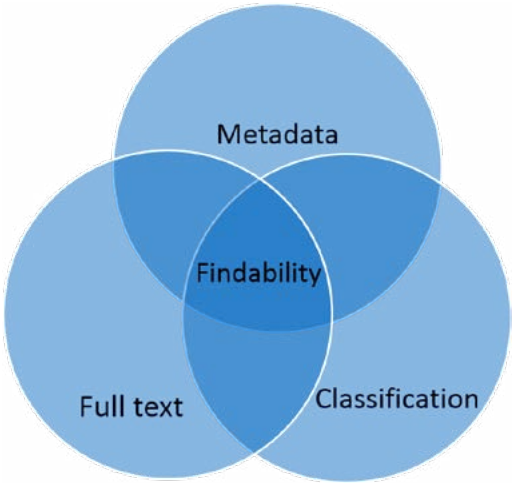
Findability

Introduction to Findability

Search and retrieval are key activities for users, and one of the primary means by which they will interact with digital information. Users will have many different ways to search for information within a particular repository. They will be able to search by:

- Metadata associated with a record, folder, etc.
- Content (full text) – for example text string.
- Browsing the folders in the classification structure or the nodes in the navigation scheme. We refer to the grouping of files as categorization or classification; while there are nuances, we treat them as interchangeable in this section. These are also referred to as taxonomies; we discuss taxonomies in more depth elsewhere in this section.

We can call these the “Three Pillars of Findability.” We address each of these in more detail elsewhere in this section, but let’s review each one briefly.



Search vs. Classification

There are two main approaches for effective access to stored information – classification or via a search engine.

Classification organizes information by storing it, logically-speaking, in collections, folders, or ‘buckets.’ This first categorization approach can be summed up as 'aggregate and organize.'

By contrast, the second approach uses the power of searching or search engines to find information, and does not recognize or have any need for aggregation. Note that this can apply to metadata or full-text searching. This second approach can be termed 'Find by raw power.'

This approach also depends on the user knowing what to search for in order to retrieve the desired records; for example, a search for “tank” will retrieve records relating to water storage devices as well as mechanized military vehicles – but would miss records relating to armoured personnel carriers (which are not generally referred to as tanks).

In a sense there is a 'tug of war' between these approaches. Each approach has its limitations and strengths.

© AIIM – 2019

Metadata

Let's start by looking at the pros and cons of searching using metadata.

Pros:

- Metadata search is very flexible. A good searcher can construct very sophisticated searches within or across multiple fields.
- Metadata searches are much easier to perform ranged searches on. For example, it's fairly straightforward to search for all invoices over a particular amount, all emails sent within a certain time period, etc.

Cons:

- Flexible can mean complex – especially if users aren't very well trained and familiar with the interface.
- Numeric searches require that the applicable fields use the applicable data type. This is often the default, but if a field is a text type as opposed to date, numeric, currency, etc. the searches will not provide the expected results.
- Any metadata search requires that the metadata be present – and correct, and correctly spelled, etc. in order to find it. Similarly, no matter how good the metadata is, it may not help a user who has made a typographical error.
- Finally, metadata requires a certain consistency in terms. If users can type anything they want into a field… they will. It's not just spelling, but also when a user enters a completely alternate term that a human can recognize as same or similar but a computer cannot. A thesaurus can be valuable in this context, but the preference should be to standardize on specific terms for specific concepts as much as possible.

Full-text Search

Here's a look at the pros and cons of using full-text search.

Pros:

- You can find any word or phrase in a document, group of documents, throughout a repository, or even across repositories and systems.

Cons – and there are a lot of cons to that one pro!

- The index being searched is what it is – complete with misspellings, alternate spellings, alternate terms for the same concept, etc.
- It's very easy for users to search incorrectly. First, most users aren't trained to search anymore – they use a search engine and call it a day. That works when you're looking for a new recipe, but not so much when you need to find all documents relating to a particular person or function. And even if they can search correctly, users make mistakes and misspellings of their own in the search interface.
- Perhaps most importantly, you can only use full-text search on things that have full text to be searched. That means office productivity documents; most (but not necessarily all) PDFs; web pages; and a few other types of information. It generally does not extend to images, audio, or video, though this is starting to change. But you generally can't search what isn't there. It also means that the text has to be indexed – this is generally automatic, but may not be instant.

Classification

Finally, let's look at the pros and cons of using a classification scheme.

Pros:

- Classification schemes can organize information for ready access.
- This allows users to understand what is included – and often what is excluded – from a particular process or domain area. Classification schemes also provide the preferred, or only, terms for a given concept.
- Classification schemes often map to how users do their work.
- A logical classification scheme will allow users to browse it to find what they are looking for – much like a bookstore.

Cons:

- Classification schemes are often not as logical or complete as they could be. Organizations and processes change, and it's hard to keep up at times.
- The more complete the classification scheme is, the more complex it is likely to be.
- Classification schemes often map to users' bad habits. It's not uncommon to see different levels of detail at the same level of a classification scheme, or the seemingly ubiquitous "Other" or "Miscellaneous" category.
- If the classification scheme isn't logical *in the eyes of the users using it*, they will get lost in it and become frustrated.

Ultimately if the classification scheme is too complex, too unwieldy, too incomplete, or too difficult to understand, users will not use it consistently.

The Thesaurus

A thesaurus can provide another way to support findability for all three pillars. A thesaurus is a list of terms and their relationships. These relationships can be more complex than a simple parent-child hierarchy.

The value of a thesaurus in the information management context depends on how it is used (and how complete and up-to-date it is). For metadata and full-text searching, thesauri can allow users to locate a particular concept using their preferred terms, rather than just the "official" one. For example, a metadata field for a location might have the value of "San Francisco", "SF", or "San Fran" on different documents. With a thesaurus and some additional setup, a user could search for any one of those three terms and find all the documents that contain any of them.

The Three Pillars of Findability

To conclude, then, each approach has its pros and cons – in isolation. What should be apparent is that the best results will come from leveraging all three approaches. If information is classified logically and consistently, with correct and consistent metadata, and a full-text indexing is possible, it should be much easier to find information and ensure that it is the *correct* information.

© AIIM – 2019

Classification Approaches

Classification Scheme – Definition

Put simply, a classification scheme is any structure an organization uses for organizing, accessing, retrieving, storing and managing its information. ISO 11179 defines classification schemes as including keywords, thesauri, taxonomies, and ontologies; for our purposes we can add file plan and records retention schedule to that list as well. Most organizations have MANY classification schemes being used at any given time; which one to use depends on who is using it and what the purpose is.

Another common and related term is business classification scheme, or BCS. A BCS is nothing more than a classification scheme which is based on an organization's business functions and activities.

Finally, many classification schemes are also referred to as taxonomies or categorization schemes.

Effective Classification Schemes

The key issue for any classification scheme is its ease-of-use and performance for the users.

If users are not happy, (and they won't be happy if there is poor ease-of-use and / or poor performance) then the whole environment will not be accepted, and the initiative will be deemed to have failed.

The way in which the BCS is designed and deployed will be a major factor in the ease-of-use and performance for the users – who, it's important to remember, may have minimal training in information management best practices.

This means that the focus needs to be on ensuring that the classification scheme is:

- Easy to use. It reflects the way users work, and, as much as possible, the terms they use.
- Concise. Terms or folders that aren't used shouldn't be added "just in case" – this will make it more confusing for users.
- Predictable. If the classification scheme is logical, users will be able to predict where to file and find information. If it isn't, users will not be able to find their information and will get frustrated.

Finally, every classification scheme will have to change over time as the business and its operations change. So there needs to be a process in place to make required updates when necessary.

You should note that the usability of any classification is affected by:

- The number of levels in the classification scheme. In general, more than four or five levels can be confusing, but ultimately the number of levels required may depend on the nature, size and complexity of the organization; and the degree of business need for speed and accuracy in control and retrieval of information.
- The user interface for the system in which the classification scheme is implemented. There are significant differences in the way that the current products in the marketplace address the requirements for this.
- The availability and quality of other retrieval tools. As discussed earlier, while the classification scheme is important, effective metadata and access to indexed full text can significantly improve the findability of business information.

Which Classification Approaches?

There are two primary considerations when determining which classification approaches to take.

The first is the principles of classification – where the primary options are functional; subject/topical; and organizational.

- **Organizational** classification schemes mirror the organizational chart. These are simple to put together, but can be difficult to maintain as the organization changes or reorganizes.
- **Subject/topical** classification schemes are often seen in law firms, where all the different practices have similar in concept, but very different in practice, information management needs. Subject or matter-based schemes are often relatively static because the topics or matters don't change structurally very rapidly.
- **Functional** classification schemes are based on the business functions and activities of the organization. These are easier to maintain than organizational over time; if the organization changes, in all likelihood it will still have an accounts payable function, a recordkeeping function, etc.

Generally, a functional principle of classification is preferred. From here on, therefore, we will consider only a functional principle of classification, that is BCSs.

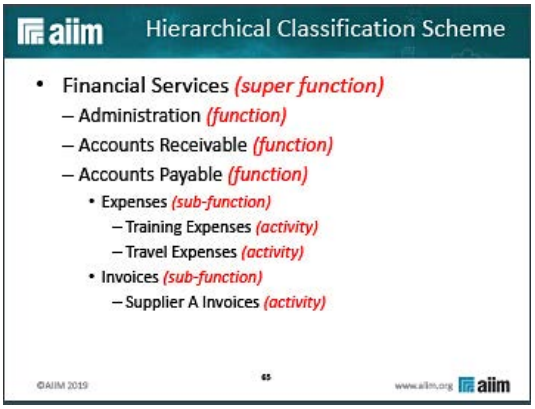
The second is deployment, where the key points are hierarchical or tree-style and keyword or thesaurus-based.

- **Hierarchical** deployments are very familiar to end users – folders, sub-folders, sub-sub-folders, and the like. This approach assumes that a particular document or record should be filed in one folder, which is not always the case. They are therefore not as powerful and flexible as thesaurus-based approaches, but they generally more usable for end users.
- **Thesaurus** or keyword-based deployments are much more flexible compared to hierarchical approaches because things can be filed according to numerous relationships using virtual folders, views, or other presentation formats. But they are significantly more difficult for end users to make sense of.

Accordingly, in this course we recommend using a hierarchical approach.

Hierarchical Classification Scheme

As you can see in this example of a hierarchical tree-structure scheme, we start with a…



Level 1 'super function' or very large area of management responsibility – in this example that's Financial Services.

This is broken down into 3 smaller functions:

Administration, Accounts Receivable, and Accounts Payable.

Each of these functions is then broken down, in turn, into smaller areas of management responsibility, known as sub functions. You can see that Accounts Payable is shown broken down into the following sub-functions:

Expenses and Invoices.

© AIIM – 2019

And you can see that each of these sub-functions is made up of activities which are performed within that area of management responsibility – that is within that sub-function. For example, you'll see that within the Expenses sub-function there are Training Expenses and Travel Expenses activities.

So, from the organization's overall Level 0 'mission,' the hierarchy first decomposes into: super functions (Level 1), then into functions (Level 2) and then into sub-functions (Level 3).

Thus, it moves from broad functional areas to narrow functional areas.

It then decomposes the smallest sub-function into 'types' of activity that make up that sub-function. The level of decomposition needed and appropriate for a particular organization in its classification scheme will depend on the number and complexity of functions it has.

Hierarchical Classification Scheme

And here's a graphical representation of what this portion of the classification scheme might look like.



Build vs. Buy

Finally, one more thought about whether to buy or build a taxonomy or other classification scheme. In practice, both have their pros and cons, and it should be considered as more of a spectrum between the two. Here are four spots on that spectrum to consider.

- **Buy.** You buy a taxonomy or classification scheme and implement it as-is. This approach is much faster to deploy than any of the others, and many industries have these available already today. This approach will help to ensure that you'll have better interoperability with any third parties, but it may not exactly fit your organization's ways of working and business requirements.
- **Modify.** In this approach you buy the classification scheme, but then customize it to meet your unique business requirements by adding new elements or removing extraneous ones.
- **Model.** In this approach you look at others' work – other publicly available schemes, how others in your industry do it (if possible), etc. The question here is really how close those other organizations are to what yours does.
- **Build.** In this approach you develop the entire thing from scratch, using internal and/or external resources. It's completely custom, so it perfectly meets your needs, but it's a very labor-intensive and demanding process.

Stakeholders for a Classification Scheme

This module will help you to Identify the stakeholders for a formal classification scheme.

Stakeholders

The first step in most implementations is to identify the key stakeholders. For a classification scheme, these would include:

- **The program owner,** in whose area of responsibility the classification scheme will be deployed and who is sufficiently senior in the organization to be able to champion the initiative.
- **Business unit managers** from those areas that will be using the classification scheme. It is their users who will be using it and, in the pilot phase, piloting the scheme so it is in their interest to ensure it meets their needs. Sometimes classifications schemes are only used by one process or function, but many times they are used by multiple groups, each of whose needs need to be taken into account.
- **Records management.** So many of these classification schemes relate to records management that their participation is a must; moreover, in many organizations that is where the expertise and experience in classification scheme development resides.
- **And IT.** For any information management program, there will ultimately be a requirement to load or otherwise implement the classification scheme in one or more applications. There will also be business logic to assign, and in most organizations one or both of these come under the purview of IT.

The Development Team

Experience has shown that the best approach to developing a classification scheme is to include a mixed core team comprising:

- **In-house staff** with experience in the development of classification schemes. As noted, this is often the records management function, but it could be anyone in the organization with the understanding and experience to ensure the classification scheme is complete, correct, at the appropriate level of detail, etc.
- **People from outside the organization,** with significant experience in the development of classification schemes – i.e., consultants. Note that this is not mandatory, particularly if the organization has staff with the necessary expertise and experience. But it is often the case that the organization believes it has more expertise than it actually does, and external resources can both speed the development process and potentially improve its effectiveness.
- And of course, **users in the relevant areas** must be involved. A taxonomy has to be usable by its target audience. Most classification schemes fail – that is, they are less likely to be used consistently – because they are too complex and difficult to use. This is often because they are developed by people with limited understanding of how to create them, or by people with limited understanding of the domain in question. A well-designed taxonomy, one that is usable, needs input from subject matter experts within the target function as well as those with background in information architecture or library science.

© AIIM – 2019



Test Your Knowledge

Question 1:

What are the issues associated with using full-text search? (select 2)

a. There has to be a text layer present and indexed.
b. Users can't navigate to the documents efficiently.
c. There may be misspellings and/or alternate spellings or terms in the documents.
d. The documents may be missing some or all of their metadata.

Question 2:

What is the preferred approach for most business classification schemes?

a. Topical
b. Functional
c. Matter-based
d. Organizational

Question 3:

a. Why is it important to have end users involved in developing the classification scheme?
b. They will be the ones using the classification scheme to do their work.
c. They have the most expertise in classification scheme development.
d. They are likely to be dissatisfied if they have no ownership of the classification scheme.
e. They understand the use cases and business needs the best.



Test Your Knowledge - Answers:

Answer to Question 1:

What are the issues associated with using full-text search? (select 2)

a. There has to be a text layer present and indexed.
b. Users can't navigate to the documents efficiently.
c. There may be misspellings and/or alternate spellings or terms in the documents.
d. The documents may be missing some or all of their metadata.

The correct answers are A, there needs to be an indexed text layer, and C, there may be misspellings or alternate spellings or terms. B refers more to the classification scheme, and D refers to metadata rather than the full text.

Answer to Question 2:

What is the preferred approach for most business classification schemes?

a. Topical
b. Functional
c. Matter-based
d. Organizational

The correct answer is B, functional, because it is more flexible compared to the others. For A and C, Topical and matter-based are both fairly specific to specific industries and based on relatively static lines of business. For D, organizational is difficult to maintain as the organization changes.

Answer to Question 3:

Why is it important to have end users involved in developing the classification scheme?

a. They will be the ones using the classification scheme to do their work.
b. They have the most expertise in classification scheme development.
c. They are likely to be dissatisfied if they have no ownership of the classification scheme.
d. They understand the use cases and business needs the best.

The correct answers are A and D. For B, this may be true in certain instances, but in most organizations end users do not have this background. For C, this is also possible, but not one of the primary reasons to have users involved.

© AIIM – 2019

Extracting Intelligence from Information

Extracting Information from Structured Applications

Introduction to Structured Data

The intent of this discussion is not to make you a database administrator. However, it is important to understand how structured data and databases work in order to manage that data effectively in the context of an information management program.

As the name suggests, structured data is data with a structure. That is, the format is well-defined and consists of a number of fields, or rows, and the values within those fields, or columns. The spreadsheet is a very simple example of this. Each complete row, combined with its values, form a single record (which may or may not be the same as a record in the context of formal records management practices).

Course ID	Course Name	Mode	Duration	Cost
AIIM01O	Intro to RM	Online	1 (day)	$50.00
AIIM01C	Intro to RM	Classroom	1	$100.00
AIIM02C	Advanced RM	Classroom	2	$200.00
AIIM03O	Intro to BPM	Online	1	$50.00
AIIM03C	Intro to BPM	Classroom	1	$100.00
AIIM04C	CIP Prep	Classroom	4	$400.00

There are many ways in which structured data can be stored, including relational databases, XML databases, and spreadsheets, but they all have a similar internal structure.

Structured Data Example

Where structured data gets interesting is that in many relational databases, there are multiple tables that relate to each other according to certain fields. This makes it easier to abstract and manage data.

For example, let's say you offer training courses and you want to create a database to store the information about your courses. While you could enter everything into a spreadsheet, if your courses are successful the spreadsheet will quickly become unwieldy. Instead, you could have a database that has one table that lists information about your courses: subject, whether it's online or in a classroom, the duration of the course, and the cost. You might have another table for student registrations: name, course or courses registered for, whether they've completed the course or not. Now, you can run all kinds of queries and reports: how much did a course make, how many students took the course, which courses did a particular student take and did the student complete them, what's the completion percentage by course or by delivery mode, and many more.

The challenge here is that if you determine that the retention for your course details is 2 years from the date of the last course offering, and your customer details are retained for 5 years from the date of last purchase, you'll lose the entire first table 2 years after the last course is offered. For the next 3 years, that Customer_Details table will be missing the data on the courses the student took because the CourseID field from the Course_Details table, and all of its data, is gone. This is one of the things that makes structured data so tricky to manage over time.

Data vs. Presentment

A common issue with structured data is that the data in databases is not really meant to be human-readable, and certainly not by the end user or customer. Take, for example, credit card statements. All of that data is in the database, and when credit card statements are generated to send or present to the customer, formatting is added so that it looks like a statement. If the statement is late, formatting might be added to turn the amount due figures red and bold; if the statement shows zero balance due, formatting might be added to include a special offer. These in turn are done using business rules and logic.

But this raises a fundamental question: what's the record? In other words, is the record just the data, or is it the data combined with how it is presented? If the latter, how can the organization ensure that it can regenerate that presentment a year or five years later if needed, given both the involvement of business logic and the likelihood that the overlay and the presentment rules will change over time? Or do you even need to?

Extract Using Native Tools

In terms of technology approaches to extracting and capturing structured data, the first thing we should note is that these are all things that would be done by database administrators or someone else in IT.

The first approach is to archive structured data using native tools. Most structured database applications and relational database systems have their own built-in mechanisms for extracting data for reporting, migration, etc.

The benefit of this approach is that, since the tools are built for that particular application or platform, the data extraction is much more likely to maintain referential integrity between all the moving parts within the system. In other words, no data is being inadvertently lost or orphaned.

The challenge is that in most cases these capabilities do not extend to other applications, meaning that the organization would have to extract data from each structured application individually. This was part of the intent of the creation of the Structured Query Language (SQL), but in practice each database management system uses its own flavor of SQL.

Extract Using Third-party Tools

Another alternative is to extract structured data using third-party tools. Some of these are stand-alone solutions, while others leverage what the main structured data application vendors know about accessing and extracting structured data to make those capabilities available to others.

The obvious benefit is the potential for the tool to be able to extract data from different applications within the enterprise. The challenge is that they tend to work well with the most well-known applications, meaning that if yours is an unusual or highly customized application, they may not work as well… or at all.

Output Capture

The last approach we offer here is to capture structured data in an output format. This approach converts the relational data within the application into a flat data stream, such as XML, or series of data objects, such as PDFs. This is often how individual statements are generated to send to customers or to post for electronic bill presentment and payment.

The benefit of this approach is that the resulting data streams or files are complete and maintain any formatting required at the time the report is run, so they are complete and readable from the customer's perspective. And if the organization is already doing this, there is no real extra cost in terms of data extraction.

But there are some significant disadvantages. First, this is analogous to printing digital records – all of the underlying functionality is stripped from the resulting data objects. If this is for legacy data, this might not be a bad thing, but organizations should be aware.

It does create some additional costs and resource requirements in terms of storing and managing the resulting streams or objects.

Perhaps most importantly, this approach doesn't necessarily address underlying performance and compliance issues, UNLESS the data that was output for capture is subsequently archived or deleted. In other words, in many organizations, this creates yet another copy, rather than serving as the copy of record.

© AIIM – 2019

Extracting Intelligence from Scanned Images

Recognition Technologies

Scanned images can be analyzed to recognize and extract text or other intelligent content, including company logos. There are a number of ways of doing this; let's start reviewing some of the different approaches to character recognition.

Optical Character Recognition, or OCR, is commonly used with printed office-type documents. Here, the scanning capabilities are capable of discerning different fonts, including bold and italic, and line and text spacing. Text can typically be extracted and copied into a metadata field or placed in a separate, but associated, text file. This data can then be used to search for a specific image or even "within" an image or PDF.

Intelligent Character Recognition, or ICR, is a sophisticated technique that enables the reading of handwritten characters, and may use the adjacent characters or context to improve the recognition rate.

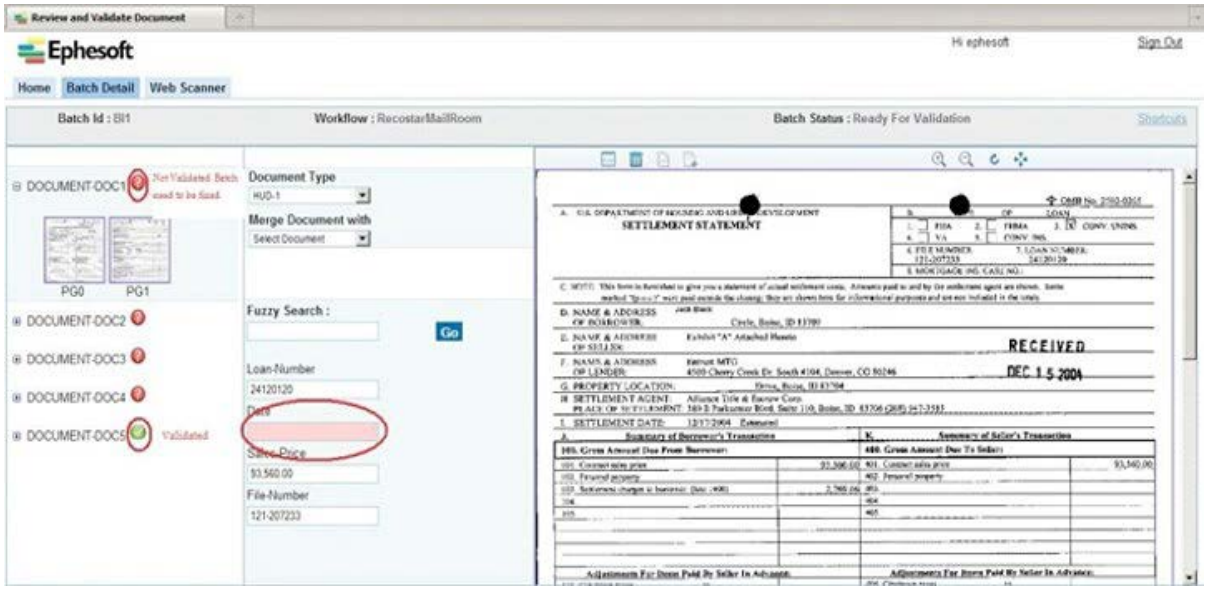
Optical Mark Recognition, or OMR, is a technique for high volume data capture typically from a marked grid on an accurately printed OMR sheet of paper. The OMR input generally translates into "yes or no" or "true/false" responses. OMR is often used on forms; another common application of OMR is in the high volume marking of school examination papers.

Handwritten Character Recognition, or HCR, is similar to Intelligent Character Recognition. HCR scanning is used to interpret poorly defined handwriting as strings of text.

Another use of scanning is the detection and interpretation of **barcodes**, to speed up data input. A form may have a barcode, or the barcode may be on an archive folder, CD case or drawing. Use of barcodes with physical assets can be very effective in ensuring that objects and metadata are matched up for search and retrieval later. Barcodes are available in a variety of formats; some 2D barcodes can store hundreds to thousands of characters of data.

Forms Recognition

Advanced OCR tools can be set up to recognize zones in order to automate forms processing activities. For example, the top right corner of a form may be a consistently structured customer address block. The zone OCR settings can zoom in on those predictable sections, identifying and capturing the information and automating the contract or form handling. The text recognition can be used not only to later full text search on the imaged documents, but text can also be extracted to pre-populate database, workflow or other structured fields.



Intelligent Character Recognition is designed to read and extract text from handwritten documents. This is best suited for constrained handprint – that is, where a field is structured for the user to enter a single letter in each box.

In short, forms processing combines many of the different recognition technologies to interpret and extract the information from the scanned form.

Today many scanning software solutions are able to recognize different forms according to their unique characteristics and apply recognition technologies in a more flexible way. In some cases, the solution may even be able to differentiate between different classes of documents and treat them appropriately, reducing or eliminating the need for batch separation.

Quality Control

It's important to note that while recognition technologies are robust and mature, they are not foolproof. The accuracy of the recognition will depend heavily on quality of the scanned image, which in turn depends on the quality of the original document. Many capture applications will include some capabilities for "cleaning up" a scanned image – deskewing, despeckling, removing lines or holes, etc. to make the image as good as possible prior to performing the recognition process.

Because of these potential issues, it's important to perform quality control on the extracted data. When first setting up a capture process, quality control should be performed against 100% of the scanned images and their relevant extracted data. As confidence grows in the process, this could be scaled down somewhat.

One of the other ways to improve quality control is to apply data masking and data validation. For example, if the recognition process is being used to recognize a currency amount, the field could be set up to require only numeric values and only in a pattern that matches the currency and numbering pattern used to express currency in that environment.

Automating Information Extraction

Throughout this section we've talked about different approaches for extracting intelligence from information. But it's important to understand the *why* as well as the what and how. There are several key benefits from automating information extraction:

- **Accuracy.** Humans make errors, and humans are very inconsistent in how they make them. Sometimes it's not paying attention and entering the wrong amount or transposing digits. Sometimes it's a typographical error. Sometimes it's selecting the wrong value from a dropdown list or the wrong date from a calendar picker. Regardless, automation can significantly reduce the number of errors.
- **Completeness.** Humans also aren't always the best and being complete – that is, ensuring that all fields are filled in (correctly). With mandatory metadata fields we can make staff enter data, but it's often not good data. Meanwhile, optional fields are not filled in at all, which drives towards making more and more fields mandatory. Automation helps to ensure that all the information is captured, extracted, completed, and managed.
- **Consistency.** As noted above, humans make errors even when their choices are limited. Without those limits, it's difficult to know how something will be captured, or filed, or what data will be entered into a particular field. Is it San Francisco, or San Fran, or SF, or some kind of location indicator code like CA-228a? Automation can significantly improve the consistency of these tasks because they occur precisely according to their rules, workflows, values, etc.

Here are some more technical benefits from automating information extraction:

- **Speed.** Automation tools work at the speed of software and can work 24/7 without a break. This generally means faster processing in addition to the other benefits.
- **Scalability.** Similarly, working at the speed of software helps the solution to scale up. Processing hundreds of documents an hour is certainly possible for software; how many staff members would be required to complete the same review?
- **Automation.** All of these benefits fall under the umbrella of automation. Automation ensures that things get done, and that they get done correctly, completely, and consistently. This does mean that the entire process needs to be thought through carefully to ensure what happens at each particular stage in the process. It does require up-front work to develop the business rules and business logic required for effective automation. And it often requires, or at least benefits from, technology solutions – but the planning and design work needs to happen before effective automation is possible. As we've explained elsewhere in this course, automating a bad process – in this case, extracting intelligence from information – simply means executing a bad process faster.

© AIIM – 2019



Test Your Knowledge

Question 1:

What are the issues associated with extracting capturing structured data? (select 2)

a. The costs associated with storing the extracted structured data
b. How to capture the data as presented to the customer
c. How to ensure data isn't lost during the capture process
d. The difficulty of finding a tool that can extract structured data

Question 2:

Which capability would be most effective for extracting hand-printed information from scanned forms?

a. Barcode recognition
b. Intelligent character recognition
c. Optical character recognition
d. Optical mark recognition

Question 3:

What are the main strategic benefits of automating information extraction? (select all that apply)

a. Increased accuracy
b. Better completeness
c. More consistency
d. Reduced headcount



Test Your Knowledge - Answers:

Answer to Question 1:

What are the issues associated with extracting capturing structured data? (select 2)

a. The costs associated with storing the extracted structured data
b. How to capture the data as presented to the customer
c. How to ensure data isn't lost during the capture process
d. The difficulty of finding a tool that can extract structured data

The correct answers are B and C. For A, while there is a cost associated with storing data, it is not as significant an issue as B and C. For D, there are many tools available to do this.

Answer to Question 2:

Which capability would be most effective for extracting hand-printed information from scanned forms?

a. Barcode recognition
b. Intelligent character recognition
c. Optical character recognition
d. Optical mark recognition

The correct answer is B. For A and D, barcodes and optical marks are not considered hand-printed information. For C, OCR works best on machine-printed information, not hand-print.

Answer to Question 3::

What are the main strategic benefits of automating information extraction? (select all that apply)

a. Increased accuracy
b. Better completeness
c. More consistency
d. Reduced headcount

The correct answers are A, B, and C. For D, while the business case for automation is sometimes made on the expectation of reducing headcount, in practice users are generally repurposed to more value-added activities rather than actually being let go.

© AIIM – 2019

Analytics and Artificial Intelligence

Use Cases for Analytics and Artificial Intelligence

Structuring the Unstructured

Let's start by defining what we mean by analytics and artificial intelligence. Making content searchable and meaningful happens when free-form text data is structured – and this is a process.

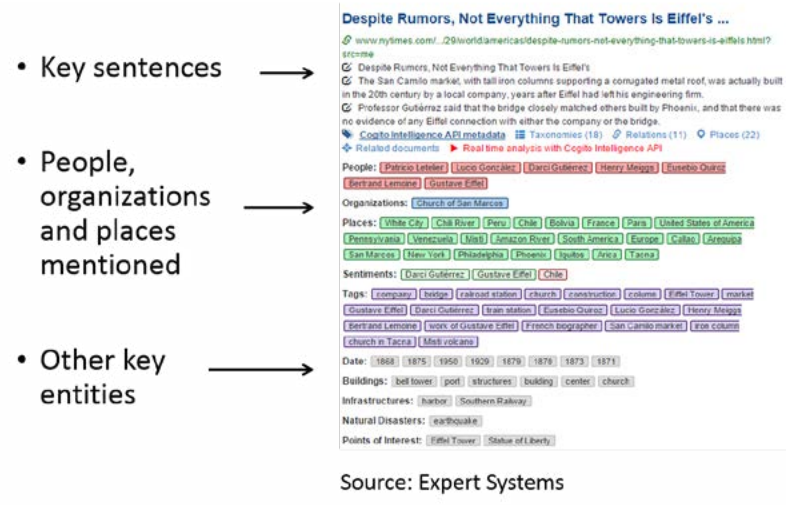
Documents, snippets of text, transcribed conversations, i.e., the text data inputs, are transformed by software – applying natural language processing techniques to transform free-format text within documents into core elements, terms and characteristics. We can then extract the meaning of these entities and use them to do things like automatically classify and route information.

These structured outputs are used to derive new insights – to include in search indexes to make document retrieval more relevant, or to create new variables that can be used in predictive analysis. The structured results can be used to score social media postings with sentiment polarity - positive, negative, neutral, unclassified, or classify content that may already exist, or to assign metadata values that describe the material without anyone actually reading it.

So it's often said that the point of text analytics is to structured the unstructured data. Of course, to do this, you need born-digital information with a text layer that can be transformed and analyzed. If there are any physical documents to be included, they need to be digitized and have recognition technologies applied to extract the document content. And it helps to have good quality, consistent metadata fields and values, meaning you may need to map alternate or synonymous terms to the preferred term for a particular concept.

Text Mining and Analytics

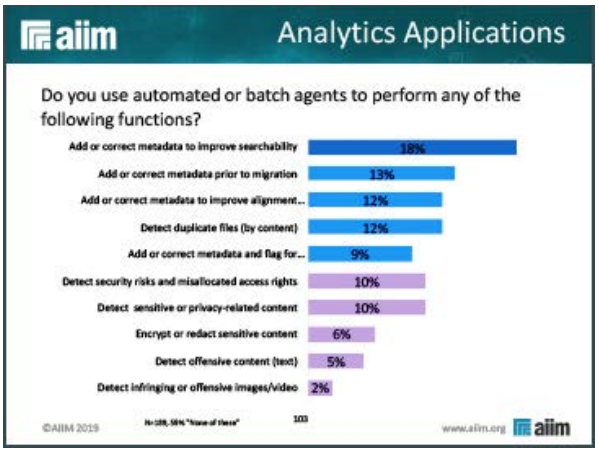
Here's an example you've seen already, in the context of metadata extraction. But you can see other things here including named entity extraction (the people, organizations, places, and other entities); topic recognition; and recognition of key sentences that could be used individually or together to summarize even a long, complex document. Simply finding these terms and topics in free text is referred to as text mining; once we have them, we can do other things like sentiment analysis and other more complex sorts of analysis.



Analytics Applications

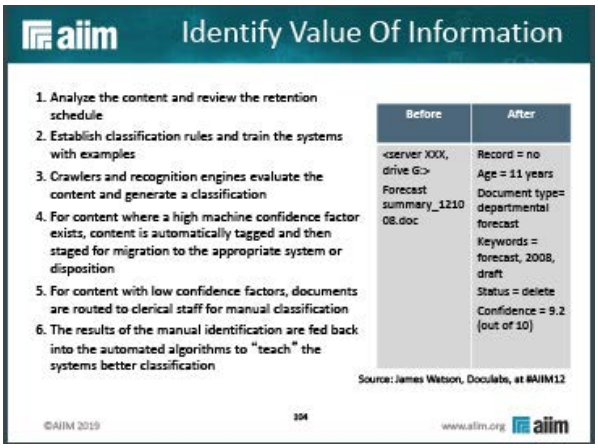
Now let's look at some more specific use cases. The first one we'll look at is governance broadly. This is still an emerging area, but already we're seeing organizations leverage analytics to add or correct metadata, detect exact or near-duplicate files, and detect sensitive types of content such as the credit card information we discussed earlier. This last one can significantly assist in safeguarding information from being breached or removed from the organization, inadvertently or intentionally.

You can see that the top three applications refer to adding or correcting metadata. The top one refers to improving searchability within the current system, and the second adds metadata prior to a migration to improve searchability within the new system being migrated to. In other words, we can use analytics to directly improve the findability of enterprise content.



Identify Value of Information

Here's an example of how we might use content analytics to determine the value of a particular piece of information. We have a document called "Forecast summary_121008.doc" located on a mapped network shared drive at G:\Sales. The traditional approach to reviewing this document is to have a human open the document, review its contents, and determine its business value moving forward.



This is OK – as long as someone actually does it. But you can't have humans manually review thousands or millions of documents on your shared drives or SharePoint or wherever your digital landfills are. In addition, different people may classify the same or similar documents differently. And the same person may classify the same document differently over time; to see this in practice, check the email categorization structure you have set up in your own email inbox.

Enter content analytics. In this approach, we would "train" the technology solution to make a better, more consistent distinction between information that should be saved and managed, and information of limited or no value. We start with these steps:

- Start by analyzing the content to determine what it is. At the same time, review the retention schedule to get a sense for what needs to be retained and what doesn't.
- Establish classification rules and train the systems with examples.
- Use crawlers and recognition engines to evaluate the content and generate a classification.

© AIIM – 2019

- For content where a high machine confidence factor exists, content is automatically tagged and then staged for migration, either to the appropriate system, or disposition.
- For content with low confidence factors, documents are routed to subject matter experts for manual classification.
- The results of the manual identification are fed back into the automated algorithms to "teach" the systems better classification.

Back to our example, this is how it would work in practice. The analytics engine finds that same document in G:\Sales. It determines the following:

- The document does not appear to be a record because it's not stored in an approved storage location for records.
- The age of the document is 11 years, which seems to be old for a forecast. This is extracted from the date the file was last updated.
- The document type is a departmental forecast, based on the naming conventions that were fed into the system.
- The system identifies three significant keywords: forecast (from the file name), 2008 (from the file name and date last updated), and draft. This last comes from the DRAFT watermark that appears on every page and which is just metadata to the system.

All of these are pulled together and, based on the rules established by humans, the document is given a status of Delete with a 92% confidence rating. Depending on how comfortable the organization is from a risk perspective, the document could be deleted automatically, left alone, or sent to a human for further review. At the very least, though, this approach would identify the most obvious cases for deletion, in bulk, and automatically.

Auto-categorization

Auto-categorization tools are often thought of as a technology to assign a lot of metadata to documents without the need for large scale human intervention, saving cost and effort. However, no tools can operate without any human intervention, and quality of results depends on the skill with which the tool is configured and monitored. In migrating very large collections of content into a metadata-and-taxonomy controlled environment these tools can be the only viable strategy for getting a lot of metadata assigned, relatively consistently to content, very quickly. Human beings have neither the patience nor the consistency to do this kind of large scale effort in a short space of time.

Auto-categorization tools can also be useful at the search end of the activity chain, especially if common search terms are factored into the auto-categorization activity.

Auto-categorization tools use three broad approaches:

Business rules. Here, a series of rules are defined by a specialist to say "given the following criteria, associate the following subject tags to the document." This is a labor intensive task and requires a deep knowledge of the content for universal use, but in context-specific cases this can be a very useful strategy.

Teaching sets. Here, the administrator compiles a sample set (a "teaching set") of documents for each concept in the taxonomy. The auto-categorization engine analyzes the content for consistent and characteristic features common to the set using a range of techniques (e.g., word frequency, terms used in combination or close to each other, structural elements), and then goes looking for other documents that match the patterns they have identified. The teaching set approach depend on having sufficient consistent document examples associated with each of your taxonomy concepts, and the auto-categorization engine can lose accuracy if the language and structure of documents relevant to specific concepts "drifts" from the initial training set over time.

Semantic analysis. This is a complex pattern sensing technique based on analyzing the contents of documents and the language patterns within them, inferring similarity relationships between documents based on that analysis, and then giving cluster labels to the documents based on rules about where meaningful labels can be found to characterize documents (e.g., in the document titles). In practice, semantic analysis works best as a tool for suggesting categories, which a taxonomy manager can then edit or modify, to connect the content clusters that have been identified with formally researched and tested taxonomy terms.

Analytics and the Knowledge Worker

Here's another look at some of the use cases that are specific to knowledge workers, but broadly applicable across industry sectors and locations.

TECHNOLOGY	DESCRIPTION
Dynamic Clustering	Groups documents by topics or concepts – easier to batch for review, faster to review (i.e., all are related)
Concept-based Categorization	Sifts through collections looking for only documents that correspond to sets of examples
Concept Search	Find me everything like this – easy and efficient, finds relevant docs, not just responsive ones
Email Threading	Which custodians matter, which emails are the most inclusive – and what's been changed
Near-Dup Identification	Finds and groups nearly-similar documents – minor edits of a master of "main" document (less to review)

Source: Laura Webster, Director, Partner Solutions at Content Analyst Company

This list comes from Content Analyst Company.

- Content analytics can apply **dynamic clustering** to allow for some interesting approaches to how documents are reviewed in bulk. This approach would allow assignment of documents based on topics to those subject matter experts with expertise in a particular topic.
- **Content-based categorization**. This technology attempts to automatically categorize documents into particular categories based on their content rather than requiring users to manually categorize documents. Again, this goes back to the consistency point made earlier.
- **Concept search**. Traditional search relies on strings of text characters, such as "center" – C-E-N-T-E-R. Traditional search wouldn't recognize similar terms such as depot, headquarters, downtown, or middle, or even different spellings such as C-E-N-T-R-E. Analytics tools can help to identify concepts based on the term as well as how it's used in context so that users searching for a particular concept using a particular term also get results or that concept that use different terms.
- **Email threading**. This was one of the original use cases for content analytics in the context of e-discovery. Who knew what, when did they know it, and who did they tell can all be extracted and understood more readily using these tools.
- **And near-duplicate identification**. It's a simple matter to identify exact duplicates with the same file size, file date/time, and even the identical bitstream. It's much more challenging to identify minor edits, or Word vs. PDF renditions, and so forth. These tools can help streamline this review process by automatically comparing versions and identifying those discrepancies.

© AIIM – 2019

Content Analytics for Your Industry

This visual was provided by IBM. The use of content analytics has proven business value in obvious use cases like voice of the customer and customer insights. Organizations are actively using analytics to review and analyze customer satisfaction across a variety of information sources and platforms ranging from customer help desk and surveys to social media.



Public safety organizations such as police departments rely on analytics to identify and analyze trends in crime rates and response. These are particularly useful for example in community policing and in longer-term initiatives such as anti-gang and anti-terrorism programs.

The promise of analytics in health care is incredible. From infectious disease outbreak and management, to conducting analysis relating to particular diseases or treatments, to improving claims management post-treatment, to even identifying trends in helping to prevent readmissions which tend to be more expensive and more acute, analytics is becoming increasingly important to healthcare organizations.

Insurance companies have used structured analytics for years to assess risk, provide underwriting, and detect fraud. But the use of content analytics promises to increase access to useful information that is not stored in a relational database to improve these processes.

And financial firms can use these analytics processes and tools to identify problematic trends and increase operational efficiency.

Issues Associated with Analytics and Artificial Intelligence

There are several main issues we'll look at regarding the effective use of AI. These are:

- Resource availability
- Training data
- Model management
- Black box AI
- Accuracy

Now let's take a closer look at each of these.

Resources Available

The first key consideration for analytics and AI is that of resource availability. This consists of two areas of concern.

- **Computing resources.** AI is extraordinarily resource-intensive because of the size of the data sets often involved and the computing tasks associated with performing the various steps involved in the process. It's only been in the last few years that the pricing has dropped, for two reasons. First, computing power has continued to increase at a rapid rate, including the ability to dedicate processors to these sorts of tasks. Second, Internet bandwidth and cloud processing architectures have made these capabilities available to a much broader audience.
- **Human resources.** All the technology in the world won't help without the human resources require to:
  - Make the models
  - Build the representative data sets
  - Process and interpret the results
  - Apply those results to the organization's actual business issues and desired outcomes
  - And all the other tasks required to manipulate and manage the data and outputs over time

There are numerous studies available decrying the shortage of data scientists, but these are only one cog in the artificial intelligence machine. Existing resources can be brought to bear, but they need training and guidance to do this sort of work effectively.

Training Data

There was a terrible tragedy a few years ago where a self-driving car (which was using AI to drive, recognize traffic or obstacles, etc.) hit and killed a pedestrian. In this case the onboard logs showed that the car saw the pedestrian, who was walking a bicycle, but did not brake. The theory is that the car hadn't been trained to recognize that particular composite – a human pedestrian and a bicycle. This raises a question regarding self-driving cards – how many different possible composites are there that could theoretically be seen on the roads?

And for AI, it raises another, similar question: how does the training set measure up to reality? The training set needs to include documents that are similar enough to be readily recognizable as invoices, or contracts, or whatever, while being different enough to be representative of the different variations on the concept. High quality data that is representative of what the organization sees on an ongoing basis is critical to the success of an AI initiative.

Model Management

Every AI process and project includes the creation of one – or many models. These models are based on reality, but they are not reality and cannot account for every single possible instance or exception. As the saying goes, "All models are wrong; some models are useful." So the models used for an AI or analytics project might be useful… at the time it's built, because it is modeling business conditions, constraints, etc. at the time it was built.

© AIIM – 2019

But things change over time. How the business operates, the legal or regulatory environment, technologies, etc. all change, and these days they seem to change quite quickly. When conditions change, the model needs to be changed as well – if it can be. This means there needs to be active monitoring of how the model is performing, and when it starts to degrade, it either needs to be updated – rules, training data sets, corrections to outputs, etc. – or it needs to be replaced with a new model, and that model needs to have the same input from the business.

"Black Box" AI

One of the key concerns around AI is that so many of them operate as a sort of "black box" where raw data goes in and intelligence, or a decision, or at least a recommendation comes out the other side – without an understanding of how one led to the other. Was there bias, implicit or explicit? Was the training data sufficiently representative? There are so many different flavors of AI that could be contributing that it becomes difficult to trust the decision. On the other hand, if the decision is explainable and repeatable, this goes a long way towards underscoring the value of the resulting information. This also means that there needs to be effective security to ensure that the data cannot be tampered with before, during, or after the analytical process.

This transparency is important for a second reason: it's required under the law, or at least under some laws and regulations. The Equal Opportunity Credit Act in the U.S. and the General Data Protection Regulation in the European Union both require that transparency and an explanation when a negative decision is made on the basis of automated analysis. And they are certainly not alone.

Finally, if you don't know what's going on with the model, it's very difficult to improve how it works.

Accuracy

The last issue is that of accuracy. There is often a misperception that analytics are, or should be, 100% accurate out of the box. This is not true and likely not possible for a variety of reasons including several we've discussed already. For some AI-based processes it may be impossible for the results to ever be 100% accurate. So dispelling this misconception is one of the first things that needs to happen as part of an AI initiative.

Now, it is true that analytics can be as accurate as humans, if not more, in some respects (and substantially less in others). Computers are uniquely positioned to make sense out of very subtle patterns, and not very good at image recognition – yet. Perhaps more importantly, even where AI gets something wrong, it will generally get it wrong consistently and repeatably. Compare this with the variety of ways in which humans can make mistakes!

But ultimately organizations need to recognize that analytics and AI are just tools, and their value to the organization will vary significantly depending on the issues we've identified here and how they are addressed. This is often one of the most significant issues to overcome in getting buy-in and continuing support for an AI initiative.



Test Your Knowledge

Question 1:

What is the purpose of analytics when applied to business information? (select 2)

a. To apply structure to unstructured information
b. To extract intelligence from unstructured documents
c. To apply formatting to structured information
d. To analyze structured information to identify potential information security weaknesses

Question 2:

What are the key resource constraints associated with an AI initiative? (select 2)

a. Invalid models
b. Insufficient computing capabilities
c. Shortage of trained staff
d. Lack of transparency

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the purpose of analytics when applied to business information? (select 2)

a. To apply structure to unstructured information
b. To extract intelligence from unstructured documents
c. To apply formatting to structured information
d. To analyze structured information to identify potential information security weaknesses

The correct answers are A and B, and in that order. For C and D, analytics has very little to do with formatting or security of structured information.

Answer to Question 2:

What are the key resource constraints associated with an AI initiative? (select 2)

a. Invalid models
b. Insufficient computing capabilities
c. Shortage of trained staff
d. Lack of transparency

The correct answers are B, computing capabilities, and C, shortage of staff. A and D are both issues, but they are not resource-related.



Enterprise Search

Search within Applications

There are as many different ways to search as there are applications and systems in the organization. But we can group them into two main categories: application search and enterprise search.



Application search is used to find information within a specific application. Almost all applications have some sort of Find or Search button that users can use to find the information located within that application – emails within Outlook, Word or Excel documents in a repository, even applications or data on your cell phone.

When you search within an application, you're only searching the information that's managed by that particular tool.

The search engine also typically understands the application security and access model. So if you were searching within a financial application, as an example, based on your login or your authentication into the system, you wouldn't be able to access any documents that you did not have the authority to view.

The Value of Application Search

There are a lot of advantages to having search within a particular application. Frequently the search is tuned for that particular content and data structures, meaning that you can get really good high recall and very good advanced search capabilities.

Search is available where the employee is actually working within a particular business application. You don't have to exit that application to pull up your search engine, and in turn search where you were working in the first place.

The results of the search might even be actionable. For example, if you search for something within your email application, you can immediately respond to that email that came up in the result without having to leave your environment.

Drawbacks to Application Search

Of course, as with every technology, there are drawbacks to application search. A search subsystem within an application is usually not as feature rich as some of the best of breed alternatives that are out there.

Search within applications can be difficult to tune beyond the default settings. If you're dealing with a search that's part of a very large or complex content management system, you might have some challenges in tweaking that or scaling that beyond its default settings.

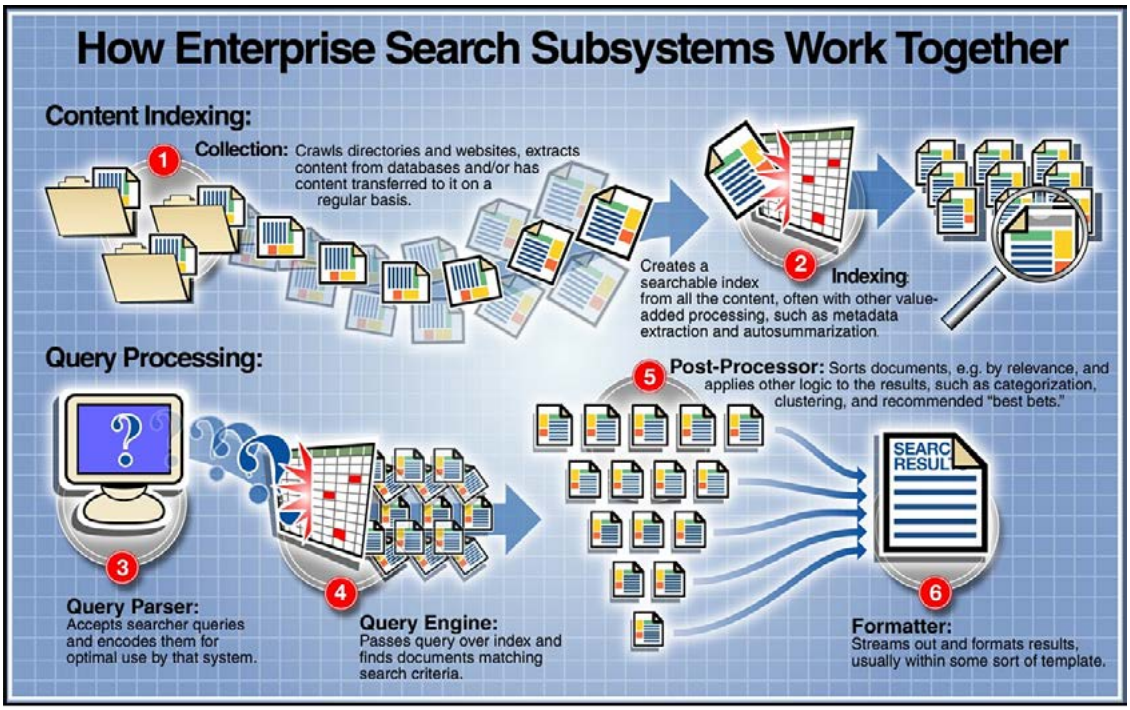
Different applications may have different search engines available. For example, in an ECM or recordkeeping system, there may be a basic search system that comes with the application, and options to upgrade to more powerful search capabilities for a fee. It's also common for the built-in search to have some limitations in terms of the amount of information that can be stored or searched, so you may need to upgrade as the amount of your information increases.

Also, by definition, application search results are limited to a single application or repository. So if you are searching for content around a particular topic, you will generally need to search each individual repository, application, or system separately in order to find all of the information relevant to that topic.

© AIIM – 2019

Enterprise Search

In contrast, enterprise search systems are designed to provide users access to information across a variety of repositories and systems. Enterprise search consists of several key steps and processes.



Source: CMS Watch

The first step is content indexing. What happens first is that a collection is created. The technology crawls directories and websites and it extracts content from databases and other repositories and has content transferred to it on a regular basis. So if one of those repositories is updated the search engine will have some sort of procedure that enables it to go in and source and index that updated content.

Once it gathers all that content, it creates an index. That is a searchable index of all the content. And oftentimes, there's other value-added processing, such as metadata extraction, and also auto-summarization. All of this happens in the background on an ongoing basis depending on the system.

When a user makes a query, it is processed by a query engine of some sort that evaluates the search criteria entered and finds any documents that match those criteria.

Finally, a formatting page is created that presents the results according to a template and established criteria – newest, most relevant, etc.

Enterprise search is sometimes also referred to as federated search. While there are differences, they are beyond the scope of this course.

Enterprise Search – Drawbacks

Enterprise search systems can make immediate and direct contributions to the financial health of an organization. As soon as people can find content more easily, you're likely to improve productivity and also just make employees generally happier. But note that the technology is deceptively complex. You need careful planning, project management, and budget controls in order to make an enterprise search system implementation successful.

If you're searching a single repository or system, an enterprise search tool is likely overkill. Make sure you don't buy more technology than you need. However, if you're in a situation where you need to extract and merge content from multiple file systems, multiple databases, and lots of application repositories, application-level search tools will not be sufficient. In that sort of scenario, you need an enterprise search tool. And you'll also likely need to do a lot of custom work on that enterprise search tool to tweak it to specifically be able to browse through and create indexes from your specific repositories.



Test Your Knowledge

Question 1:

Which solution will be most effective for searching within a single system or repository?

a. Enterprise search
b. Federated search
c. Application search
d. Centralized search

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Which solution will be most effective for searching within a single system or repository?

a. Enterprise search
b. Federated search
c. Application search
d. Centralized search

The correct answer is C, application search. For A and B, enterprise and federated search are designed to search across multiple repositories. For D, centralized search is not a technology solution.



Digitalizing Information-intensive Processes

Introduction

In this domain we focus on how to streamline and automate information-intensive business processes.

The domain begins with a review of business analysis: discovering and reviewing existing business processes and the role of the business analyst in a process improvement initiative.

We review how to construct a flowchart using best practices, and identify the limitations associated with flowcharts.

We describe approaches for reviewing, troubleshooting, and improving existing business processes using a variety of techniques such as parallel processing and task queuing.

Finally, we outline key use cases and considerations for leveraging newer process automation technologies including robotic process automation, case management, and decision management.

Introduction to Business Process Management

Digitalizing Business Processes

Organizations that work at "the speed of paper" are increasingly being rendered non-competitive and irrelevant. We work and live today in an age where ubiquitous broadband internet and ever-more-powerful mobile devices combine to let workers work where, when and to some extent how they want to.

But this means more than simply scanning paper and leveraging email in support of digitization initiatives. Rather, we think it calls for a fundamental rethinking of how we approach key business outcomes. It requires that information be digitized and that work processes be developed, or reworked, with a focus on digital-first and automation wherever possible. And it requires that organizations adopt a posture of agility and responsiveness rather than one of passivity and reactivity.

Reasons for Process Change

The Case for Process Change

The reasons for changing business processes and improving them are fairly obvious. These high-level, strategic reasons for wanting to change, and wanting to look at our business processes and improve them can be broken down into the following three categories:

- **The need to survive.** This is often masked from many people in the organization and is really understood only by senior management. It is not always apparent, but this may be the single biggest reason for business process change.
- **The need to grow (or grow further).** Although companies don't always need to grow, growth is what most companies strive for. In order to do that, the organization may need to reduce some inefficiencies, save some money, and improve its processes to better deliver value to its customers.
- **The need to stay relevant.** This is increasingly important for small businesses and for businesses in areas that are ever changing - for example, technology. While what the organization is doing now may be working well today, technology changes quickly and it's important to be prepared for change while it's still a choice. Too many organizations wait until they are irrelevant before making needed changes – just ask Blockbuster. So staying relevant is in a real sense the same as surviving.

You need to be aware of which reason or combination is driving a particular project because it will impact how the project is shaped and driven.

© AIIM – 2019

Here are some more specific reasons for undertaking process change:

- To reduce costs associated with a particular process
- To reduce the time it takes to complete a process
- To improve customer satisfaction and focus on the customer
- To grow revenues
- To explore new markets or simply to look at new opportunities

Times change. And that's one of the biggest reasons there's now a resurgence in business process management. Globally firms are reaching a point where they don't feel they can eke out any more efficiencies from their current processes. At the same time, they understand that new entrants are coming into the marketplace with dramatically different business models.

In the insurance industry, for example, companies such as Progressive have developed an online Internet service that has had an obvious impact on older insurance companies and their business model. Examples like that exist throughout industry and throughout the world. New competitors come online and throw a "spanner in the works" to existing business models. Times change and organizations have to change with them.

Key Drivers for Process Change

There are three main drivers for process change.

- **Changes to the marketplace.** As we noted earlier, these are changes that can render an organization or even an entire industry irrelevant. Consider the Fortune 500 – the list of the largest companies in the U.S. by total revenue. Since the original list was published in 1955, more than 400 of the corporations on that list no longer exist, at least in the same form. Competitive pressures in the marketplace don't just come from other competing organizations, but also from changes in how consumers use and perceive the marketplace.
- **Changes to legal or regulatory requirements**. Organizations are required to comply with these or face penalties and other liabilities. One of the more pressing issues today is privacy and protection of personal data – in just the last few years we've seen a number of changes to privacy regulations around the globe, requiring any organization that collects or processes personal data to review and potentially overhaul any process that relies on it.
- **Finally, changes to technology can displace previous technologies very quickly.** Netflix and others won over Blockbuster for many reasons, but a large part had to do with the ability to stream media and movies directly to consumers courtesy of ubiquitous broadband internet, ever-increasing data speeds, and mobile device capabilities. But something as relatively simple as moving from "wet-ink" to digital signatures can significantly streamline business processes.

Strategic Considerations

We end this section with a discussion of strategic considerations. As you consider whether to change a particular process, it's important to keep these considerations in mind.

First, every process touches another process. Very few processes are completely self-contained. This means that it's important to understand that changes to a process will likely cause issues to related processes, and to understand the impact of those changes. This is one of the most overlooked aspects of the planning phase, but it can cause significant unforeseen complications.

Next, technology and automation approaches cannot fix bad processes by themselves. A bad process that is fully automated is simply a more efficient bad process. It's important to look at the process itself to see whether it has issues, can be streamlined, etc. before starting the technology discussion.

Finally, changes to processes will require changes to behavior. Management consistently underestimates the amount of change management that will be required in order to effect change. Sometimes this manifests as the idea that if staff don't simply get on board with the changes, they will be fired. But you can't fire everyone.

Selecting a Process to Change

This module will help you to distinguish among different business process scenarios and determine which are most suited for change.

Away From vs. Go To

There are two broad categories of reasons for change. There are “away-from” and there are “go-to” reasons for change that are somewhat self-explanatory.

“Away-from” reasons for change:

- People don’t want to be fined or go to prison.
- They don’t want the auditor to shout at them and tell them that if they don’t sort things out, he’s going to close their workshop down.
- They don’t want to be on the front page of the newspaper or on the news for all the wrong reasons.

“Away-from” reasons are only drivers when the heat is intense and when there’s a real problem. Although “away-from” reasons are extremely important reasons for business change, we have to be realistic about them. They will not gain the support in the long-term that “go-to” reasons for change will.

“Go-to” reasons for change are those that excite people:

- Making more money
- Making customers happy
- Doing things more quickly

Those kinds of things can engender support over a longer period of time. And those kinds of reasons will always have more staying power.

Identifying Processes to Change

In order to determine whether a process is suitable for change, we need to start asking questions in an analytical manner.

“Why does this happen?” This is the most important question in any kind of business process improvement activity and one that people don’t ask enough. So when somebody tells you, “This is how something works. And this is the way it works,” you need to say, “Why?”

What will happen however is you will find that sometimes we do things because something else doesn’t work or because there’s some other trigger point. We need to:

- Be clear when the start and end points of our process are
- Be clear why things are actually happening within the process
- Spot exceptions.

There are no examples anywhere in the world of 100 % perfect processes. There will always be some exceptions.

Our focus initially should be on the normal, standard process – what’s done 70% or 80% or 90% of the time – and make sure we get that right. Exceptions will require a lot more analysis, but remember it’s easy to get sidelined by exceptions in the early stages of trying to identify elements of a process. We need to capture the normal process.

Is this actually a process I’m looking at here or is this really a task in its own right, and actually part of a larger process? A lot of people will not be able to answer that question for you. You need to make that decision when you understand more about the context of the activity that you’re looking into.

Are there duplications of data or effort? The stereotypical example is of a name and address being placed into a database by multiple departments in different ways and in different data structures.

© AIIM – 2019

Before Making ANY Changes

It's important to note that we don't want to change processes just to change them. Rather, we change them because we need to – as we mentioned earlier, in order to stay relevant, or grow, or even just to survive. That means we need to identify the specific reasons why we need to change – and what the status quo is that we need to change from.

Similarly, it's important to identify the value of making the change – and of not making it. For example, if it costs an organization $20 per invoice to process them manually, and $10 to process them in a more automated fashion, then every invoice costs the organization $10 more than it should. If your organization is processing 1,000 invoices per month, that's an extra $10,000 in costs every single month.

These are both "Why?" questions that need to be addressed before getting into the "What" and "How" and "When" questions – those will be addressed once the process to be changed is determined.

Finally, it's important to validate the scope of the potential change. That could impact everything from technology selection, to stakeholders, to the amount of change management required. Moving from ink signatures to digital signatures is a change; completely automating the accounts payable process using digital capture, robotic process automation, and digital signatures is a much bigger, more impactful, and certainly more expensive change.

It's also important to do some preparatory work to ensure that the process is changed appropriately. First, the organization needs to identify any specific legal or regulatory requirements or constraints that may be why the process is how it is. A process that seems very complex, or that has redundant steps, may be so because a law or regulation requires it to be.

Next, we need to look at the process holistically and how it impacts other processes. Changing your invoice processing process from manual to a more automated, digitally-based approach could also impact the audit, how those invoices are filed, etc. It's also important to understand who the stakeholders are for those processes and what their goals and needs are, so that you can determine what the impact of changing your process could be for them.

And you have to understand the current process before you make changes to it. This means that when you select a process to change, one of the first steps should be to document the existing process. This likely means talking to staff, reviewing existing documentation such as standard operating procedures, and creating a flowchart of the as-is process. This will make it easier to review the process with stakeholders and identify the parts of the process that need to be updated.

Process Scenarios

This module will help you to compare and contrast different business scenarios, such as routing, workflow, and business process management.

Business processes and activities can be grouped into different scenarios in order to make more sense of them.

This section looks at scenarios for scoping ad hoc processes and production as well as basic routing and workflow to fully integrated business processes, and the differences between them.

Ad Hoc vs. Production

One of the most important distinctions within process management is to understand the difference between ad hoc and production type processes. Ad hoc processes are those that are typically the norm in a collaborative environment. For example, a staff member occasionally scans a document, files it, and possibly distributes it to a wider audience.

There's a high degree of individual judgment involved in ad hoc process. Typically, these processes are either somewhat unique or seldom used. The emphasis here is on the constraints, rather than rules or directions.

Production processes are repetitive. These processes are run again and again using a standard operating procedure. A team of specialists scanning, indexing, and processing documents (for example, checks or paper forms) are driven by very closely mapped business processes. They do the same thing and obey the same rules every time. There's little or no individual judgment involved in a production process.

Functional Process Scenarios

Process scenarios can be broken down into three divisions. These divisions map very closely to the kinds of technologies you'll use to automate your processes.

At the simplest level is routing. One step up is workflow. It's more complex and includes parallel activities. The highest level and most complex and sophisticated of all business process management is integrated BPM.

Simple Routing

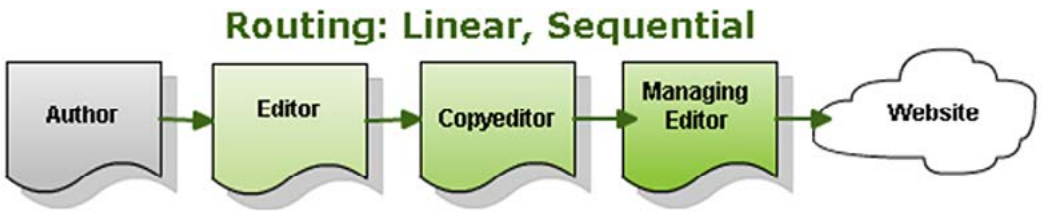
Routing is the most basic type of process. It's linear. It's a simple matter of moving something from one place or person to another. In routing, the tasks are carried out in a sequential order – one, two, three, four. For many people, this is as sophisticated as they want to get. For example:

- An author drafts an article.
- It goes to somebody for review.
- Then it goes to a supervisor for approval that it's ready for publishing.
- And once the supervisor approves it, it's published.

It's the same linear route every time. This works if you have a very simple, publishing type process for content, for example. But the reality is that in more complex business processes you're going to overtax such a system very quickly. Consequently, basic routing is fine for simple approval flows, but not for much else.

Routing Example

This image illustrates the routing example and the same sequential process every time – author, edit, copy edit, manage, publish. This is not full process management or something that you can use to manage or to plot a business process.



Workflows

Most process improvement work involves true workflows - processes with multiple dependencies, sub-processes, parallel processes, decision points, etc.

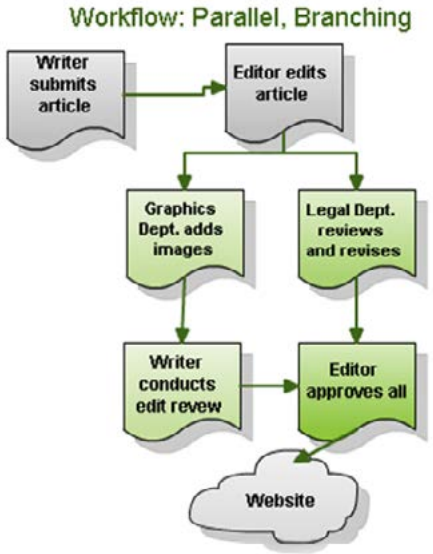
Because this level enables the capture and design of more complex and sophisticated business process activities, we can apply technology to make more subtle process improvement changes.

© AIIM – 2019

Workflow Example

This example shows the same publishing activity as more of a true workflow.

The editor can send the document off in parallel to graphics and to legal. When they're finished, the editor can review it and ultimately pull it back together, ready for approval and publishing. This layer of sophistication is not available with simple routing.



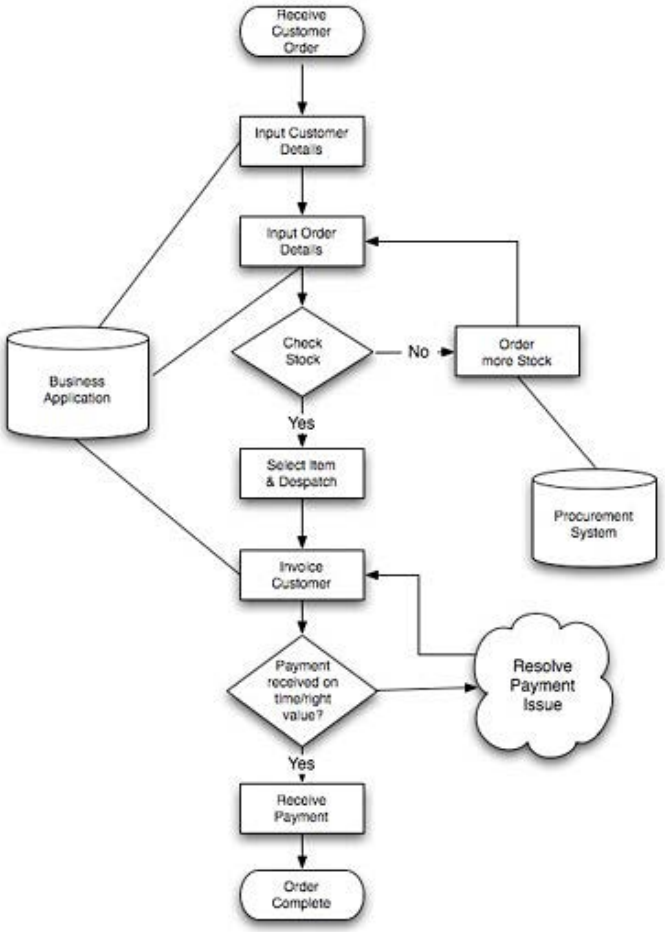
Business Process Management

Full BPM business process management encompasses an even greater level of sophistication and allows for data-to-data, or system-to-system integrations alongside the human activities within our process.

Business process management is also used when coordinating across multiple processes, systems, or organizations.

Business Process Example

As you can see in this example, people do the work, and sometimes provide input to systems (for example, procurement or a business application) for additional processing.





Test Your Knowledge

Question 1:

What are the primary reasons organizations need to consider process change? (select 2)

a. To stay relevant
b. To ensure customer satisfaction
c. To provide goods more cheaply
d. To grow

Question 2:

Which reason for change has the most staying power over time?

a. “Away from”
b. “Stay relevant”
c. “Go to”
d. “Status quo”

Question 3:

The sales department needs to create a contract for a large project. The project will require legal approval and project management planning. Which process scenario would be most effective for this purpose?

a. Automated
b. Business process management
c. Workflow
d. Routing

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What are the primary reasons organizations need to consider process change? (select 2)

a. To stay relevant
b. To ensure customer satisfaction
c. To provide goods more cheaply
d. To grow

The best answers are A, to stay relevant, and D, to grow. For B, process change might be needed to support customer satisfaction, but it won't ensure it. For C, not all organizations produce goods or have their price as a primary objective.

Answer to Question 2:

Which reason for change has the most staying power over time?

a. "Away from"
b. "Stay relevant"
c. "Go to"
d. "Status quo"

The correct answer here is C, "go to." For A, we noted at the start that these reasons only last while the pain lasts and then they are forgotten. B, stay relevant, is more of an "away from" reason. D, "status quo," involves no change at all.

Answer to Question 3:

The sales department needs to create a contract for a large project. The project will require legal approval and project management planning. Which process scenario would be most effective for this purpose?

a. Automated
b. Business process management
c. Workflow
d. Routing

The correct answer is C, workflow, because of the need to have multiple departments and approvals involved. For A, automated is not really a process scenario, and you can't really automate legal review yet. For B, business process management is probably too complex for this scenario, and for D, routing is not flexible enough.

Introduction to Business Analysis

Here is the definition of business analysis from the International Institute of Business Analysts.

> "…the set of tasks and techniques used to work as a liaison among stakeholders to understand the structure, policies, and operations of an organization, and to recommend solutions that enable the organization to achieve its strategic goals."

So clearly business analysis is the process of identifying business needs and finding solutions to business problems.

Business analysis is a catalyst or key enabler for identifying business opportunities, building a business architecture framework, and determining the best solutions for the enterprise. Analysis allows a practitioner to understand how an organization ticks and facilitates improvement by identifying business needs and finding solutions to business problems.

Understanding the business context is important before making changes to processes.

Why Business Analysis?

So why talk about business analysis? Because you can't plan the destination without knowing where you start. Here this means understanding the need for change and the impact of that change, not just in the particular process itself but in the other processes it touches.

In fact, you should never attempt to change a business process, the way people work, or the way things work without first analyzing what the impact of that change will be. Many people think they understand the techniques of analysis, but few actually do. There are many people who are called business analysts in the workplace who don't actually have a solid grasp of these techniques or understand the context they should be used in. And most projects fail, not because of technology, but because of a lack of insight, a lack of stakeholder support or a lack of planning. These are considerations a business analyst must address to do his or her job properly.

The Benefits of Business Analysis

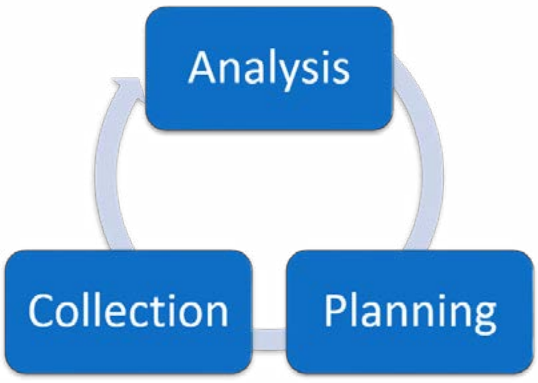
There are a couple of specific benefits that can be gained from a formal business analysis. First, it helps to ensure that processes align to the organization's overall goals and objectives, and where they don't, to determine how to move them in that direction.

Similarly, it can help to ensure alignment of technology solutions to those overall goals and objectives – strategically and operationally. This includes the identification and prioritization of business requirements for new or upgraded technology solutions.

And it can be very helpful in identifying inconsistencies, bottlenecks, and exceptions. These can be within a particular process; across multiple processes, particularly where one hands off to another; or even where similar processes are performed differently by different functions within the organization.

The Process of Analysis

This image sums up the process of analysis. You plan, you collect, you analyze, and based on that analysis, you repeat the process. You try and do this in a consistent manner, following standards where appropriate, ensuring that you do things in a repeatable manner.



Your analysis process should be consistent, structured, and repeatable. These three things together equal a methodology. When you hear the word "methodology," it simply a consistent, structured, and repeatable way of doing things.

We don't spend a ton of time on specific methodologies in this course because different organizations will use different methodologies according to their own business needs. The more important consideration is to pick a methodology, or synthesize one from the many available approaches, and then stick with it consistently.

© AIIM – 2019

What is a Business Analyst?

The role of the business analyst is multifaceted. In order to be effective, business analysts:

- Must become familiar with the enterprise or the organization they are working within
- Gather requirements by collecting information from people and from systems
- Analyze those requirements
- Make recommendations to move the organization forward

A business analyst has another role to play, and that is to facilitate communication between the business, the solution team, and the IT department often, or simply facilitating communication between various departments or individuals. The goal is to ensure that everybody understands from their perspective what kind of changes and recommendations are being made.

A business analyst analyzes business activities. They provide *objective* analysis based on fact gathering.

A business analyst provides a bridging mechanism ensure that the current business process activities and the future proposed activities can be achieved. The analyst has to provide documentation, analysis and an understanding of the process that can be understood and used by different stakeholders, be they business or technical or different departments and groups.

To do this, you must apply proven analysis methodologies and techniques, and stay consistent in your activities throughout.

Business Analyst Activities

A business analyst's three main activities focus on: requirements planning, requirements analysis, and recommendations.

Requirements planning assesses what information needs to be gathered and the tools and techniques for gathering it. During this phase the analyst identifies stakeholders and determines the information capture methods that will be used.

Requirements analysis succinctly defines the current process. The current process is sometimes called the as-is situation. It is important for the analyst to communicate findings about the requirements in a way that everyone can understand. Typical tools are flowcharting, modeling, and documentation. Thorough communication at this stage helps create a context or rationale for the recommendations.

Recommendations are built on well-defined requirements. Recommendation must offer a functional design that will focus the organization on systematically moving forward. Recommendations take the not good enough 'as-is' situation to a more ideal 'to be' future that will succeed.

Caveats and Pitfalls

Let's wrap this discussion up with a couple of caveats and pitfalls.

Never underestimate the value and importance of business analysis. So many people think this is a very generic and simple thing to do. But the reality is that many people are not undertaking proper business analysis and they pay the cost for it. Proper business analysis may well be the largest success factor in your project.

Never underestimate the need for trained and competent analysts, and good analysts are difficult to find. It doesn't matter what you call the role – business analyst or another title. It is the skills and the techniques of analysis to bring about process change are important.

And it's important to manage stakeholders' expectations, including management and staff alike. Many projects fail because insufficient attention was paid to management's lack of interest, employees' push-back, and so forth. The overall corporate culture is important; so too are the individuals that work within it.

The Benefits of Process Automation

There are a variety of expected benefits from automating a business process including financial and non-financial benefits and the benefits of consistency and reporting.

Financial Benefits

The entire point of business process management as a discipline is to streamline and automate wherever possible so that business processes are as efficient as they can be. Sometimes this is directly and financially quantifiable; in other cases, it's more of a background benefit: harder to measure but no less real for that.

Here we see some of the more tangible, financially measurable benefits from automating business processes.

- **Time savings.** If a manual, paper-based process takes 2 hours to complete, and automating it reduces that time to 15 minutes, that's an hour and 45 minutes saved per. That means the same number of employees can complete many more transactions in the same time period.
- **Cost savings.** Time is money, and time saved is money saved.
- **Reduced error rate.** Every time a human touches a process, it's an opportunity to make a mistake: forgetting to include something, accidentally entering a wrong value, etc. Fewer errors means time savings means cost savings; it also reduces the likelihood of poor customer satisfaction, increased liability, and so on.
- **Consistency of process execution and process outcomes.** If you automate a process end-to-end, you can ensure that no steps were missed, everything is there, all the metadata has been filled in, etc.

Other Operational Benefits

These are harder to quantify financially but can be measured and are certainly important as well.

- **Increased visibility.** You can't manage what you don't measure, and manual processes are often hard to measure meaningfully. In an automated process there will be metrics and reporting that show what's going on with the process as a whole and with the individual instances of that process.
- **Workload balancing.** Automated processes can be set up to feed into work queues. From here the next task could go to the next available operator, or it could be assigned to the best available resource for processing. If a process has some exceptions that causes it to slow down, the queue can automatically move to the next available resource so there is no bottleneck behind this one.
- **Better allocation of the workforce.** This takes into account that many steps in a given process are mundane ones like looking up data in one system and adding it into another. These don't require creativity, insight, or even though most of the time and certainly aren't particularly fulfilling to the employees performing them. Automate those processes and those tasks and free up staff to do more creative and innovative things.

Strategic Benefits

Finally, we can look at some more high-level, strategic types of benefits of process automation.

- Improved governance. A process untouched by human hands can't be wrecked by them, whether inadvertently or intentionally. Automated reports show what they show and have the data to back them up. And the automation rules and outcomes can serve to document the process and its execution over time to support compliance requirements.
- Increased customer satisfaction. In the digital age, customers don't want responses at the speed of paper, and they don't want manual processes that could provide them incorrect information because something was missed or processed incorrectly.
- Deeper insights based on reporting. Again, it's hard to get good, meaningful data about manual processes. But automated processes can provide a wealth of information about how a process is working, how issues with a specific transaction arose and were dealt with, and so on.
- Increased accountability. Similarly, automation tools can provide insight about how particular staff members or departments or shifts are getting on, where there might be quality or performance issues, etc.

© AIIM – 2019

Information Gathering Approaches

The collection of information involves a series of different activities. The first thing that needs to happen is to identify the key stakeholders for the process. This will certainly include the process owner, who is a business manager, and the process participants.

It should also include other key stakeholders who are impacted by the process: other consumers of the outputs of the process; other processes impacted by that particular process; and specialty information management functions within the organization such as IT, legal, and records management. These latter need to be consulted because there may be aspects to the existing process that are the way they are because of specific records or legal concerns or technology decisions or limitations.

Ways to Gather Information

One of the more important ways to gather information about a process is to review what documentation already exists about that process. A particular activity may seem chaotic, with no process in place. However, that doesn't mean there isn't a documented process somewhere. That doesn't mean that for audit purposes, there is a documented process which does exist, but which nobody follows.

You may interview people one-to-one or on a group basis. You may run workshops, for example, where you can gather information. In some companies, it's better just to talk to one or two people. In others, it's better to talk to lots. It depends on the culture.

Another way is to shadow, or observe, people as they perform their tasks. This lets you see how the work is actually being done, as opposed to what the documentation might say. But there's a risk that people will do their work differently because they are being observed. Moreover, you need to identify the "right" person to shadow, and depending on the initiative, there could be a number of them.

Brainstorming is a technique for gathering information. In the hands of the right facilitator, it can be a useful tool for extracting information and key ideas.

One of the best ways of collecting information is simply to present how you think it works and get feedback. You have found a lot of information. You already have some knowledge. You may have reviewed documentation. You may have made some observations. Now it is time to test and confirm. So you would restate how you understand the process to the people doing it and the other stakeholders we identified, and then ask people how close this is. It is a very good way of getting people to refine and clarify elements of a process.

And finally, you should repeat as necessary. Gathering information often requires iteration as exceptions are remembered, other steps or stakeholders identified, and information is verified or shown to be inaccurate or incomplete.





Test Your Knowledge

Question 1:

What is the benefit of using a formal business analysis methodology?

a. It ensures that the project will be completed successfully.
b. It ensures that the analysis process is consistent, structured, and repeatable.
c. It provides a mechanism for users to be involved in the process improvement project.
d. It provides a list of business requirements for technology solution providers.

Question 2:

How does process automation improve process governance?

a. Processes can be executed consistently and transparently.
b. Staff can be reallocated to more interesting or variable tasks.
c. Process monitoring tools provide increased accountability.
d. Processes can be substantially streamlined to reduce costs.

Question 3:

When evaluating an existing process, who should be targeted to gather information from? (select all that apply)

a. The process owner
b. Process participants
c. External consultants
d. Specialty information management functions within the organization

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the benefit of using a formal business analysis methodology?

a. It ensures that the project will be completed successfully.
b. It ensures that the analysis process is consistent, structured, and repeatable.
c. It provides a mechanism for users to be involved in the process improvement project.
d. It provides a list of business requirements for technology solution providers.

The correct answer is B. For A, a formal methodology can support a project, but it cannot guarantee its success. For C, while this is true, it's not as direct a benefit as B. D could be an outcome from using a methodology but it is not a direct benefit.

Answer to Question 2:

How does process automation improve process governance?

a. Processes can be executed consistently and transparently.
b. Staff can be reallocated to more interesting or variable tasks.
c. Process monitoring tools provide increased accountability.
d. Processes can be substantially streamlined to reduce costs.

The correct answer here is A, processes can be executed consistently and transparently. Process automation can allow for all of B, C, and D, but these aren't process governance issues as much as they are process automation benefits.

Answer to Question 3:

When evaluating an existing process, who should be targeted to gather information from? (select all that apply)

a. The process owner
b. Process participants
c. External consultants
d. Specialty information management functions within the organization

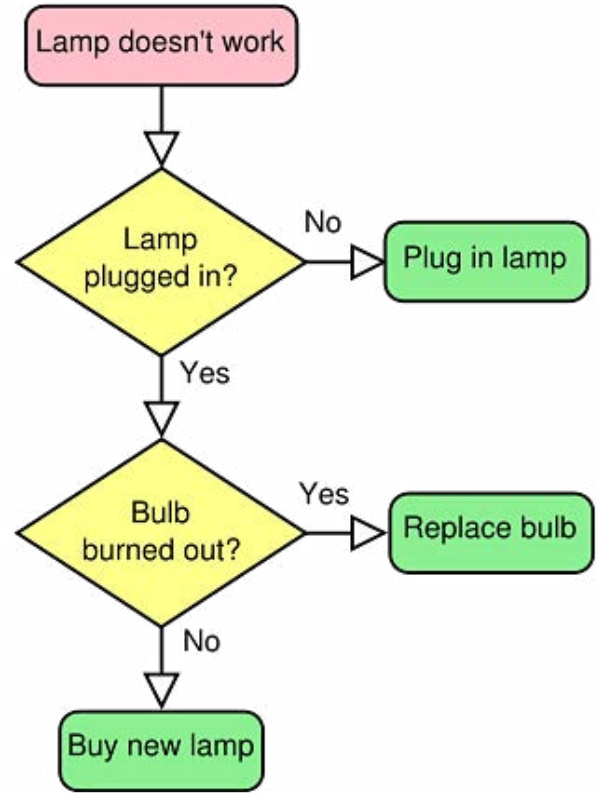
A, B, and D are all correct; A and B should be fairly obvious but it's also important to talk to those specialty functions to ensure their concerns are addressed. For C, external consultants may be helpful, and may even be the ones gathering the knowledge, but they won't know about the existing process until they gather information.

Flowcharts

Introduction to Flowcharts

Basically, you can analyze or you can design a new business process in many different ways, simply by using documents or sketches or whatever. But the normal accepted way is to use a flowchart. And it's not just for designing new processes, but also for helping you to visualize existing processes.

Here's an example of a process to troubleshoot a lap that does not work. So we start at the start: The lamp doesn't work. We can check whether the lamp is plugged in, and if not, plug it in. Next, we can check whether the bulb is burned out, and if so, replace the bulb. If those two steps don't get the lamp to light, we can assume the lamp is broken and we need to go buy a new lamp.



This is a very simple example – you could easily add a number of other steps to check like breakers, power to the house, is the lamp plugged into a switched outlet and is that switch on, etc.

Flowcharts – Definition

Here's a more formal definition of a flowchart.

It's a graphical representation of the sequence of activities, steps, and decision points that occur in a particular discrete process.

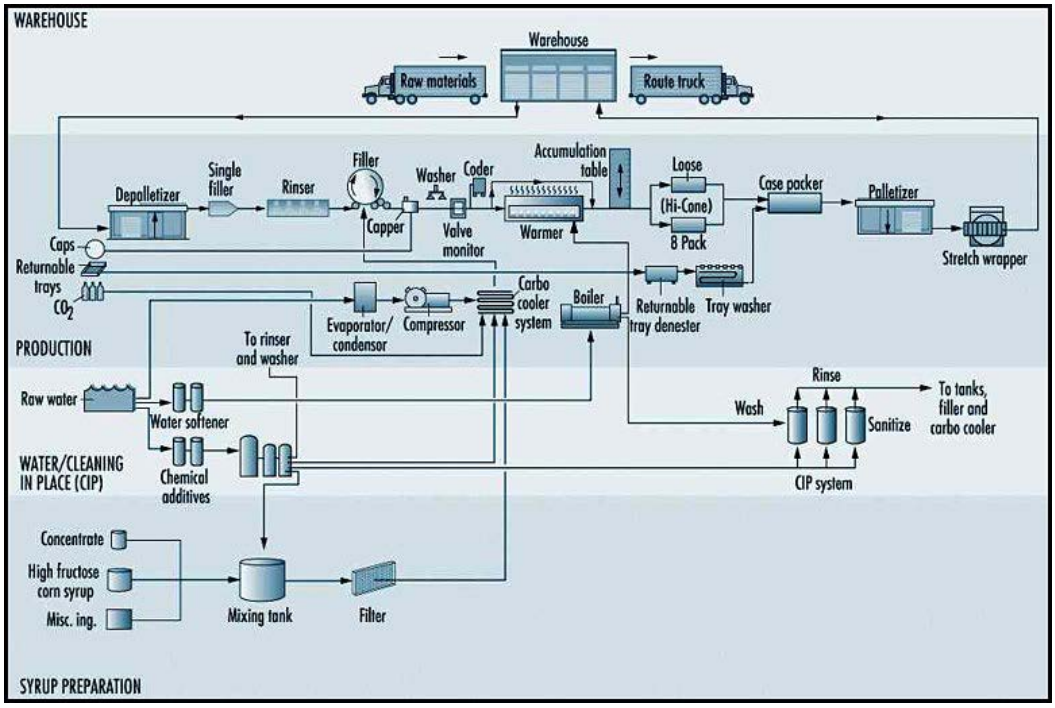
We emphasize the word 'sequence'. And that's really what flowcharts are all about, showing us the *sequence* of events in a process.

Another Flowchart Example

The example we gave you before of the light bulb was about as basic as you could possibly have for a flowchart.

Here's another example of how a flowchart can be used – and this is a very different kind of flowchart. This flowchart actually represents a number of different processes, so we could almost consider this a meta-process of sorts. This flowchart depicts the entire process of production of syrup for soft drinks, from preparation to distribution. It uses a variety of icons to denote different steps in the process.

© AIIM – 2019



Why Flowchart?

Flowcharts are designed to explain the sequence of tasks within a process graphically. The point of a flowchart is to provide a simple means of communication so that technical, business, or people from different departments and different perspectives can actually understand the same process information.

We use flowcharts to identify when we're using analysis techniques, bottlenecks, and loops, so where work is being held up inefficiently, or where work is being undertaken that's completely unnecessary.

In fact, it's often the case that when business analysts look at complex processes, they find that there are tasks which are being duplicated within a process. And a flowchart is a very good way to go through a problem analysis if you know there are inefficiencies, to spot bottlenecks and loops, et cetera, and then to provide a blueprint so that you can actually develop a better process moving forward.

We also want to identify variations in process activities, so we can see where various departments or workers or systems interact, and see how they're processing things in different ways, and possibly share efficiencies across roles and across process steps.

Symbols and Functions

Most flowcharts rely on four standard symbols:

- Terminator
- Process
- Decision
- Connector

In addition, there are standard symbols for depicting the process flow and representing different functions or groups as participants in the process. Let's look at each symbol in a little bit more detail.

Map Elements



The terminal symbol – represented as circles, ovals, or rounded rectangles. They usually contain the word "Start" or "End," or another phrase signalling the start or end of a process, such as "submit inquiry" or "receive product."



The process symbol is represented as a rectangle. It shows that something is performed.



The decision symbol is represented as a diamond or rhombus showing where a decision is necessary, commonly a Yes/No question or True/False test. The decision symbol is peculiar in that it has two arrows coming out of it, usually from the bottom point and right point, one corresponding to Yes or True, and one corresponding to No or False. (The arrows should always be labelled.) More than two arrows can be used, but this is normally a clear indicator that a complex decision is being taken, in which case it may need to be broken-down further or replaced with the "predefined process" symbol. Decisions can also help in the filtering of data.



An arrow coming from one symbol and ending at another symbol represents that control passes to the symbol the arrow points to.



Swim lanes within the map denote the involvement of different people or functions within the process.

Flowchart Best Practices

Here are some basic flowcharting best practices.

Every flowchart should start with a trigger. These are often written as a verb-noun phrase, e.g., “Submit expense report.”

The flowchart should have a defined end, and because of the possibility of multiple decision points, may have multiple ends. In our example earlier of the lamp, there are 3 end points (the green terminators).

All decision point paths should be resolved, by connecting to either another step or an end point.

For right-to-left languages like English, flowcharts should generally be developed from top left to bottom right.

Flowcharts are generally more readable if they use standard symbols and use them consistently. Where different symbols are used, a key or legend should be provided.

Finally, the flowchart should be kept as simple as possible. If it starts getting larger than a single page, it should be broken into multiple flowcharts. One way to do this is to identify the most common sequence within the primary flowchart and then develop additional flowcharts that show major variations. Another way is to break up major sequences into their own flowcharts – for example, an insurance claim process could be broken into separate processes and flowcharts for claim filing, claim investigation, and claim resolution.

© AIIM – 2019

Flowchart Limitations

lowcharts can be very helpful in getting stakeholders to understand what's going on in a process. But they suffer from a number of significant limitations that can limit their usefulness.

Flowcharts only show the graphical sequence of tasks within a process. While this is important to understanding the process, it's only one aspect. Flowcharts generally don't show things like durations, costs, etc. This means that in reviewing the as-is process depicted in a flowchart, the team may be changing or eliminating short tasks and leaving the longer ones in. Or they may be taking out steps that are necessary to support regulatory requirements. They may even be adding steps that will create bottlenecks or otherwise make the process less effective.

Here's an example. An airline wanted to improve its customers' experience in checking into their flights. It created a flowchart that includes all the relevant tasks: verify the identity of the customer, confirm the booked flight, check the customer's bags, and issue the boarding pass.

As part of the process review, the airline decided to add a step in the process to upsell the customer to a better seat (at additional cost). Adding the step to the flowchart was simple, but the flowchart didn't capture that that specific step takes an average of 30 seconds for customers to read, understand, check seat possibilities, decide to purchase or not, actually pull out a credit card and use it to make the purchase, receive the printed receipt, and move on with the process.

Had they known that, they could have foreseen that adding that step would effectively almost double the transaction time per customer, and they could have been prepared for lines out the door, or perhaps only added that step for the web or app-based check-in rather than at the busy airport ticket counter.

More Flowchart Limitations

There are some other limitations with flowcharts to consider.

- Given the limited amount of information that can be displayed in a single flowchart, it is often difficult to capture dependencies to other workflows – if you did, the flowchart would take up pages and pages and be very difficult to read.
- The complex logic found in many business processes can be difficult to represent in a simple two-dimensional flowchart due to variations within and exceptions to the "standard" process.
- Similarly, if you try to pack all those exceptions and variations into the flowchart, it will become incredibly difficult to read and follow, defeating the purpose of creating it in the first place.
- Finally, there is a tendency in some organizations to keep working on the flowchart until it is perfect. In the meantime, things may have changed in the actual process or its operational environment, so that the flowchart is actually outdated by the time it is considered complete. It might be more helpful to get 80% of the way there and continue to iterate as the to-be process is being developed.

Flowcharts vs. Modeling

As we noted just now, flowcharts are good for visualizing the process, but they lack underlying data. A process model takes the sequence of events that you've created in your flowchart and extends it further.

A good model can map dependencies. It can make us aware of related flows and other processes which may be connected to ours. And it can add data intelligence to the steps, to show detailed information about what's actually occurring within each process such as duration or wait states.

This also allows for some sophisticated simulations where we can run the process according to the set durations – or a range of durations – and get a better understand of the logic, ensure the process is efficient, identify any potential bottlenecks, etc. This is often the most significant to process modeling because often bottlenecks are subtle – some combination of this step taking slightly too long, which makes that one take slightly too long, which backs up the entire thing.



Test Your Knowledge

Question 1:

What is the purpose of a flowchart?

a. To improve the decision-making process
b. To explain the sequence of a process graphically
c. To document business requirements for an information management solution
d. To allow real-time monitoring of workflows within a business process

Question 2:

What are the key limitations associated with flowcharts? (select 2)

a. They don't include durations, costs, or other process-related data.
b. They are poor at mapping dependencies to related workflows.
c. They are too complex to understand for users without specialized training.
d. They only support a limited number of tasks and decision points.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the purpose of a flowchart?

a. To improve the decision-making process
b. To explain the sequence of a process graphically
c. To document business requirements for an information management solution
d. To allow real-time monitoring of workflows within a business process

The correct answer here is B, to explain the sequence of a process graphically. A flowchart can document the decision-making process, but it can't directly improve the process itself. For C and D, these are done in other ways and don't really relate to flowcharts except indirectly.

Answer to Question 2:

What are the key limitations associated with flowcharts? (select 2)

a. They don't include durations, costs, or other process-related data.
b. They are poor at mapping dependencies to related workflows.
c. They are too complex to understand for users without specialized training.
d. They only support a limited number of tasks and decision points.

The correct answers are A and B. For C, the value of flowcharts is that they are easy to understand – at least if they are done properly. For D, this is not true – though as the number of steps and decision points increase, the readability and clarity of a flowchart does tend to decrease.



Troubleshooting and Improving Existing Business Process

Troubleshooting an Existing Business Process

Troubleshooting involves asking the right questions. Here are several direct questions; this list is by no means exhaustive.

- Is the process flow correct, efficient, effective, and adaptable? There's a natural aesthetic to good process design. Do things flow in a natural order or does it look like it has evolved over time? An elegant process is immediately identifiable. We can look at a process map and know whether or not it's well designed and fluid.
- Is there evidence of obvious defects in the process? Often these may be flagged up and people are already using workarounds to make a defective process work.
- Do we have unnecessary sign-offs and approval activities in the process? Sometimes these are unnecessary steps and can be removed.
- Who is accountable for the process? If so, who is actually in control?
- What metrics are there, are they sufficient and informative to define the expected outcomes?
- Is there one standard way of working, or does each agent do things differently?
- Are all the steps really necessary? Some may be historic, "we always do this," and some may be lingering to keep someone busy, but not adding value.
- Are we asking for data we don't need? Forms (both electronic and paper) gather a lot of information. Do we actually need it all?
- Are the hand-offs between various process elements and tasks smooth? Do we find that when things are handed off from one group to another, that information needs to be continually rechecked and revalidated?
- Is there a good communication path between process steps? Do people understand why they're getting the information? Are they communicating their needs back up the chain?
- Do the people undertaking the tasks understand and follow the documented process? Perhaps the process is not followed is because it is out of date or too complicated.
- Are there major bottlenecks in the process, areas where things slow down? Are there elements that take up a disproportionate amount of time?
- Is there clear and logical input and output to each step? Or is it fuzzy? Do things just happen? Is there a clear logical?
- How else could this process step be accomplished? Is there an obvious different way it could be done? Sometimes improvements are evident right up front, sometimes brainstorming ideas can reveal alternatives.
- Could we do something upstream to eliminate the step? Gathering, checking, and approving information once might bring improvements down the line. Think about upstream elements rather than downstream, i.e., elements earlier in the process.
- How do exceptions impact the overall process? How much effort is expended in dealing with exceptions? It may not be much, or it may be enormous. But we want to look at exceptions at a step-by-step level, because they can have a disproportionate impact on different elements of our process.

Bottlenecks – Definition

A bottleneck is a point in a process where the flow is impaired or stopped altogether.

An approval process could be a bottleneck. For example, the overall time expended on a particular activity from beginning to end might be three weeks, with two weeks spent waiting for an approval as it sits on somebody's desk. With quicker approval times the process could all flow much quicker.

© AIIM – 2019

Another simple example for a delivery truck: delivery to the customer causes delays. Investigation reveals the person notifying the customer is sick, so no notification was given to the customer. The customer warehouse has only one forklift truck, so the unloading takes longer thus resulting in the delivery truck having to wait causing delays elsewhere for other deliveries!

Bottlenecks are the main issue in inefficient processes. They can cause a lot of problems in terms of lost revenue, dissatisfied customers, wasted time, poor quality products or services and higher stress in team members.

Some bottlenecks are obvious just by looking at the flowchart; typical bottlenecks occur at approval tasks. In other cases, the bottleneck might be more subtle and hard to identify just by looking at a flowchart; in those cases, a process model or simulation might be the better option. A process model might show for example that the bottleneck only occurs at a certain volume or when a task exceeds a certain amount of time.

It is important to balance the flow rather than just increase the capacity. Without proper analysis you might simply move the bottleneck elsewhere.

Questions to Ask – Bottlenecks

When we see *bottlenecks* in our maps, we want to ask the following kinds of questions:

- Were people aware of this bottleneck? Typically they are. And, if they were aware of it, why did it continue to exist? There may be a good reason or it may be that something is fundamentally wrong in the process.
- Is it always a bottleneck? Is it always a problem or is it sometimes a problem? If only sometimes, is that because it occasionally gets overloaded? Maybe we need to start thinking about batch processing and better managing the workload at that bottleneck phase, or understanding more about peak periods of activity.
- Can some elements of what's happening at this bottleneck be moved upstream? Could they be done earlier, and thereby lighten the workload at this point?
- Can the task be split into multiple smaller tasks and parallel processes? If one person or system is overwhelmed, is there some way of splitting the task into different elements and involving other resources so that it gets done quicker and makes better use of the resources?
- Can the bottleneck be resolved via better process balancing? Typically in production environments, bottlenecks can be resolved via better process balancing. We need to manage the flow of activities more efficiently throughout the process to ensure that workload is fair and efficient at all points.
- Ultimately, we may need to reallocate resources - not an easy thing to do. But if that's the way to resolve it, you need to spot that upfront and deal with it as soon as possible.

Bottlenecks deal with resources, time, volume, complexity, and politics and structure. Many times, we'll find that an approval process, for example, that has become a bottleneck is there because a senior manager decides it's there and it's almost like a job fulfillment activity. Changing that approval flow could be difficult, regardless of how inefficient it is.

Addressing Bottlenecks

The capacity of a series of tasks is determined by the lowest capacity task in the sequence, and this constraint indicates where the current bottleneck is. Note the use of the word 'current', because fixing one bottleneck might just reveal another.

We can address bottleneck by increasing the process capacity by creating a parallel process. Or we can add additional resources to increase capacity at the bottleneck. For example, an additional machine can be added in parallel to increase the capacity. This is often the simplest fix and should be the first option tried.

We can also minimize non-value adding activities that could include transport, rework, waiting, walking between tasks, testing and inspecting, and support activities.

We can increase our flexibility to deal with workloads by outsourcing certain activities. Also, flexibility can be enhanced by postponement, which shifts customizing activities to the end of the process.

Approvals process steps are by far the biggest causes of bottlenecks and this should be eliminated, if they are redundant.

The final point is about anticipating peak loading times. For example, if there is a surge in activity at month-end you might clear work-in progress to free-up resources at month-end, or draft in extra help for a few days each month.

Hand-offs

Hand-offs are when the task passes from one person to another. We need to look at how this hand-off process works in practice. Good practice suggests we need a smooth transition from one process or activity to the next. One example might be the end-of day or shift procedure, whereby patient case notes are handed over smoothly to incoming shift. Another example is where the workflow passes to another agent, and all of the information is available in a shared folder, such as web page or document editing we mentioned earlier.

If hand-off is not smooth there is an impact on the next process, by definition, so we should analyze what is occurring, what causes a disconnect, what the impact is, and then determine the options for remedy.

Hand-off points are also points of vulnerability, when something could fall down and not be noticed, or incorrect assumptions might be made. “I thought you had sent it out?”

It follows therefore that we want as few hand-offs as possible, within reason.

Best Practices

When you work with bottlenecks and hand-offs be sure to follow these best practices:

Identify all hand-offs and bottlenecks as a priority. Keep asking “why are we doing this, why now,” etc. Look at each step that could be contributing to the bottleneck and any redundant tasks.

Any improved process needs to plan around or eliminate current bottlenecks and minimize hand-offs. If you can’t do that, then there will need to be many procedural checks to be put into place to ensure the smooth operation of the process.

Be aware that by just shifting a bottleneck and dealing with it doesn’t mean it’s gone all together. Quite often the problem will pop up elsewhere. Bottlenecks may need continuous attention.

Planning Workflow Routes

This module will help you to determine how to plan routing of tasks or information using a workflow/BPM system using different approaches such as deadlines or time stamps, parallel processing, or sequential processing.

Routing Methods

Routing information means moving a task or an activity or a piece of information from point A to point B.

First, we must consider the human elements and the roles in the process, analyze these to improve the process, and examine the context of the role or the individual. Next, we need to see how is this role is operating, its ability to operate effectively, and the tools that it is making use of with what resources. We then can start to make decisions as to what is the appropriate routing method to use:

- role-based system
- relationship-based system,
- group
- dynamic

Now let’s look at each concept.

Role-based Routing

First, let’s discuss role-based routing. This type of routing assigns tasks to job functions - for example accounts payable clerks. We could even break that down further – accounts payable clerks for names A to M and accounts payable clerks for names N to Z.

This typically is considered the most practical approach, particularly in mid-size or larger organizations because it covers most eventualities. For example, if somebody’s sick or if somebody leaves, the process doesn’t come to a standstill because tasks can be assigned to other people in that same role. This is a practical example because this happens every day in organizations (and also underscores the importance of cross-training for smaller ones).

© AIIM – 2019

As we consider the role-based route, we must ensure that our roles are effectively mapped and defined. We often find that this is not the case. You will need to define and map roles on your project or empower other people to do so in the project. For example, if a worker is an accounts payable clerk, this doesn't mean that that tasks he undertakes on a day-to-day basis have been properly mapped and documented. However, those in this accounts payable role are the only people who understand what they do, and they'll need to map and define specific roles.

Group Routing

Group routing includes distinctly individual roles that are combined for specific tasks. Such tasks include reporting or reviewing and feedback. We make use of parallel processes with group routing, and potentially combine this with sequential processes. In a moment, we'll use an example to show you this.

Dynamic Routing

You can dynamically route individuals, roles, or groups. In most business processes, various routing steps will be based on changeable, variable parameters. Typically, rules determine where the task needs to be routed to next. For example, if you need approval for a purchase over $50,000 dollars, this will need to be routed to one place. If your purchase requires approval of over $100,000 dollars, then this will need to be routed to another place.

These approvals are automatically triggered, and they are therefore dynamic. The rules often are more complex than this example. Let's say we're dealing with claims and there are many different variables that our rules engine (assuming we're using one) could trigger the next stage. This rules engine can route the next tasks based on the outcome or the triggers of various rules. Dynamic routing doesn't replace the roles we have discussed. It is a way of understanding how we might move tasks forward in a dynamic and rules-structured way.

Parallel Processes

Parallel processing is the number one method for increasing process efficiency. Long drawn-out activities can often be collapsed down into a parallel stream, and dramatically reduce the time taken to undertake the activity. It is quite a common feature of processes.

In "as is" processes, elements of the process not making the best use of resources or duplicating effort can be easily identified. Parallel processing is the first option to consider every time to resolve potential bottlenecks.

An example might be in call-center operations. In order to respond to 200 incoming calls per hour, how many operators will you need? Clearly one is not enough.

Approaches to Parallel Processing

There are three primary approaches to using parallel processes – which one to use will depend on the specifics of the process, the resources available, etc.

- **Specialization.** In this approach the process is split up into specific tasks for different roles and assigned using group routing. In a supermarket, individuals are tasked with operating the checkout lines, or restocking shelves, or running the accounting department.
- **Duplication.** This approach uses the roles-based routing approach and simply creates more instances of the process running in parallel. An example here might be opening more checkout lanes at the supermarket. However, once all the lanes are open, the system in this case is running at maximum capacity.
- **Time shifting.** This approach also uses the roles-based routing approach, but instead of adding more lanes, we move some of those tasks to different times. So a shift of cashiers works in the morning, a different shift works in the evening.

For the latter two in particular, we might combine these using analytics about store historical usage patterns and determine that we need more checkout clerks right after work and on the weekends, and relatively fewer in the middle of the morning on a weekday.



Test Your Knowledge

Question 1:

What is considered the most significant issue for an inefficient business process?

a. Decision points
b. Clusters
c. Handoffs
d. Bottlenecks

Question 2:

Why is role-based routing considered the most practical approach to task routing?

a. It costs less compared to group or dynamic routing.
b. It lets someone else in the same role perform the process when necessary.
c. It takes into account unique variables that change the routing dynamically.
d. It provides the most options for full process automation.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is considered the most significant issue for an inefficient business process?

a. Decision points
b. Clusters
c. Handoffs
d. Bottlenecks

The correct answer is D, bottlenecks. For A, the existence of a decision point isn't necessarily an issue. For B, clusters are groupings of tasks close together, but they don't prove that the process itself is inefficient. For C, handoffs can also lead to issues, and in fact to bottlenecks, but they are not issues per se.

Answer to Question 2:

Why is role-based routing considered the most practical approach to task routing?

a. It costs less compared to group or dynamic routing.
b. It lets someone else in the same role perform the process when necessary.
c. It takes into account unique variables that change the routing dynamically.
d. It provides the most options for full process automation.

The correct answer is B, it allows tasks to be routed to other people in the same role for example when someone is out sick. For A, there is no direct correlation between the approach and cost. C is a better description of dynamic routing. For D, all of the approaches provide similar options for process automation.



Selecting the Right Process Automation Solution

Tool and Process Scenarios

Earlier in this course, we've touched upon various scenarios that you need to be aware of when you're thinking about business processes and business process management tools options, for example, forms-centric versus process-centric approaches to BPM, routing, workflow BPM, sort of hierarchy, the difference between ad hoc and product systems.

You need to be thinking about how different technologies map to these different scenarios that you may be facing. Let's look at each one of these in a little bit more detail, so that we can help you narrow down the field to the right set of choices for you.

Forms vs. Process-based Technologies

BPM tools fall into two basic categories: a forms-based approach and a process engine-based approach.

Simply stated, a forms-based approach tool can generate a form that can be used to route information from A to B, or maybe A-B-C. Forms-based tools also can be used to build simple business applications in order to quickly prototype ideas and organize the prototypes to see if they're compatible. These kinds of tools typically work with a messaging infrastructure (e.g., email systems and app front-ends) to improve information movement within the process.

The process engine-based approach to BPM is more complex. These tools are designed for close integration with other applications. Process engine-based tools include powerful tracking and management capabilities and provide support for much more complex process situations. The tools for this approach usually require a dedicated software system for directing, monitoring, and modelling information. Some tools are BPM specific, with plug-ins for major software vendors, and others form part of a business suite or industry-specific suite. Over the years some bigger software companies have acquired BPM software and expertise to integrate BPM into their own products.

Routing, Workflows, and BPMS

As we've previously discussed, routing is the movement of information or a task from one place to another. Routing is linear; it moves content or a person from A to B. Routing focuses on particular tasks – when A is complete, B can begin. Basic routing is linear process management.

Workflow is more expansive. A workflow system incorporates flexibility into tasks. It's more than just simply moving things from A to B to C to D. A workflow system allows processes to occur in parallel or split formation. A can go to one, two, three, four, five simultaneously, and then link up with B. This design responds to more complex situations and allows an organization to save time and increase productivity.

Full business process management systems encompass an even greater level of sophistication and allow for data-to-data, or system-to-system integrations alongside the human activities within a process. Business process management systems are also used when coordinating across multiple processes, systems, or organizations.

Case Management

As a reminder, case management consists of content and process management capabilities that allow a case to be managed as a single unit from inception to completion and disposition. The overall process is repeatable, but case management is generally not transactional; rather, individual cases are unique within the bounds of the process.

Transactional Content Management

Finally, just a brief discussion of transactional content management. Like some of the other tools we've discussed in this course, transactional content management fits under the umbrella of process improvement technologies. However, while most other tools focus on the steps in the process, transactional content management focuses on the content of a document-centric transaction.

Transactional content management is a good fit for processes that focus on documents and that generally require some sort of review and approval tasks. Examples might include insurance claims, contracts, accounts payable, invoice processing, etc. Documents are guided through the workflow to the ultimate approval, and then they are retained in a repository of some sort along with an audit trail.

© AIIM – 2019

Which to Select?

Here's a summary with recommendations for when to consider each tool. If your process:

- Is simple and linear, consider an ad hoc approach (no tool, or a messaging tool) or routing.
- Is more complex, but still operates within a single process area, department, and system, consider a workflow solution.
- Crosses processes, systems, and/or organizational boundaries, consider a full BPM system.
- Focuses on document-centric transactions that generally require approvals, consider a transactional content management system.
- Focuses on cases comprised of multiple documents that need to be processed and managed as a unit, consider case management.

Pitfalls and Caveats

It is very easy to pick the wrong technology with BPM. Clearly, there are huge differences between them, and you've got to match those against your needs. We cannot stress that enough.

Scalability and power and really amazing systems can seem attractive, and may be important, but usability is much, much more important. If people find something easy to use and amenable to the way they work, they are more likely to use it. The most powerful system in the world isn't of much value to anyone if it isn't used properly – or at all.

Think about it this way – a BPM tool really is a communications device. So it needs to talk to all interested parties, whether that be a data source in the system or an individual user. Often, they can't inter-relate, or they can only talk in a limited fashion. If a robust, powerful, expensive IT system cannot communicate to get the process working, then it's a failure.

The reality is that workflows are seldom standardized workflows or fixed business processes. Technologies tend to assume that they are standardized and that everything is idealized. But there will be exceptions. Make sure the tools you go with can adapt and embrace the reality of many exceptions.

Process Metrics

The key to effective monitoring is to identify and capture key metrics. We can use those to either improve the existing process or to modify the process routes dynamically, for example by redirecting work to an area that isn't quite as busy as some other areas.

Metrics can be captured in real time. But they typically will not represent all the available data. The monitoring has to focus in on key elements that tell a meaningful story about the process. We have to be very clear about what we're gathering, why we're gathering it, and what we're going to do with it.

Levels of Reporting

The levels of reporting or even monitoring that one might want to undertake typically involve something that we might want to report via group or department on the particular activities in progress that they're making in relation to the process.

We may well want to look at the status of the process. Is the process active? Is it inactive? Is it blocked? Is there a problem? We might want to look at the status of a particular job that's running through this system. But, remember we likely have many instances at any one time and many jobs active. We might want to check on a particular job and see its status. We may want to drill down to individual tasks and see how they're performing.

We likewise might want to be looking at how much time particular activities take. We could look from beginning to end, or at a particular task stage, or the volumes that we're currently handling, and how those volumes are being managed by various individuals or groups, etc. There are many different facets to the way we might want to report. And you can see here that you could possibly slice and dice this kind of approach in many different ways to meet your specific need.

But two key things stand out here. We can either report on the basis of time meaning how much time it's taking to do things. Or we can report on the basis of volume, how much volume is currently flowing through. We can probably report on a combination of both.

The key point is that we can drill down to any level we want, whether that's at the task level or just the overall process. There are lots of levels of reporting. The facilities are there for us to design the reporting to match our specific needs.

What to Monitor?

What are we actually going to monitor? Here are the general categories – volume, costs, revenues, general tasks and processes, user workload, and jeopardy reporting – where things are going wrong.

Volume

Volume reporting – how many jobs are currently active? How many are at what stage? Those are the kind of key metrics we're looking for. With volume, how many are on-time? How many are overdue? When did they start? When were they completed? What patterns are now starting to emerge? Can we use those patterns to really start working out where we can improve or where we can reallocate work in a cost-effective way? How many jobs are typically active at any one time?

And perhaps most importantly, can we use this information to predict future activity? Growth rates can have a significant impact on processes that outgrow the capacity of the tools and staff executing them, especially when the growth rate is higher than expected.

Costs

Everybody's concerned about costs. How long did it take to execute a customer order? How long did it take to resolve a query? How long did it take to dispatch goods? How long did it take to reorder and have delivered replacement stock?

By monitoring costs, we can provide analysis as to whether specific cost-related activities need to be built into the process. If we're monitoring activities, we spot, for example, where approval levels are working well or proving to be a bottleneck, or proving to be unnecessary. If there is an approval step in any kind of process flow, and the approval is always yes or always no, then it probably shouldn't be there.

Time versus costs, must be accurately factored into the process. Model, time, and cost taken together can provide a rich picture as to what's going on.

General Tasks and Processes

Often, we're going to want to track particular tasks or general tasks and processes. We may want to track the status of specific tasks at any one time, or tasks that have become mission critical, or that have been highlighted as being a problem area, or that we want to study for future improvement. We want to know where things are in the process. The information that we're getting here has to be highly accurate or it's of no value. So we want it in near or real time.

If we have that kind of information, even at a generic level, we can quickly get indication of the length of time left to complete process. If something's being held up at a particular stage, then we can get some indication of how long it's going to be left to complete a process. That kind of information can be very valuable in customer support situations where we know roughly how long things should take and can determine whether they are processed in a timely manner.

User Workload

A very popular method of monitoring is to focus in on particularly users. For example, how much work has Alan done today? Or how much has this role undertaken this week, today, this hour, this month? And by working that out and by extracting that data, we can get some idea of how long it takes him to complete a particular task. And we can then manage future tasks accordingly.

We can also of course take that information and compare it with the activities of other individuals or positions within our organization and see if there's a comparison. Maybe it's taking a lot longer or it's been done more quickly, more efficiently. We can use this, for staff training or staff improvement activities.

The kind of questions we're going to ask though are typically, how many overdue or backlog tasks do they have? And can any of that work be reallocated?

© AIIM – 2019

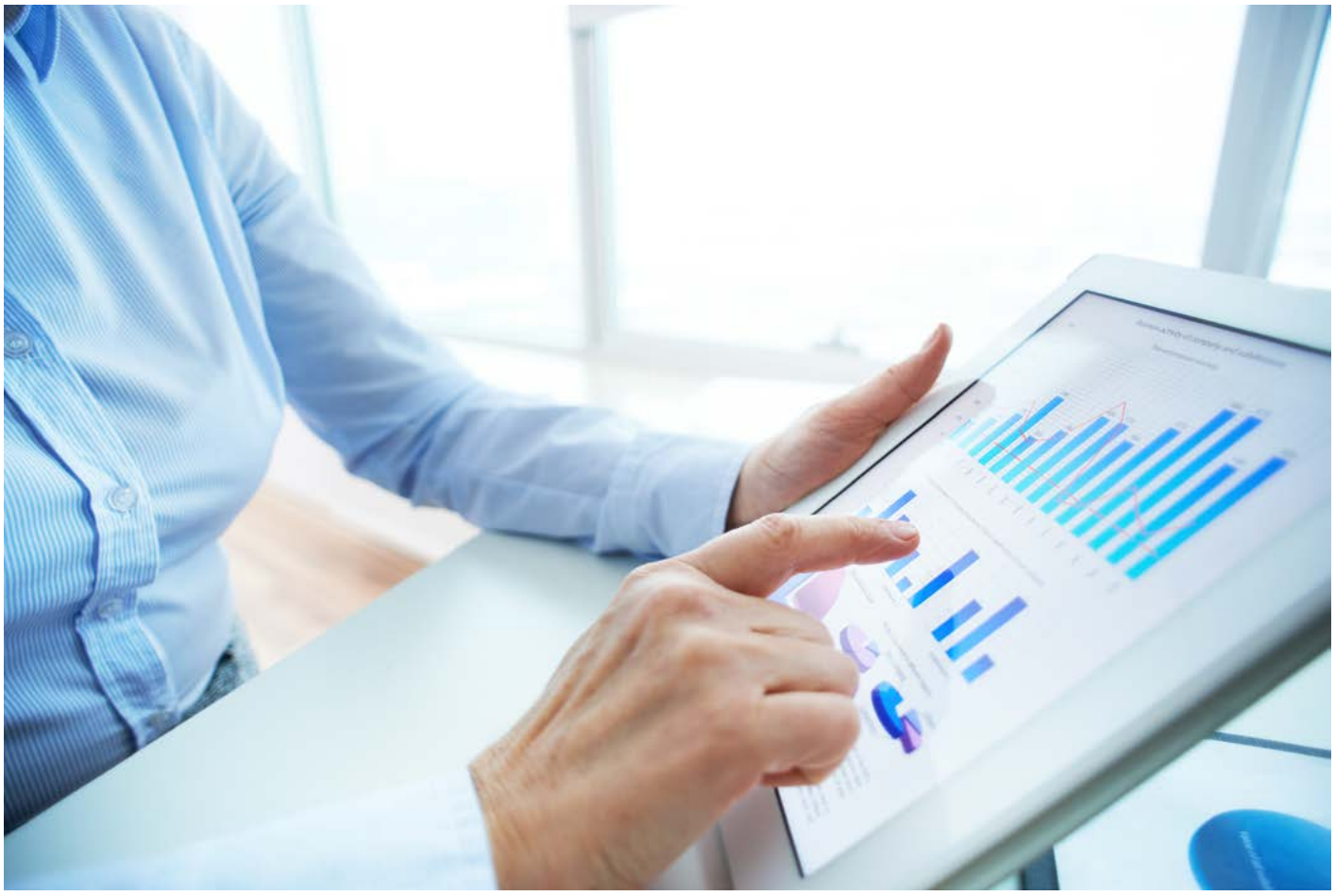
Jeopardy Reporting

Jeopardy reporting provides information about any kind of tasks that are late and may jeopardize the outcome of future elements in the process.

These are the kind of reports that give senior management a look at what's going on, and an understanding of what needs to be prioritized so that they do not become late. They enable management to answer these questions. For example, what is actually causing problems? Can we determine which ones are starting to fall behind? Is there some kind of pattern as to why we're falling over at this particular stage each time? Is it maybe something that's occurring upstream in the flow that's impacting the downstream?

Maybe even getting it down to an individual level – can I identify problems with the actual resources themselves? We could have a serious backlog and it simply turns out that the individual is sick, is not there, and the work's just backing up. Or we have somebody who's incompetent, who needs more training or maybe needs reallocating to a different role.

Jeopardy reporting can be a very powerful management tool in its own right.





Test Your Knowledge

Question 1:

What elements should be considered when determining which process automation solution to use? (select all that apply)

a. Types and nature of content within the process
b. The complexity of the process
c. Whether the process crosses systems or organizational boundaries
d. The legal or geographic jurisdictions in which the organization operates

Question 2:

Which type of metrics are used to identify and analyze the efficiency of individual staff members within a particular process?

a. Exception reporting
b. User workload
c. General tasks and processes
d. Key performance indicators

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What elements should be considered when determining which process automation solution to use? (select all that apply)

a. Types and nature of content within the process
b. The complexity of the process
c. Whether the process crosses systems or organizational boundaries
d. The legal or geographic jurisdictions in which the organization operates

The correct answers are A, B, and C. D is certainly an important consideration for many information management-related issues, but it would not determine which process automation approach to follow.

Answer to Question 2:

Which type of metrics are used to identify and analyze the efficiency of individual staff members within a particular process?

a. Exception reporting
b. User workload
c. General tasks and processes
d. Key performance indicators

The correct answer here is B, user workload. The other answers are all valuable metrics, but they focus on other things than individual staff member efficiency.



Robotic Process Automation

Introduction to Robotic Process Automation

So what is robotic process automation, or RPA? The Institute for Robotic Process Automation and Artificial Intelligence offers this definition:

"Robotic process automation (RPA) is the application of technology that allows employees in a company to configure computer software or a "robot" to capture and interpret existing applications for processing a transaction, manipulating data, triggering responses and communicating with other digital systems."

It goes on to argue that RPA provides dramatic improvements in accuracy and cycle time and increased productivity in transaction processing while it elevates the nature of work by removing people from dull, repetitive tasks.

RPA is generally used for high-volume, repetitive tasks such as looking up data in one system to do something with it in another.

RPA Capabilities

RPA tools work much the same as their human counterparts: they log into systems, search for information, copy and paste information, complete form fields, and so on. They essentially take the screens that workers look at, and key data into manually – and the RPA tool fills in those same fields on those same screens by automatically pulling the relevant data and moving it to the right location. That alone can be transformative to a business process activity.

The RPA tool is scripted/programmed/instructed to perform each repetitive task - it is told specifically what to do each time. This means that once built and deployed, the RPA tool will not change. Therefore, in turn any data sources and destinations the RPA needs to interact with must be highly structured and unchanging. To put it another way - RPA tools don't deal with quirks, errors, exceptions, or the normal mess of human interactions well at all – they deal with repetitive structured interactions where the same thing happens every time. This means that where there is need for flexibility, judgement, or unexpected conditions, a human needs to be involved.

Most RPA tools do not alter existing systems or infrastructure, so there isn't a significant technical challenge to getting started. Rather, RPA tools operate as virtual workstations substantially similar to the ones used by human staff. Similarly, RPA tools are not replacements for underlying business applications. As we stated earlier, they simply automate the manual tasks of human workers.

Types of RPA

As anyone that has ever worked in a clerical situation, filled in a mortgage application, or submitted an insurance claim will know, things are never quite as clean, efficient, and structured in the real world as people like to claim they are. Life is messy.

That's why RPA tools often require ongoing assistance. Hence, RPA tools can be said to fall into two categories: attended and unattended.

Attended RPA tools are more like virtual 'assistants' than true virtual workers.

This is a very common kind of RPA tool; it lives in a human worker's desktop and helps that worker by essentially copy and pasting data between systems when triggered by a programmed event. It automates the regular drudge work of key entry, while the human worker gets on with other tasks that require flexibility or creativity.

In contrast, there is a class of RPA tools we call unattended. Those that don't require human assistance and intervention.

They run unseen and simply take on work tasks that are added to their virtual work queue. This are for the work that never changes, where there are few if any exceptions.

Neither that attended or unattended approach is right or wrong – it simply depends on what you are trying to achieve.

But, as a general rule of thumb, the more exceptions there are to a task, the more attended an RPA tool needs to be.

© AIIM – 2019

Limitations of RPA

Effective though they may be in many use cases, RPA tools are limited in scope and function.

RPA tools are not going to work well if there are too many types of formats, processes and applications to handle, as the effort required to train, deploy, and manage (potentially multiple) robots means it is may not be worth the effort.

Similarly, if your processes often generate complex exceptions, RPA probably isn't a fit.

If your processes change frequently, RPA is definitely not a fit.

Finally, it's worth noting that although the underlying technology of RPA tools is not particularly new, the application of that technology to real world business problems is new and immature.

Use Cases for Robotic Process Automation

RPA by Business Process

Here's a general idea of how to determine whether RPA would be effective for a particular business process. If the workers in that process:

- Perform structured, repeatable, computer tasks, such as data entry
- Make simple decisions based on predetermined rules
- Need to access multiple systems to accomplish a task
- Need to search for, aggregate, or update information

That process is a good candidate for RPA.

The kind of processes that match that set of criteria makes for a long list of RPA target areas, from data entry to automatic formatting. It can include uploading and moving data or creating messages and undertaking basic quality checks.

Hence, we commonly see RPA tools used for basic data entry and validation, for automating formatting, to automatically upload/download/import and export data, etc.

In other words, these are common tasks that translate to lots of real-world business process activities.

RPA Candidate Process by Industry

Here we can see a list of processes in financial services that are already seeing RPA used to automate tasks – everything from credit card sign ups to direct debit cancelations and standing order modifications. All processes that run the same way every time – the same systems, the same steps, the same tasks, the same data every time – are good candidates for RPA.

Financial Services

- Account audit requests
- Branch risk monitoring
- Check approvals and processing
- Claim repairs
- Credit card signup
- Customer communication
- Customer complaints
- Customer onboarding
- Deductible payment

- Direct debit cancellations
- Dispute resolution
- Employee data management & integration
- Fraud account closure
- Internet application processing
- Lost/stolen card management
- Marketing campaigns
- Modify address details
- Modify direct debit details
- Payments terms administration
- Personal account closures
- Personal loan applications
- Regulatory compliance reporting
- Statement reconciliation
- Wire transfer

This is only a summary of some of the possibilities. It's far from a comprehensive list of the types of work in financial services that RPA tools can be deployed. They all meet the criteria of repeatable, standardized processes and tasks with minimal complexity in exceptions.

Here are even more example areas that RPA can bring benefit to in the worlds of banking and insurance – these lists can go on and on.

Banking

- Account cleansing
- Account reconciliation
- Accrual support
- Audit support and validation
- Credit card orders processing
- Delinquent loans notification
- Fixed asset amortization
- FX accounting
- Month-end close
- Mortgage approval
- Refinancing entries
- Report generation
- Tax reporting

© AIIM – 2019

Insurance

- Accounting data mismatch
- Bulk payments
- Compliance reporting
- Deceased notification
- Funds transfer
- Marketing campaign support
- Payments transfer
- Pension auto-enrolment
- Policy document data transfer
- Redemptions processing

But take a look at them – month end closing, mortgage approvals, pension automation, etc. Yes, they are repetitive dealing with the same type of data, systems and tasks repeatedly – but they are far from inconsequential. These are business critical activities: activities that if they go wrong can be disastrous for an organization or its customers. We mention this as it is all too easy to conflate 'simple' technology with 'unimportant' activities. RPA tools are involved in running major business activities with high risk and value associated to them.

RPA in Purchasing

So let's dig even deeper into what people are doing in the real world today with RPA – the following two case studies have been anonymized but they are based on actual and active real world RPA deployments today.

The first is in purchasing. We are using this as a use case that is common across any large organization in either the public or private sector. Even mid-sized firms struggle with this kind of situation.

Purchase orders come in from all corners of the globe. Each supplier has their own format and ways of doing things – but fundamentally they are all purchase orders. Today the challenge is that the firm needs highly skilled and knowledgeable staff to make sense of all the incoming POs – and to enter all this data into multiple business applications. It is slow, error ridden, and ripe for improvement.

In this case the firm deployed an unattended RPA system – one that can identify the relevant data in each PO based on predefined rules that have been built by knowledgeable staff members. Once the RPA system identifies the relevant data, it auto populates that data into the relevant business applications.

This technology firm has seen a lot of success using RPA for their purchasing processes. It no longer runs Monday to Friday 9-5; instead, it runs 24/7, and have really seen a big improvement in the time it takes to process a PO. Finally, and of interest for this course – the implementation time for the RPA system was only seven days. In their eyes at least, this was a very quick fix and quick return on their investment.

RPA in HR

Similarly, here is a multinational firm that was struggling to manage the onboarding of new recruits. Due to the nature of the work this firm undertakes, recruitment involves a lot of cumbersome background checks – across different countries and different languages. The process has become so cumbersome that it was taking months to onboard a successful recruit. In practice many good hires had found other work before the checks were complete, and in more cases than anyone would like to admit some recruits were engaged before they had been properly vetted – in a few cases with negative results.

This firm deployed an RPA tool that extracted PII (personally identifiable information) from the recruits' applications – data like their name, address, date of birth, etc. That data was then automatically run against multiple databases in parallel, generating reports on the candidates and quickly enabling the prioritization of some candidates for further checking.

With the RPA tool running, the 2-3 month process dropped to under two weeks. Problem candidates were quickly identified, and now, no candidates are ever recruited without proper background checks. Interestingly in this case the firm did not reduce headcount; rather they deployed the existing staff to other tasks.



Test Your Knowledge

Question 1:

What is the benefit to using robotic process automation capabilities?

a. RPA allows for complete automation of all tasks within a process.
b. RPA is highly flexible and can add address exceptions and variations very efficiently.
c. RPA automates low-value tasks, freeing up employees for higher-value ones.
d. RPA can extract meaning from unstructured documents.

Question 2:

Which process would be most likely to benefit from RPA capabilities?

a. Contract negotiation
b. Account closure
c. Annual performance reviews
d. Manufacturing quality control

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the benefit to using robotic process automation capabilities?

a. RPA allows for complete automation of all tasks within a process.
b. RPA is highly flexible and can add address exceptions and variations very efficiently.
c. RPA automates low-value tasks, freeing up employees for higher-value ones.
d. RPA can extract meaning from unstructured documents.

The correct answer is C, RPA automates low-value, repetitive, mundane tasks that do not require judgement or decisions. For A, RPA can't automate any task that requires flexibility or judgement. For B, RPA is not at all flexible. For D, RPA doesn't interact with documents or data at that level – it's simply task-level automation.

Answer to Question 2:

Which process would be most likely to benefit from RPA capabilities?

a. Contract negotiation
b. Account closure
c. Annual performance reviews
d. Manufacturing quality control

The correct answer is B, account closure, because that will follow the same process every time. For A and C, these two processes require much more flexibility and judgement. For D, manufacturing quality control is beyond the scope of what RPA can directly do.



Case Management

Introduction to Case Management

Let's start with a definition of what case management is. This is AIIM's definition: a "case" is any project, transaction, service or response that is "opened" and "closed" over a period of time to achieve resolution of a problem, claim, request, proposal, development, or other complex activity. It is likely to involve multiple persons inside and outside of the organization, with varying relationships to each other, as well as multiple documents and messages.

Case management consists of content and process management capabilities that allow a case to be managed as a single unit from inception to completion and disposition.

The overall process is repeatable, but case management is generally not transactional; rather, individual cases are unique within the bounds of the process.

How Case Management Works

Case management may be described as managing bundles of content rather than individual documents or images. A "case" is a compendium of information, processes, advanced analytics, business rules, collaboration, and sometimes social content that relates to a particular interaction or issue involving a specific party like a customer, supplier, patient, student, or defendant. Case management solutions are designed to manage all of this to help drive more successful, optimized outcomes – even as they also attend to and secure the individual bits of material contained therein.

Case management applications, depending on industry vertical and specific use case, may manage, route, process, archive, or dispose of a "case file" or group of related documents as an aggregate – not as individual items. Cases are often handled as transactions, forwarding groups of related forms, applications, evidence, or supporting documentation throughout an audited, traceable business process.

Incident Reporting and Tracking

If this sounds like Enterprise Content Management (ECM) to you, you are largely correct, but there are differences. Perhaps most notably, case management applications include functions like incident reporting and investigation management. These capabilities involve entire processes unto themselves and can encompass active or work-in-progress documents of all kinds that ultimately will need to have content management principles applied to them. As such, they require specialty care and tracking not only up front as information is captured and analyzed, but after the fact as well, when remediation steps are taken. So they represent a distinctly different, though related, aspect of content and information management.

Case Management and Workflows

Finally, a workflow or business process needs to take place to move the case to its outcome. Within CRM systems, an alerts-style functionality will frequently exist against a given customer log. In traditional document management systems, workflow may involve moving a given document through each process stage – probably in a serial manner. Neither is appropriate to a case management scenario where the focus is "the case" or collection of information, not a customer or single document. The process outcome is the successful resolution of the case. The participants may need to respond against given deadlines, and those involved in managing the case need to see progress reporting and action monitoring against the case.

Adaptive Case Management (ACM)

The Workflow Management Coalition (WfMC) defines adaptive case management (ACM) as "information technology that exposes structured and unstructured business information (business data and content) and allows structured (business) and unstructured (social) organizations to execute work (routine and emergent processes) in a secure but transparent manner."

© AIIM – 2019

It argues that ACM involves three distinct paradigm shifts:

- It deploys the organization structure, process structure, and acts as a system of record for the entities and content involved.
- It enables non-technical business users to create/consolidate processes from business entities, content, interactions, and rules.
- It moves the process knowledge gathering phase of the process lifecycle into the actual process execution phase.

The "adaptive" part is that a case manager can make new cases that build on structure of previous cases through some form of copy or reuse of the earlier cases without needing any special skills. Over time the case manager will adapt the system to their own style of working without needing the help of any specialist. In traditional case management applications, regular patterns are expected, defined, and pre-programmed. But in ACM, knowledge workers adapt the system to their needs as they work.

Another way of thinking about ACM is to contrast it with BPM. BPM brings content into and through processes, while ACM brings processes to the content of the case. In the context of ACM the process may not even be fully defined but is determined as the case progresses.

Use Cases for Case Management

Vertical Applications

As previously noted, case management applications, depending on industry vertical and specific use case, may manage, route, process, archive or dispose of a "case file" or group of related documents as an aggregate – not as individual items. Cases are often handled as transactions, forwarding groups of related forms, applications, evidence or supporting documentation, and even communications throughout an audited, trackable business process.

Case management is not "one size fits all." A case management application will often have to conform to the norms and constructs that pertain to a particular vertical market. Case management in a law firm has a very different set of features, workflows, taxonomy and metadata requirements compared to case management used by investigators in an insurance company.

Some examples of vertical case management applications might include:

- Legal casework
- Insurance claims
- Police investigations
- Healthcare
- University admissions

Horizontal Applications

The second common type of case management application can address issues that cut across multiple markets, so-called "horizontal applications." These "horizontal" use cases are those that exist in one form or another in many organizations. Examples of these might include:

- HR onboarding
- Vendor and contract management
- Correspondence management
- Service requests
- Complaint resolution

Whether looking at vertical or horizontal applications, processes that can benefit from case management capabilities will generally involve multiple documents and/or communications that need to be managed as, well, a case. There will be a process involved with a start and finish, but the middle will be significantly less structured and require more judgement and flexibility.



Test Your Knowledge

Question 1:

What is the primary difference between BPM and case management?

a. BPM is rules-based, while case management does not use rules.
b. Case management is more suited for transactional processes than BPM is.
c. BPM is able to handle exceptions better than case management.
d. Case management is more flexible for less-structured processes.

Question 2:

Which of these would most benefit from case management capabilities?

a. Invoice processing
b. Criminal investigations
c. Training development
d. Performance evaluations

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the primary difference between BPM and case management?

a. BPM is rules-based, while case management does not use rules.
b. Case management is more suited for transactional processes than BPM is.
c. BPM is able to handle exceptions better than case management.
d. Case management is more flexible for less-structured processes.

The correct answer here is D, case management is more flexible for less-structured processes. For A, case management does rely on rules. For B, BPM is much more suited for transactional processes than case management is. For C, case management, especially adaptive case management, is generally able to handle exceptions better than BPM.

Answer to Question 2:

Which of these would most benefit from case management capabilities?

a. Invoice processing
b. Criminal investigations
c. Training development
d. Performance evaluations

The best answer is B, criminal investigations. For A, invoice processing is generally too transactional in nature. For C and D neither training development nor performance evaluations would generally require moving around a number of different documents through a process as a discrete unit.



Automating Governance and Compliance

Introduction

In this domain we focus on some of the information management policy and process elements, particularly as they apply to the information lifecycle.

The domain begins with information governance and setting up strategies and structures to support effective information management and ensure its alignment to business goals and objectives.

Next, we review best practices for records management – ensuring information is trustworthy and authentic throughout the lifecycle until disposition occurs.

We take a look at information security tools and techniques that should be part of the information professional's toolkit.

We spend some time on privacy and data protection – how to safeguard personal and sensitive data, how to build privacy into work processes and technical solutions, and what happens in the event of a data breach.

We take a brief look at eDiscovery – the steps involved and some of the challenges associated with producing information in response to a legal or regulatory request.

Finally, we end this domain with a review of key digital preservation issues and strategies.

Automating Governance and Compliance

One of the key challenges of effective information management is that users don't want to do it. That is, they already have jobs and often information management is seen as administration, or overhead, or busy work, etc. And even if they did, it's now what they are trained to do. So to be effective, organizations have to embrace the approach of "streamline and automate." To paraphrase Einstein, governance and compliance tasks need to be made as simple as possible (but no simpler). And they need to be automated to the maximum extent possible – ideally users have no idea that things like information security and records management are even happening.

Many of the tasks outlined in this domain support all of the approaches to intelligent information management we've previously discussed. Good governance and records management support innovation, and the customer and employee experience, and executing processes nimbly. Privacy can certainly impact how employees and customers engage with the organization – or elect not to. And all of the tasks here help the organization to minimize risk. But there's a reason that one is listed last – in too many organizations these tasks and disciplines are either ignored completely, or come at the expense of effective business outcomes. Organizations need to balance the need for compliance and risk management with the need for them to achieve their core goals and objectives.

© AIIM – 2019

Introduction to Information Governance

Data and Information Stewardship

This quote illustrates the key gap that exist in most organizations. Most organizations have put in place financial systems to manage their financial assets. They have put in place HR systems to manage their employees and the information associated with them. And they have often put in place ERP systems to manage their physical assets. They also acknowledge that the way in which they handle information is critical to their ability to compete. However, there is often a gap between matching this importance to effective information management systems. The concept of stewardship of information assets can help organizations understand what needs to be done in terms of putting in place systems to manage their information assets.

Here, we further expand the concept of stewardship. The key concept is that a Steward is a "keeper of the flame" in terms of data and information quality and availability. The core concept is that an awareness of information stewardship provides a framework to think about information governance in a context that reflects the broader well-being of the organization. Too often, the way information is managed is purely a function of the needs of an individual department in an individual process. Information stewardship provides a framework to think more broadly about this question.

Stewardship Responsibilities

This check list from the U.S. Geological Survey (USGS) is a good example of the components of an effective information steward. The information steward:

- Is accountable for information integrity and quality
- Creates standards and business rules
- Specifies and standardizes business terms
- Communicates with business representatives to identify, and ensure that, information meets their needs
- Establishes information security requirements
- Ensures that documentation is development and maintained, including documentation related to metadata
- Validates that organizational and regulatory/compliance access and security requirements are being met
- Participates in the internal team guiding the governance effort
- Is an active proponent of best practices

Stewards and Custodians

Both the business and IT have critical roles to play with regards to information governance. Too often organizations assume that this function can be totally delegated to either IT or to records management or compliance. The concept here is an important one. The business must assume responsibility for stewardship of the organization's information assets. IT must assume responsibility for the framework in which this is done. It is critical that this set of responsibilities be as seamless as possible and as easy to implement as possible from the perspective of the individual knowledge worker.

The Data Governance Council

One last point to make. We just noted that business is responsible for the stewardship of information and data, and IT has custodianship for the systems that information and data resides on. Both need to be represented in strategic decisions. The way to do that is to set up a data governance council comprised of both business and IT representatives. This council will own the overall data and information management framework throughout the enterprise, and should be comprised of senior managers that have authority, responsibility, and accountability within the organization.

Information, System, and Process Inventories

One of the most important foundations of an information management program is the inventory. Simply put, you need to know what information you have before you can manage it. As we've discussed throughout this course, poor information management can lead to significantly increased risks, whether from compliance, eDiscovery, or data breaches; much more importantly, it leads to poorer quality decisions, reduces organizational agility, and ultimately impedes the business of the business.

In the context of this course, the inventory is a systematic process used to discover, identify, and review organizational information assets. In this module we propose three types of inventory. First, the organization needs to conduct an inventory of all of its systems, particularly those that create, store, or manage information.

Then, the organization needs to inventory its information. This inventory will start with the systems identified in the systems inventory, but will often need to include other systems and locations that IT either does not know about, or does not control, such as offsite storage and hosted, cloud-based, or third-party applications.

Finally, the organization should inventory its processes with a focus on information flows.

Now let's look at each of these in a bit more detail.

The Systems Inventory

The purpose of the systems inventory is to identify all the systems in the organization, and specific information about each system in a particular context.

The idea behind a systems inventory is not new – it has long been a good practice for any organization to know about its assets and to manage them effectively, and the first step to do so is to understand what those assets are.

We can consider systems as the assets of IT, and an entire discipline and technology market has grown up around IT Asset Management (ITAM). Similarly, when it comes to eDiscovery, judges and attorneys have been counseling organizations for years about the need to develop an ESI (electronically stored information) data map.

In both cases, and for our purposes as well, the systems inventory starts by identify all the systems in the organization:

- Enterprise-wide systems, such as email or network file shares.
- Departmental- or process-specific systems such as financial, sales automation, customer relationship management, and the like.
- Even more specialized systems that might only be used by a few people or very infrequently.
- And legacy systems. These are particularly difficult to track – and particularly important if they still store information with business value.

For each system, the inventory should identify the broad kind(s) of information that are stored in the system, with an eye towards the business use of that information. In other words, it's more valuable to know that this file share contains engineering or sales or training information than that it includes PDFs, scanned TIFF images, and PowerPoint files.

It's also important to understand certain characteristics about these systems.

- How old is the system and where is it in its lifecycle? That is, is the system a current version and/or still supported by the vendor?
- Is it customized or integrated with any other systems?
- Where is it physically located? This is often a significant issue for multinational organizations, and governmental entities, because of privacy and data protection concerns.
- Who owns the system (and therefore the data on it)? IT is a custodian, but ultimately the business is the steward and owner of the information on those systems.

© AIIM – 2019

Before we move on to the information inventory, just a brief note on the systems that likely won't be on the systems inventory, generally because IT simply doesn't know about them. These include:

- "Rogue" or "shadow IT" systems. These may or may not be supported in some contexts, but the extent of their use, and sometimes even their existence, is unknown to IT because IT isn't provisioning or supporting them. Common examples include commercial file sharing applications such as Box or Dropbox and personal email accounts and other communications technologies.
- One-off systems. Again, IT generally wasn't involved in implementing these and so doesn't know about them. Common examples include Access databases, Lotus Notes applications, single-seat applications such as authoring tools, etc. The information from these "systems" is often stored on a local computer, which has significant implications for security as well as for disaster recovery and business continuity.

We include these because they are a frequent source of concern for recordkeeping, compliance, risk management, and eDiscovery – and because, since IT doesn't know about them, they can't provide much support or assistance for them. It also means that the inventory team will need to find another way to uncover them.

The Information Inventory

With the systems identified and inventoried, we turn to the second type, the information inventory. The primary purpose of the information inventory is to identify the organization's information holdings in some detail. This can include both records and non-records and both physical and digital holdings. In this course we will focus on digital holdings.

The inventory serves to identify how information is created and used by the organization and by individual groups and users.

And it identifies where all that information is stored. This is especially important for digital information. With physical records it's usually pretty easy to tell where they are stored and when things start getting out of control. But with digital information, too often it's "out of sight, out of mind."

The inventory also helps to identify inconsistencies in how information is created, stored, accessed, and managed.

The objective of the information inventory should be to locate, identify, and describe ALL the organization's information holdings. Once this is done, the organization can analyze the findings to identify obsolete or duplicate records or documents and begin the process of consolidating them and/or disposing of them.

It can be used to identify those holdings that are critical to the continued operations of the organization and manage them in the manner most likely to ensure their continued access and availability in the event of a disaster.

And it can be used to identify current and future storage requirements.

The Process Inventory

The process inventory is used to identify the processes within the organization. This inventory can be used to identify processes and sub-processes within a particular functional area as well as those that cross them. In this context, the focus is on information-centric processes and how information flows into them, through them, and ultimately out of them.

The process inventory is important to assess data protection needs associated with particular processes. We'll look at that in more detail later. Similarly, it can be used to assess compliance requirements around recordkeeping, information security, and other regulatory needs.

Finally, the process inventory is also helpful in conducting process improvement activities. We look at that in more detail elsewhere in this course.

Information Characteristics

The inventory should include at a minimum the following types of information:

- Whether a particular piece of information exists in more than one format. rendition, or version, and if so, which one(s) are considered to be the copy of record. Only one should be the definitive record, but it is entirely possible during the inventory that several will be identified as such. This will be sorted out as part of the analysis.
- The particulars of the media – is it paper, microform, some other physical format such as core samples, physical media such as CDs, or digital? For digital information, the inventory should also capture file format and software characteristics – for example, Microsoft Word 2007 or Corel WordPerfect X4.
- The physical and/or logical locations. CDs, paper, and other physical objects have physical locations; for purely digital files stored on servers, the document's physical location may be difficult to determine and so a logical location may make more sense.
- The time span of the information in question. This is particularly important for legacy systems and locations such as decommissioned applications, old network file shares, etc.

The Information Lifecycle

Here is some additional information to gather as part of the inventory. This relates more to whether and how the information is managed. For example, records in a digital recordkeeping are considered to be managed; documents stored on a user's computer or personal laptop arguably are not.

The inventory should attempt to determine whether particular documents are being managed as records, and if not, whether they might be candidates for such management.

It should also note basic creation data including when the information was created and, if updated, when it was last updated. Scanned images will most likely be created and updated the same time or in very close proximity; Word documents and PDFs might be updated over a period of time. A database may be active for years.

Who Uses Information

Some of the user-related information that should be gathered in the inventory include:

- Who created or received the information? This could be a named user, a role, or a department. For information received from another location, the recipient should be noted.
- Who uses or accesses the information?
- How is the information used? In other words, is it reference material for customer service, is it synthesized with other information, are actions taken upon it?
- Who manages the information (to the extent it is managed) on an ongoing basis? In other words, let us say that an organization scans all its invoices and further that the scanning department is separate from accounting. In this case the scanning department creates it, but it is the accounting department that uses it and, in all likelihood, should be the one managing it.

© AIIM – 2019

The Business Context for Information Management

Initial Assessment

The organization needs to conduct an initial high-level assessment to determine the need for information governance. It is important to identify the challenges the organization faces and understand what is doable within a particular time frame for an organization. Examples of tangible business problems could include a failed audit, a data breach, or the need for improved data quality for risk-management purposes.

This will result in defining the initial scope of the program.

Organizational Review

The very first step of the assessment is to conduct an organizational review. This includes a couple of different considerations.

The operating model. Is the organization in the public sector or the private sector? If the former, what sort of organization is it: local, state, or federal government; primary education; university; and so forth. If the latter, is it publicly held or privately? Or is it some other type of group such as a non-profit foundation, a trade association, a union, etc.?

The geographic model. The locations and jurisdictions an organization operates in are perhaps the most significant factor in its information governance environment because laws and regulations will be different in every jurisdiction. In addition, any organization that operates in multiple jurisdictions will have to comply with the requirements specific to each of them. This becomes very complex – and important – given the recent trend towards increased privacy and data protection regulation around the globe.

The industry sector context. Some industry sectors operate in very unique ways, and all of them have their own operating best practices and regulatory regimes.

Contextual Review

The next step is to review the organization's regulatory environment. Here we outline the elements that will inform the regulatory environment.

At the top are statutes, regulations, and case law with the force of law that describe regulatory requirements.

Next are any mandatory standards of practice, which might be found in some industry sectors but not others.

Next come voluntary codes of conduct and codes of ethics which the organization has agreed to abide by.

It is also important for the organization to adhere to its own rules and procedures.

And we end with community expectations. This may seem a bit odd at first glance, but consider that public sector organizations are considered to have a primary responsibility to their citizens and an expectation of openness, transparency, and accountability even where there are no specific statutes. For private sector organizations, there is an expectation that the organization will not harm the public, or its shareholders, or its employees, even where that harm is simply bad publicity.

For multinational organizations, all of these may apply differently in different jurisdictions. For example, privacy is treated significantly differently between the U.S., Canada, the European Union, etc. Similarly, data protection, legal systems and processes, legality of digital signatures, etc. are all likely to have different implementations or expressions in different jurisdictions. Information professionals do not need to be experts in all of these, and certainly not across the entire world, but they do need to understand that the rules may be different in different places and to confer with local subject matter experts.

The next step is to look at the organizational context of the work process in question. This involves identifying the work processes to be reviewed and whether they take place within a particular department, across departments, or even cross organizational boundaries.

The review should determine whether the process is centralized or decentralized and who is responsible for the performance of the process – both at the individual staff/transaction level and from a management perspective.

The review concludes by describing the framework that will be used to identify and break down functions, processes, and transactions and how they relate to each other. The actual analysis will be done as part of the functional analysis which we will describe next.

Assess the Current State

Consider using a maturity model to assess the organization's current state of information management. Maturity models are a useful way to measure the information management capabilities of the business for several reasons:

- They provide a quantifiable score, which can be used to establish a baseline and trending over time
- They are often developed by consultants, associations, or other third parties with no direct connection to a particular organization or its issues
- They provide a roadmap for improvement: At level one, the behavior is X; to move to level two, the organization needs to implement behavior Y
- They help to prioritize among many competing potential initiatives

As one consultant notes, "The true outcome of a maturity model assessment isn't what level you are but the list of things you need to work on to improve[1]."

There are many of these in the market today; which specific one to use depends on what you're trying to measure – strategy and governance; execution of particular information management processes and disciplines; how to monetize your information assets; and others. The list provided here is a starting point, not an exhaustive list.

Identify the Business Impact

Once you've determined the current state, the next step is to identify the business impact of that. In other words:

What are the business implications of the current state?

- Are we receiving complaints by customers or staff?
- Are we losing opportunities due to poor information management?

What are the cost implications of the current state?

- Does it make our processes more inefficient and increase our operating costs?
- How high are our eDiscovery costs?
- How much of our storage costs are spent storing content that we could get rid of?

What are the risks of the current state?

- Do we have problems meeting compliance requirements?
- Have we identified any information security, privacy, or data protection issues?

[1] Martin Fowler, https://martinfowler.com/bliki/MaturityModel.html

© AIIM – 2019



Test Your Knowledge

Question 1:

What are the responsibilities associated with data stewardship? (select 2)

a. Deliver quality data to business owners
b. Specify business terms and quality needs
c. Provide business requirements
d. Create the business data model

Question 2:

What is the benefit of conducting an information inventory?

a. It identifies the key stakeholders for an information management initiative.
b. It provides an understanding of the overall information management environment.
c. It describes the business requirements for an information management solution.
d. It ensures that information is kept as long as required and then disposed of appropriately.

Question 3:

Why is it so important to understand everywhere the organization operates?

a. Information management-related costs vary by geographic location.
b. Language issues can present significant challenges to the IM program.
c. To minimize the difficulties of scheduling meetings across time zones.
d. Legal and regulatory requirements are different in different jurisdictions.



Test Your Knowledge - Answers:

Answer to Question 1:

What are the responsibilities associated with data stewardship? (select 2)

a. Deliver quality data to business owners
b. Specify business terms and quality needs
c. Provide business requirements
d. Create the business data model

The best answers here are B, specify business terms and quality needs, and C, provide business requirements. For A and D, it is the IT custodian delivering quality data and creating the actual data model for the business.

Answer to Question 2:

What is the benefit of conducting an information inventory?

a. It identifies the key stakeholders for an information management initiative.
b. It provides an understanding of the overall information management environment.
c. It describes the business requirements for an information management solution.
d. It ensures that information is kept as long as required and then disposed of appropriately.

The correct answer here is B, it provides an understanding of the overall information management environment. A is an important but separate step that often precedes the inventory. C and to some extent D are outcomes of the inventory.

Answer to Question 3:

Why is it so important to understand everywhere the organization operates?

a. Information management-related costs vary by geographic location.
b. Language issues can present significant challenges to the IM program.
c. To minimize the difficulties of scheduling meetings across time zones.
d. Legal and regulatory requirements are different in different jurisdictions.

The best answer is D, legal and regulatory requirements are different in different jurisdictions. For all three of the other answers, A, B, and C, these are things to keep in mind as needed but they won't have nearly as much impact on an information governance initiative as the legal and regulatory requirements.

© AIIM – 2019

The Information Governance Program

Information Governance Roles and Responsibilities

Governance is more than policies and procedures. In order to be effective, the governance model has to have the support of senior management, business unit managers, supervisors and line managers, end users, and many other specialized roles such as IT, records management, and legal. Accountability for good governance therefore has to exist at every level and within every department of the organization.

In this section we will identify the roles required to provide the governance structures and accountability required to support them. We should make two points at the outset. First, these roles do not necessarily equate to job titles or descriptions. Some roles will be held by many individuals, while other roles might be combined within one person.

Second, in general these roles will not be specific to any particular process or repository but will extend to governance of most or all types of information within the organization: other repositories, other information objects, and other instruments.

Program Owner

The program owner acts on behalf of the organization to ensure that the program aligns with the overall organizational strategy. The owner is personally accountable to the board and senior management for delivering the benefits of the program.

The owner will work with the project sponsors and program steering committee to determine which projects are needed to establish and/or further the program and to prioritize among them.

The owner owns the business case for the program. There are many different ways to develop and implement an ERM program; each of these may require different policies and procedures, different processes, different roles and responsibilities, and even different technical architectures depending on whether the program is being implemented from a business-driven approach, a customer-driven approach, or a risk-driven approach.

For many organizations, the program owner will be the CIO, general counsel, the head of the governance, risk, and compliance (GRC) function, or wherever organizational responsibility lies for governance and compliance.

Business Unit Manager

The business unit (BU) managers are responsible for their own departments, divisions, agencies, or other types of business units.

In the context of the IM program, the BU manager is responsible for communicating the benefits of effective information management to the business unit staff. The benefits will come from the business case initially; the BU manager will need to translate those benefits into something the employees can relate to.

The BU manager is also responsible to the organization for ensuring that employees of that BU comply with the requirements of the program. That means that the BU needs to have some way to measure and audit compliance and needs to put a plan into place to address any gaps that appear as the program continues.

Finally, the BU manager is responsible for ensuring that staff members are trained on the components of the program, including policies and procedures. That training might be developed and provided by HR, a third party, or someone within the BU, but it is important that it be delivered initially and that staff receive periodic refresher training.

Information Technology

Information technology, or IT, plays a significant role in the overall governance effort. We should note that IT here could refer to internal IT staff, whether centralized and enterprise-wide or decentralized at the business unit level, or it could refer to an outsourced IT staff.

IT installs, configures, and maintains the systems that make up the technology portion of the information management system. These systems include but are not limited to the applications that create documents and records, any archiving applications, the backups for each of those, and any IM or ERM solutions in place, including hardware and software.

IT also has to ensure that any solutions recommended for implementation, whether related to IM or any other technology-enabled process, fit within the overall organizational IT architecture. For example, some IM applications solutions only work with Oracle databases; this would likely meet with resistance if the organization relies on an otherwise Microsoft SQL Server-based IT architecture.

Records Management

The records management (RM) function is another key role for an IM program. First, it is often the RM function that creates the classification structures used to classify organizational information resources; those structures should be used or reused within the IM program – for example, by having matching folders created in users' inboxes, or by having IT integrate the office productivity suite with an IM repository that uses those classification structures.

Records management does the research to understand the unique records retention and compliance requirements of the organization. Records management provides input as to the value of that Information according to legal or regulatory requirements that must be complied with.

Records management drafts records management policies and procedures for the organization, including file plans, retention schedules, disposition instructions, and others.

Finally, records management should review the IM environment to ensure that it supports those retention requirements. By environment we mean people, processes, and technologies. For example, SEC 17a-4 requires that certain documentation be stored on non-erasable, non-rewritable storage. A records manager at a financial services firm in the U.S. would know that this particular requirement applies to the organization. The records manager may or may not know which technologies support this requirement, but could identify it *as* a requirement and have a discussion with IT as to how to satisfy it.

Legal

And finally we come to legal. The legal function is an important contributor to the governance of the IM program; indeed, many of the requirements for governance are related to legal issues including compliance and discovery.

The legal function is the most likely to receive notice of any requirements to hold and/or produce Information, including subpoenas, notices of intent to pursue legal action, legal holds, and the like.

The legal function will conduct or guide the discovery process should it be required. This discovery process includes imposition and release of legal holds; identification and retrieval of potentially relevant information; de-duplication of that information; review of that information for privilege, work product, and confidentiality; and production of relevant information to opposing counsel, inside or outside counsel, auditors, or some other requesting party.

Records management generally drafts records-related policies and procedures; however, it is generally legal (or in some instances senior management) that actually signs off on policies and procedures, thereby clearing the way for their implementation.

Governance Structures

In addition to the roles we've already described, there are a few other roles/structures that can significantly impact the success of an information governance initiative. These include:

- The steering committee
- One or more Centers of Excellence (CoE)
- One or more Communities of Practice (CoP)
- Coordinators

Now let's look at each of these in more detail.

The Steering Committee

Establishing a steering committee for an IM initiative can provide an effective assurance and advisory framework for the project. Generally, the steering committee is chaired by the program owner or a delegate. It is comprised of senior managers who have the responsibility and authority for complying with the initiative and who can make decisions about strategic direction.

© AIIM – 2019

The steering committee is designed to ensure the participation and support of senior departmental managers in the successful delivery and implementation of the project through the provision of policy and priority guidance, risk management, issue resolution, effective change management, and fiscal management and accountability.

Issues beyond the mandate of the steering committee are escalated to the program owner, who can escalate project or program-related issues to more senior management if necessary.

It is possible for an organization to have multiple steering committees for different topics if they are warranted.

The Center of Excellence

A center of excellence (CoE) is a governance structure or body that provides leadership, guidance, and best practices around a particular technology, process, or skill. They often serve as the repository of knowledge about a particular topic. As such, they are generally comprised of subject matter experts on that topic.

This is a formal governing body, and its members are assigned to it, ideally in a full-time or significant capacity.

Centers of excellence offer a number of benefits to the organization:

- They can evaluate new technologies, practices, and standards for their applicability to the organization.
- They can determine standards. For example, a SharePoint center of excellence might define what branding is acceptable within the organization's SharePoint environment, or develop a process for site provisioning. This helps to ensure consistency across the organization.
- Members are by definition experts in a topic and can serve as internal consultants on that topic.
- They can develop and deliver, or at least identify, training and reference works that address a particular topic.
- They can identify appropriate metrics and measurements relating to a particular topic and make specific recommendations for improvement.

An organization could have multiple centers of excellence for different topics.

The Community of Practice

A community of practice is similar to the center of excellence insofar as it is focused on sharing of knowledge. However, communities of practice tend to be much more organic in their development and growth. People generally join and participate in communities of practice because they want to, not because it's part of their job description.

The formality of a community of practice lies on a spectrum that depends on the interest of the participants and the availability of support from the organization. A given community of practice will generally last as long as it is useful and providing value to its members.

Like CoEs, communities of practice provide opportunities to share knowledge among practitioners – however, this sharing isn't limited to best practices, but often includes lessons learned and even horror stories. In the community of practice, some will learn through sharing, while others will learn more passively and simply listen to others' experiences. In their 2001 article, "Communities of Practice and organizational performance," Lesser and Storck argue that communities of practice can improve organizational performance in four ways:

- Decrease the learning curve of new employees
- Respond more rapidly to customer needs and inquiries
- Reduce rework and prevent "reinvention of the wheel"
- Spawn new ideas for products and services

Communities of practice also frequently identify issues that should be raised to the center of excellence (or even the steering committee) for resolution.

One example of a community of practice that may be familiar to many students is the local AIIM chapter, or other professional, trade, or interest-based organization. Participation is not required, but participants often gain significant benefits from networking with their peers, learning from them, sharing their own insights and stories, and encouraging mutual discussion and feedback.

An organization could have multiple communities of practice for different topics.

Coordinators

The last group we want to introduce consists of individuals who are subject matter experts in their own role, process, and department. They are salespeople, and clerks, and engineers, HR staff, IT staff, etc. They have also, however, been trained on records management, or ediscovery, or other information governance-related processes and act as a liaison between their particular work process or department and the broader information governance program. This means raising issues from their work area and sending them up to the records team as well as disseminating information from the team to their area.

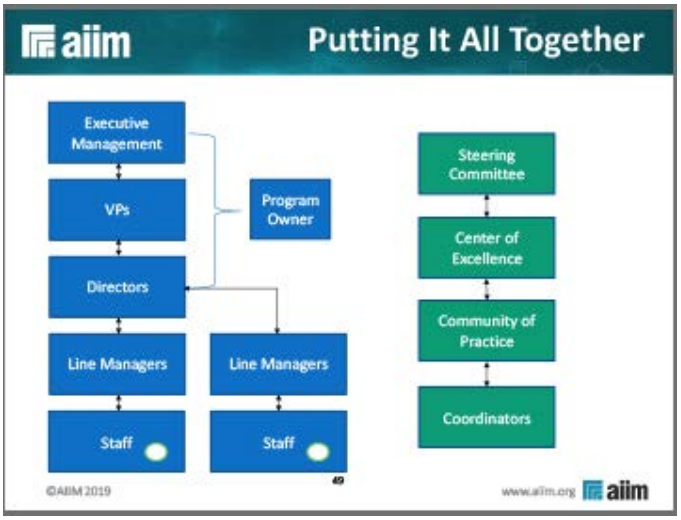
These roles might be called coordinators, or liaisons, or advisors, or any of a number of other titles, but their importance to the success of an information governance initiative cannot be overstated.

Coordinators can offer first-line support on minor issues, such as how to find particular information, whether something is a record or not, where to store something, etc. They are often part of the team to conduct a systems and information inventory. They may be delegated the responsibility to perform transfer or destruction tasks in accordance with the records program. And they are available as a resource to their team for any other information governance-related issues.

Putting It All Together

So what does all this look like? Here's an example of how these different groups and roles might interact on an ongoing basis in an organization with high information governance maturity.

On the left you see an abstract of a typical organization chart: executive management, VPs, directors, managers, and staff.

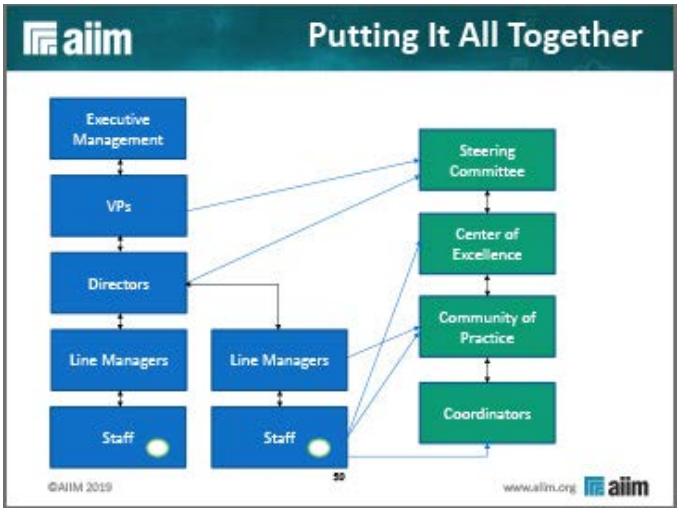


On the right, you see the governance structures we've outlined.

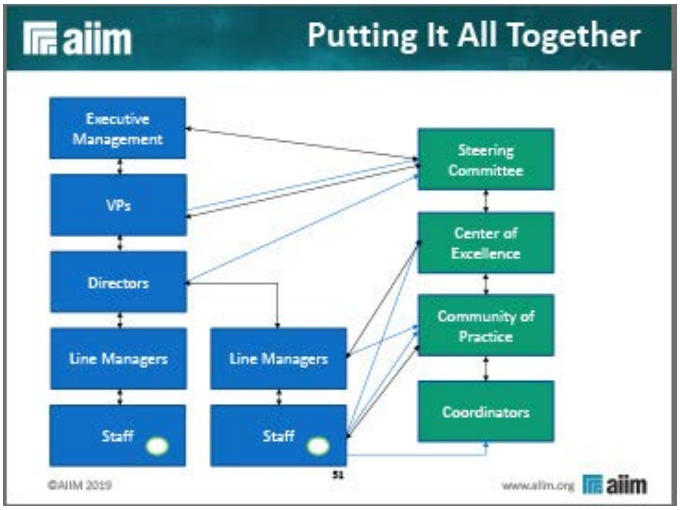
The potential program owner for the IG program could be the CIO, general counsel, or a VP or even a director depending on how the organization is structured. This diagram doesn't show records, IT, legal, etc. but they would certainly be present in the org chart and might also participate in the center of excellence and/or community of practice depending on their focus.

© AIIM – 2019

Here is the model again, with blue arrows added to denote participation.



We now add black arrows to the model to denote flows of information between the governance structures and the organizational structure.



Here's how it would typically work:

The steering committee meets regularly but infrequently (perhaps quarterly) to address issues raised by either executive and senior management or the center of excellence and communicates back to them as needed.

On an ongoing basis, the center of excellence raises issues to the steering committee and provides guidance to the community of practice and to the organization through line managers.

The community of practice meets regularly and frequently (perhaps monthly or even weekly depending on interest) to learn through sharing and raise issues. If the CoP cannot address those issues, they are raised to the CoE.

Coordinators (shown by the white circle within the Staff box in the org chart) participate in the CoE and raise issues there. On an ongoing basis they perform their primary function while also serving as the immediate point of contact for IG-related issues and questions.

Recommendations

While all of these governance structures contribute to effective take-up of information governance processes, there is a certain amount of prioritization that should be done. Of the four, arguably the most important is the steering committee. An effective steering committee can go a long way towards resolving longstanding issues around process and information silos and help to resolve conflicts between competing priorities. The next most important probably depends on your culture, but coordinators can go a long way to "operationalizing" effective information governance processes.

Do not force communities of practice. If there is a need, people will participate. Provide resources – a meeting room, the time to actually get together, a place for accessing white papers and presentations, etc. – but they work best when they work more organically.

Centers of excellence requires a certain amount of organizational maturity if they are to have the authority to set standards and guidelines and promote or enforce compliance with them. They also require the strong backing of management. If they do not have that, they cannot be effective and standards will be seen as guidelines and options at best.

And above all, governance requires communication. We discuss this in more detail in the Implementation section, but the better the communication – messages tailored to appropriate audiences, at appropriate times, using appropriate and engaging communications methods – the more likely the initiative is to be successful.

Evaluating the Existing Information Governance Strategy

Any new initiatives for Information Governance (IG) should start by identifying or reviewing what is already in place.

- Does the organization have any structures, roles and accountabilities for managing information?
- Are there any policies, objectives and strategies in place for managing information?
- What are the current information governance capabilities, understood in terms of resources and knowledge (e.g., capital, time, people, processes, systems and technologies)?
- What are the existing information systems, information flows and decision-making processes (both formal and informal)?
- What are the relationships with, and perceptions and values of, internal stakeholders and the organization's culture?
- Are there any standards, guidelines and models adopted by the organization for managing information?

If there is no existing IG program, the first step to starting one should be to get senior management support for an IG initiative. This should be focused not just on compliance, but also on improved business outcomes. We address the full process for developing an IG program elsewhere in this course.

To Gain Support for Information Governance Program

And what is then the secret for getting stakeholder support? The answer is communicate, communicate, communicate. This is not just telling them but listening to their priorities and concerns. Remember we have two ears and one mouth. Plan accordingly – 2/3 of the time should be spent listening, 1/3 of the time is spent explaining and asking questions.

Our job is now to engage stakeholders and demonstrate the importance of introducing or updating an Information Governance framework. They need to see and understand the benefits for the organization, but remember that you will get even more support if they see personal benefits. This could be better access to information, better support of mobile devices, etc.

© AIIM – 2019

Symptoms of Poor Information Governance

As you evaluate the existing IG program and all the components involved, it's also important to look at areas where the information governance seems to be lacking. For example:

- Information is kept beyond its usefulness to the organization. This increases costs and potential legal liabilities, and can significantly increase the risks associated with a data breach or other information loss or disclosure.
- Responses to inquiries take too long: from a customer service perspective, from an internal operational perspective, and even in terms of responses to legal or regulatory requests.
- The organization stores too much redundant, outdated, and trivial information. This could include things like personal files, a folder called "1999 Forecasts," a folder called "Bill's Files," and so forth.
- There is significant uncertainty as to whether a particular document is the correct version, the most up-to-date version, or whether it's a copy, etc.
- There is information or information-related systems that seem to have no specific owner – or in some cases multiple owners such that nobody takes responsibility for it.
- The use of personal devices – flash drives, smart phones, personal email – to access corporate systems is uncontrolled or ungoverned.

Reviewing the Information Governance Program

So why review the information governance program? As we've noted repeatedly, you can't manage what you don't measure. Operational measures are important in many contexts, but periodically senior management needs to evaluate the program to determine whether or not it's actually successful. This is for several reasons:

- Failure to follow the program can put the organization at significant legal risk for failing to comply with applicable laws and regulations.
- It will also increase organizational issues and risks.
- And it can cause a loss of reputation and trust if the organization is not following its policies and the adhering to the laws and regulations it needs to. This is a significant ethical dimension that too many organizations don't consider until something bad happens.

Indeed, all of these reasons support the need for a comprehensive information governance program in the first place.

Evaluating the Information Governance Program

Program performance evaluation – in other words, monitoring and auditing of the program, its objectives, policies, and processes – is a critical information governance program component.

The purposes of carrying out performance evaluation are to:

- Establish managerial responsibility and senior management oversight
- Provide tailored communications at all levels
- Monitor and audit program effectiveness
- Provide incentives and impose discipline
- Respond appropriately to violations
- Periodically assess risk and modify accordingly
- Ultimately ensure that the program is meeting the organization's needs – and senior management's expectations

Taking this to a much higher perspective, our focus is to establish ways to determine the effectiveness of all information governance elements.

For example, do the policies work, do people follow them, does this benefit the organization?

Performance evaluations will therefore help us identify incidents of non-compliance. They also help to ensure that any issues identified are addressed, and that appropriate corrective actions are taken and are effective in addressing those issues.

Information Governance Program Management Review

Senior management buy-in and support is absolutely critical for your information governance program!

Top management needs to review the organization's information governance program, at planned intervals, to ensure its continuing suitability, adequacy, and effectiveness.

The management review should consider:

- The status of actions from previous management reviews.
- How well is the program performing against defined targets, e.g., perceived information quality.
- The results of monitoring and audits.
- Changes in external and internal issues that are relevant to the program. This could be changes to business, new opportunities with the help of technologies, or new or modified regulations.
- Information on the performance of records processes and systems, including trends in nonconformities and corrective actions; monitoring and measurement evaluation results, and audit results.
- Opportunities for continual improvement.

The results of the review are the activities used to improve the information governance program and take advantage of continuous improvement opportunities.

Elements to Review

Here are some of the most important areas that require regular audits:

- Does staff follow our policies? Do they do as they are told?
- Are the policies correct? Do they create problems for staff or customers?
- Does staff even know what the policies are? Did the training and communication work?
- How effective are the systems that staff use? What works/doesn't work?
- How satisfied are staff with information access, information quality, information sharing, etc.?

And try to have a consistent reporting format over time to allow you to see trends and get a true picture of the effectiveness of the Information Governance program.

© AIIM – 2019



Test Your Knowledge

Question 1:

Which governance structure will have the greatest impact on the information governance strategy?

a. The steering committee
b. The center of excellence
c. The community of practice
d. The coordinators

Question 2:

Which elements should be reviewed when starting an information governance initiative? (select 2)

a. Existing information management-related tools and systems
b. Existing structures, roles and accountabilities for managing information
c. Existing folder structures on the network file shares
d. Existing information governance training presentations

Question 3:

What should senior management review to ensure proper oversight of the information governance program? (select 2)

a. The status of corrective actions from previous reviews
b. Any changes in internal and external issues that could impact the program
c. Help desk queries and technical support requests relating to the program
d. The business rules and business logic used to enforce information governance policies



Test Your Knowledge - Answers:

Answer to Question 1:

Which governance structure will have the greatest impact on the information governance strategy?

a. The steering committee
b. The center of excellence
c. The community of practice
d. The coordinators

The correct answer is A, the steering committee, which is made up of senior managers with the authority and accountability to make strategic decisions. B, the center of excellence, will support and implement the strategy. C, the community of practice, and D, the coordinators, will be the ones following the strategy.

Answer to Question 2:

Which elements should be reviewed when starting an information governance initiative? (select 2)

a. Existing information management-related tools and systems
b. Existing structures, roles and accountabilities for managing information
c. Existing folder structures on the network file shares
d. Existing information governance training presentations

The best answers here are A, existing tools and systems, and D, existing structures, roles and accountabilities. For B and C, folder structures and training should be reviewed, but this will happen later in the IG initiative.

Answer to Question 3:

What should senior management review to ensure proper oversight of the information governance program? (select 2)

a. The status of corrective actions from previous reviews
b. Any changes in internal and external issues that could impact the program
c. Help desk queries and technical support requests relating to the program
d. The business rules and business logic used to enforce information governance policies

The best answers here are A, status of corrective actions, and B, changes to internal or external issues that could impact the program. C, help desk queries, should be reviewed, but not necessarily by senior management. Similarly, D, business rules and logic are important and should be evaluated but again not by senior management.

© AIIM – 2019

Personal Data

Introduction to Personal Data

Let's start out with a basic definition of personal data. This comes from the European Union's General Data Protection Regulation (GDPR). Article 4 of the GDPR says that:

> 'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

We use the definitions in the GDPR frequently in this and other discussions of privacy because it's so broad and applicable to so many organizations. But this is not intended to be solely a discussion of GDPR; there are many other countries and jurisdictions that have or are contemplating creating privacy and data protection laws. You will see that many of them are quite similar, at least in how they define personal data.

Different jurisdictions refer to it by different names – for example in the U.S. Federal law refers to personally identifiable information, or PII, and personal health information, or PHI. But the idea is the same.

What is Personal Data?

As you can tell, this is a very broad definition – and additional recitals in the GDPR broaden it by including things like:

- Data from devices
- Data that can be combined to indirectly identify individuals
- Pseudonymized or anonymized data, where that data can be combined with other data to identify individuals

In fact, a major theme in privacy regulations and definitions is that it's not necessarily a single piece of data, but combinations that lead to data being personal. There are many people around the world named "John Smith" and many Starbucks coffee shops. But when you combine the name of a person, the name of a company, the location of a branch of that company on Tower Road in Denver, Colorado, and a picture of that person with the title, "Employee of the Month," you get a very specific identifiable individual.

This definition comes from the Personal Information Protection and Electronic Documents Act (PIPEDA) of Canada. Here is the official definition from the Office of the Privacy Commissioner of Canada:

> Under PIPEDA, personal information includes any factual or subjective information, recorded or not, about an identifiable individual. This includes information in any form, such as:
>
> - age, name, ID numbers, income, ethnic origin, or blood type;
> - opinions, evaluations, comments, social status, or disciplinary actions; and
> - employee files, credit records, loan records, medical records, existence of a dispute between a consumer and a merchant, intentions (for example, to acquire goods or services, or change jobs).

Sensitive Data

The GDPR goes further to identify what it calls "special categories of personal data." In Article 9, it notes that:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

These can be processed under certain circumstances, but clearly there could be significant ramifications to an individual whose data in these areas is breached.

In many jurisdictions financial data is also treated as sensitive data due to the harm that can be caused by its loss.

When is Personal Data Not Personal Data?

So when is personal data not personal data? The GDPR notes that properly anonymized data – that is, data with no personally identifiable elements – is by definition no longer personal data. Similarly, data can be pseudonymized – while the data is similar structurally, it employs alternate descriptions such as "Person 1." Proper pseudonymization is difficult, though, and even without an individual's name, it leaves open the possibility that enough data exists to combine into a specific person's identity.

Personal data that has been encrypted is still personal data, of course – but encrypted data can't be read and is generally considered to render the organization immune from the penalties associated with data breaches.

Finally, many types of personal data are only personal when they can be tied to a specific individual. So as we noted earlier, the fact that a person named "John Smith" exists in the world is not personal data in isolation. It's when we mean *that specific* John Smith that it needs to be properly safeguarded. A picture of a person with the face clearly shown is likely to be personal data; a face in a large crowd that isn't readily identifiable likely wouldn't be.

Strategies for Protecting Personal Data

The first thing we'll look at is how to approach data protection strategically.

Assess your risk. Before you can protect data, you need to know what you have and where it is. That means you need to conduct an assessment and an inventory. We discuss these in more detail elsewhere in this course.

Minimize the data you collect. Most data protection regulations recommend or require this anyway. If you don't collect it, you don't have to protect it or worry about it getting compromised.

Segregate that data. Don't put all personal data in one place – rather, keep just the data you need for a particular activity or function.

Review how you share information with third parties. This includes parties processing data on your behalf as well as suppliers, partners, etc. This also means ensuring that anyone you do share with manages the personal data appropriately.

Consider pseudonymizing or anonymizing data where possible. If done properly these can reduce or even eliminate the risks of a data breach.

Dispose of data. You should keep personal data as long as there is a business or legal reason to do so. Once that information is no longer needed, it should be disposed of in accordance with your records management program. Again, if you don't have it, you don't have to worry about it.

It's also important to think about what happens in the event that personal data is compromised. Depending on the nature of the data, the breach, and the regulatory environment, organizations may have a positive obligation to report breaches to customers and/or regulatory oversight entities.

Data Protection Practices

Next, we can look at how to operationalize data protection practices.

This starts with policies in each of the areas we just listed, along with accompanying procedures, guidance, etc. to tell employees how to comply with the policies. These policies might also address the use of personal data on mobile devices, websites, social media, and other third-party sites, etc.

Information-centric processes should be reviewed to evaluate potential data protection concerns and identify corrective action.

A common approach in information security is that of "least privilege" – that is, employees should only have access to the information they need to perform their jobs. This is especially important when it comes to personal, sensitive, or confidential types of information. This helps to minimize exposure of that data, whether inadvertent or intentional.

It's also important to monitor and review data protection-related processes and practices regularly. Many times, when breaches occur, they are much worse than they needed to be because the organization didn't know it had been breached. Moreover, regulations change, operational processes change, where an organization does business changes and these changes need to be taken into account in terms of how they impact the data protection program.

© AIIM – 2019



Test Your Knowledge

Question 1:

Which types of personal data would be considered sensitive and require even more strict protection from disclosure? (select 2)

a. IP address
b. Social media handle
c. Medical information
d. Financial account information

Question 2:

Why is it important to monitor and review data protection practices? (select 2)

a. To compare the cost of compliance to the cost of a data breach
b. To ensure that no breaches have occurred and that the program is effective
c. To take into account changes in the operational or regulatory environment
d. To determine whether data can be anonymized effectively



Test Your Knowledge - Answers:

Answer to Question 1:

Which types of personal data would be considered sensitive and require even more strict protection from disclosure? (select 2)

a. IP address
b. Social media handle
c. Medical information
d. Financial account information

The best answers are C, medical information, and D, financial account information. A and B are both personal data in many privacy regimes, but they are not generally considered to be sensitive information because their disclosure doesn't cause the same amount of harm that disclosing medical or financial information would.

Answer to Question 2:

Why is it important to monitor and review data protection practices? (select 2)

a. To compare the cost of compliance to the cost of a data breach
b. To ensure that no breaches have occurred and that the program is effective
c. To take into account changes in the operational or regulatory environment
d. To determine whether data can be anonymized effectively

The best answers here are B and C. For A, monitoring and reviewing data protection practices won't really provide those costs. For D, monitoring would not support that determination.



© AIIM – 2019

The Privacy and Data Protection Strategy and Assessment

The Privacy Assessment

In this section we offer several different assessments. The first set of assessments include the business, technical, and data processing assessments. These help to determine the starting point for any privacy-related initiatives and serve as a baseline.

Building on this, the Privacy Maturity Model provides a framework of industry-recognized best practices against which an organization can assess its current practices, identify gaps, and build a roadmap for improvement.

Finally, a privacy impact assessment should be undertaken any time new processes or technologies are implemented which would impact privacy and data protection requirements.

Let's look at each of these in a bit more detail.

The key starting point for a privacy initiative is to assess the current state of the organization. That is, it's important to know what the organization does, and what it has.

Elsewhere in this course we describe a general approach to business and technical assessments for an information management initiative. We outlined a number of specific elements to include. As a reminder, the business assessment includes things like:

- Business drivers – what is the purpose for the initiative?
- A maturity model and assessment – we'll look at a privacy maturity model shortly.
- A review of the regulatory environment.
- An inventory and subsequent review of the processes within the organization, particularly those that create, receive, or use data that would fall under the privacy or data protection umbrella.

The technical assessment includes things like:

- An inventory of all systems within the organization and identification of the data created, stored, and managed therein.
- An inventory of how information flows into, through, and out of the organization.

The privacy compliance research company Nymity suggests that there is another assessment to be done – how data is processed within an organization. In a recent white paper, "Does GDPR Article 30 Require a Data Inventory?," they argue that, while traditional data inventories such as described above are good, they are challenging – and not sufficient for the GDPR. Instead, they suggest the creation of a data processing inventory, which would identify how and why information is processed by a data processor. This would also include identifying all third-party processors of personal data and ensuring that they adhere to the same processes and practices required of the organization.

Privacy-related Inventories

We discuss the systems, information, and process elsewhere in the course, but in brief, the intent is to identify all of the systems in the organization, identify the information they store, manage, or otherwise interact with, and identify the processes by which information flows into, across, and out of the organization.

This also means distinguishing between corporate and personal data as corporate data such as a corporate address is not generally required to be treated the same as personal data.

Privacy Maturity Model

Maturity models offer a useful mechanism for assessing practices within an organization. In the privacy field, one such model was developed by the American Institute of Certified Public Accountants (AICPA) and the Canadian Institute of Chartered Accountants (CICA), with additional contributions from ISACA, an international professional association focused on IT governance.

The Privacy Maturity Model is based on the Generally Accepted Privacy Principles developed by AICPA and CICA. The ten principles are:

- **Management.** The entity defines, documents, communicates and assigns accountability for its privacy policies and procedures.
- **Notice.** The entity provides notice about its privacy policies and procedures and identifies the purposes for which personal information is collected, used, retained and disclosed.
- **Choice and consent.** The entity describes the choices available to the individual and obtains implicit or explicit consent with respect to the collection, use and disclosure of personal information.
- **Collection.** The entity collects personal information only for the purposes identified in the notice.
- **Use, retention, and disposal.** The entity limits the use of personal information to the purposes identified in the notice and for which the individual has provided implicit or explicit consent. The entity retains personal information for only as long as necessary to fulfill the stated purposes or as required by law or regulations and thereafter appropriately disposes of such information.
- **Access.** The entity provides individuals with access to their personal information for review and update.
- **Disclosure to third parties.** The entity discloses personal information to third parties only for the purposes identified in the notice and with the implicit or explicit consent of the individual.
- **Security for privacy.** The entity protects personal information against unauthorized access, both physical and logical.
- **Quality.** The entity maintains accurate, complete and relevant personal information for the purposes identified in the notice.
- **Monitoring and enforcement.** The entity monitors compliance with its privacy policies and procedures and has procedures to address privacy related complaints and disputes.

The Privacy Maturity Model consists of five levels, similar to other maturity models:

- ad hoc
- repeatable
- defined
- managed
- optimized

There are a total of 73 GAPP criteria, grouped into the 10 GAPP principles from the above section. Each criteria can be measured against these five levels. So for example, in Management, criterion 1.1.0 deals with privacy policies, which are described as follows: “The entity defines and documents its privacy policies with respect to notice; choice and consent; collection; use, retention and disposal; access; disclosure to third parties; security for privacy; quality; and monitoring and enforcement.”

Each of the five levels provides a proscriptive statement that describes the extent to which the entity complies with the criterion. This allows the organization to evaluate its privacy and data protection practices thoroughly against industry-defined best practices and identify areas to focus on for improvement.

Privacy Impact Assessment

Another assessment tool is the privacy impact assessment, or PIA. The Privacy Assessment Impact Framework project defines the PIA as “a systematic process for evaluating the potential effects on privacy of a project, initiative or proposed system or scheme and finding ways to mitigate or avoid any adverse effects.”

© AIIM – 2019

According to the PIAF, the concept of a PIA has been around since the mid-1990s. The U.S. Federal E-Government Act of 2002 requires "that all federal agencies conduct a "privacy impact assessment" (PIA) for all new or substantially changed technology that collects, maintains, or disseminates personally identifiable information (PII), or for a new aggregation of information that is collected, maintained, or disseminated using information technology."

Similarly, Article 35 of the European Union's General Data Protection Regulation, or GDPR, requires a data protection impact assessment. "Where a type of processing, in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data."

This assessment should include at least:

- A systematic description of the operations
- The purpose for the processing
- An assessment of the necessity and proportionality of the processing
- An assessment of the risks to the rights and freedoms of the data subjects
- The measures that will ensure the protection of personal data, taking into account the rights and legitimate interests of data subjects and other persons concerned

While this description comes from the GDPR, and the previous one from the U.S. Federal government, these are broadly applicable to any organization and for any process that involves the handling of personal data.

Privacy by Design

Wikipedia defines privacy by design as an approach to systems engineering which takes privacy into account throughout the engineering process, rather than attempting to add privacy protections on at the end of the process. It comes from a report issued in 1995 by the Information and Privacy Commissioner of Ontario, Canada, the Dutch Data Protection Authority, and the Netherlands Organisation for Applied Scientific Research. It has been incorporated into the General Data Protection Regulation (GDPR) in Article 25.

As with security and other elements, the intent is that privacy be considered as systems, processes, etc. are designed, rather than being added at the end or as an afterthought.

Privacy by design is more than just software and systems engineering; it includes legal and regulatory elements, process considerations, and other elements of the organization's governance framework.

The primary author of this approach, Dr. Ann Cavoukian, the Information and Privacy Commissioner of Canada, developed a reference guideline that further describes the 7 principles in the context of fair information practices.

- Proactive, not reactive; preventative, not remedial. Systems should be designed to anticipate privacy invasive events before they happen and prevent them.
- Privacy as the default – private data should be protected automatically, and with specific safeguards such as minimizing the amount of data collected and minimizing how it is shared.
- Privacy embedded into design – as an essential component of the design, not an add-on.
- Full functionality, not zero-sum. Privacy should not come at the expense of the goals of the process or solution – or vice versa.
- End-to-end security – lifecycle protection. Privacy and security should be in place all the way through destruction of the information where applicable.
- Visibility and transparency. This includes accountability, openness, and compliance. Policies should be human-readable, not filled with 50 pages of legal terms and definitions.
- Respect for user privacy. The system should be designed from the user's perspective and should empower the user to make fully-informed choices and consent.

Privacy by Design Applications

Privacy by design has traditionally been thought of in the context of technology and IT systems design and engineering – implementation of encryption, severing or masking personal identifiers, access controls, and the like.

But it can also be extended to what Dr. Cavoukian refers to as accountable business practices. This is the idea that businesses that take their customers' – and employees' – privacy and the protection of personal and sensitive data seriously will be more trusted. This in turn will lead ultimately to more customers, more revenue, more profit, etc. It actually becomes a competitive advantage to take privacy seriously.

And privacy by design extends into the physical design of workspaces. As an example, many hospitals, pharmacies, and medical clinics require very personal and sensitive information from their patients and customers. One change many of them have made is to have some distance between the clerks and the patient they are interacting with, and the rest of the patients waiting to check in, so that those conversations take place out of earshot of other patients. Another example she cites is the practice of ensuring that file cabinets or file rooms that hold personal or sensitive data remain locked when not in use.

Privacy by Design in Practice

What this section really comes down to is that organizations should protect personal and sensitive data proactively. That means assessing existing privacy practices, privacy safeguards in solutions, and the other considerations we've already mentioned here and ensuring that privacy is considered up front rather than as an afterthought – or not at all.

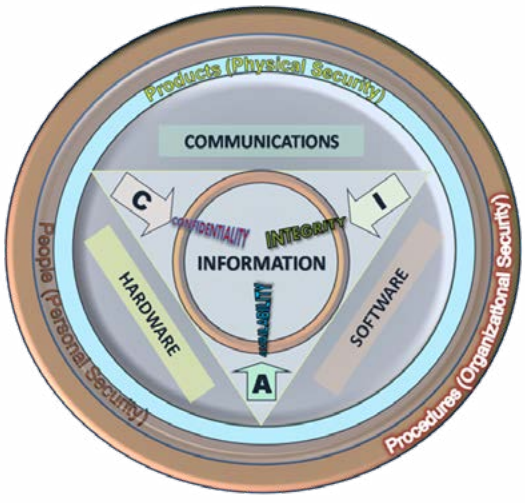
Organizations should review the seven Principles listed earlier, and every person involved in privacy in the organization should consider how they would want THEIR personal and sensitive data collected, managed, used, shared, and protected.

Information Management Security Tools

Information security is the practice of defending information from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording, or destruction.

There are many models for information security, but a common model is the CIA circle. CIA stands here for:

- **Confidentiality** – prevent disclosure of information to unauthorized individuals or systems
- **Integrity** – maintain and assure the accuracy and consistency of information over its entire lifecycle
- **Availability** – ensure it is available when it is needed



The CIA Circle: Confidentiality, Integrity, and Availability
Source: Wikipedia.org – Information Security

© AIIM – 2019

Role-based Security

In role-based security, a role is a business or system function undertaken by one or more individuals in an organization. It may also be a group of individuals requiring the same access rights to perform some business functions. Role-based security allows access rights to be assigned to users based on the roles they perform in their business function, rather than on their individual identities. In this model, access rights are assigned to a group or role and NOT to an individual. Individuals are then assigned to the group or role and *inherit* the access rights of the *role*. A user can belong to many groups of course.

The key reason for bringing this up is that while roles-based security is often set up by IT, a system administrator, or a related role, it is the business process owners that know what roles need to be set up for a particular system or process and what those roles need to be able to do in order to perform their jobs.

Annotation

Annotations and redactions can be defined as additional information to be provided for, and in conjunction with, a particular information object. In the physical world, these might be sticky notes, scribbled notes in the margins, or even other documents stapled, taped, or clipped to the original. The additions are linked to the original, but do not generally make permanent changes to it. Where notes or redaction are involved, they might be done on a duplicate copy (and we will discuss this in further detail shortly).

In the digital world, we have a similar requirement for functionality: the annotation or redaction must be linked to the original document in a way that can be preserved, and it must not make changes to the original document.

Redaction

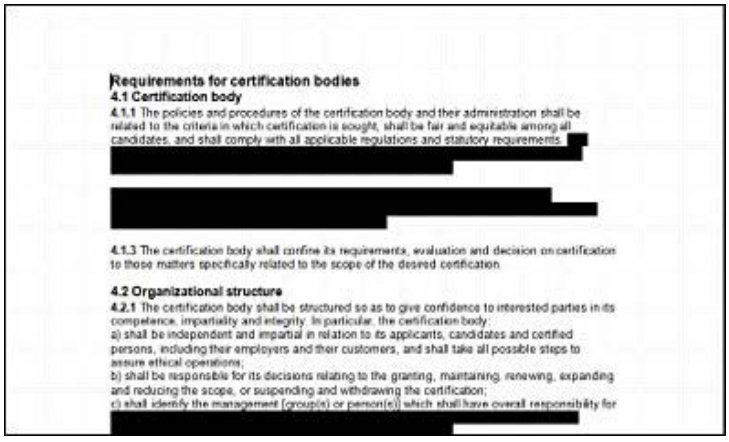
Administrators sometimes need to publish, or make available, content which contains information which is still sensitive. The information could be sensitive for any number of reasons – it is classified information, it contains personally identifiable health or financial information, etc. However, it is still required to be produced, for example to comply with a Freedom of Information Act type of request or as part of a legal action.

For this reason, administrators need to be able to first remove the sensitive information, without affecting the underlying content. This process is referred to as redaction.

Redaction Concerns

Redaction can be a powerful security tool if done properly. The problem is that it's very easy to do it *improperly.*

Consider the graphic shown here. The redactions show up as black lines that cover the text underneath. You could make a similar looking image using a highlighter set to black. You could turn some of the black text that's still visible to white. In both cases, however, the most cursory review will allow a curious user to see what's under the so-called redaction. Similarly, sometimes redaction is done by creating large black, white, or other box shapes over the text to be blocked. But again, this is not really redaction and the information underneath the shape is still there.



Requirements for certification bodies

4.1 Certification body

4.1.1 The policies and procedures of the certification body and their administration shall be related to the criteria in which certification is sought, shall be fair and equitable among all candidates, and shall comply with all applicable regulations and statutory requirements.

[redacted]

4.1.3 The certification body shall confine its requirements, evaluation and decision on certification to those matters specifically related to the scope of the desired certification.

4.2 Organizational structure

4.2.1 The certification body shall be structured so as to give confidence to interested parties in its competence, impartiality and integrity. In particular, the certification body:

a) shall be independent and impartial in relation to its applicants, candidates and certified persons, including their employers and their customers, and shall take all possible steps to assure ethical operations;

b) shall be responsible for its decisions relating to the granting, maintaining, renewing, expanding and reducing the scope, or suspending and withdrawing the certification;

c) shall identify the management (group(s) or person(s)) which shall have overall responsibility for

[redacted]

The appropriate way to redact is to use redaction tools – and be trained on their proper usage. The tools themselves are very mature, and generally solid in ensuring that there is no way to access the underlying text.

It's also important to ensure that there is auditability and traceability for any redactions and any changes to them. It is common for the party receiving redacted documents to challenge the redactions in kind and in amount. The organization needs to be able to demonstrate what was provided to the requestor and the specific reasons for any redactions made.

Digital Rights Management

Digital rights management, or DRM, is a term referring to any of several technical methods used to control or restrict the use of digital information on digital devices with such technologies installed. The media most often restricted by DRM techniques include music, visual artwork, movies, and printed material that needs to be restricted in circulation.

DRM works as a set of technical measures that can prevent protected information from being shared, transmitted, saved, printed, etc. These technologies are invariably proprietary and based on encryption. If an attempt is made to use protected information in a way that is not authorized, the information remains encrypted and secure. This sounds good – for example, an organization could use DRM to limit the access to any information it shares with third parties. Emails could be made to expire after 30 days. Documents could be sent in a way that doesn't allow them to be printed. And so forth.

However, DRM raises a number of issues.

- Because it is proprietary and based on encryption, if the system is compromised, often the decryption mechanism goes away, rendering all DRM-protected information permanently encrypted and therefore inaccessible.
- DRM is often tied to particular hardware and/or software environments. This makes it difficult to share protected information with third parties such as customers, suppliers, opposing legal counsel, or regulatory agencies.
- For the same reason, it may be difficult or impossible to migrate protected information from one system, format, or storage media to another.

DRM can be useful for information with a relatively short usable lifecycle, but may present significant issues to information that needs to be accessed by third parties or over extended periods of time.

Encryption

Encryption is the process of applying a complex transformation to a digital document so that it cannot be re-rendered in an application in a readable form unless the corresponding decrypting transformation is applied. Encryption can be applied at three points: on a document in motion, on a document in a container, and at a document at rest.

- **Documents in motion.** When documents are sent as attachments, a copy of the document is made, attached to the communications message, and transmitted. The original document remains in its original location. In this case, the attached copy can be encrypted as can the message if supported – it doesn't matter because the original remains unchanged.
- **Documents in containers.** Containers here refer to any sort of repository. Almost all information management systems today provide for encrypting the data containers. That is, everything that is in the system is encrypted by default. So in this case there is no need to encrypt the original. Many, but not all, information management systems *can* ingest encrypted documents, but it can cause issues including not being able to access it over time.
- **Documents at rest.** These are documents not stored in repositories, but rather on flash drives, your computer's desktop, networked file shares, etc. Encrypting these documents is dangerous because if, or perhaps when, the password is lost, the documents will be permanently encrypted. There are many reasons why it is better to store information in repositories; this is just one more – and a very important one at that.

For documents in containers and at rest that get encrypted, encryption might be OK in the short term, but it presents significant long-term access issues. Consider: how will you ensure the documents can be decrypted, say, 50 years from now? How do you pass down the passwords without their being compromised? Can you really believe that decryption mechanisms will be the same 50 years from now?

© AIIM – 2019



Test Your Knowledge

Question 1:

Which assessments should be conducted at the beginning of a privacy-related initiative? (select 2)

a. The functional requirements assessment
b. The information inventory
c. The business assessment
d. The training needs analysis

Question 2:

What is the key outcome associated with Privacy by Design?

a. Privacy is embedded into systems and business practices.
b. Privacy is added to core systems functionality before they are rolled out to everyone.
c. Personal data is only shared with others inside the organization.
d. Personal data collection practices are designed for effective anonymization.

Question 3:

Who should determine the roles and access levels required for a particular function?

a. Information technology
b. Legal staff
c. Business owners
d. Information security



Test Your Knowledge - Answers:

Answer to Question 1:

Which assessments should be conducted at the beginning of a privacy-related initiative? (select 2)

a. The functional requirements assessment
b. The information inventory
c. The business assessment
d. The training needs analysis

The best answers here are B, the information inventory, and C, the business assessment. A, the functional requirements assessment, would be handled much later in the process, as would D, the training needs analysis.

Answer to Question 2:

What is the key outcome associated with Privacy by Design?

a. Privacy is embedded into systems and business practices.
b. Privacy is added to core systems functionality before they are rolled out to everyone.
c. Personal data is only shared with others inside the organization.
d. Personal data collection practices are designed for effective anonymization.

The best answer here is A, privacy is embedded into systems and business practices. B is the traditional “bolt on privacy as an afterthought” approach that serves organizations very poorly. C, limiting the sharing of personal data to those inside the organization is not material to privacy by design. For D, anonymization can be one technique that supports privacy by design, but it’s only one of many considerations.

Answer to Question 3:

Who should determine the roles and access levels required for a particular function?

a. Information technology
b. Legal staff
c. Business owners
d. Information security

The correct answer is C, business owners. For A and D, IT or information security might be the ones to set up access levels in a particular system, but they would not determine the roles and access levels required. B is simply incorrect.

© AIIM – 2019

Privacy and the Information Management Program

A good information management program supports good privacy practices. As with most programs, we can review this in the context of people, processes, and technologies. We should also note that whether we're talking about information management or privacy management, technology can help the program by providing tools, allowing for automation, etc. but it is no magic solution by itself.

People and Privacy

The first area to consider is people – that is, how do we get people to do effective information management in support of the privacy program? It starts with answering the question, "What's in it for me?" or WIIFM (say: whiff-em). Different groups will have different goals and objectives. While all of these should roll up to the organization's goals and objectives, there will be different focuses and tasks involved.

Next, the organization needs to put effective information governance into place. This includes but is definitely not limited to:

- Policies. There need to be effective, consistent, understandable policies in place and employees need to be trained on them and held accountable for following them.
- Procedures. Employees will also need more in-depth procedures and supporting resources like job aids, guidelines, and checklists to ensure they are complying with the policy.
- Controls. There should be controls put in place that describe how tasks will be accomplished, and audits to ensure they are being accomplished effectively.

And of course employees throughout the organization and at all levels will need training. That training will differ depending on role, level in the organization, etc. but it needs to happen and it needs to be refreshed periodically.

Privacy and Information Management Processes

Throughout this course we have identified a number of information management-related processes that also directly support effective privacy practices. These include:

- Inventorying the organization's systems, information holdings, and processes. If you don't know what you have, it's very difficult to manage it effectively.
- Capturing information into a repository so it can be managed effectively including being able to find it, access it, and trust it.
- Retention and disposition. It has always been a best practice to keep information as long as it has business or regulatory value, and then get rid of it in accordance with the organization's records program. Similarly, a number of data protection regulations around the world require organizations to keep personal data only as long as required to meet business needs and then get rid of it.

Privacy-enhancing Technologies

And here are some of the technologies that can help to support privacy and data protection practices. We discuss many of these in more detail elsewhere in this course.

Access controls. Where possible, this should be centralized using a single sign-on (SSO) approach, such as Microsoft Active Directory. This helps to ensure consistency within the access control scheme. It also makes it easier to grant, and revoke, access when a user leaves the organization or changes roles within it. Access controls also allow for more granular control, so that access to sensitive information can be limited to only those with appropriate need.

Encryption. Encryption of information can significantly improve its security and prevent unauthorized disclosure in the event of a breach. As we discuss elsewhere in this course, it does present some challenges, especially to long-term access. It is also important that this be a centralized and standardized service, rather than allowing everyone in the organization to encrypt however they see fit.

Redaction. This allows the organization to make available information required, for example, under Freedom of Information Act-type laws and regulations, but to hide from disclosure sensitive information in accordance with that Act. Redaction can be problematic if not done correctly.

Repository. One of the key benefits of using a digital recordkeeping system or other content repository is that the repository provides many of these capabilities – access controls, encryption at rest, redaction, audit trails, etc. Records should always be stored in some sort of a repository to ensure that they can be protected and managed according to the organization's legal, regulatory, and operational requirements.

Monitoring and data loss or leakage prevention. These tools can review emails and other types of information and information stores to detect and prevent transmission of particular types of content to unauthorized persons inside or outside the organization. For example, these tools could determine that an email attachment is actually a spreadsheet that contains credit card numbers. Once this has been detected, the tool could block transmission, notify a manager, or even set up an incident to track the attempt.

Privacy and Architecture

Architectural considerations are not technologies in themselves, but rather a way of architecting information management solutions to take into account performance, security, and privacy and data protection considerations.

For example, an organization has offices in the U.S., Canada, and Germany. Each of those three countries has its own data protection requirements. So when the organization acquires an information management solution, it might need to architect it such that there are separate data stores in each country, potentially with different access controls and business logic to meet those particular requirements. This could be considered part of a much larger concept called "privacy by design."

Privacy and Automation

Automation can significantly streamline and speed up processes of all types and throughout the organization. One of the ways it does this is to minimize human errors, whether this is because something was done incorrectly or because it wasn't done at all and should have been. Automation also allows for setting conditions to ensure things happen the way they are supposed to. Similarly, many of the issues associated with data breaches are the result of improperly set security, or phishing, or other issues associated with human error. This isn't to say that humans don't need to be involved, and supervise, and audit, and so forth. But to the extent that the opportunity for human errors is removed, this is a good thing for information management and for privacy.

Privacy and Retention

There are many ways for privacy and data protection issues to arise. But they are often much worse than they need to be because of poor records management practices. So one of the key ways records managers can contribute to effective privacy is through the records retention schedule.

As discussed earlier in the course, the retention schedule provides a listing of the records the organization owns along with instructions as to how long to keep them and what to do with them at the end of that period. Retention periods and disposition instructions are based on a number of factors including legal and regulatory requirements, operational needs, and the value of the records to the organization.

Periodically the records manager should review the retention schedule to ensure it meets current legal, regulatory, and operational requirements. At the same time, managers should not be keeping records beyond the stated retention periods "just in case" – this can significantly increase costs, potential liabilities, and in the event of a data breach, the number of records that are exposed. If there is a legitimate business reason for keeping records longer than the statutory minimum, those requirements should be included in the retention schedule.

At the same time, in some jurisdictions there are mandatory maximum retention periods, after which relevant information must be destroyed. Different jurisdictions take different approaches to this question when applied to records or archives. The key point here is that there needs to be regular, ongoing communication between the records program and the privacy and data protection program.

© AIIM – 2019

Data Sovereignty

The notion of data sovereignty is that data that is collected is subject to either the laws of the jurisdiction in which it is collected, or the jurisdiction where its targets are located, or both. This is an important topic given global companies and their thirst for information about their customers and users, and the surge in privacy-focused legislation designed to give users control over their personal data.

Consider the following examples:

- Under the USA PATRIOT ACT, the U.S. Federal government is allowed to access any information physically located within, or accessible from, the United States. This has resulted in many countries developing or updating guidance for how personal data is stored to disallow it from being stored in the U.S. At the same time, the U.S. Federal government requires that Federal data be stored in the U.S. as do Germany, France, and some other countries.
- In 2013, the U.S. Department of Justice asked Microsoft to grant access to a Hotmail email account for a user based in Ireland. Microsoft refused, citing European data protection regulations. The Department of Justice sued on the grounds that Microsoft, an American company, clearly had access to the data from within the U.S. and won the case. In 2016 Microsoft sued the U.S. government on other grounds and the case was ultimately settled with a change to Federal policy about these types of searches.
- Facebook, Twitter, and companies like them have data centers all over the world and need to comply with data protection regulations in every jurisdiction in which they do business. Similarly, nearly every cloud provider has the same approach and faces the same issues.

Location, Location, Location

The first elements to determine are where your data is stored and where your data is collected from. This will help to determine your specific regulatory requirements. For example, under the European Union's General Data Protection Regulation, or GDPR, their data protection rules extend to any organization processing the personal data of "data subjects" inside the EU, and to any organization established in the EU regardless of what data they collect. So if you market your goods or services to individuals in the EU, you are likely subject to the GDPR's regulations.

Data Sharing

Next, do you share data with third parties such as customers, suppliers, or regulators? If so, you need to be aware of the data protection requirements where you are and where those third parties are.

Similarly, do you use outside processors such as data collection services, cloud-based applications or services, apps, etc.? The same concerns would apply.

Policies

If you do share outside the organization, you need to have policies that address that because most data protection regulations put the onus for compliance on the organization collecting the data irrespective of how that data is collected and processed. In other words, you can't blame a breach or privacy incident on your third-party processors or organizations you shared with – it's your responsibility. That means you need to make sure that they can comply with the same data protection requirements.

This also means reviewing your business rules around everything from website cookies, to privacy statements, to opt-in forms, to any specific workflows or process automation and integration that may be in place.

From a governance framework perspective, organizations need consistency, but also need to comply with the specific regulatory requirements in the jurisdictions in which they operate. That might mean that the organization should take a decentralized approach to governance, wherein the main governance framework is developed at the organization's headquarters or home office and individual tweaks or changes are made in each jurisdiction.

The alternative is to pick the most restrictive regulatory regime and ensure that all policies and activities adhere to that. That may not be feasible, however, depending on the specific jurisdictions because of the potential for contradictory requirements. Regardless, the organization needs to spend some time and effort reviewing its policies and its overall governance framework.

Architecture Considerations

Finally, one of the things to consider if you're a multinational organization or need to share outside your organization is how to architect your technology solutions such as email servers, collaborative tools, or other information management solutions to ensure that they comply with applicable data protection requirements. Different privacy and data protection requirements can significantly impact the design and implementation of these systems. This will generally require a distributed architecture; this also means that IT needs to know the specifics around where particular types of data, especially personal data, are stored.

A distributed architecture could involve having completely separate and distinct solutions in each geographic jurisdiction that are not connected to each other at all, so there is no way for data to be shared. For example, the organization could implement individual enterprise content management solutions in each office.

It could also be set up as a single distributed or federated solution as part of a farm of sorts. Individual servers are set up in each jurisdiction, but overall governance and rules frameworks are centralized.

And it could be set up as a cloud-based architecture – public, private, hybrid, etc. – with due diligence performed to ensure that the applicable data centers meet the organization's regulatory requirements.

Responding to a Data Breach

Sometimes it seems like there is a new data breach every day. A data breach occurs when an unauthorized entity gets access to personal or other sensitive data held by an organization. This could be because the organization was hacked; often times it occurs through human error or a misconfigured system; sometimes it is even done by disgruntled employees or former employees who retained access to enterprise systems.

Regardless of the reason, and regardless of the underlying purpose, a data breach can cause immense harm to an organization:

- **Financial or legal liability.** In many jurisdictions, failure to properly safeguard personal data can result in fines, lawsuits, or other legal ramifications.
- **Loss of trust in the organization,** which in turn could lead to loss of customers and business.
- **Operational impacts,** including the time required to respond to the breach, notify regulators or those affected, etc. as well as the potential for loss of data that may then need to be recreated.

The Data Breach Process

The data breach response process needs to include these steps. The order is not as important as that each step is followed.

- **Discover the breach.** This sounds obvious, but if the organization isn't regularly reviewing its security logs, its access controls, etc. a breach could go on for some time.
- **Once the breach is identified, it needs to be contained** to prevent further loss of data from the breach. This will depend substantially on the nature of the breach and could involve anything from changing access accounts and passwords, to implementing a patch or hotfix, to completely shutting the system down.
- **Assess the risk.** This should include identifying what was compromised; the number of accounts or amount of data potentially affected; what kind of harm the breach could lead to; and how to minimize that harm.
- **Notify regulators and affected individuals.** Different jurisdictions have different requirements for who has to be notified under what circumstances and in what timeframe, but these requirements have become significantly more stringent in the wake of so many high-profile breaches of massive amounts of personal data. For example, under the European Union General Data Protection Regulation (GDPR), many breaches require notification within 72 hours of discovery of the breach. This needs to be coordinated with legal, customer communications, risk management, etc. Organizations will often set up an incident or breach response team that includes all of these stakeholders to ensure everything is properly addressed.
- **Post-mortem analysis**. This starts with the actual incident itself – how did it happen, how can future events be prevented, how can the potential harm of a breach be mitigated, etc. It also includes the response process – what went well, what went poorly.

© AIIM – 2019



Test Your Knowledge

Question 1:

Which information management process most directly supports an effective privacy program?

a. eDiscovery
b. Retention and disposition
c. Contract management
d. Version control

Question 2:

Why is data sovereignty important in the context of privacy and data protection? (select 2)

a. Compliance costs are different in different jurisdictions.
b. It can impact how technology solutions are designed and implemented.
c. Organizations have to comply with the laws and regulations where they operate.
d. It prohibits the use of cloud-based services for processing of personal data.

Question 3:

Who should notify regulators about a data breach?

a. The person who discovered the breach
b. The information security group
c. The incident response team
d. Any employee who knows about the breach



Test Your Knowledge - Answers:

Answer to Question 1:

Which information management process most directly supports an effective privacy program?

a. eDiscovery
b. Retention and disposition
c. Contract management
d. Version control

The correct answer here is B, retention and disposition, because it is a privacy best practice to get rid of information when it is no longer needed. For A, eDiscovery may be affected by an effective privacy program or the lack thereof, but it doesn't support it the way retention and disposition do. C and D are not directly relevant to the privacy program.

Answer to Question 2:

Why is data sovereignty important in the context of privacy and data protection? (select 2)

a. Compliance costs are different in different jurisdictions.
b. It can impact how technology solutions are designed and implemented.
c. Organizations have to comply with the laws and regulations where they operate.
d. It prohibits the use of cloud-based services for processing of personal data.

The best answers here are B, it can impact the solution design, and C, organizations have to comply with the laws and regulations in every jurisdiction in which they operate. For A, compliance costs may be different in different jurisdictions, but this is not directly related to the idea of data sovereignty. For D, data sovereignty considerations need to be considered when using cloud-based services, but they don't prohibit their use per se.

Answer to Question 3:

Who should notify regulators about a data breach?

a. The person who discovered the breach
b. The information security group
c. The incident response team
d. Any employee who knows about the breach

The correct answer here is C, the incident response team because it pulls from all the stakeholders to ensure the response is legally and technically accurate. A sometimes occurs, especially if that person has reported it to the organization and received an insufficient response, but the organization has no control over that. For B or D to report it would likely result in incorrect or incomplete information being transmitted and could cause additional liability for the organization.

© AIIM – 2019

Introduction to Records Management

ISO 15489, the international standard on records management, defines records as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business."

The first point to note is the word "information." Note this word is *not* document, and most definitely not *paper* document. A record could, in principle, be in any form or format we can think of, so long as it conveys information. This means that electronic, paper, and other physical records are clearly included. So are audio, video, instant messages, and so on.

The word "maintained" is quite significant here. It indicates that it is not enough to 'capture' records. They have to be stored, and managed properly once stored. This includes disposing of them when they are no longer needed.

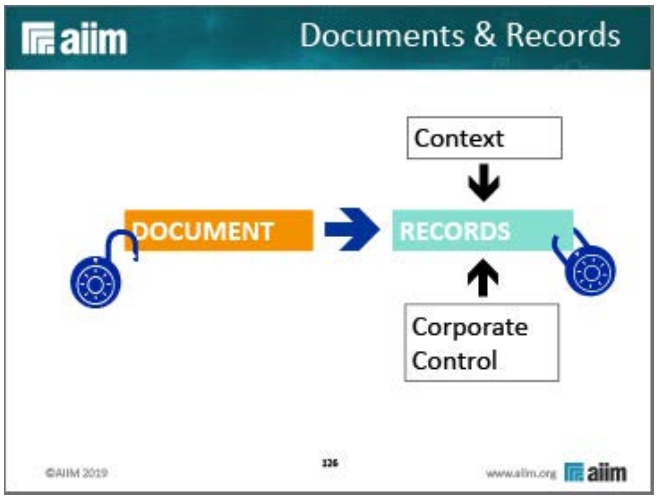
The idea of "evidence" is more difficult. For a record to be good evidence, say in a court case, there must be no doubt that it is complete and unchanged.

Finally, "pursuance of legal obligations or in the transaction of business" indicates the two reasons for which records need to be kept.

Documents and Records

All records are documents, but not all documents are records. Here is what happens when a document does become a record:

A document, by definition, is not necessarily controlled. That means that it can be changed by suitably-authorized people – shown here by being 'unlocked'. At some point it becomes a record. The key point is that, at that time, it is protected against change. This can be as a result of the need to keep records, according to the definition, "as evidence…". It is also common sense that we do not want anyone to be able to change a record, ever. This is indicated on the diagram by the record being locked.



At the same time, it is stored 'in context', that is, it's put in a particular system and a particular container or folder within that system. Metadata is assigned; so is access control.

Also, the record passes into corporate ownership – that is, it no longer 'belongs' to an individual, but instead it 'belongs' to the organization, and is managed by the records manager. This process is generally referred to as "declaring" records.

The Purpose of Capturing Records

According to ISO 15489, the purpose of capturing records is to:

- Establish a relationship between the record, the creator, and the business context that originated it. Records must be trustworthy, reliable, and authentic – and a key consideration for that is to capture the record in its context as close to the event it documents as possible. Much of that context is captured in the form of metadata, which we discuss in more detail elsewhere in this course.
- Place the record into a controlled environment. This ensures that the record and its metadata cannot be altered, tampered with, or deleted inappropriately and helps to demonstrate the trustworthiness, authenticity, and reliability required to ensure its evidentiary value.
- And link the record to other related records. Whether paper or digital, records are often created and related to other records – for example, all personnel files. The capture process provides a mechanism to establish the relationship between THIS record and other similar records that can then be managed in similar fashion.

The bottom line is that by capturing a record, it allows the organization to manage the record much more effectively than if it is simply retained on a file share or on users' computers.

While all those reasons are true and valid, they are more specific to records management. But there are other reasons for capturing records that support business goals and objectives.

First, capturing records enables innovation. Most approaches to capture result in some centralization of access. This means that employees know where to go to find a particular record. It also means that they can trust that that version is the correct, most current, approved one. At the same time, access to the record can be shared.

That sharing can also enrich customer experiences, both internally and externally. Knowing where records are and that they are complete and correct makes it easier for customer service staff to respond to queries efficiently and effectively. Records could even be made available through a self-service website, portal, or app.

Finally, capturing records helps to minimize risk and protect information assets. We can set up access controls to ensure that only authorized individuals have access to records, and only to the records they should have access to. This is much, much easier to do when records are stored in a secure repository. We can also set up security to ensure that records cannot be printed, or downloaded, or emailed outside the organization.

Once a record is captured, we can configure the system to ensure that it cannot be modified, edited, or deleted except under very specific circumstances. And we can minimize risk further by ensuring that records that have no further business, legal, operational, or historical value to the organization are disposed of in accordance with the records program.

Deciding What to Capture

Next you need to decide what to capture – that is, what types of information rise to the level of a record and need to be managed more formally? First, as we noted in our definition earlier, records support business decisions and transactions. It is not generally necessary, nor advised, to capture everything for many reasons including cost and liability.

Whether or not a particular document is captured and declared as a record should be based on its content, not its format or whether it is physical or digital. If the content meets the definition of a record, it should be captured and declared whether it is a Word document, a PDF, or some other file format.

You will need to decide what format to use to capture it. Generally speaking, digital records should be captured in their native format to preserve as much of their functionality as possible; it is also preferable to use open, standards-based formats such as PDF where that is feasible.

Many digital formats use proprietary formats and data structures, which can significantly impact the ability to capture and manage them over time. In addition, some types of content are inherently dynamic and may never really be "complete" until they are set to read-only or can change how they are rendered based on the device. This is great for consuming content on a variety of devices and platforms, but can significantly complicate the recordkeeping process.

© AIIM – 2019

Best Practices for Capturing Records

Records are declared and managed as records in order to safeguard their evidentiary value and ensure they remain admissible. That evidentiary value includes that the records are reliable and trustworthy. There are many different models for admissibility across the world, but they generally agree that in order to be admissible, a particular record should meet these requirements:

- The record was created as part of the normal course of business. There are other records like it and the organization can demonstrate the process for creating them.
- It was created in a timely fashion – at or very near the time the event or transaction occurred.
- It was created by a person with direct knowledge of the event or transaction.
- And it hasn't been changed, altered, or deleted since it was captured. This can be demonstrated through the use of access controls that prevent those changes, and audit trails that confirm the record hasn't been altered since its capture and declaration.

Automation can significantly support this process. If you have a workflow for creating or receiving a document, it is generally straightforward to determine when that document is complete and add a step to declare and capture it into a target repository.

Records and Non-records

What Should You Capture?

So what should you capture? First, it depends on the organization. Different industries and sectors have significantly different legal and regulatory requirements for keeping records. A multinational oil and gas firm will have substantially different recordkeeping needs than a small sole proprietorship scanning service bureau operating in one U.S. state. There is no one-size fits all definition.

That said, we can say that records have value. ISO 15489 defines records as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business." We capture records to ensure that we can maintain that evidentiary weight.

So generally we can approach the determination of what to capture as follows:

- **Legal.** Is there a specific legal or statutory requirement to keep a particular record for a particular period? These are very common in highly regulated industries.
- **Fiscal.** Do the records document financial obligations either to or from third parties?
- **Operational.** Are the records in question needed to complete other, non-financial transactions or activities?
- **Historical.** Are the records themselves of some particular historical value to the organization? These tend to be a very small percentage of corporate records, but a larger and sometimes significant percentage of governmental records holdings.

We should also note that the decision whether to capture a record or not depends on its content and the value of that content to the organization, NOT on whether it is physical or electronic, or what file format it's stored in. This is a fundamental tenet of records management: a transaction sent via email holds the same business value as the exact same transaction conducted through Word, or via fax, or by postal mail, and the resulting record should be managed according to the same basic principles.

What NOT to Capture

There are also types of information that probably don't need to be captured.

- **Redundant –** all the multiple copies and versions of a particular piece of information.
- **Outdated –** information that has been superseded or that has met its business purpose and is no longer needed.
- **Trivial –** things that are business-related, but barely. Examples might include announcements about minor promotions or organizational anniversaries, or the unofficial "time off tracker."
- **Not business-related.** Think of all the vacation photos, MP3 files, and all the myriad other types of personal information that probably shouldn't be stored on organizational servers, but is.
- **References and templates.** These need to be managed, but they are generally not managed as records; instead, they are kept until superseded.

These types of information only serve to make the valuable information harder to find and should not be retained at all, or should be destroyed on a regular basis. There are a number of approaches to "clean up" these types of information; we discuss them in more detail elsewhere in this course.

It's important to note, though, that whether you think a particular type of information has value or not, it should not be destroyed without reviewing with the records management, legal, or compliance function. "Old" information that hasn't been touched in years could still have a mandatory retention requirement or be on legal hold and getting rid of it can result in significant liability for the organization.

Records vs Non-records

With this in mind, you should now be clearer on what is, and what is not, a record. If you are new to records management, some of the following may need a little getting used to, especially if you associate 'anything that is in a file' with the idea of 'record'. In fact, many items that people keep, or file, are definitely not records.

Typical examples of records:

- Financial documents
- Policies and procedures
- Meeting minutes

Typical examples of non-records:

- Press clippings
- Invitations, for example to someone's going-away party
- Announcements of promotions

This much is straightforward enough. But then there are some kinds of information where it is less clear, such as memos, internal email, and drafts. Here, the position is that some of these should be records – but then some should not be. Different organizations, under different regulatory regimes, and with different corporate policies, will have different ways to separate these.

The point of the distinction is that because records document business and legal obligations, they need to be managed more formally in order to preserve their evidentiary value.

Vital Records

Another important records-related concept is that of the vital record. This refers to that subset of records that are critical to the organization's mission or existence, or are required in the immediate aftermath of a disaster to begin the recovery process.

Loss of these records may cause irreparable harm to the organization. Because of that, vital records require exceptional protection, such as multiple redundant backups, offsite storage, or storage in fireproof vaults. They are also periodically reviewed to ensure they remain worthy of this exceptional (and more expensive) protection.

© AIIM – 2019

The Benefits of Automating Records Management Tasks

There are a number of benefits organizations can realize from automating records management tasks. In some ways these benefits are similar to those associated with any type of process automation: speed, consistency, completion, etc. But there are some other reasons to pursue automation of records management tasks that make it particularly important to organizations – and to individual employees.

Increase Ability to Scale

The first reason to consider automation is that it provides better ability to scale to the increasing volume, variety, and velocity of information created or received by the organization, and thus the increase in records as well.

Consider this scenario: An organization has 1 million files on its file share. Its employees have not taken the time to manage this information appropriately, so it's a bit of a digital landfill. At the same time, per certain regulations, it cannot delete files without confirming that they are not records. At only 10 seconds per file to make this determination, that's 10 million seconds, or 2,777 hours, or over a person-year of time just to review them for 10 seconds each. So that's not likely to happen either. The net effect of this is that the organization has declared 1 million files – and more every day – to be *permanent records*.

Enter automation. We can leverage business rules, metadata, and content analytics to determine that half of what's on the file share is copies. Maybe 25% of it is old, outdated records that should have been disposed of long ago. That's 75%, or 750,000 files, that can probably be gotten rid of sight unseen provided that records management and legal understand and agree.

Improve Findability

Another significant benefit of automation is that we can improve the findability of enterprise information. Information that is stored on an employee's computer, or in email, or even in a file share, is very difficult to find even for the employee that stored it there, and practically impossible for anyone else. And getting employees to understand a complex taxonomy may not make it better – and may even make it worse!

But we can set up automation based on existing metadata such as that something is an invoice, or what company it is from, or when it was issued or paid. We can classify it automatically based on those properties and a workflow step that indicates it's been paid – or that it hasn't been.

If we don't have that metadata, we can use automation approaches such as analytics to determine that a particular document is an invoice. We can use recognition technologies and/or analytics to understand the contents of the document, and extract that metadata from the document and then put it into that workflow.

And for all the other files already stored in the repository, or for those that are being migrated from one system to another, we can leverage automation techniques to convert metadata from one data structure to another, fill in missing metadata, construct metadata fields and options based on folder structures, and much more.

All of these contribute to users being able to find their information more efficiently, and to trust in what it is that they find.

Reduce Human Error

The next thing to consider is human error. Highly motivated, well-trained humans can be very accurate in determining what to capture, because they are the ones doing the work and they know what information they capture, receive, process, and manage in support of that work.

But humans make mistakes. In the paper world transposing two digits of an invoice when it's being filed can cause that invoice to be lost almost as if it were shredded. With digital files and manual data entry, humans could inadvertently lose things much more efficiently!

Worse, humans are consistently inconsistent. That is, when a business rule is not correct, the application could make a lot of mistakes, but they will be the same mistake and, once the error is corrected, the mistakes go away. Humans make a variety of different errors that can be difficult to track down.

Enter automation. Again, things like business rule, metadata, and analytics can be used to extract meaning from documents automatically; put them on a workflow; at the appropriate point in the process, capture the final document as a record and apply metadata and security; and apply the appropriate retention based on the type of record it is. This allows the organization to better meet its compliance requirements because the rules can be shared and their outcomes can be reproduced.

And they can support more complete and consistent capture of records – that all elements or pieces are present and complete, that it has been reviewed and approved by whomever it is required of, that it actually was captured and to where, and that all of that is tracked in an audit trail.

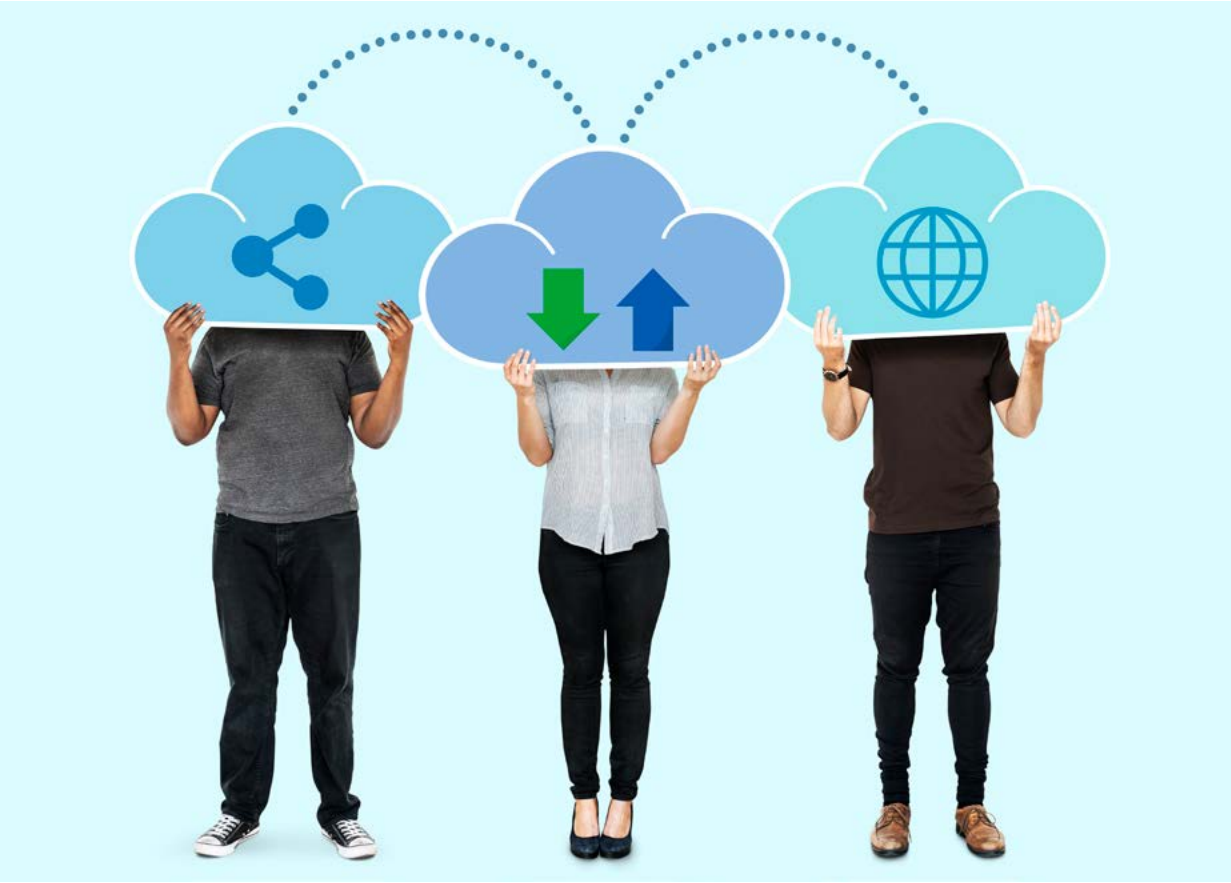
Finally, they can support more consistent and timely disposition of those records as well at the end of the lifecycle. This requires that users and management trust that the automation is performing as expected, and that that trust is supported by verification and periodic auditing. But organizations with sufficient maturity may be able to move from the "ask permission" model of disposition to a bias towards disposition – where records are dispositioned unless there is a specific reason not to such as a legal hold (and "just in case" wouldn't count).

Reduce the Burden on Users

Finally, and arguably most importantly, automation is important because it reduces the burden on users. Users really don't want to be records managers, for a couple of reasons:

- They don't have time. Every organization wants to do more with less, and every minute spent "doing records management" is a minute not spent selling, or buying, or managing, or engineering, or whatever it is they do for their primary job.
- They don't have the interest. Records management is incredibly important, but if you're a doctor or engineer or salesperson, you want to do those things, not things that aren't related to them or that are considered overhead or administrative tasks.
- They don't have the expertise. In many organizations, staff get a short training on records and information management when they join and perhaps an annual refresher. In other organizations they get no training at all. We mentioned earlier that staff are the best positioned to know what information is important, but it does not follow that they know what to do with it to safeguard it over time.

But we can leverage automation capabilities and approaches to streamline and automate many of these tasks. We can have these users leverage their expertise to help design workflows and business rules, and to help train analytics tools to be more effective, and to review the results for accuracy. This will help users to be more efficient in their own jobs while still ensuring that records management and compliance requirements are being met.



© AIIM – 2019



Test Your Knowledge

Question 1:

What type of content should be captured and declared as records?

a. All content created as part of a business process
b. Content that is stored in standards-based formats
c. Content that documents a transaction or decision
d. Content that is being reviewed and edited

Question 2:

Which of these is most likely to need to be managed as a record?

a. A signed and fully executed contract
b. An email message noting this month's employment anniversaries
c. A project management template
d. 9 of the 10 copies of the disaster recovery plan found on the network

Question 3:

How does automation of common records management tasks benefit individual employees? (select 2)

a. It allows them to focus on core business tasks.
b. It provides evidence of the amount of work they have performed.
c. It ensures that all information is captured directly into the records repository.
d. It improves their ability to find and trust the information they need for their jobs.



Test Your Knowledge - Answers:

Answer to Question 1:

What type of content should be captured and declared as records?

a. All content created as part of a business process
b. Content that is stored in standards-based formats
c. Content that documents a transaction or decision
d. Content that is being reviewed and edited

The correct answer here is C, content that documents a transaction or decision. For A, many processes create content that does not need to be declared and managed as records including drafts, references, and working notes. For B, the format is immaterial to the decision as to whether something should be managed as a record. And for D, content that is being edited is not ready to be declared as a record; rather, it should be declared if the final published or approved content meets the definition of a record.

Answer to Question 2:

Which of these is most likely to need to be managed as a record?

a. A signed and fully executed contract
b. An email message noting this month's employment anniversaries
c. A project management template
d. 9 of the 10 copies of the disaster recovery plan found on the network

The correct answer here is A, a signed and executed contract. B is an example of a trivial document. C is an example of a reference or template. D is an example of redundant copies.

Answer to Question 3:

How does automation of common records management tasks benefit individual employees? (select 2)

a. It allows them to focus on core business tasks.
b. It provides evidence of the amount of work they have performed.
c. It ensures that all information is captured directly into the records repository.
d. It improves their ability to find and trust the information they need for their jobs.

The correct answers are A, it allows them to focus on core business tasks, and D, it improves their ability to find and trust information. For B, records management automation would provide only indirect evidence and for only some employees. For C, automation can help, but you would not want all information to be captured into a records repository – just the records.

© AIIM – 2019

Retention and Disposition

Determining Retention Requirements

We have already discussed how a record is defined in the context of records management and that records might be stored in a number of different ways. By defining something as a record the organization has decided that it must be retained, so that it is available and accessible to the organization and can be referred to if necessary in the future.

Depending on the structure of an organization's systems for managing records, the nature of the record, and policies on content of a recordkeeping system, records might be retained:

- In a digital recordkeeping system
- In some other content repository, such as an ECM system or a file sync and share application
- In a folder on a shared network drive
- Within specific software applications in relation to structured data
- On physically separate electronic media such as audio or video cassettes or CDs/DVDs; or
- On paper or in some other physical form, such as samples

Since this course focuses on digital records management, we won't spend any time on paper/physical records retention, instead focusing on the others and, particularly, the first one.

Retention Periods

The intrinsic nature of records varies. Some records might hold great importance long into the future and need to be retained, perhaps indefinitely. An example of this might be an important historical record like the Magna Carta or the U.S. Declaration of Independence. In the case of such an important record, it might be stored in a number of different ways, including both the original document and other physical and digital copies. These are often referred to as permanent records, indefinite records, life-of-organization records, etc.

Other records might be important - but their importance might not so be enduring. Consequently, such records will not have to be retained indefinitely, but will instead be retained for shorter periods. At some point these records will no longer have to be retained since they hold no further relevance. Depending on the type of record and the specific contents, a record might be kept for a very short period, or decades or more.

And retention needs to be applied consistently – to all records across the organization and all formats, paper and digital. It is the content that determines the retention, not the format.

How to Determine Retention Periods

In considering how long a record needs to be retained, the factors that should be considered are:

Legal requirements – is there a statutory period that the record must be retained for? Laws or supporting regulations may require certain records to be retained for specific minimum periods of time. This is the most important one, because destroying records before they meet their legal or regulatory obligations can significantly increase the organization's liability.

Statutes of limitations – these are legal requirements as well, but from the opposite perspective. That is, they describe how long a particular offense can be prosecuted in a particular jurisdiction. Organizations should keep relevant records until the statute of limitations for that type of offense has expired. For example, in the U.S., individual Federal income tax returns can generally only be audited for 3 years. Individuals should keep their returns for at least 3 years, and as long as they are not committing tax fraud or tax evasion, should get rid of them once that retention period has expired.

Industry standards – within particular industries there might be commonly accepted standards for the retention of particular types of record. An organization could adopt this standard or decide to vary it, though to have shorter retention periods compared with industry standards would not reflect well on the organization in the event of litigation.

Operational – whether or not a record is covered by minimum legal requirements, the organization may decide to retain it for a longer period of time due to operational needs. This should be a positive decision agreed to by the business, records management, legal, etc. and not simply through neglect or lack of ability to make a decision.

Administrative – at what point will the information contained in the record cease to be relevant to the ongoing operation of the organization? Where a record is not covered by legislation or commonly accepted standards, then the continuing relevance will need to be examined, and an appropriate retention period determined.

Historical – Some records may have enduring historical value beyond their legal, operational, or administrative value. These records tend to document a unique event in the organization's history. Public sector entities have a more significant obligation to document how they accomplish their missions and so tend to generate more historical records.

Not surprisingly, this is a complex subject. In any one industry or sector there is always more to learn and always room for debate. Centers of excellence for records management such as national, state and corporate archives are often well placed to provide specific advice.

Retention Considerations

There are a couple of other things to consider when identifying or reviewing retention periods.

Sometimes a record may have multiple retention requirements. For example, let's say that a record is required to be kept for four years at the state level, but six at the Federal level. The longest retention period is the one that should be applied, because if it's only kept for the shorter period and then destroyed, it will cause liability if the Federal agency needs it. So in this case it should be retained for six years.

The other consideration is that whatever the retention period, and wherever it is in that process, a legal hold supersedes retention and disposition. If the legal hold is received the day before the record is scheduled for disposition, it must be retained. Organizations that fail to put disposition on hold can be subject to steep sanctions for destroying evidence. We discuss this in greater detail later in this course.

One final note: all organizations have – or should have – backups. These are used in the event of a disaster to recover some or all of the organization's information. Some organizations include them as part of the records program, some do not. In either case, backups should be retained as short a period of time as necessary to fulfill the organization's disaster recovery and business continuity requirements. After that, they should be destroyed, recycled, however the organization approaches this. And as we just discussed, in the event of a legal hold this recycling or disposition might need to be put on hold as well.

The Value of Retention – and Disposition

As we have seen, part of the process of declaring a record involves determining how long to retain it. At the same time, however, there is a tradeoff between effective retention and over-retention. Sometimes it seems like the reasoning is that if having information is good, having more must be better. Other times it seems that when it is time for disposition, individual managers determine that they need to keep their information longer "just in case." This can significantly increase the risk of liability to the organization in the case of legal or regulatory matters or audits. It also increases the liability associated with a data breach – the more information is kept, the more will be exposed in the event of such a breach.

Rather, organizations should dispose of records in accordance with their records policy and program. Doing so will minimize the risks associated with keeping information too long. The combination of the policy, retention schedule, and documentation of disposition will allow the organization to document that it kept what it was supposed to, as long as it was supposed to, and that it disposed of the information as part of the normal course of business, all of which will help to reduce risk.

At the same time, disposing of information will also help to manage the costs associated with storing that information, including the costs of managing that information over time, which helps the organization be more efficient. And by reducing the amount of information stored, users and customers will find their particular records more efficiently as well. We will address disposition in more detail shortly.

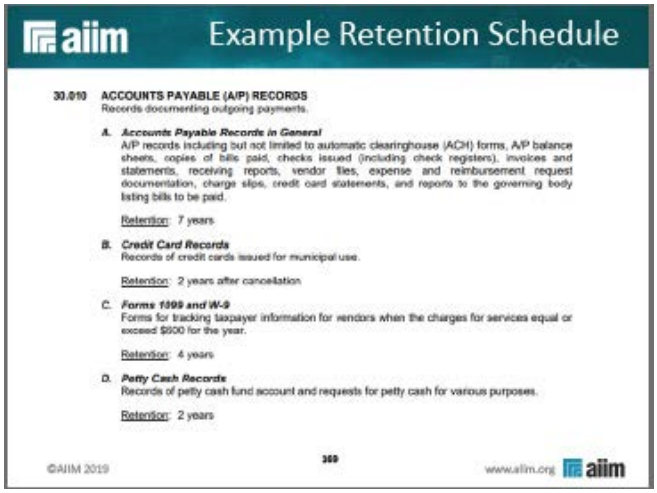
© AIIM – 2019

The Purpose of the Retention Schedule

A central part of records management is making sure that records are kept for as long as they are needed, but for no longer. ERM systems, and records management more generally, achieves this by using retention schedules. MoReq, the Model Requirements for Managing Electronic Records, defines retention schedules as:

> "A set of instructions allocated to a *class* or *file* to determine the length of time for which its *records* should be retained by the organisation for business purposes, and the eventual fate of the *records* on completion of this period of time."

This definition tells you that retention schedules are essentially instructions that tell a system how long to keep records. We should also note that in MoReq2, the preferred term is "retention and disposition schedule;" however, the instructions not only tell the system how long to keep the records, they also tell it what to do with them at the end of the lifecycle.



Retention Schedules

Depending on the size and nature of an organization, the volume of records that need to be managed will vary. Many organizations will be dealing with many hundreds of thousands of records and sometimes millions of records each year. It would be extremely time consuming if an organization needed to give consideration to each individual record in relation to how long it needs to be retained for.

Consequently, organizations adopt rules relating to the retention of records. This is documented in a retention schedule, which identifies a number of options for retention, based on the intrinsic nature of a particular class of records, and determines on the basis of the nature of those records how long they must be retained.

An important benefit of a retention schedule is that, when properly applied, it will ensure consistency in the treatment of records that share common characteristics. That is, the retention schedule identifies and manages groups of records, not individual records.

Retention schedules will be specific to a particular organization and will need to be managed and applied by that organization in relation to its own systems for records management. That said, there will often be retention requirements that apply to all organizations within a given sector due to legal and regulatory requirements.

There is one other consideration to take into account – the high volumes of records that need to be managed and to ensure consistency of treatment.

Retention schedules operate on the basis that records are not unique. In other words, you can identify documents with shared characteristics, and group them on the basis of these shared characteristics, or a common intrinsic nature. These groupings are often referred to as records series.

The benefit of this is that you can apply the same retention schedule for each group of records, and manage that group of records as a unit for the purpose of managing it over its lifecycle.

Retention Schedule Considerations

Another way to look at a retention schedule is that it serves as a list of all the records the organization should have. If the organization has something that is not on the schedule, a determination needs to be made whether it can be gotten rid of, or whether it should be added. Similarly, if something is missing from the inventory that is on the schedule, the organization needs to decide whether it needs to be located, or is no longer applicable and can be removed.

Finally, a retention schedule should not be used as the primary taxonomy for non-records-manager end users. This is because they are designed to support records retention and disposition, which may require breaking like things into different categories. For example, an organization may handle retention differently for expense reports depending on the source: a staff employee, a manager, executive management, or an outside consultant or contractor. But the accounting clerk that processes them is probably looking for all the expense reports from this month or that relate to a particular project or office. Different roles search in different ways and may require different taxonomies.

The Elements of the Retention Schedule

Example Retention Schedule

Here's an example of a retention schedule. This example comes from the Colorado State Archives Municipal Records Retention Manual. The records identifier includes a series code, 30.010, and then the individual records series, shown here as A through D. Each series has a title and descriptor and the retention period: 7 years for A, general accounts payable records; 2 years for B, credit card records, and so on. In the series shown here, anything which is not permanent is destroyed when the retention period is complete and assuming there are no legal or other holds in place, so no other disposition instructions are supplied.

Retention Scheduling

Once a retention period has been determined, a further complexity in relation to retention is to identify when the retention period begins and when it should end. There are two primary approaches to determining this: time-based, and event-based.

For some records there may be a set period for retention that relates to the creation event. For example, legal requirements to retain financial records will define a set period that a purchase invoice that has been paid needs to be kept. The retention period might in this case be related to the end of a particular financial year plus a set number of years beyond that. For example, invoices might be kept until the end of the current fiscal year + 2 years. An invoice created in fiscal year 2019 would be kept until the end of fiscal year 2021.

In other circumstances, the retention period might only be triggered once a specific event relating to that record occurs. An example of this might be when a case is closed. Once there is no longer any activity relating to that case, a set period for retention would be triggered. Thus, while the period for which the record has been retained in its 'live state' will vary from case to case, the period of retention once the case is closed will be the same.

A further complexity is where the retention period relates to a specific event occurring in the future. For example, in relation to pension entitlements, records that relate to that entitlement will need to be retained by an organization until the person with that entitlement reaches a particular age. This event is specific to the subject of the record and is in the future. These are more difficult to automate because so many computer systems want a particular date. If you are currently employed, you don't know your date of separation so you can't enter that into a digital recordkeeping system.

The bottom line to all of this is that the organization has to be consistent in how it retains information, and more importantly, how it disposes of it. An organization that has a policy and retention schedule it doesn't follow is actually in worse shape than one with no records program at all.

© AIIM – 2019

Disposition

Disposition is defined by ISO 15489 as the "range of processes associated with implementing records retention, destruction or transfer decisions which are documented in disposition authorities or other instruments."

In other words, disposition and even destruction are accepted phases in the records management lifecycle provided they are conducted in accordance with the records program.

The Benefits of Disposition

Before we get into the details of how to destroy records, we should identify why to destroy them.

Keeping everything for forever is expensive. Whether records are stored in physical or digital formats (or some combination), there is a storage cost incurred. Paper records take up space – and that space is either owned, leased, or in effect rented from a third-party records storage service company. Electronic records are still stored on physical media; even though they take up less space, the storage media itself isn't free, and the cost to manage that storage can be quite expensive.

It also takes time to find records. The larger the haystack, the more time it will take to find something. Even the best classification schemes and indexes take time to wade through – and if your search returns 150,000 hits, how useful is that? Now consider that a significant portion of the hits in the haystack are old expired records (or even non-records) that could have been gotten rid of.

At least in the U.S., the courts have held that there is no legal obligation for organisations to keep "every scrap of paper" or to keep everything forever. In the absence of specific laws or regulations requiring retention of something for some period, the organization is generally left to its own judgment to determine how long it needs to keep something.

The bottom line is that destroying records when they are no longer valuable can reduce risk to the organization: the risk of not being able to find something when needed, the risk of inadvertent disclosure, and the risk of spending more than is necessary in storage. But it has to be done consistently and in accordance with a records program in order to ensure it will not be called into question later.

Disposition Principles

In other words, disposition is an accepted phase of the records lifecycle. Whether this refers to transfer to another organization, or the ultimate destruction of the records, when documents and records no longer have any business value, they can, and should, be disposed of at the end of their lifecycle, in accordance with the established records program.

Records Disposition

At some point in time, governed by the retention period that has been applied to a particular record, the retention period will end. This signifies that the record no longer needs to be retained. Action then needs to be taken in relation to the disposition of the record. There are four main options in relation to disposition. These options are:

Review – A record that has reached the end of its retention period may be subject to a process of review. This means that the record is flagged within the ERM system, and no further action is taken until the record has been reviewed and a decision taken on further action. The advantage of this option is that it does provide for an opportunity for the content of a record to be examined, and an assessment made of any continuing value to the organization.

Transfer – A record that has reached the end of its retention period may be transferred, or exported, to an archive. This means that the record is removed from the 'live' ERM system to an electronic archive. This is likely to be a more relevant option in the public sector, where the concept of maintaining public records is more prevalent. However, some private sector organizations also maintain a company archive.

Destruction – A record that has reached the end of its retention period can be destroyed. The ERM system will automatically identify when the retention period of a particular record has expired and delete the record.

Destroying Digital Records

For electronic records the method of destruction will be determined by the medium on which the record has been retained, and whether the medium itself can be destroyed, such as a flash drive or CD.

Where the record is maintained within an information management system, the record will be deleted in accordance with how that system works – there will be an audit trail that tracks when the record was destroyed and who approved it.

Files not retained in a repository are easy to delete – in fact that's one of the issues with that approach – but even deleting may not be enough. Many times, when files are deleted from e.g., a file share, they are simply moved to a recycle bin of sorts. Even if you "double delete" or select to permanently delete, generally what's happened is that the system pointers to the file have been removed, but the raw data is still there and still recoverable through forensic technology. So if you have significant security concerns that a given file not be recoverable by any means, there are ways to address them, but they are significant more complex and expensive.

Destroying Physical Media

Magnetic and other physical media can be destroyed in a variety of ways. Magnetic tapes can be demagnetized, though the process for doing this is a bit more in-depth than simply sliding a refrigerator magnet across the tape.

CDs can be shredded, and any physical media – CDs, hard drives, entire computers – can be destroyed using industrial pulverizers. Your organization may not have these, but there are a number of companies that can offer this service for a fee.

We can also recycle physical media. This applies to both paper and other physical types of records. Recycling is not just good for the environment – in some jurisdictions it is the law.

Some organizations may try to recycle hard disks by sending them, or the computers that contain them, to be recycled either for their base materials or as-is to worthy organization such as schools and charities. The data on the hard disk must be completely removed prior to that recycling taking place or the data may be recoverable.

Document the Destruction

While the destruction of the record in these cases takes place outside of an information management system, it is advisable that details about the record and its destruction are contained within the system. This allows the system to identify that a record has reached the end of its retention period and that action needs to be taken.

If this approach is not taken, then separate logs of such records would need to be kept if effective records management is to be maintained. In the absence of these, it is likely that the organization will lose sight of these records, and that appropriate action in relation to their retention is not taken at all, or is taken in an inconsistent manner.

When physical records are destroyed, many organizations will log that using certificates of destruction. These can be quite detailed and will generally include the record series that were destroyed, the method used to destroy them, and when they were destroyed. These are then stored as records themselves and may be kept either as long as the longest series they document, or as permanent records.

Digital records work a bit differently. When digital records are in a formal recordkeeping system, the system will document various events that occur in relation to the records in an audit log. At a bare minimum, the system should keep a record of the destruction – that is, a record of the fact that a record used to exist, who authorized its destruction, and when it was destroyed. Audit logs themselves should be considered records and should have appropriate retention and disposition as reflected in the organization's records retention schedule.

Finally, as we have discussed previously, many organizations choose to outsource day-to-day responsibility for storing records to a third party. This is quite common with physical records but is starting to grow as a practice for digital records as well. In either case, the third party may offer destruction services. In the event that the third party performs the destruction, documentation of this fact is critical and will almost always take the form of the aforementioned certificates of destruction.

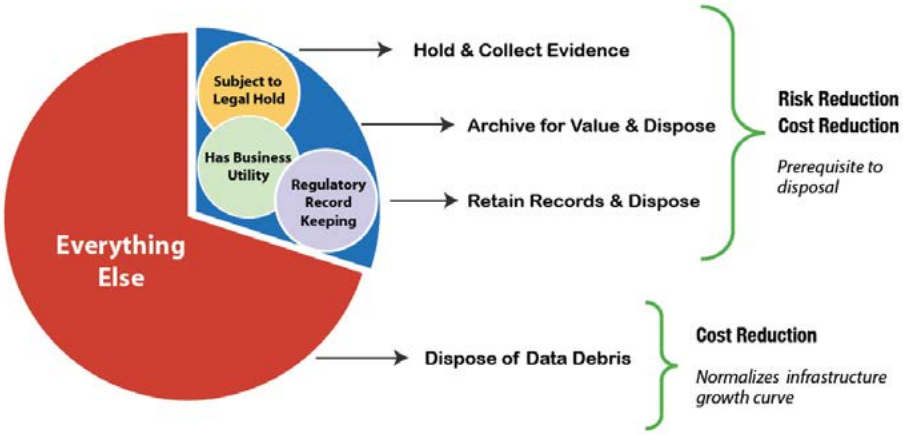
© AIIM – 2019

A Note about ROT

The Compliance, Governance, and Oversight Council (CGOC) conducted a survey in 2012 regarding information holdings. What they found was that for the average organization:

- 1% of organizational information is subject to a current legal hold
- 5% is retained as records
- 25% is relevant to ongoing business operations
- Up to 69% of an organization's information has no business value

This last "bucket" includes information that is redundant, outdated or obsolete, and/or trivial (ROT). This information still uses scarce organizational resources to store and manage it and can significantly increase the cost and potential liability associated with over-retention.



Source: https://www.cgoc.com/resources/information-lifecycle-governance-leader-reference-guide

These numbers will vary across organizations and even types of organizations – for example, in the public sector the "important" information is probably a higher percentage due to open records types of laws as well as the need to manage some information for historical purposes. The broader point is that every organization has a significant percentage of information holdings that has absolutely no business value but which still consumes organizational resources. Those holdings should be reduced and minimized to the extent possible.

Defensible Disposition

One of the keys to effective records management is to have a bias towards disposition. In other words, most records programs require a series of approvals prior to dispositioning records, including getting approval from, at a minimum, legal and from the business owner of the records in question. This means that if anyone in that process doesn't approve of disposition for any reason, the records are retained longer than required. This is especially true, and especially important, for most digital recordkeeping processes.

A better approach is to develop the process in such a way that business users and legal are notified that records are up for disposition, but that the disposition process will take place unless they have a specific reason as to why it shouldn't. Instead of asking for (and waiting for) permission, the records team follows the established policies and processes. This does require more information management maturity on the part of the organization, and business owners and legal have to trust that records management has followed the process effectively.

This is often referred to as "defensible disposition" and it can significantly reduce the amount of information stored that has no business value to the organization. If the organization is able to set the bias towards disposition, policies and even automation can be applied to ensure that non-records are disposed of efficiently. This would for example allow the organization to apply disposition rules to folders on networked file share locations or even to users' individual computers.

This seems like a very fine distinction, but it can make a dramatic difference in the efficient disposition of records at the end of the records lifecycle.



Test Your Knowledge

Question 1:

An organization has determined that certain records have two different retention periods, 6 years and 10 years. How long should the records be kept?

a. Indefinitely, until the two periods can be consolidated
b. 6 years
c. 10 years
d. 16 years

Question 2:

What are the primary purposes of the retention schedule? (select 2)

a. To allow users to locate their records by records series
b. To identify the systems that will be used to store digital records
c. To describe how long to keep records and what to do with them after that
d. To list all the records in the organization

Question 3:

What are the primary elements of the retention schedule? (select all that apply)

a. Records identifiers
b. Disposition instructions
c. Classification values
d. Retention periods

Question 4:

Which statement about disposition is the most accurate?

a. Business information should never be deleted because it could be of value in the future.
b. Disposition is an accepted practice that reduces organizational risk.
c. Disposition is most effective when it's performed by end users.
d. Once business information has no value, it should automatically be deleted.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

An organization has determined that certain records have two different retention periods, 6 years and 10 years. How long should the records be kept?

a. Indefinitely, until the two periods can be consolidated
b. 6 years
c. 10 years
d. 16 years

The correct answer here is C, 10 years, the longest of the two retention periods. A is simply incorrect. B, 6 years would result in the organization having increased liability for 4 more years. D, 16 years is adding the two together, which would increase the organization's liability for 6 more years.

Answer to Question 2:

What are the primary purposes of the retention schedule? (select 2)

a. To allow users to locate their records by records series
b. To identify the systems that will be used to store digital records
c. To describe how long to keep records and what to do with them after that
d. To list all the records in the organization

The best answers here are C, describe how long to keep records and what to do with them after that, and D, to list all the records in the organization. For A, users will often not even know their records series identifiers and would not generally use them to locate their records. For B, the retention schedule has very little direct bearing on the systems that will be used to store digital records.

Answer to Question 3:

What are the primary elements of the retention schedule? (select all that apply)

a. Records identifiers
b. Disposition instructions
c. Classification values
d. Retention periods

The correct answers are A, B, and D. While retention schedules sometimes are used as classification schemes by non-users, this is generally not the best way for users to search.

Answer to Question 4:

Which statement about disposition is the most accurate?

a. Business information should never be deleted because it could be of value in the future.
b. Disposition is an accepted practice that reduces organizational risk.
c. Disposition is most effective when it's performed by end users.
d. Once business information has no value, it should automatically be deleted.

The correct answer here is B, disposition reduces organizational risk. A, keeping everything forever, will significantly increase organizational costs and risks. For C, the challenge here is that like other records-related tasks, end users generally won't do it for a variety of reasons. D is a good choice – but could be dangerous as some information might be under legal hold, even though it has no other business value.

eDiscovery

Legal Holds

The legal hold is used to indicate that there are occasions when the legal process clashes with retention and disposal of records.

The idea behind the legal hold is that a records manager or administrator can – and indeed must – prevent any modifications to, as well as the destruction of, certain records. The reason is that these records are required, or may be required, in a legal case. When an information request is pending the "duty to preserve" is established and the organization should not destroy or otherwise remove pertinent records.

As soon as the notice is received by legal, the legal hold needs to be communicated to all affected parties, which could be anywhere in the organization depending on the nature of the legal matter at hand. Legal will identify potential holders of information pertinent to the matter, and will communicate to them the types of information that should be put on legal hold.

Legal Holds and the Information Lifecycle

In normal circumstances records can be reviewed and disposed of when the predefined retention period matures, for example after two years. Legal hold is a change in record status that takes precedence over retention and disposition. An administrator then needs to set the legal hold parameter within the recordkeeping system. The system is thus prevented from disposing of these records until the parameter is removed. Some legal cases can run for many years, so it is important that pertinent records are not affected or destroyed simply because that 2-year period has elapsed. At the end of the legal process, including any appeals, the legal hold needs to be removed and normal disposition activities resumed.

Importantly, whether something has been captured, declared as a record, etc. generally does not matter for the purpose of the legal hold. Rather, the key questions are whether the information exists, and whether it is potentially relevant. This means legal holds could apply to records, documents, personal files, backup tapes, or anywhere else information is stored. It is important to seek clarification with the legal team.

In addition, nobody in the organization should attempt to convert file formats, encrypt records, etc. unless explicitly directed to by legal. Documents should be locked, secured, etc. to ensure they cannot be edited or deleted, inadvertently or intentionally. Copying should also be avoided without legal's approval because changes in file metadata can compromise the document or record.

The system should provide some mechanism to implement and clear a legal hold. If this is not implemented the organization must identify an alternative procedure to ensure compliance. In either case the organization must include the use of legal holds in its governance policies and procedures.

© AIIM – 2019

Collecting Information fromThird Parties

This module will help you to define the issues associated with collecting information from sources not owned or controlled by the organization, such as personal devices and commercial social media platforms.

"Outside" Sources

In the discovery process there is sometimes a need to collect information from sources outside the organization. These could include:

- **Third parties** – parties who are not directly part of the organization.
- **Employees' personal devices** – particularly smart phones but sometimes laptops or other devices. We would include personal apps such as personal email, personal Box or One Drive accounts, etc. here as well.
- **Commercial social media platforms** – even for an official organizational account, that data is stored outside the organization and subject to the platform's rules. Traditional Software-as-a-Service and other cloud-based solutions are not necessarily included here, because while they do store data outside the firewall, there is a much stronger contractual relationship that should ensure access when necessary.

Collecting the Information

There are two main issues associated with collecting data from these outside sources that are unique as compared to the rest of the discovery process – and they are different for each of the groups we noted. The first is actually collecting the information at all.

- Third parties: It's not their legal matter, so it may be difficult to get them to actually do it within the scope of the matter. They may need to be subpoenaed to get them to do what is necessary. This would definitely be done through the legal team.
- Personal devices. Which of your employees is going to give up their phone, laptop, etc. and let you go through it? But they may have to, legally. This is a big reason why organizations should resist allowing employees to commingle work and personal devices and apps. We've talked about the push towards Bring Your Own Device/App (BYOD and BYOA) elsewhere, but this can be a significant issue.
- Finally, social media brings a couple of issues. First, there may be legal prohibitions against collecting directly from e.g., Facebook. The platforms will generally comply with requests from law enforcement, but if you're simply an organization engaged in a legal matter, they generally won't (or can't) assist directly. There are many ways to target social media, but they almost all involve going to the individual whose account is being targeted and getting or forcing them to provide access.

Authenticating the Information

The other major issue is authenticating the data once it's received. Authentication is always a major issue, and in the case of data provided by others, it becomes even more important. In the age of Photoshop, metadata and the chain of custody are important considerations.

- **Third parties.** This will depend significantly on what and how they produce and the chain of custody. Again, this should be conducted by legal.
- **Personal devices.** Getting data off the device can be complex and may require specialized tools and skills; once the data is off it needs to be safeguarded to preserve its evidentiary weight.
- **Social media platforms.** Again, some of this is an outcome from how the data is collected. If it can be downloaded directly this will be easier to authenticate in some ways. Screenshots are almost never reliable courtesy of Photoshop and other image editing tools; but content on social media might be able to be corroborated through interviews, others' pages, etc.

The eDiscovery Process

Triggers for Discovery for Disclosure

Whether the question is one of discovery or disclosure, the process is generally triggered in the same way, by a request for information. In the case of disclosure this could be a Freedom of Information Act, Open Records Act, or Sunshine Act request or its equivalent; it could also be a formal notice of audit. For a discovery request it could be triggered by a notice of dispute or a preservation letter.

Whatever the triggering event, response to, and compliance with the request is not optional; a failure to respond or respond adequately can result in sanctions. We will examine sanctions later in this section.

Often the responding party will send a response to the other party and/or the legal authority asking for a narrower request, and some negotiation takes place to determine exactly what will need to be placed under litigation hold and what will ultimately need to be produced.

Duty to Preserve

The general concept behind the duty to preserve is that an organization has to preserve information relevant to a case as soon as it knows about litigation as noted on the previous section.

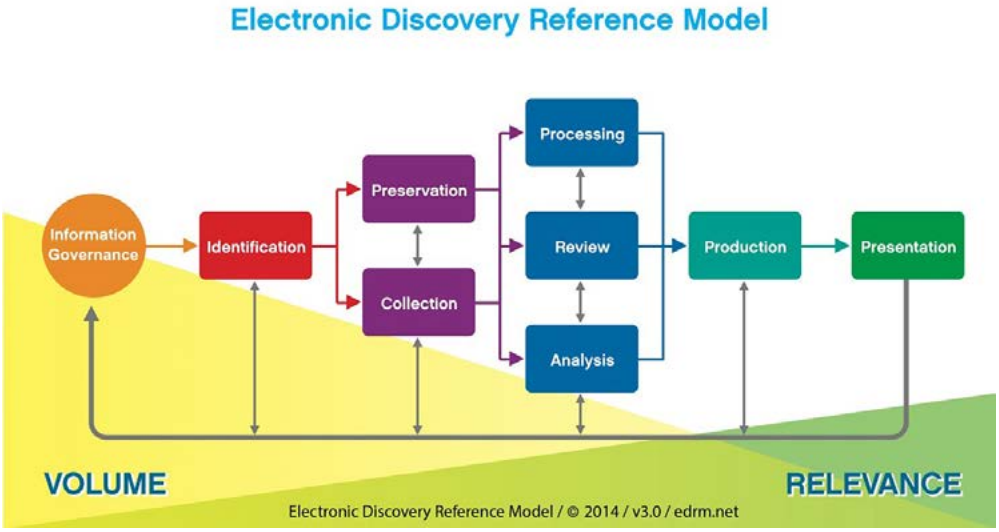
But the organization also has to preserve information if it has reason to believe that it will be sued. For example, an employee is accused of sexual harassment and is terminated. Both the alleged victim of the harassment and the terminated employee may have grounds for a lawsuit. The organization could reasonably anticipate litigation and might be well-served by preserving any email or other traffic between the two individuals and their formal chain of command as well as the entire inbox and even any computers or other resources of the terminated employee.

Another oft-cited example of this, which doesn't relate to email but is illustrative of the principle, is a U.S. based court case, Testa v. Wal-Mart, 144 F.3d 173 (1st Cir. 1998). This case involved an injury to a delivery person (Testa) at a Wal-Mart. Testa verbally threatened to sue, and an employee included that verbal threat in an internal report. By the time Testa filed a formal lawsuit, Wal-Mart had destroyed its records relating to the incident pursuant to its records retention policy. The courts found that when it destroyed the records it had received notice of a potential lawsuit and that the records would have been relevant.

The bottom line is that as soon as the triggering event happens, the responding organization has a positive requirement to preserve any information related to the matter in question. This is often something that is worked out as part of the discovery process – one side will assert that the other should have anticipated litigation and preserved more, and the other will assert ignorance of anything prior to formal notification.

Electronic Discovery Reference Model

The Electronic Discovery Reference Model is a framework for developing, selecting, and evaluating electronic discovery-related products and services. It was developed by more than 125 organizations and released into the public domain in May 2006. The project also includes work on an XML schema to allow e-discovery products to interoperate, and metrics relating to the e-discovery process.



© AIIM – 2019

This section describes the electronic discovery process from start to finish using the ERDM standard. The current model includes nine steps, from the baseline of information governance through the discovery process to presentation of the material to the court. We will look at each of these steps in more detail in this section.

The goal of discovery is to provide everything that is relevant – and nothing that isn't or is privileged. This is depicted in this graphic by the yellow and green triangles. At the start of the process, there is a lot of information and not a lot of relevance to a particular case. If the organization has good information governance, the volume will already be lower and the relevance higher. As the process unfolds, the idea is that the volume of information will decrease and the relevance will increase.

Information Governance

The first step in the model really begins before any litigation commences or is anticipated. organizations that have an effective information management and information governance program will enjoy significantly reduced costs for discovery. When the organization knows what it has, and that information is kept for as long as necessary but no longer, and users are trained on and follow the program, discovery is less burdensome, less costly, less disruptive, and more effective.

Identification

In the identification phase, the organization must determine what information it has that is potentially responsive. An organization that has a strong records management program that includes electronic records will find this a much easier process than one with no records program or one that only addresses paper and other physical records. This step also involves identifying potential witnesses who can testify as to how different systems work on a day-to-day basis, whether any backup or archival processes exist and how *they* work, and any other relevant considerations.

Preservation

As we noted earlier, the duty to preserve refers to the positive requirement to keep things that are needed as evidence - or potentially needed. Once the duty to preserve commences, the organization must impose a legal hold that prevents anything related to the matter from being altered or deleted subject to the limits of the technology. If backup tapes fall within the scope of the litigation hold, the tapes must not be recycled or destroyed.

And once the case is over, the hold should be terminated; that fact should be communicated to the relevant parties, and the information should return to the normal records management framework including destruction where applicable.

However, there is another issue associated with most types of electronic documents: the risk that the ways the system and associated processes work can incidentally delete information required to be maintained. Example: the digital recordkeeping system crashes and there is no backup or the backup was done improperly. This is akin to paper being lost in a fire or flood. How backups work and are tested is IT's job, but information professionals should ask whether backups and the backup process are being reviewed regularly.

Collection

The next step in the discovery process is collection of potentially relevant data. This means copying or moving that data from the various locations and data stores and storing them in a centralized location where they can be further processed and reviewed. This helps to ensure the process is defensible; at the same time, the collection should be targeted narrowly so that time isn't wasted collecting and processing extraneous information.

The next consideration is accessible versus inaccessible information. Email messages in the message store, Excel spreadsheets in an ECM system, and all of the information in the ERM system are all considered to be accessible. Information on backup tapes is generally considered to be inaccessible. This difference is important because at least in the U.S. organizations may not have to produce information from inaccessible sources unless there is no other source of the information. Backup media can still covered by a litigation hold, but it is often not required for litigation.

Collection also requires preservation of the chain of custody, so as to ensure that the documents collected cannot be challenged as to their authenticity. The process should be documented and, where possible, captured in an audit trail.

The bottom line is that collection methods for digital information should be legally defensible, efficient, auditable, and targeted.

Processing

The next step is processing. Processing can be used to automatically reduce the amount of data to be reviewed, for example by de-duplicating documents that are identical or by filtering (or "culling") documents that do not fall within a specified date range.

One of the tasks involved with processing is to determine what output will be required for litigation and then process the data to achieve that output.

Review

The next step in the process is the review phase. The review ensures that documents that will be produced are responsive to the matter and that all documents that are responsive are produced. It also looks for sensitive material which is responsive but may need to be redacted or marked as confidential. And it reviews documents to ensure that they are not privileged under the attorney-client privilege or as work product.

Reviewers may also need to review near-duplicates. During the processing phase it is pretty straightforward today to identify and cull exact binary duplicates using software. However, changing the file format, such as from WordPerfect to Word; changing the data format, such as from email to image; or even changing a single space or period all result in a non-duplicate document. Versions of documents also present this issue, which almost always requires human intervention today to ensure that all versions are understood.

Analysis

The analysis step takes the information that has been processed and reviewed and begins to draw conclusions about it. The body of documents is searched, key documents and individuals are identified, and timelines are established. There are a number of specialized tools that can assist in this process. Analysis can also turn up additional individuals and timeframes that were not identified at the outset during the identification phase.

Production

The production phase is where the responsive materials are actually produced and made ready for the requesting party. During the meet and confer, the parties agree as to what format will be produced. Generally speaking, information has to be produced either in its native format, or in the format in which it is most commonly used and accessed. With few exceptions electronic information cannot be produced by printing it to paper or its digital analogs like TIFF or PDF because of the amount of information lost in the conversion.

The other consideration is what to use as the mechanism for access. Will the data be delivered to opposing counsel, or will it be made available through a hosted environment? This latter approach is increasingly popular particularly for complex legislation involving several geographically disparate parties.

Presentation

The final step is presentation of the material in court where it can be challenged. The information may be presented by a technology expert or by someone familiar with how the materials are created (which certainly may not be the same thing). Opposing counsel may challenge how the materials came to be created, how they are managed, whether they can be altered or not, what would be required to alter them, and so forth. As so much of this depends on the nature of the case and the nature of the materials, students are encouraged to seek out additional information from their legal counsel or through any of a number of electronic discovery-related resources available online.

© AIIM – 2019

Failure to Produce

What happens if you cannot produce the information requested? A failure to produce could provide grounds for sanctions, including fines, adverse inference instructions, or default judgment. The key is to make a good-faith effort to produce what is requested.

If there is information that cannot be produced, the first step is to admit that fact to the court and identify why it cannot be produced – for example, that the tape in question was recycled inadvertently or the backup failed. If the information doesn't exist, it doesn't exist. That may not get you off the hook, but it's better to be honest than to dither and hope that it doesn't come up – because it will.

The organization must document any steps that were taken to remedy the issue. That may mean pulling information from backup tapes. It must produce what *can* be produced, with detailed information as to what wasn't produced and why. And it must certainly produce what it can in a timely fashion. A company that produces quickly, shows that it has exercised diligence, and has taken steps to address gaps for current and future litigation may still be held accountable for those gaps but it is likely to be less painful than if it attempts to cover-up or simply pleads ignorance.





Test Your Knowledge

Question 1:

A legal hold has just been placed for invoices relating to a certain project. What step should you take regarding the invoices?

a. Convert them to a standardized file format.
b. Copy them to a flash drive, then write protect it.
c. Set access controls to prevent editing or deleting them.
d. Delete them and notify legal that they have been disposed of.

Question 2:

How can you authenticate data provided by third parties? (select 2)

a. By setting up collection systems with passwords and access controls
b. By establishing and safeguarding the chain of custody
c. Through analysis of the metadata associated with the data
d. By capturing screenshots from the devices and accounts requested

Question 3:

How does information governance affect the discovery process?

a. It makes production of relevant information faster and easier.
b. It ensures that only relevant information is produced.
c. It safeguards attorney-client and other privileges.
d. It provides protection against disclosure of personally identifiable information.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

A legal hold has just been placed for invoices relating to a certain project. What step should you take regarding the invoices?

a. Convert them to a standardized file format.
b. Copy them to a flash drive, then write protect it.
c. Set access controls to prevent editing or deleting them.
d. Delete them and notify legal that they have been disposed of.

The correct answer is C, set access controls. For A, converting the files makes changes to them that could call their authenticity into question. Similarly, for B, copying the files makes changes to the file metadata that affects their evidentiary weight. D will likely expose the organization to significant legal liability.

Answer to Question 2:

How can you authenticate data provided by third parties? (select 2)

a. By setting up collection systems with passwords and access controls
b. By establishing and safeguarding the chain of custody
c. Through analysis of the metadata associated with the data
d. By capturing screenshots from the devices and accounts requested

The correct answers are B, chain of custody, and C, metadata. A, secure collection systems, could ensure no further changes once data is collected but it wouldn't address the authenticity of the data itself. For D, as we discussed, screenshots are unreliable and could be edited.

Answer to Question 3:

How does information governance affect the discovery process?

a. It makes production of relevant information faster and easier.
b. It ensures that only relevant information is produced.
c. It safeguards attorney-client and other privileges.
d. It provides protection against disclosure of personally identifiable information.

The correct answer here is A, it makes production faster and easier. For B, it can help to get to the relevant information but it cannot guarantee it. For C and D, again good governance can help but it cannot guarantee that privilege or disclosure against personal information are protected.

Digital Preservation

Digital Preservation Risk Factors

Almost all physical records require little or no technology to allow humans to extract the information from them. You can pick up an old record, and whether it is written or printed on clay, papyrus, paper, or anything else, you can make out what it says. There are exceptions of course – microfilm, audio, and video recordings all need some sort of technological assistance, but these are all relatively recent developments, and are in the minority.

By contrast, digital records need technology to allow humans to understand them. This includes a lot of different elements:

- Servers
- Disk drives
- Network
- PC
- Screen
- Operating system
- ERM system software
- And so on

The other issue is that software and hardware evolve rapidly, and compatibility issues can start to occur in as little as a single upgrade.

Digital Preservation Issues

You may by now have realized that digital preservation can be broken down into three key issues:

- Storage media obsolescence
- Media degradation
- Format obsolescence

Now let's look at each in a little more depth.

Storage Media Obsolescence

The first problem we'll consider is storage media obsolescence. This refers to the simple truth that storage media, and the devices to read them, tend to fall out of fashion quite quickly. This picture shows a selection of obsolete computer storage media. You can see:

- A short length of punched paper tape
- A single 80-column punched card
- A selection of 5¼ inch floppy disks
- Two different formats of tape cartridges
- An 8-inch floppy disk, the first widely-used floppy disk format
- A 3½ inch 128megabyte rewritable magneto-optical disk



So for each of these, and many others not shown, the key question is where would you find the hardware to read that media today? And it's not just the 5¼ floppy drive itself, for example, but also a suitable power supply, the software drivers required for your current computer to recognize it, the cables to connect it to your computer, and even a hardware port that accept those cables.

© AIIM – 2019

Media Degradation

Media degradation is a straightforward issue. This problem arises because over long periods of time, storage media degrade – that is, they undergo physical and chemical changes. These changes can affect the information stored on the media, resulting in the loss of some or all of the data. This deterioration can be minimized by effective storage techniques; we'll review these shortly.

Format Obsolescence

Finally, the most significant problem: format obsolescence. The problem arises because, like storage media, many file formats quickly fall out of fashion; and worse yet, those that do not are superseded by new versions. Format obsolescence is the most pressing and difficult problem to solve for digital preservation for a number of reasons, but they all come down to how to read the data stored on the media. And the more proprietary or complex the file formats, the more challenges they present to long-term access.

Wherever possible, organizations should use standardized file formats; if those do not meet their needs, they should look at formats with significant market share as they are more likely to be supported over time. Standardized archival formats like PDF/A should be considered for files that need to be retained for significant periods of time.



This illustration shows a selection of software packages, most of which have faded away, but some of which have been superseded by more modern versions.

Storage Media Obsolescence

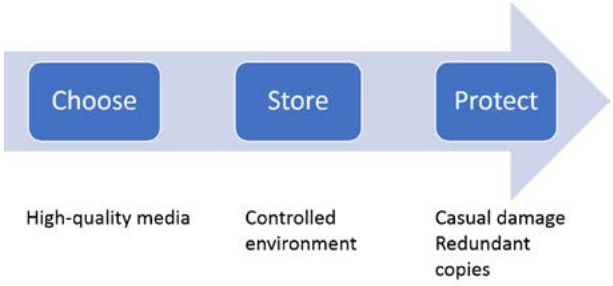
To address storage media obsolescence, there are three main elements.

- Copy information from older media to newer media while the data is still readable.
- Select standardized storage media whenever possible.
- Select current storage media. This means moving from older technologies, of course, but it also means waiting until technology is in general usage. There is no way of knowing whether a technology or solution will survive in the marketplace until it does, and no organization wants to migrate millions of records to the latest format only to have that format fail in the market a year later.

Media Degradation

To address media degradation, consider these approaches:

- **Choose** high-quality media designed for enterprise use, rather than inexpensive media found in consumer office supply stores.
- **Store** media appropriately. For older optical media, this means storing it its case or sleeve. For all media, it means storing it in relatively climate- and temperature-controlled environments.
- **Protect** storage media against casual damage. Again, cases, caps, or other appropriate storage apply here. Don't leave storage media in your car in the heat of summer or in your pocket to be washed. Also consider making multiple copies and storing one copy in a separate location.



Format Obsolescence

The third problem is format obsolescence, that is the risk of not being able to use a file format because of the evolution of software. This is the most difficult problem to solve. There are a number of approaches that could be considered; the three most common are:

- Technology preservation
- Emulation
- Migration

Technology Preservation

Technology preservation is simple to explain. To preserve access to the records that you want to keep, you simply keep a copy of the hardware and all the software needed to access it; and you carefully maintain for as long as you need to maintain access to the records.

First, you may think it is easy enough to keep a couple of PCs – and indeed it may be. But their ongoing maintenance will not be easy. If something goes wrong with one of the circuit boards, or a DVD drive, for instance, it will be difficult or practically speaking impossible to get it mended. And, perhaps ironically, a prolonged period without use is likely to cause some mechanical components, such as DVD drives, to fail.

Second, software maintenance will become progressively more difficult. Preserving obsolete software will mean preserving the ability to fix it if and when it causes problems. This is difficult to arrange over a long term.

Of course, the technology preservation "collection" would grow rapidly too – you would have to keep not just one computer or computer network, but rather one of each uniquely different configuration. You can imagine that the hardware and software involved would, over decades, constitute an unmanageably large and expensive collection.

And finally, it is not clear how records would in practice be accessed if access means using decades-old hardware and software. Would future users have to be trained in decades-old technology? Would they have to know how to operate software of all ages?

Ultimately this is not a viable approach; we include it because organizations still do it. And it works, after a fashion… until something breaks.

Emulation

Emulation is a technique which uses software to make older applications run on newer hardware. For example, old ECM applications that ran on MS-DOS or Windows 95 could be run from Windows 10 or Mac.

One of the key benefits to this approach is that the underlying digital records themselves remain unchanged. This is significant in terms of the trustworthiness and provenance of those records.

However, writing an emulator requires significant expertise with the target environment(s). And as those environments have gotten more complex, the emulators have become more complex to write as well.

And as systems age, documentation and expertise in how to use them becomes increasingly difficult to locate. Consider: Where would you find someone in your organization who knows how to use the Commodore operating system, or MS-DOS, or any operating system other than recent versions of Windows or Macintosh?

Migration

The last approach we will review is migration. In this approach, old formats on old media are converted, or migrated, to new formats and new media. For example, old records created in WordPerfect and stored on CDs are converted to recent Microsoft Word versions and stored on hard disks or solid state disks.

Migration requires a strategy and a detailed process to ensure it goes smoothly.

There will be records that are difficult to migrate – because they are proprietary, because they are complex, or because they are simply not as well-known or understood.

That means that the process needs to include some way to confirm the accuracy of the migration process, at the process level as well as for individual records. It's critical to ensure that the migration didn't cause any loss of functionality, or, if it did, to document those issues.

© AIIM – 2019

In some cases, there may be interactions between records, or between components of records, that serve to complicate the migration further. For example, a CAD drawing may be comprised of many different drawing files which are assembled and rendered as required. In order to migrate this record, the system would need to recognize this fact and locate all of the external reference files. This could be especially problematic if the files are not all stored together.

Finally, you cannot migrate something if you don't know what it is – actually, if you don't know exactly what it is. So the process needs to account for unknown formats.

Despite the above issues, migration is the approach considered most likely to be used, and most likely to succeed, in the general case when preserving access to electronic records. It is widely viewed as the most practical, well-understood and realistic method.

The Digital Preservation Strategy

This definition for the digital preservation strategy comes from The National Archives of the UK.

The aim of your digital preservation strategy should be to achieve consistency in the management of digital records. Your strategy should identify the actions required for active preservation and provide:

- A formal means of accepting records, including an agreed standard for file formats and levels of description for the records. The bias here should be towards the use of open, standard formats for files and storage media.
- A secure process for transferring records into storage and then managing them appropriately (including integrity checks).
- A way of mapping processes to capture the descriptive information into a searchable database linked to the records to allow them to remain discoverable.
- A formal means of providing the content of the preserved records to users in the most appropriate format for the content of the record.
- A rigorous system for monitoring the preservation activities that can produce usable audit data.

Immediate Actions

The first rule is: "Know your holdings." You will not make the right decisions on digital preservation unless you know about the records you hold. There are six main things you need to understand:

- What file formats your records are in, and what storage media they are stored upon
- How the mix of file formats may change in the future (to the extent that this is known or can be predicted)
- The physical storage conditions of the records and all their backups
- How long they have to be retained
- Their value or importance to the organization (or to whoever else has an interest in them), and how to determine that value over time
- How information is used today – and may be reused tomorrow, often in unforeseen ways

These six steps form the foundations for almost any preservation activity. Anyone wanting to make a start, or indeed wanting to confirm that no action is needed, should start here.

Next Steps

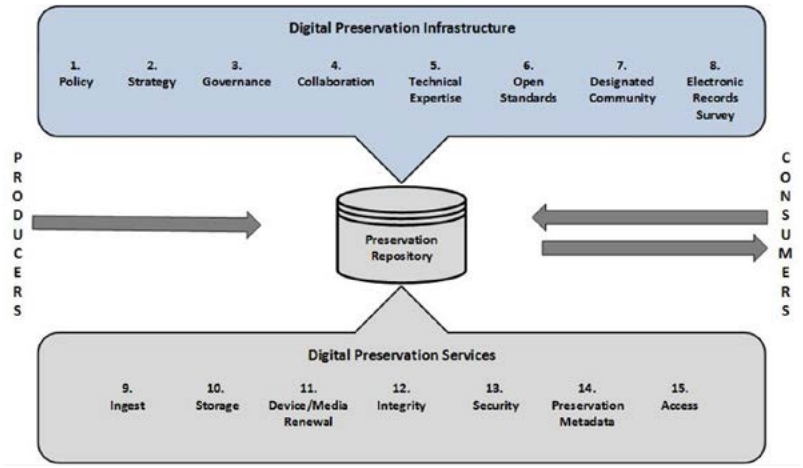
Next, it's important to assess the capabilities of systems that are currently used to store digital information over the long term and identify gaps. There are a number of assessment tools available including:

- Library of Congress – NDSA Levels of Preservation
- Digital Repository Audit Methods Based on Risk Assessment (DRAMBORA)
- Audit and certification of trustworthy digital repositories (ISO 16363:2012)
- Digital Preservation Capability Maturity Model (DPCMM)

Let's look at this last tool in a bit more detail.

The Digital Preservation Capability Maturity Model

The Digital Preservation Capability Maturity Model (DPCMM) was developed by Charles Dollar and Lori Ashley to allow organizations to assess their digital preservation practices and compare them to best practices as outlined in ISO 14721, Open Archival Information Systems, and ISO 16363, Audit and Certification of Trustworthy Repositories.



Source: Charles Dollar and Lori Ashley

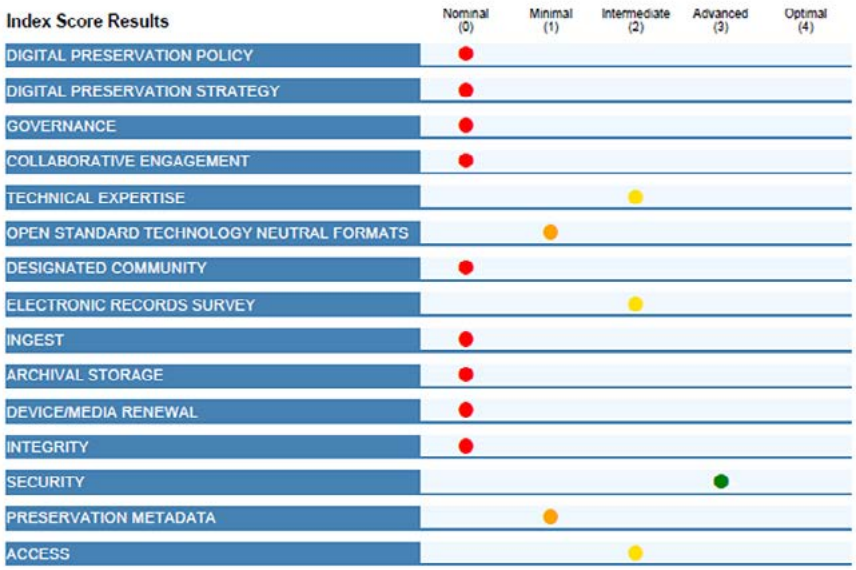
The model consists of 15 components, grouped into two categories:

- **Digital preservation infrastructure.** These components focus on planning and support for the initiative.
- **Digital preservation services.** These components focus on the specific actions required to effect digital preservation within the organization.

Each component is scored from 0 to 4 and then the aggregate score is used to determine the overall maturity of the organization's digital preservation practices. Each level of each element offers specific guidance on the target capabilities.

The DPCMM Assessment

And here's the report generated using the DPCMM assessment. You can see that this particular organization has a very low level of digital preservation maturity across the entire framework.



© AIIM – 2019

Next Steps

Once you identify your risks and gaps, the next step would be to develop a strategic approach to digital preservation. Preservation actions are rarely very urgent, but are likely to be costly in many cases. So, a measured, strategic approach is called for. The strategy may involve collaborating with other organizations or departments within the organization to spread costs, initiating research internally, and a range of different actions. The sorts of actions a strategy might contain include:

- The adoption of standard file formats, having chosen one or several preferred file formats which are best suited for your organization's likely future needs. Note that this might imply migrating all newly-received records into preferred formats on receipt, or retrospectively migrating existing records to these preferred formats, or both.
- The adoption of preservation standards, including standards for processes, such as ISO/TR 18492, and file formats including but not limited to ISO 19005, PDF/A, and ISO 32000, PDF.
- The creation of – or subscription to – a technology watch function. The idea of a technology watch function is that it routinely scans the technology scene with a view to spotting when technologies you use may be about to become obsolete, and also to evaluate new technologies and file formats in terms of their preservation potential.
- And finally, the organization should integrate the preservation strategy into the overall information governance framework.

Example Digital Preservation Strategy

In 2006, the Online Computer Library Center, or OCLC, outlined a concise digital preservation strategy consisting of four stages.

- Assess the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications. For example, engineering drawing formats such as AutoCAD often include multiple views and layers, multiple externally referenced (and pathed) drawings, and complex three-dimensional models. Similarly, an Excel spreadsheet can include formulas, calculations, links to other worksheets and workbooks, and other complex and proprietary capabilities. Both of these are leaders in their respective markets, so they are less risky in the short term compared to lesser-known and –used formats, but over extended periods they are at significant risk.
- Evaluate the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied. Some file formats may be readily convertible to more open or standards-based file formats – for example, PDF or TIFF to PDF/A. Others will be much more complex as noted above. Regardless, the strategy should also indicate which preservation approaches to use such as emulation or migration.
- Determine the appropriate metadata needed for each object type and how it is associated with the objects.
- Take whatever actions are required to ensure access to the content – now and over time, by following the strategy.

OCLC, and similar archival-focused organizations, offer a wealth of detailed resources for planning a comprehensive digital preservation strategy.



Test Your Knowledge

Question 1:

What is the most effective approach to minimizing digital preservation issues?

a. Keep a copy of all hardware and software used to create digital records.
b. Select standard formats whenever possible.
c. Convert non-standard formats to PDF.
d. Make copies of important digital files and store them offsite.

Question 2:

What actions should be taken first to identify the organization's digital preservation issues? (select 2)

a. Identify current file formats and storage media.
b. Convert all digital files to standards-based formats.
c. Determine the value and importance of existing information.
d. Create a technology watch function.

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What is the most effective approach to minimizing digital preservation issues?

a. Keep a copy of all hardware and software used to create digital records.
b. Select standard formats whenever possible.
c. Convert non-standard formats to PDF.
d. Make copies of important digital files and store them offsite.

The best answer here is B, select standard formats whenever possible. This applies to media as well as file formats. A is the technology preservation approach which as we noted has very limited value over time. C, converting non-standard formats to PDF, may be useful in some circumstances but many formats don't lend themselves to that conversion. And D, making copies and storing them offsite, does not address the actual preservation issues and in fact makes them worse by proliferating additional copies.

Answer to Question 2:

What actions should be taken first to identify the organization's digital preservation issues? (select 2)

a. Identify current file formats and storage media.
b. Convert all digital files to standards-based formats.
c. Determine the value and importance of existing information.
d. Create a technology watch function.

The best answers here are A, identify current holdings, and C, determine their value and importance. B, convert to standards-based formats, may not even be feasible but would certainly not be the first step. D, create a technology watch function, is also helpful but would generally happen much later in the process.



Implementing an Information Management Solution

Introduction

This domain focuses on the general processes for selecting, designing, and implementing an information management solution. It follows the implementation process from start to finish:

- Developing the information management strategy
- Making the business case for information management
- Identifying and prioritizing the business requirements for the solution
- Designing and implementing the system
- The change management required to ensure uptake of the system

All implementations will follow these steps to varying degree. That said, there are a lot of differences in the details of a particular solution. For example, an enterprise content management (ECM) solution will definitely have different requirements compared to an engineering drawing management solution; and a cloud solution will have some different steps compared to an on-premises solution.

It's also important to note that many of these steps would apply to other, non-technology-related information management projects as well. If you're developing or updating an information governance framework, a policy, a business process, etc. there's still planning and strategy, a business case, requirements, design and implementation, and change management involved. Similarly, if you're developing a center of excellence, or a steering committee, it still requires many of those steps as well.

Finally, we try to underscore throughout this section that when implementations fail, it's generally because of people, not the technology itself. What we mean is that technology projects generally fail either because of insufficient planning, or because users can't or won't use it because they aren't trained, they don't understand the solution, it's too sudden or radical a change, etc. That's why change management – training, awareness, communications, organizational culture, etc. – are so important to a successful implementation initiative.

The Benefits of Intelligent Information Management

Most organizations and departments have a mandate to continuously improve operations. A *conventional change agenda* involves better tools and technology, better behaviors, and better processes, all focused on generating better efficiencies and improved productivity.

Ultimately, AIIM believes digital transformation is *more than conventional change*. Digital transformation is about doing things *differently* – and doing *different* things as well. And different not just for the sake of being different, but in support of the key strategic objectives facing every organization in the age of digital disruption. In other words, digital transformation is not about incremental process improvement. Digital transformation is about using information in brand new ways.

We believe that digital transformation will allow organizations to:

- **Enable innovation –** thereby bringing new products and services to market more efficiently
- **Enrich customer experiences,** including personalization and support for privacy requirements
- **Execute processes nimbly and on-demand,** through process automation
- **Engage the next generation of employees,** to ensure access to information anytime, anywhere, on any device
- Minimize the risks and costs of information-intensive processes

This means that every organization is on – or should be on – a digital transformation journey. In recent AIIM research, we found that 79% of organizations realize that they must transform into a true digital business in order to survive.

© AIIM – 2019

The Strategic Benefits of Intelligent Information Management (IIM)

As the currency that fuels and funds the journey, information is an organization's most valuable asset.

This also means that, as the outline above suggests, the focus for information management is broader than simply reducing information-based costs and risks. While this is important, it is insufficient. Rather, organizations need to focus on how to effectively monetize their information assets, directly or indirectly, to move the organization forward. Information management must become a business enabler.

Intelligent Information Management – Definition

We believe that there are four capabilities that are required to meet the challenge of digital transformation. We refer to these collectively as Intelligent Information Management:

- Modernizing the information toolkit
- Digitalizing core organizational processes
- Automating compliance and governance
- Leveraging analytics and machine learning
- Now let's look at each of these in more detail.

Modernize the Information Toolkit

In a 2013 research note, Gartner argues that "Traditional IM technologies and approaches are under pressure from exploding volumes and diversity of information. Big data forces will push the information infrastructure in most organizations to the breaking point, leading to challenges and disruption in the business. Silos, poor-quality data and conflicting semantics are the norm. A modern information infrastructure helps organizations describe, organize, integrate, share and govern information assets independently of applications and use cases…."

We're not saying throw out all your legacy systems and information. What we are saying is that those legacy systems present significant challenges to the organization. These include, but are not limited to:

- The real costs to support those systems in terms of licenses, maintenance, and support
- The IT costs associated with them
- The individual end user costs in terms of finding and using information stored within them
- The challenges associated with accessing information across discrete siloes of information

Instead, organizations need to focus on modernizing their information infrastructure, with an emphasis on:

- Flexibility – configuration rather than customization.
- Loosely coupled systems rather than tight integration between them.
- Automation in support of common user-centric and management tasks.
- Mobile - and cloud-enabled, and ideally designed with mobile as the preferred platform.
- Rationalizing and streamlining the existing environment. Where legacy systems can be decommissioned, they should be. The information stored therein should be appraised for its continuing business value and any legal or regulatory compliance requirements, with a bias towards disposition.
- And balancing legitimate governance needs and issues with the overall business goals and objectives. This means that when evaluating a cloud-based solution, the discussion should include security, where data is physically located if applicable, etc. but with an emphasis on how to make a solution work rather than focusing purely on the risks.

Digitalize Core Business Processes

Too many organizations find their operational processes and efficiency stifled by too much paper and manual tasks. Unscalable, out-of-date legacy systems, and a continued dependence on paper-based information, result in inefficient manual tasks and error-prone data entry, which cause expensive and duplicative re-work.

According to AIIM research from 2017:

- 51% of organizations cite too many manual business processes as a significant limitation to operational efficiency
- 59% cite manual data (re-)entry and search as major shortcomings to their business processes
- And 32% are struggling with non-digitized (paper-based) processes

Organizations that work at "the speed of paper" are increasingly being rendered non-competitive and irrelevant. On the one hand, organizations need to "digitize everything that moves" – as early in the process as possible. On the other hand, every effort should be made to "digitalize" processes by working digitally as much as possible: born-digital documents stay digital; their flow into and through the organization is streamlined and automated; and how they are managed and ultimately disposed of is automated as well.

Automate Governance and Compliance

The amount of digital information and digital records present in most organizations, and the continued explosion in the rate of growth of them, means that manual information management processes no longer work… if they ever did.

Consider that many organizations today still use network file shares like a "G:\ drive" to store, share, and collaborate on documents, some of which will become records. Users store and organize these documents in the way that makes sense to them in their work context. And they save them for long periods of time "just in case." Now fast forward five years, and those file shares might have tens of thousands to millions of documents on them for a larger organization.

Records management best practices suggest that none of those documents should be deleted until they are reviewed to make sure they aren't records, they aren't applicable to any open legal or regulatory issues or audits, etc. But where do you find the time, and the staff, to review 1 million digital documents, many of which are the same document in different versions, renditions, or simply identical duplicates? At just 15 seconds to open, scan through, and make a decision about each document, that's nearly 4,200 hours of staff time or more than two full years to get through them all. And 15 seconds is absurdly low. And in the meantime, more and more documents are being created and shared.

Perhaps more importantly, users don't want to be information managers. They want to do their assigned job tasks with minimal interruption for administrative tasks.

Because of these two issues, the only way for organizations to ensure they are in compliance with the things they need to comply with is to automate governance and compliance. Policies and procedures have to be turned into business rules that can be automated.

Leverage Deep Learning

Finally, intelligent information management requires that organizations leverage deep learning and machine learning. These are both subsets of artificial intelligence that can extract not just data but meaning and insights from structured and unstructured information. There are already a number of use cases for how deep learning can benefit organizations, ranging from automatic recognition and translation of text in images, to image searching, to fraud and threat detection.

Machine learning has been available for a number of years, but only recently have powerful computing capabilities come down enough in cost to support widespread deployment. Deep learning is still emerging as a set of capabilities available to organizations. In either case, it's important for organizations and records managers to know that, while these capabilities are starting to become available and can be quite powerful in some contexts, they are still not 100% perfect – at least, not yet. Organizations should evaluate these tools with a focus not on perfection but whether they can be leveraged to improve how information is managed and how better information can support ongoing business processes.

© AIIM – 2019

The Impact Areas of an Information Management Initiative

You will now look at the key aspects of the organization that need to be considered and planned for when implementing information management. These include:

- Development of information management processes and procedures across the organization
- Introduction of new ways of working
- Take-up of the new environment by users
- Changes in business processes
- The importance and cost of change management
- And the typical benefits achievable as a result of IM programs, in both small and large organizations

As we go through them, you might find it helpful to pause the presentation to write down your initial thoughts about the impact of your own IM-related program in each category.

Impact – the IM Framework

The first impact to look at is the development of the information management framework. The framework is designed to do a few things:

- Outline the overall goals for effective information management throughout the organization
- Identify major areas of responsibility for information management, e.g., the roles of management vs. end users vs. specialists
- Support consistent information management practices across the entire organization
- Identifies relevant governance requirements, for example privacy practices or mandatory recordkeeping requirements

As you might expect, development of this framework is itself a significant effort that will require participation and resources from across the organization. We address these elements and others in more detail elsewhere in this course.

Impact – Ways of Working

For most users of a new information management system, the ways of working represent the biggest change introduced by the program. For example:

- Users must learn to save information into an information management repository.
- They need to learn which information goes in which repository.
- And if the user wants to share information with colleagues, they can send them a link to the version in the IM repository, not an email attachment.

There are many such examples. The program should develop process models for routine activities that involve the relevant elements of information handling, and review them against the capabilities of the IM system.

Impact – Take-up by Users

Now consider how a move towards IM can benefit organizations of different sizes. The benefits of an IM-related program are strongly linked to the level of its adoption and use across the organization.

That is to say, if the scope of the program is extensive, availability is good, and the users accept the new environment, this will lead to better information and content management. This can provide greater productivity benefits from information sharing and the re-use of information

Under a new IM-related environment, many everyday office activities may well be very different from how the users do their work currently, and change tends to cause resistance, unless changes are introduced sensitively.

One particular irritation to end users may be that they are required to enter an array of new metadata attributes whenever they save their work. This can be reduced by attention to good design and configuration, and verified during demonstrations and piloting a new system.

The key is always to consider the end users *first*. If they hate the system, or feel they are being overlooked, they won't co-operate and the whole program will fail.

Impact – Benefits

In another module, we explored the benefits of intelligent information management. In brief, they include:

- Being able to rationalize and modernize information infrastructure
- Protect information assets, including customer data
- Take advantage of cloud and mobile
- Digitize core business processes and workplace
- Automate governance and compliance
- Leverage deep learning and machine learning capabilities

Taken as a whole, this really means that effective information management will help organizations to enable and support business goals and objectives; enable monetization of information assets in some cases; and reduce information-related costs and risks. And as these processes and tools spread throughout the organization, the benefits to the organization will grow.

Impact – Business Processes

Information management offers the opportunity to replace slow and cumbersome paper-based business processes with much faster and more secure electronic ones.

Many processes can be automated by using workflow with an IM system. This may be one of the most fruitful applications within IM; many processes that used to rely on paper forms, the internal mail, and an army of clerks can now be done via the network. Examples within your organization might include travel claims, hotel booking, transport arrangements, and timesheets. The only people who need to be involved in these types of transaction are the originator of the transaction and the approver. And both can complete their parts of the process on-line with no paper forms required.

Automation of business processes is regarded by many as a major cost benefit to expect with IM, but few organizations can claim this as a mature capability. It is not often properly exploited.

Impact – Change Management

As indicated already in this course, change management should be a high priority in your IM program. Far more programs fail because of inadequate resourcing for change management than fail for technical reasons.

The keys to successful change management are:

- Managing communications
- Providing relevant training, and at the right time
- Providing support

We will discuss this in more detail later in the course.

© AIIM – 2019



Test Your Knowledge

Question 1:

What are the primary strategic benefits associated with improved information management? (select 2)

a. Reduced risks and costs
b. More accurate analytics
c. Improved innovation
d. Faster implementations

Question 2:

How will improved information management have the greatest impact on the organization? (select 2)

a. Reduced information-related costs
b. Faster implementation of information management solutions
c. Changes to how users do their work
d. Less need for training and communication



Test Your Knowledge - Answers:

Answer to Question 1:

What are the primary strategic benefits associated with improved information management? (select 2)

a. Reduced risks and costs
b. More accurate analytics
c. Improved innovation
d. Faster implementations

The best answers here are A, reduced risks and costs, and C, improved innovation. Better information management may lead to more accurate analytics and faster implementations, but not necessarily, and in both cases these are more tactical outcomes than strategic benefits.

Answer to Question 2:

How will improved information management have the greatest impact on the organization? (select 2)

a. Reduced information-related costs
b. Faster implementation of information management solutions
c. Changes to how users do their work
d. Less need for training and communication

The best answers here are A and C. Improved information management practices in and of themselves will have little impact on how long a given solution takes to implement – that's more a function of the particular solution and the scope of the implementation. And improved information management will not reduce the need for training and communication – to the extent it affects it at all it will likely increase the need for it.



© AIIM – 2019

The Information Management Strategy

The first major step in the development of an information management implementation plan is the creation of an IM strategy.

A program strategy outlines the types of activities needed to fulfill the mission of the program. An effective strategy and business case will clearly articulate the vision for the IM deployment, provide direction and an action plan. The IM program strategy will form the basis for how your organization will fulfill the business objectives.

The IM program strategy is an important foundation. It is a communication vehicle, where all stakeholders agree on the vision for IM program and how it supports organizational goals. Do not invest significant research and analysis, however, at this early stage. The strategy is meant to be a high-level document in which the proposed IM environment is outlined. The specifics of the actual IM deployment will evolve over the life of the project, as business and technical requirements are assessed.

The IM program strategy is created at the request of an executive manager or governance board before the program progresses. Stakeholder approval and executive endorsement is a key step. No IM program should begin without this endorsement. Lack of endorsement runs the risk of the work beginning without appropriate funding or management support.

Three Key Areas to an IM Strategy

Three key areas need to be explored to define the IM program strategy. Our goal is to transition from the current state, to a desired future state, and that will take time, tools, change management, and execution plans.

First, identify your organization's current state. This will help in setting baselines as part of success measurements and identify what is needed across the organization.

Second, do a review of the organizational environment. Identify the behaviors that need improvement, understand what support will be available to the IM team for various project phases and activities, and document what tools and processes are currently in place. Documenting this insight into the current environment will help get clarity on the scope of work and change needed to achieve the third element – the desired future state.

This final stage of strategy definition is to document the future state. Make this a collaborative, iterative activity, defining the model with ongoing consultation of stakeholders.

Developing an IM Strategy

The effort needed to develop an IM program strategy will vary from organization to organization. Larger organizations or those with complex practices and processes will require a larger effort.

Organizational culture is a factor. How long does it take to get things discussed and approved? Another important factor is the scope and nature of the program. The amount of stakeholder consultation needed will affect the effort. Buy-in to the program objectives, scope and success factors will be important to achieve goals. It is better to consult stakeholders at all stages, and to communicate findings regularly and clearly in order to ensure ongoing support and funding.

Resource allocation is the final factor we'll review. Ideally, the IM program strategy will be produced by a single person, or a small team, to ensure continuity and consistency. A range of skills are needed, such as good understanding of the organizational culture and business, the ability to consult and elicit agreement, familiarity with the technological possibilities, and understanding of the potential outcomes.

Five Steps to Developing an IM Program Strategy

The elements to be described in the strategy should summarize the needs of the organization, as well as the constraints that will be faced. Let's now step through the elements of a strategy that should be reviewed, documented and communicated to stakeholders.

First, articulate the business vision. The business vision will help shape what the future state could look like after delivering the IM program. Interviews with executives or other internal stakeholders should capture their business objectives, and express an understanding of the competitive forces, market position, customer and revenue cycles, supplier ecosystems and company success factors.

Second, outline the critical success factors that will determine how the program can be assessed.

Third, identify the key performance indicators (KPIs) that will be the measurements and data used to assess how well the critical success factors are met. As we've discussed elsewhere in this course, KPIs for an information management strategy could include reductions in storage costs, reductions in time or cost to particular business processes, increased innovation, and many others.

Fourth, establish success measurements so that tangible progress can be measured on a regular basis. Showing clear and realistic reporting and communication plans to stakeholders will improve buy-in.

Finally, the strategy should outline the drivers for change. What imperatives are compelling the business to improve its handling of content, encourage better collaboration or communication, or raise productivity levels? Sources of this information will be senior management interviews, external industry or market research, internal business or marketing plans.

Define Key Stakeholders

After meeting with the executive sponsor, key stakeholders should be identified and added to the interview list.

As with any initiative, it is vital to gain the support of key stakeholders for success. This means making the business case to key stakeholders, so they understand its importance to the organization as a whole and to their particular sphere of influence. It also means ensuring that their concerns are identified and addressed as part of the overall project. And it means ensuring that the business drivers and benefits have been articulated in a way that makes sense to them.

One way to do this is to put stakeholders, or their representatives, on the project team. Many of the stakeholders may be too senior to participate actively, but their interests should be represented by someone from their business unit. This will also help to ensure identification and resolution of conflicting priorities within the implementation program and among the overall project portfolio.

If the stakeholder is not "officially" on the project team, ensure that they can be heard during the process, in order understand their roles and how they will be impacted, or wish to be impacted by the project.

Management Commitment

Consider the levels of commitment and roles of middle and senior management in an IM-related program.

Implementing an IM-related program is often a strategic decision, certainly if it is large or critical for the organization. It is possible that it will impact on every person working in the organization. As such, top level support and commitment is essential. If this is not clear at the outset, it is likely to undermine later stages in the program and may in fact trigger the first signs of failure.

Management must lead by personal example. If for example a senior manager is seen to ignore the IM-related working practices, it may well be interpreted by users as a sign that the organization is not really serious about the new environment. If the ways of working are mandated for everyone, this must include managers all the way through to the top. Users tend to resist change: if they think it is optional, they won't do it.

Information Management Roles and Responsibilities

Information governance roles can be classified into three main categories.

The first category is that of information owners. This includes setting policy, establishing governance, matching any Information management requirements to the needs of the enterprise, etc.

Then there are administrative roles that support both the user community and the broader organization. One specialist role is that of the information manager; another might be the administrator of a content management or digital recordkeeping application if one is present. These roles act as stewards of information and ensure its availability.

Exploiting the information involves direction from the business; these roles might include business managers, knowledge management, audit and compliance, and other groups within the organization that leverage Information assets to accomplish the goals and objectives of the business.

Around all of this, and possibly extending outside the organization, there are the users.

Note that the roles are described as though they are individual roles, but in fact they are often combined within a number of existing roles, depending on the existing set up and size of the organization.

© AIIM – 2019

Program Owner

The program owner acts on behalf of the organization to ensure that the program aligns with the overall organizational strategy. The owner is personally accountable to the board and senior management for delivering the benefits of the program.

The owner will work with the program steering committee to determine which projects are needed to establish and/or further the program and to prioritize among them. The owner also owns the business case for the program.

For many organizations, the program owner will be the CIO, general counsel, the head of the governance, risk, and compliance (GRC) function, or wherever organizational responsibility lies for governance and compliance. This is because the information management program should be enterprise-wide in nature, rather than many initiatives that are conducted at the business unit level, and because many organizations are addressing Information management out of concern about litigation and compliance rather than specific operational concerns.

Business Unit Managers

The business unit (BU) managers are responsible for their own departments, divisions, agencies, or other types of business units. In some cases, they may be responsible for a process rather than a specific department. They are certainly responsible to senior management for the conduct, outputs, and performance of those business units or processes, and for everything that goes on within them.

In the context of the IM program, the BU manager is responsible for communicating the benefits of effective Information management to the business unit staff. The benefits will come from the business case initially; the BU manager will need to translate those benefits into something the employees can relate to.

The BU manager is also responsible to the organization for ensuring that employees of that BU comply with the requirements of the program. That means that the BU needs to have some way to measure and audit compliance and needs to put a plan into place to address any gaps that appear as the program continues.

Finally, the BU manager is responsible for ensuring that staff members are trained on the components of the program, including policies and procedures. That training might be developed and provided by HR, a third party, or someone within the BU, but it is important that it be delivered initially, and that staff receive periodic refresher training.

Information Technology

Information technology, or IT, plays a significant role in the overall information management initiative. We should note that IT here could refer to internal IT staff, whether centralized and enterprise-wide or decentralized at the business unit level, or it could refer to an outsourced IT staff.

IT installs, configures, and maintains the systems that make up the technology portion of the IM system. These systems include, but are not limited to, the applications that create documents and records, any archiving applications, the backups for each of those, and any other information management solutions in place, including hardware and software. In the case of cloud-based solutions, IT still will likely play a role in selecting, provisioning, and supporting those systems.

IT also has to ensure that any solutions recommended for implementation, whether related to IM or any other technology-enabled process, fit within the overall organizational IT architecture. For example, some IM applications solutions only work with Oracle databases; this would likely meet with resistance if the organization relies on an otherwise Microsoft SQL Server-based IT architecture.

Records Management

The records management (RM) function is another key role for an IM program. Records management does the research to understand the unique records retention and compliance requirements of the organization. Records management provides input as to the value of that Information according to legal or regulatory requirements that must be complied with.

Records management drafts records management policies and procedures for the organization, including file plans, retention schedules, disposition instructions, and others.

Finally, records management should review the IM environment to ensure that it supports those retention requirements. By environment we mean people, processes, and technologies. So for example, then, SEC 17a-4 requires that certain documentation be stored on non-erasable, non-rewritable storage. A records manager at

a financial services firm in the U.S. would know that this particular requirement applies to the organization. The records manager may or may not know which technologies support this requirement, but could identify it *as* a requirement and have a discussion with IT as to how to satisfy it.

Legal

The legal function is an important contributor to the governance of the IM program; indeed, many of the requirements for governance are related to legal issues including compliance and discovery.

The legal function is the most likely to receive notice of any requirements to hold and/or produce Information, including subpoenas, notices of intent to pursue legal action, legal holds, and the like. If a regulatory agency requests information of the organization, that request is also likely to be directed to the legal function. Legal, in turn, must communicate any requirements for holds on disposition, archiving, recycling of backup tapes, etc. to the relevant parties in the organization; because of the scope of electronic information creation and usage, this would probably include records management, IT, and business unit managers and may include all of the users in the organization.

Records management generally drafts records-related policies and procedures; however, it is generally legal (or in some instances senior management) that actually signs off on policies and procedures, thereby clearing the way for their implementation.

Business Users

And of course we have the business users – the ones who are actually doing the work of the organization and who are expected to comply with the information management program requirements. One of the key points to remember is that the business is in the business of the business. That is, every organization has a mission or charter of what it exists for. Users are hired and paid to do their jobs, and while IM is and should be a part of that, engineers have to engineer, sales people have to sell, etc., so some attention needs to be paid to the balance between primary job responsibilities and IM responsibilities. A great way to address that is through automation; we discuss that in much more detail throughout this course.

Support Structures and Roles

As we've discussed earlier, there are some other support structures and roles that can offer significant support during and post implementation of an information management solution. As a brief review, these include:

- **The steering committee**, which is comprised of key stakeholders or their representatives and determine the strategy for the initiative.
- One or more **Centers of Excellence (CoE)**, which consist of subject matter experts that provide guidance and best practices.
- One or more **Communities of Practice (CoP)**, which consist of super-users and other interested users and that provide peer-to-peer support in a particular area.
- **Coordinators**, who serve as liaisons between their individual business functions or processes and the IM groups.

Other Stakeholders

Other stakeholders include customers, partners and suppliers, and even the general public for example for government agencies. All of these stakeholders might be currently impacted by the organization's information management practices and will almost certainly be impacted by any changes to the IM program. This raises an interesting question: how do you capture THEIR needs and requirements for the program?

One way to do this is to include on the project team internal users who interact most often with those external users, such as sales, customer service, or technical support. Another way is to simply ask them, such as by sending out a survey or contacting selected customers and partners. And there are certainly other ways. The important point is to note that the IM program will impact them as well, at least indirectly, and the organization does not want to implement the program in such a way as to adversely impact them.

© AIIM – 2019



Test Your Knowledge

Question 1:

What is the purpose of an information management strategy?

a. To describe the business requirements for an information management solution
b. To identify the key stakeholders for an information management program
c. To gain executive buy-in for an information management initiative
d. To assess the organization's awareness of information management best practices

Question 2:

Which roles are key stakeholders for an information management initiative? (select all that apply)

a. Records management
b. Business unit managers
c. Local government agencies
d. Information technology



Test Your Knowledge - Answers:

Answer to Question 1:

What is the purpose of an information management strategy?

a. To describe the business requirements for an information management solution
b. To identify the key stakeholders for an information management program
c. To gain executive buy-in for an information management initiative
d. To assess the organization's awareness of information management best practices

The primary purpose of the IM strategy is C, to gain executive buy-in for an IM initiative. While it will list the key stakeholders, that is not its main purpose. A and D are generally outcomes of the strategy.

Answer to Question 2:

Which roles are key stakeholders for an information management initiative? (select all that apply)

a. Records management
b. Business unit managers
c. Local government agencies
d. Information technology

The best answers here are A, B, and D. Local government agencies would not generally be a key stakeholder for a specific IM initiative in most organizations. This may vary for particular entities, especially highly regulated ones or those where local government imposes strict compliance requirements, but RM, business unit managers, and IT will always be key stakeholders for an IM initiative in every sector and organization.



© AIIM – 2019

The Information Management Assessment

The Organizational Assessment

This assessment will generally include a look at the business and operational environment; the regulatory environment; and the organizational culture.

The organizational assessment is an early and important step in developing or updating an information management program. It's difficult to figure out how to get to the end goal of effective information management if you're uncertain as to where you're starting from. The assessment helps to ensure that key stakeholders, the project team, etc. understand the starting point for the initiative and the environment in which the initiative will take place. It also helps to identify the most critical issues to address and to prioritize among competing needs.

Complete Key Assessments

The first step in any process analysis is to review the context in which the organization operates. This consists of two steps: a review of the regulatory environment and a review of the operational environment.

Here we outline the elements that will inform the regulatory environment. At the top are statutes, regulations, and case law with the force of law that describe regulatory requirements.

Next are any mandatory standards of practice, which might be found in some industry sectors but not others.

Next come voluntary codes of conduct and codes of ethics which the organization has agreed to abide by.

It is also important for the organization to adhere to its own rules and procedures – in fact, within information management generally there is an argument to be made that it is better to have no policy at all than to have one and not follow it.

And we end with community expectations. This may seem a bit odd at first glance, but consider that public sector organizations are considered to have a primary responsibility to their citizens and an expectation of openness, transparency, and accountability even where there are no specific statutes. For private sector organizations, there is an expectation that the organization will not harm the public, or its shareholders, or its employees, even where that harm is simply bad publicity.

The essential information to collect during this business assessment phase will include the data necessary to rate the organization's current standing in these key areas: the information flow and delivery across the organization, the maturity levels of information and infrastructure systems, information processes, and people skills in the org chart.

The information flow and delivery assessment maps how content moves across the organization, through all phases of the content lifecycle. It should include content originating inside and outside of the organization and maps inputs, outputs, systems, content and people. The deliverable is a document to capture this flow and delivery.

The information and infrastructure maturity level assessment help ascertain how engrained information management practices are within the business and IT functions. The maturity model is a useful tool for this analysis, and the output should be a mapping of strengths, weaknesses and the gap between current and desired states.

Understanding the Need

The first step in the business assessment is to understand your organization. It's necessary to find out what the needs are, what's good, what's bent and what's broken. This is often mistakenly referred to as a gap analysis. This organizational scan is a fact-finding mission first; the analysis will come after you have completed the assessment.

One of the tools we can use for this assessment is a "Maturity Model." A maturity model describes the behaviors, practices and processes of an organization related to a given discipline; in this case, information management. A maturity model can provide several useful inputs or lenses through which to view your organization. It can provide a way to articulate your starting place, a shared vision of how to evolve, a framework for prioritizing actions, and a way to define what improvement means for the organization.

A maturity model can be used as a benchmark for comparison and as an aid to understanding – for example, a comparative assessment of this organization versus another, or compared against itself over the years. Performing a maturity model review every regularly is a common practice to measure one's progress over time.

Identifying the Gap

The scoring levels described in the previous section will be assigned to the current or "As is" state. This indicates where the organization is today with regards to that particular dimension. The second half of the dimension assessment is to determine where the organization NEEDS to be.

These scores are decided by the team inside the organization and should be led by those working on developing the IM program. It is a self-assessment, where the participants are guided through the process. This is ideally done through a series of workshops, one to gather the ratings and a second to validate what the facilitator gathered and interpreted from the workshops.

The difference between these two "scores" is the gap. Your next step is to identify the level of effort required to close the gap. This is a true "Gap Analysis."

Complete Key Assessments

The information process assessment should document the business process for key activities focusing on information management. Techniques can include interviews, reviews of procedures or flow diagrams. The outcome should be a documented high-level information process reflecting the current state, using flow charts or process models.

Finally, the people skills and organizational structure assessment should determine if the team and the organizational structure is able to deliver on or support the project, compare resources and expectations. The outcome should result in capacity document and skills matrix, and be an important input to future training plan development, hiring plans, contractor qualifications, or potential technology acquisition.

The Technical Assessment

As with the organizational assessment, you need to know where you are before you can determine technology needs. This assessment includes a review of the existing technical environment, what systems and data stores are present, and their stages in the systems lifecycle.

What is a "technology assessment?" This second phase of our implementation lifecycle will concentrate on the technical aspects of the information management deployment strategy. The goal of technical assessment phase is to develop a technical strategy to complement the business strategy. It identifies the elements of the existing technical architecture and the extent to which they impact or will be impacted by the information management initiative including any specific constraints. It ensures that the business requirements are tied to the information infrastructure – and that the infrastructure will meet the business needs of the organization.

We'll step through a range of activities to determine and document the technical requirements needed to support the established business requirements. Ultimately the technical assessment and the business assessment are key activities to complete before beginning design and development activities because they will directly impact those activities.

Business Drives Technology

Business drives technology, not the other way around. While technology should be a means to find ways to do business, become more efficient and creative and grow more quickly, it is the goals and objectives of the organization that need to be front and center during planning. An iterative approach to gathering both business and technical requirements can help ensure that this interplay between the two is assessed and captured.

Most technology projects fail – meaning they didn't accomplish the expected business outcomes on time and on budget. But the simple fact is that most of them fail not because of a specific failure of the technologies involved. Rather, it's generally due to one or more of a lack of planning, not having the right people involved, ineffective business processes, or a failure to understand technology and its limitations.

© AIIM – 2019

Identify Applications

The first step in the technology assessment is to identify all of the applications in the organization that create or manage information. These applications can be grouped for consideration into five categories.

- **Enterprise.** These are applications that are generally provisioned by a central IT staff and made available to the entire organization. The most well-known example of this today is probably email; these might also include customer relationship management applications, financial software, enterprise resource management, and the like.
- **Business unit.** For larger organizations these might be similar to the enterprise applications – for example, a multinational corporation might have email servers in every country so as to not run afoul of privacy and data protection regulations. These might also be applications that are particular to that business unit because of what it does. For example, a research-focused business unit may have different applications than the product development business unit.
- **Departmental.** These are generally specific to a particular department. For example, in municipal government, the court departments might have a docketing applications, while the engineering departments might have CAD and project management applications.
- **Stand-alone or end-user.** Every user has some of these, whether they are simple desktop clients like browsers and PDF creation tools, or more complex applications required for their day-to-day use like project management or CAD applications. In many organizations, the records management program and most of its instruments are created using simple database applications like Microsoft Access or even spreadsheets like Microsoft Excel.

Finally, it is important to consider integrations from two perspectives. First, if there are existing integrations in place, it is necessary to review them and the impacts on them of any additional technology changes. Often applications are integrated using custom code, and any change to any of the applications could result in errors or the integration failing altogether. Second, this is also an opportunity for the assessment team to determine whether integrations between information management systems and other systems might be beneficial. If this is the case, the team would need to identify those requirements as part of the requirements analysis phase.

Review Application Functionality

Once all the applications have been identified, the next step is to examine their functionality to determine how they interact with information-related processes.

Some of these applications can be used to create documents or records. Microsoft Word, for example, would certainly fall into this category, as would the email system (for at least some messages), scanning applications, and many of the line of business applications we identified earlier.

Many applications are used to access and retrieve information, and again we must examine this in the light of the information management program. It could be that there are compliance and data protection requirements that must be implemented in the application to ensure that unauthorized users cannot access certain types of information.

Many of them store information, either in their own proprietary repository, a relational database, or using a combination of the file structure and security to mimic a repository. The points to identify here include who has access to the file storage area, who can change that access, what users with access can do to the files in the storage area, whether controls can be set up to prevent records from being changed or deleted, and so forth.

Identify Other Considerations

Other things to look for in the technology assessment include:

- **Data structures and formats.** Does the application create potential records in standard formats like XML, or popular proprietary formats like Microsoft Office, or does it create them in highly proprietary formats that will require significant effort to read if the company goes out of business?
- **Stage in the technology lifecycle.** Is the system held together with bailing wire and chewing gum, or is it a current or recent release? If it is a version that is no longer supported, running on a version of Windows that is no longer supported, that's an indicator that it may be time to replace it.
- **Vendor and/or channel support.** Some vendors are very good at providing support, either directly or through their channel partners; others provide very minimal support. This can be a significant reason to change to a vendor that can support the application – or to select one for the initial implementation.

- **What is the cost of maintaining the existing application vs. changing to another application?** The cost will certainly include the hard costs of software licensing, annual maintenance, potentially hardware, and even integration or customization costs. And it will certainly include costs of software maintenance, upgrades, and the like, which could be dramatically different (higher or lower) between the existing application and a potential replacement. But it will also include soft costs of having staff to implement it, support it, and maintain it. An existing system might be more expensive to keep running than a new one, even taking into account the costs associated with the change.
- **Usability issues.** The best software application in the world is no good if it's too complex for the users to use, or too slow and cumbersome, or too buggy. A "just good enough" application that does what it does really well might be better than the comprehensive suite that doesn't really play well together.

The Deployment Timeline

The final wrap-up task in our technical assessment phase is to begin defining the business and technology capabilities and building a timeline for deployment. There may be some work that can start now, rather than waiting for the RFP process to be completed. For example, are there IT upgrades that can proceed? Can a retention or disposition schedule be reviewed now, independent of product?

Items to review should include: hardware, software, the network environment, database, any existing information management solutions, and the development environment used internally.

Identify the requirements that can be met by current systems, or by minor modifications to current systems. If portions of the project can be accomplished without procuring new technologies, especially with existing in-house resources, it is a good use of time to proceed.

Evaluating your Existing Information Management Systems

Information Management Systems

Every organization has information management systems. Sometimes these are as simple as networked file shares, but most organizations have at least one content repository, and even for small organizations the average is 4-6. For a larger organization, just the content repositories may number in the dozens, and when you factor in versions and variants across locations, it quickly adds up.

There are other systems that can store content such as enterprise resource planning (ERP), financial systems, customer relationship management systems, and others.

And we have to take into account cloud-based solutions as well.

Evaluating your Information Management Systems

It is very common for organizations to decide to improve their information management efforts, and to immediately begin the process of identifying, researching, and selecting an information management system. But as we just noted, most organizations already have something in place. So the question has to be asked – why doesn't the organization simply use what it has? There are any number of reasons for this:

- Changes in the overall technical strategy
- Fitness for purpose
- Cost
- Stage of lifecycle

Now let's look at each of these.

Change in Strategy

Changes in the overall technical strategy. Cloud is a significant example of this – if the organization plans to move to the cloud, the incumbent on-premises solution won't meet that need irrespective of any other considerations. Similarly, if the organization has decided to change for example from IBM to Microsoft, or from a proprietary vendor to an open source model, the incumbent system's capabilities may be irrelevant.

© AIIM – 2019

Fitness for Purpose

Absent a change in strategy, the organization needs to review its systems to determine their capabilities and how they are being used. After all, it may be that the solution doesn't meet the organization's business requirements. Maybe it never did. Or maybe the organization's requirements have changed.

However, this is something to consider carefully to ensure that those gaps are real and not simply issues of perception. Many organizations for example complain that their systems "don't work" because they "can't find their documents" – but a closer examination shows that the system is configured improperly, or metadata is not applied correctly or consistently (or at all), or users were never taught how to search the system effectively. It could be that the system is being used in a way it wasn't designed or intended to be used.

The bottom line, though, is that if the existing solution truly does not meet the organization's overall business needs, it needs to be replaced.

One more point. Sometimes organizations believe that they don't need a new solution because they have networked file shares they can use for information management. File shares are not generally considered effective information management solutions for a variety of reasons including security, lack of metadata support, findability, and many others. If this is all an organization has, it should strongly consider researching, selecting, and implementing an information management solution that will more effectively meet its business needs.

Costs

Sometimes the issue is a perception of cost – that a new system will be cheaper to use than the incumbent. This is especially true when organizations consider moving to a cloud-based solution. While it is often true that a cloud-based solution costs less to acquire and implement up front, over time the cloud solution may cost just as much as a comparable on-premises solution. This also ties into the overall technical strategy as noted earlier.

It's important to take into account how much it would cost to replace the incumbent system with a new one – and it's not just acquisition costs, but costs to train users, support, and administrations. A new system might require upgrades to client machines, or servers, or other software such as the underlying relational database if applicable. And it's the productivity costs associated with the disruption caused by implementing a new solution.

Finally, organizations should take into account the costs required to migrate any data from the incumbent solution to the new one. This can be a significant cost and effort.

Stage of Lifecycle

Finally, and relating to several of the issues we've identified, it's important to consider the stage of lifecycle of the incumbent solution. If the solution is no longer supported by the vendor, it's probably time to replace it. Similarly, if it's only supported by integrators or other third parties, it's only going to become more expensive to support over time. Again in this case the IM capabilities of the incumbent solution don't really matter here – it needs to be replaced.

Developing an Information Management Program Roadmap

The program roadmap is essentially a high-level overview of the information management program. It provides a list of projects within the program that will need to be completed for the organization to reach its desired end state, and some semblance of prioritization among those projects. There are dozens of projects that can be used to develop a comprehensive information management program. We will list some of these shortly; a longer list is available as an additional resource.

The projects required are identified as the result of the assessments. Many organizations will already have completed some of these projects. Other organizations may be at the very beginning of the information management journey. Different organizations will have different priorities according to the resources available to them, their particular needs, and where they believe they can make the greatest impact the soonest.

The program roadmap is designed to provide a long-term view. For larger organizations a 3- or 5-year roadmap is pretty typical. How long the roadmap will take to complete will depend on the organization's needs and resources and how important (and therefore prioritized) the projects are.

Finally, the program roadmap provides a "path forward" for the organization so that as circumstances, priorities, and budgets change, the organization knows where it is in the roadmap and what it can do to move forward.

Types of Projects on the Roadmap

We list here a number of information management major project areas (Assessment and strategy; Information governance; Information architecture; Technology procurement and implementation; and Cleanup and remediation) with individual projects listed in the next couple of sections that would typically be found on the roadmap. Again, if the organization has already completed one or more of these recently, they don't need to be added to the roadmap; if they haven't, or if it's been several years since the last iteration of that project, their inclusion may be warranted.

This should not be considered an exhaustive list; rather, these are among the more important projects to ensure effective information management. Each of these topics is covered in more detail elsewhere in this course. We will also provide a more complete list as an additional resource.

Assessment and Strategy Projects

The first area we will look at is strategy and planning projects. Projects in the roadmap might include:

- Conduct a business assessment
- Conduct a technology assessment
- Conduct system, information, and/or process inventories
- Develop the business case
- Develop the program roadmap

Information Governance Projects

Next, there are a number of information governance-related projects. These could include:

- Develop or update an information governance (IG) framework
- Develop IG roles and responsibilities
- Develop governance frameworks for information management technology platforms
- Develop or update records management program components
- Develop or update privacy and data protection program components

© AIIM – 2019

Information Architecture Projects

Many organizations can benefit from undertaking one or more of these information architecture-related projects. These could include:

- Develop or update taxonomies
- Develop or update metadata models
- Remediate existing taxonomies and metadata

IM Technology Projects

There are a number of projects that could be done with regards to information management technologies. These could include:

- Identify information management (IM) requirements
- Procure and implement an IM solution
- Integrate the IM solution into other solutions
- Develop a federated IM approach
- Migrate information from one IM system to another

Remediation Projects

Finally, we note some remediation and cleanup projects. These could include, but are certainly not limited to:

- Decommission legacy systems
- Decommission legacy user information stores
- Conduct cleanups, such as of file shares, enterprise file syncing and sharing solutions, or even active information management solutions





Test Your Knowledge

Question 1:

Which elements should be included in the organizational assessment? (select all that apply)

a. Review of the information flow and delivery across the organization
b. Review of people skills and organizational structure
c. Review of defects and issues reported during the beta process
d. Review of the regulatory environment

Question 2:

Which elements should be included in the technical assessment? (select 2)

a. Determination of the technology lifecycle stage for existing applications
b. Review of application functionality as it applies to information management
c. Review of defects and issues reported during the beta process
d. Determination of the cost of replacement systems

Question 3:

Which would be good reasons to implement a new information management system regardless of the existing system's capabilities? (select 2)

a. When the existing system is more than 3 years old
b. When users don't like the existing system
c. When the existing system is no longer supported by the vendor
d. When the organization changes its technical strategy

Question 4:

Which of these would be most helpful for an organization planning the next 3-5 years of their information management initiative?

a. The project charter
b. The program roadmap
c. The request for proposal
d. The change management plan

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Which elements should be included in the organizational assessment? (select all that apply)

a. Review of the information flow and delivery across the organization
b. Review of people skills and organizational structure
c. Review of defects and issues reported during the beta process
d. Review of the regulatory environment

A, B, and D should all be reviewed as part of the organizational assessment. For answer C, it is important to review defects and issues from the beta, but it will happen much later in the process as part of the actual implementation.

Answer to Question 2:

Which elements should be included in the technical assessment? (select 2)

a. Determination of the technology lifecycle stage for existing applications
b. Review of application functionality as it applies to information management
c. Review of defects and issues reported during the beta process
d. Determination of the cost of replacement systems

The best answers here are A and B. The beta will occur much later in the implementation process, as will determining the cost of potential replacement systems.

Answer to Question 3:

Which would be good reasons to implement a new information management system regardless of the existing system's capabilities? (select 2)

a. When the existing system is more than 3 years old
b. When users don't like the existing system
c. When the existing system is no longer supported by the vendor
d. When the organization changes its technical strategy

The best answers here are C and D, when the system is no longer supported or when the organization changes its technical strategy. The age of the system is a factor to consider but is not itself sufficient reason to change. Similarly, just because users don't like it doesn't necessarily mean that it should be thrown out; rather, it may be a training, configuration, or other issue that bears further investigation.

Answer to Question 4:

Which of these would be most helpful for an organization planning the next 3-5 years of their information management initiative?

a. The project charter
b. The program roadmap
c. The request for proposal
d. The change management plan

The best answer here is B, the program roadmap. The project charter is specific to each project, and for most organizations any particular information management project shouldn't last 3-5 years. The request for proposal would relate to a particular solution selection and implementation project and might not be needed for all information management initiatives. The change management plan is important, but not for the extended term planning of the overall initiative.



© AIIM – 2019

The Business Case for Information Management

A business case is a structured proposal used as a tool when seeking the approval of an initiative within the organization. The goal of the business case is to get a decision that authorizes investment in the initiative. This approval is given by an executive manager or management team. It is typically prepared for and delivered to a person, group or committee who is authorized to make funding decisions. Most importantly, it outlines the expected benefits to the organization of the new solution,

A business case outlines the business need and describes the required investment. It puts this need into a strategic context and includes all the necessary information that executives would need make an informed decision whether to proceed or not. It should include the available options, an analysis of each option and a recommended decision.

Why is a Business Case Needed?

So, why is a business case needed? The basic reason is because experience has shown that an information management program with a business case has a much higher chance of achieving success. This is because a good business case is used for several important purposes.

Successful organizations focus on implementing an overall plan or business strategy. It is therefore important to show how the information management program supports this strategy. Most organizations have documented their strategy for success and described the most important aims and objectives they seek to achieve. Often organizations look 3 or more years in the future.

Research has shown that successful programs of business change have strong support from senior managers in the organization. The business case is used to set out a compelling case for change and will help convince senior managers to support the program.

The business case also exists as an important means of communication. It is used to communicate important information, in a consistent way, to people in the organization, as well as to those working on the program. If all stakeholders and team members have the same understanding of what the program seeks to achieve, they will work more effectively together.

Important decisions need to be made during the program as new information becomes available. The decisions will often require approval from individuals or groups such as the chief executive, the finance director, or an investment committee. The business case is updated with new information. The business case can then be used to provide clear recommendations for making decisions to the individuals or groups responsible for them.

Finally, the business case acts as a baseline against which the actual costs and benefits of the program are assessed so as to measure the degree of success achieved. The business case contains information on the expected costs involved in implementing the information management program, as well as a summary of the expected business benefits. After implementation, actual costs spent, and actual benefits achieved can be compared with the expected costs and benefits in the business case. If the results are the same as or better than expected, then the program can be judged a success.

Elements of a Business Case

The business case should include at a minimum:

- A description of the existing situation. What is the problem to be addressed?
- A proposal to change the situation. This might include technology but will almost always involve a change to the existing ways of working.
- The benefits of making the changes. These could be quantifiable, non-quantifiable, or intangible.
- The cost of the proposed changes – the actual financial impact of the “to-be” scenario.
- And any other potential disadvantages of making the changes, such as short-term productivity drop-offs.

Challenges to Producing the Business Case

We consider first the challenges associated with calculating financial benefits.

The first reason is that complex changes in working practices mean it is difficult to measure their impact. Factors that result in additional complexity include the following:

- The changes in working practices are likely to affect a significant proportion of staff, which means a wide scope for assessing benefits.
- The changes are also likely to affect a number of activities or tasks performed, again giving a wide scope. The changes will simplify some tasks performed and lead to a saving in staff time. The changes could make other tasks more complex and result in additional time spent to complete the task.
- The tasks may not be completed on a frequent basis and the total time spent on them may not be known.
- The impact of simplifying a task may result in a saving in time and effort elsewhere in the organization, which is more difficult to identify. All these factors make it difficult to calculate the overall impact of the changes and means that there is a high level of uncertainty over any results calculated.

The second reason to consider is that the information base is fragmented and incomplete. For instance, the information used to calculate information management benefits is usually not reported on a regular basis. Information regarding time spent on particular tasks needs to be estimated. For example, estimates can be made from discussions with informed users or through a specific exercise to measure a sample of tasks as they are undertaken.

The next reason is that, in many cases, benefits are dispersed thinly across a wide area of operations. Small savings in time may be made in a particular area within the organization. These may only become significant when applied to many members of staff across the whole organization. It will be difficult to demonstrate after implementation that the benefit, although real, has been realized. Organizations may be reluctant to consider these types of benefit.

Information management programs have their own particular characteristics, and these give rise to specific challenges for producing business cases when assessing benefits, costs, and volumes.

Significant attention is paid in business cases to those benefits that can be calculated in financial terms. As shown earlier this information is required to produce a cost benefit cash flow. Benefits are difficult to identify and calculate when producing an IM business case because the business changes are complex and the information required is hard to gather.

It is also the case that in most organizations, content management is a cost center rather than a profit center. This is true whether the organization is in the private sector, and therefore concerned with profitability, or public sector, where the “profits” are rather public benefits.

There is a wide variation in the cost information that results from the procurement process and provided by suppliers. It is often difficult to understand the reasons for the variation in software, hardware and implementation services costs.

Costs also depend on knowing key information concerning both storage and transaction volumes, which are used to calculate the cost of hardware. This can also be difficult.

© AIIM – 2019

The Costs of an Information Management Initiative

Every information management initiative has a cost associated with it, whether that initiative is more focused on technology selection and implementation or the development of an information governance framework. But often organizations will make the business case for the initiative based solely on acquisition costs – that is, how much does the software and/or hardware cost? But this only tells a part of the story.

In order to get the complete picture, we need to consider three main areas of cost:

- Acquisition and customization costs
- Operating costs
- Change management costs

This type of analysis will ensure that the organization is able to do a consistent comparison of different alternatives; ensure the internal cost of employees, as well as external contractors are assessed; and can look at the project over a multi-year deployment and operational cycle. This also allows the organization to calculate the payback period, or return on investment period, and the internal rate of return.

Acquisition Costs

The first step is to outline the acquisition costs of a new or expanded information management solution. Software licenses are only one small part of the bigger scope of the total cost of ownership. If evaluating different models, look carefully at the types of user licenses. Verify if there are different costs for administrator, read-only, concurrent or other types of roles. There may be optional modules, such as records management or privacy, that really are needed to compare apples to apples across solutions. Ensure that any third-party products recommended by the integrator or vendor are also included. This includes integrations or connectors to work with other line of business applications.

Next, assess if any hardware investments are needed. This could include servers, but also updates to desktops, laptops, scanners, or other devices. If one IM product demands a particular hardware environment, ensure this is reflected in the overall cost planning and comparison.

Consider other infrastructure dependencies. Some cutting-edge systems may not run on older versions of operating systems, databases or office suites. Ensure these infrastructure dependencies are also accounted for.

Determine if any assistance is needed during the vendor selection or Request for Proposal (RFP) phase. External expertise can help fast-track the process and share what other companies like yours have done. Also, budget for time and any travel resources needed for internal staff to do a full vendor review.

Operating Costs

Next, review the ongoing support and maintenance costs for the proposed solution. Most traditionally-licensed products will have recurring maintenance fees in the 20-30% range. This is typically not discounted at the same rate as up-front license purchases, so get clarity from your solution provider. Also be certain to account for ongoing maintenance for any third-party products needed.

Cloud-based and open source information management options may not have the up-front license costs, but will have these support and maintenance costs charged on a monthly or annual basis. It is important to know how much each support and maintenance offering costs for the response time and service level you need from the vendor. A same-day bug resolution commitment will typically mean a more expensive annual contract than one that promises a one-week bug fix commitment.

Assess the backup, recovery, security and disaster recovery costs. What kind of redundancies or fail-over infrastructure is needed to meet the needs of your business?

What is the cost of ongoing stewardship, release management, periodic code audits and assurances such as escrow? Once again, this could be dramatically different for vendors with different license or support and maintenance models.

And perhaps the single biggest expense associated with many information management deployments, the cost of staff personnel required to implement, configure, administer, manage, and maintain the solution. For example, if you need to implement an ECM solution that uses Oracle, you may need both an ECM administrator and an Oracle database administrator.

Change Management Costs

The next item to consider is the cost of customization and/or integration. Balancing the cost of each feature, integration or advanced capability with its ability to meet your identified business objective can be a trade-off. A full understanding of costs, along with clear business requirements, will help make those decisions and set priorities.

It's also important to look at future upgrades. If an extended test cycle is demanded in your environment, consider what that means for products with faster release cycles. Cloud-based or SaaS solutions may have very little cost other than user refresher training.

What is the cost of migrating current content to the new information management system? This comparison across vendors can help determine what content should be migrated, and what can be decommissioned and disposed of in line with corporate recordkeeping policies.

Finally, don't neglect training for end users, administrators or developers. Look at in-person vs online training options for both product and business-rule education. Ongoing communication, coaching and continuous improvement programs need to be included in the overall cost analysis.

Build vs. Buy

One more note. This module, and this course, assumes that most organizations will be buying (or, in the case of a subscription model, leasing) their information management solutions. Buying commercially available software is almost always preferable to building your own for a number of reasons:

- Commercial software is being used successfully by other organizations – that is, you're not the test case.
- Commercial software has been tested, both by the solution provider and by those other organizations.
- If you build your own, you need the necessary resources to build it and test it, either on staff or via a third party, and maintain it over the lifecycle of the solution.

Once you factor in all of those costs, it's likely at least as cheap to buy as it is to build – and since the staff costs to build and maintain it are likely the highest costs, buying may be cheaper than building.

You should only build your own software if it is necessary to meet highly unique requirements that are not currently being met in the marketplace. And even then it might be better to work with a solution provider to see if they can build it as a custom solution – they may be inclined to do so at a discount to earn the business or if they believe they can turn it into a viable product.

If you're looking at integrating an information management solution with another one, the option to build is a bit more favorable, because you're likely doing that to meet your unique needs. However, many solutions are already integrated with others using pre-built connectors, so it makes sense to do some due diligence.

Information Management Metrics

There are three main types of business benefits that we need to consider.

Some benefits can be quantified in financial terms and are used in the business case cost-benefit analysis. These benefits are often associated with improvements in efficiency or how to produce the same output with fewer resources. Characteristically, for these benefits, revenue or savings targets are set, together with a date by which the revenue increase or savings will be achieved. Progress in achieving the benefits can be monitored after implementation.

But, there are other types of benefits that cannot easily be measured in financial terms. They can be "quantified in non-financial terms." Some such benefits can be highly-valued by the organization. For example, faster times for retrieval of information is highly desirable. It can be quantified in terms of the average time currently taken compared with the predicted time within the proposed IM environment. However, it is not clear how the time saving can result directly in a financial saving or benefit.

In addition, there is a final group of benefits that simply cannot be quantified directly. These are known as intangible benefits. They are also important for differentiating between alternative IM environments and can be assessed on a subjective basis. One approach that can be used is to undertake a survey with users - for example, to identify what proportion of managers felt that the IM environment had led to better decision-making. Another approach used is to identify an indirect or 'shadow' measure. For example, job enrichment could be tracked by measuring staff morale, or even absenteeism or the level of staff turnover.

© AIIM – 2019

The most usable metrics will be those that provide meaningful, quantifiable results; all metrics should align with the overall goals of the organization.

Financial Benefits of an Information Management Program

We start by identifying the potential financial benefits of implementing an IM program. Traditionally information management has been more of a cost center focused on risk reduction. Where financial benefits are realized, historically they have been focused on cutting costs. For example, an organization may be holding significant amounts of information in paper format onsite, taking up valuable real estate with files and file cabinets. Getting rid of information with no more business value, and digitizing the rest, could free up that space for a more valuable use. Similarly, by keeping information only as long as needed and then getting rid of it, organizations can reduce their offsite storage and retrieval costs.

Better information management can help to reduce costs associated with audits or legal matters, though this is not as direct an impact.

And better information management could lead to a reduction in staff, which saves money on labor costs – in theory. However, as a practical matter, it is more likely that those staff will be repurposed, for example by converting file clerks into scanner operators. So while this is often part of the business case, it rarely materializes.

Non-financial Benefits

So far, we have considered where financial benefits can be achieved. The two other categories of benefits identified are those quantified in non-financial terms and those that are non-quantifiable or intangible. Whether benefits can be quantified or not will partly depend on the organization systems and processes, so there is no hard and fast rule.

In the following section you will be introduced to a range of benefits where an indication will be given of whether the benefit can expect to be Quantified in non-financial terms (capital Q) or is generally expected to be non-quantifiable (capital NQ). Organizations should seek to realize these benefits and set targets to achieve them. As described earlier, measurable targets can also be set for the Non-quantifiable benefits by using indirect measures or ‘shadow’ metrics that provide a guide that the benefit is being achieved. Suggestions will be provided as each benefit is described.

For ease of discussion, the benefits are grouped into three related areas of

- improved information sharing and access
- improved decision-making
- improved management of information

Information Sharing Benefits

First consider the benefits from information sharing and access.

There will be no lost documents, as no correctly-indexed document is ever lost. In every paper-based environment a small percentage of documents and records are lost, and finding misfiled paper documents is both expensive and slow. This can be quantified, for example, as number or percentage of documents lost.

An information management environment allows any authorized user to find any record, easily and quickly. Also, an IM environment typically allows content searching (for text records) and may include a thesaurus which allows searching with synonyms and related terms. This can be quantified, for example, as the average time taken to find a document.

There will be faster information retrieval, since an IM environment allows the quick and efficient retrieval of information stored in it. Typically, any document needed can be found and retrieved within seconds. This can be quantified, for example, as the average time taken to retrieve a document.

Information can be shared efficiently. Documents can be distributed to recipients, and unlimited numbers of users can refer to the same document simultaneously, all without the photocopying cost which paper documents would require. This is non-quantifiable but can be measured indirectly through the time taken for a related task.

If required, information can be retrieved at any time – 24 hours a day, 7 days a week (so long as this is recognized as a requirement at time of procurement). This is quantified frequently through a service standard which defines the percentage of time the system will be available.

Remote retrieval is facilitated. If a suitable infrastructure is available, documents can be retrieved from or added to the system from remote locations (for example, international offices, supplier premises, and workers' homes). This is non-quantifiable but users can be surveyed to determine whether remote retrieval is easier.

With simultaneous access, an unlimited number of authorized users can read, refer to or print any digital content at any time. Non-quantifiable, but the time taken to undertake a related task may be measured.

Finally, users will find their jobs easier and more fulfilling as they are able to search for and obtain information faster and more easily. Non-quantifiable, but may be measured indirectly through a survey of job satisfaction or the level of staff turnover.

Decision-making Benefits

Now let's look at the benefits from improved decision-making.

A higher weighting as evidence given to content in litigation, can minimize the likelihood of criminal penalties and/or regulatory fines. With an IM environment, documents presented as evidence will be complete, and completely credible. By comparison, collections of paper documents are often incomplete, unreliable, or lost entirely. Whether a particular file is complete cannot be quantified. The benefit may be quantified indirectly through measurement of percentage of litigation cases won. It may even be quantified as a financial benefit because an organization with an IM environment may be able to negotiate more favorable terms with its insurers.

Many benefits of an IM environment combine, to facilitate the taking of decisions which are completed more quickly, and based on better information. These improvements in decision-making can be extremely valuable, though the value is difficult to quantify. Non-quantifiable but may be measured subjectively through a survey of management opinion.

A properly-configured IM environment provides information on which users can rely. Files are up to date and complete. Non-quantifiable but could be measured indirectly by the percentage of external requests for information (for example arising under the requirement to provide information under Freedom of Information legislation) that can be satisfied immediately.

Information Management Benefits

Finally, benefits from improved management of information can provide:

Increased confidence from comprehensive audit trails and access controls. In a properly configured information management solution, a correctly-filed document can never be lost. Non-quantifiable but can be measured indirectly by sampling the completeness of audit trails.

By default, backup copies of the content in the information management solution will be kept. As long as an effective business continuity plan is in place, these backups can be used to recover content from any disaster which affects the information management environment. Non-quantifiable but can be measured indirectly as the forecast time taken to recover from a disaster.

An IM environment can be configured to provide exceptional awareness functionality by circulating new or incoming content according to specified rules. These rules can specify that new documents can be made known to all recipients simultaneously, to unlimited recipients in unlimited locations, or to a selective readership. Non-quantifiable but can be assessed through a survey of users.

An IM environment implemented to the expected standard will keep a complete audit trail for every piece of content it contains. This will far exceed the level of control possible over paper documents, and may provide a valuable resource in the event of audit or other investigations. Non-quantifiable but can be measured indirectly by sampling the completeness of audit trails.

And perhaps one of the most interesting benefits, improved innovation. It stands to reason that better information management will help the organization to know what information it has, and to leverage that information to speed up or otherwise improve the innovation process. This is very difficult to quantify directly but could be measured indirectly by surveying management.

© AIIM – 2019

Issues with Benefits Realization

There are a number of challenges to realizing the benefits of an information management program.

First, in almost every organization content management is a cost center. It is a cost of doing business rather than an opportunity to increase profitability.

One benefit identified for an information management program is the savings of space. This can be a significant cost if the organization is leasing expensive office space, but if the information management initiative simply results in an empty warehouse that is not reused, it is difficult to see any real savings. Similarly, if the business case is made that an organization will reclaim 10% of its office space, it is difficult to determine whether the reclaimed space will be put to good use.

Staff time savings can similarly be difficult to see. There are two challenges here. The first is that too often managers calculate staff savings by considering the as-is, where a given process takes, say, 20 minutes, and the to-be, where the process will take 15 seconds. This is a false comparison because most transactions do not come in every 15 seconds like clockwork, and the organization may not have the number of transactions required to make the comparison meaningful. The second issue is that even where those savings can be identified, the organization still has to determine what to have staff do with the time saved.

There are also significant change management and take-up issues, such that we devote a section to change management later in the course. It may be that on paper the organization saves significant staff time, but in practice the users are not trained, or find the system too complex, or are sufficiently resistant to the new system and processes as to slow their take-up and acceptance.





Test Your Knowledge

Question 1:

What are the challenges associated with producing a business case? (select 2)

a. Benefits are dispersed across the organization
b. Finding a consultant that understands your particular industry
c. Ensuring that ROI is achieved in less than 12 months
d. Comparing costs consistently across different solutions and models

Question 2:

Which elements should be included in the total cost of ownership for an enterprise content management solution? (select 2)

a. The cost of researching the market
b. The cost of server and client licenses
c. The cost of any customizations or integrations
d. The cost of downtime during implementation

Question 3:

Which of these is a financially quantifiable metric for an information management program?

a) Reduction in costs associated with offsite storage and retrieval
b) Reduction in time required to retrieve a particular document
c) Improved employee satisfaction
d) Improved information security

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What are the challenges associated with producing a business case? (select 2)

a. Benefits are dispersed across the organization
b. Finding a consultant that understands your particular industry
c. Ensuring that ROI is achieved in less than 12 months
d. Comparing costs consistently across different solutions and models

The best answers here are A, benefits are dispersed across the organization, and D, comparing costs consistently across different solutions and models. Finding a consultant doesn't really have anything to do with the business case. ROI is a complex question: many information management projects achieve an ROI substantially shorter than 12 months, while others are expected to take years. In other words, the 12 month limitation is somewhat beside the point here.

Answer to Question 2:

Which elements should be included in the total cost of ownership for an enterprise content management solution? (select 2)

a. The cost of researching the market
b. The cost of server and client licenses
c. The cost of any customizations or integrations
d. The cost of downtime during implementation

The best answers here are B, the cost of server and client licenses, and C, the cost of maintenance and upgrades. The cost of researching the market is likely to be small and diffused across the various solutions researched, and the cost of downtime during implementation will be difficult to quantify.

Answer to Question 3:

Which of these is a financially quantifiable metric for an information management program?

a. Reduction in costs associated with offsite storage and retrieval
b. Reduction in time required to retrieve a particular document
c. Improved employee satisfaction
d. Improved information security

The best answer here is A, reduction in costs associated with offsite storage and retrieval. The reduction in time required is quantifiable but difficult to accurately measure in financial terms. Improved employee satisfaction and improved information security are both difficult to quantify directly at all.

The Role of Requirements in an Information Management Initiative

There are many different definitions of business and systems requirements, but they all have a common theme, which is to capture the needs of business stakeholders in a manner which enables successful program outcomes.

For the purposes of this module, we will define information management-related business and systems requirements as "statements which identify a capability, feature, or qualitative attribute which has business value for the user."

Types of Requirements

There are two primary types of requirements to be concerned with. First, user requirements, often referred to as business requirements, define high-level abstract requirements that specify system behavior, not the technology required to perform that behavior. For example, users need to be able to find and access their information on demand.

System requirements, in contrast, are more specific and generally testable. These can be further broken down into:

- Functional requirements. These describe the functionality to be provided at significant detail. For example, "the information management system shall store records on non-rewritable, non-erasable media."
- Nonfunctional requirements, which relate to aspects of the system that are not directly related to a particular function of the system. For example, "the information management system shall respond to most simple queries within 3 seconds."
- Domain requirements are specific to the environment in which the system will operate rather than the users or processes it is used for. For example, a domain requirement for a strongly Windows- and Microsoft-focused organization might be that the system run using Windows and Microsoft SQL servers, or be compatible with specific versions of Windows, Internet Explorer, and/or Office.

The key here is that buyers, including those on the project team, need to ensure that they understand the business/user requirements and then work with suppliers and consultants to determine which functional requirements and capabilities will best satisfy those needs.

The Purpose of Requirements

Business requirements serve to document the needs of all stakeholders.

They allow the organization to ensure that the objectives of the system to be developed and delivered match up with stakeholders' target outcomes and ensure that it meets all the stakeholders' needs.

And they enable the project and program to be scoped, planned, and delivered effectively. Good requirements make it easier for vendors to design a solution that meets the organization's needs and to accurately estimate the time and resources required to implement it. They also make it easier for the organization to understand what is to be delivered and implemented.

The Importance of Requirements

Effective statements of business and systems requirements are important, because their early definition often have the most influence on a project's success. It can be difficult to fix or change direction on requirements decisions, once the evaluation and procurement processes have begun. They will drive the key solution design decisions and technical specifications, which in turn determine the size and nature of implementation activities.

According to research from Accenture, up to 85% of information management system defects originate in the original business and systems requirements which are produced. Any defects in a system lead to necessary rework or fix activity.

Also, according to Accenture, across the software development industry, on average 45% of project costs are attributable to rework. This further reinforces the fact that business and systems requirements have a very large degree of influence on any information management implementation initiative.

Significant amounts of project cost and effort can be saved or reduced through effective business and systems requirements. It is therefore a worthwhile investment to put additional effort and rigor into this phase of a project.

© AIIM – 2019

The Price of Perfection

One challenge many organizations face in implementing an information management solution – or any other complex software or IT initiative – is that there is a focus on getting all the requirements documented, completely and perfectly, and being unwilling to move forward on the initiative until the requirements are "complete." This is a challenge for a number of reasons.

First, this can lead to a project management condition called "paralysis of analysis" where the requirements process gets progressively more detailed to the point where no vendor in the industry would be able to address all of them perfectly.

It also results in the requirements process taking a significant amount of time – which is not something the organization is going to realize much value from. But more importantly, the odds are very good that if the requirements take too long to define, they will end up being outdated due to changes in the industry, business processes, or the technology.

Similarly, many organizations will reason that if the solution supports customization or integration, that it's a good thing to do it – and if some customization is good, more must be better. Besides the usual issues associated with maintaining custom code over time across multiple solutions, this causes a more immediate issue in that it delays progress. This is due to the need to identify and document those requirements, the need to build the customization to meet them, the need to test and deploy the customization, and so forth. It also increases the cost of the overall solution for many of the same reasons.

Therefore, it's almost always better to implement – something, anything – and then optimize it later when time, resources, and circumstances allow.

Gathering and Analyzing Requirements

Preparing Requirements

There are 5 main stages in producing an effective set of business and system requirements for an information management initiative:

- Plan
- Gather
- Analyze
- Document
- Agree

Each of these individual stages can be iterative by themselves with several passes required at each stage. It may be necessary to gather some additional requirements, which might involve revisiting stage 2 or conducting some additional analysis as part of stage 3.

We'll now look at each of these steps in more detail.

Planning

The first step in the process is planning. In this step the requirements team identifies stakeholders and users who will be impacted by the information management solution, and determines which approaches to use to elicit information. Based on those approaches, they may also schedule interviews or workshops, develop surveys, etc.

The business vision and high level business case will already have provided an understanding of the information management needs from key stakeholders and it is the task of the analyst to turn this vision into a more tangible set of requirements that are solvable by existing capabilities, modifications to those capabilities, or the acquisition of new technical capabilities.

It is also useful to create a common set of definitions, or a glossary, early in the process of creating business and systems requirements. This helps achieve common understanding across the various parties involved and also helps crystallize common thoughts and concepts.

There are several requirements specification standards available for certain classes of information management solutions. Standards can be useful as a starting point, as well as a baseline for terminology and vocabulary. Don't, however, fall into the trap of relying on standards as a way to short-cut true requirements gathering.

Gathering Functional Requirements

Functional requirements are those that provide the actual functionality required to accomplish the stated business requirements. For example, in the U.S., some financial services companies have to store certain types of information on non-erasable, non-rewritable media. Some organizations want solutions that comply with the Content Management Interoperability Standard (CMIS), which allows them to exchange and use information between a variety of information management systems. Most organizations need a variety of digital recordkeeping capabilities. And so on.

There are a number of approaches available for identifying and gathering requirements, including:

- **Research** – into what other similar organizations have done, capabilities of different solutions and solution providers, and industry best practices
- **Surveys** – sent to end users to determine how they work and what they need in order to accomplish their job tasks
- **Workshops** – involving small groups of users, either by function or cross-functional teams, to determine how they work and what they need
- **Interviews** – individual sessions with key users to understand their roles and requirements
- **Prototyping** – developing mockups, screenshots, or storyboards to get users to visualize what the proposed system can do and how

Non-functional Requirements

It's also important to consider non-functional requirements. These include:

- Solution look and feel, or ease of use of the system, for end users, administrators, and developers
- Support services, company viability, reputation, and understanding of your vertical industry
- Vendor or integrator approach used to integrate components, and with other enterprise applications and content repositories
- The level of customer satisfaction for each product
- The size of the user community, and the tools and forums used to support this community
- The Total Cost of Ownership or TCO. This means that you should look at not just the cost of the software licensing, but any other costs associated with maintenance, training, services, documentation, and integration, over a three- to five-year window
- The availability and quality of partners and any user groups or communities
- The investment made by the supplier in continuous R&D, for innovation, customer, partner, and community satisfaction

These will need to be determined by researching individual providers, reviewing analyst reports, and/or talking to existing clients and references.

Analyze Requirements

Once the requirements have been gathered, the team can analyze them to determine their applicability and priority.

The first step is to meet with the key stakeholders and share the initial requirements document with them. One of the key outcomes of this is to separate real requirements from stated requirements. Often, in the early stages of engagement with key stakeholders and other parties, people will tend to state unrealistic needs. It is the role of the analyst to help guide input from the interested parties and separate stated requirements from real requirements.

This is also the time to prioritize the requirements to ensure that any mandatory requirements are included. Less important requirements may be useful to distinguish between options or serve as a future implementation phase.

© AIIM – 2019

If well-documented, this process will provide a useful audit trail from stakeholders' initial comments through to final, agreed upon and prioritized requirements. This is another useful outcome of the business and systems requirements process.

Document Requirements

Documenting the requirements includes not only noting the prioritization of requirements in line with agreed ratings, but also to ensure a method is used to track the source of each requirement. Specialized software tools can be used for this, but often simple tables or spreadsheets can be used for more basic needs.

Version control will be important as the requirements gathering process occurs, tracking the input, decisions and deletions of duplicate or irrelevant items. In regulated industries such as life sciences, or ISO certified companies, this requirement tracking or version control process may already have supporting procedures or templates.

Agreement on Requirements

The final stage of requirements gathering is the agreement phase. A formal but simple sign-off process should be used to ensure that the appropriate signatures are obtained, noting the time, role and title. Often, a simple register can be added to the overall requirements document to track the agreements. Sign-off at this stage is important in order to move to the more advanced technical assessment tasks with the core team on board with the current direction.

It is likely, however, that later versions of this requirements document will need to be created. As the project progresses through later technical assessment tasks, vendor selection begins, and as pilot systems are evaluated, requirements may change. Something that is only a "nice to have" in early days may emerge as "mandatory" later in the project cycle.

If any changes are required, or any key stakeholders object to the document circulated for approval, further iteration may be needed to reach consensus, and to produce a document with minimal errors or inaccuracies.

Once the requirements are finalized and approved, they can be used as part of the request for proposal (RFP) process.





Test Your Knowledge

Question 1:

Why are business requirements so important to the success of an information management initiative?

a. They ensure that the solution has been tested effectively.
b. They represent the needs of business stakeholders.
c. They document adherence to industry standards.
d. They demonstrate compliance with applicable regulations.

Question 2:

What is the role of standards in the requirements development process?

a. To provide a complete list of requirements for a given type of solution
b. To ensure that a proposed solution is certified as meeting the organization's needs
c. To determine the most appropriate solution for the organization
d. To serve as a starting point and guide to common terminology

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

Why are business requirements so important to the success of an information management initiative?

a. They ensure that the solution has been tested effectively.
b. They represent the needs of business stakeholders.
c. They document adherence to industry standards.
d. They demonstrate compliance with applicable regulations.

The correct answer is B, they represent the needs of business stakeholders. Testing is either done by the vendor, or done as part of acceptance testing much later in the implementation process. Business requirements may reflect industry standards and applicable regulations, but they do not document or demonstrate adherence to them.

Answer to Question 2:

What is the role of standards in the requirements development process?

a. To provide a complete list of requirements for a given type of solution
b. To ensure that a proposed solution is certified as meeting the organization's needs
c. To determine the most appropriate solution for the organization
d. To serve as a starting point and guide to common terminology

The best answer here is D, to serve as a starting point and guide to common terminology. For A, standards have to be broadly applicable and most organizations will have at least some unique requirements, meaning the standard is incomplete by definition. For B, while some requirements standards include certification, this merely certifies that the solution adheres to the standard, not how well it meets a particular organization's needs. And requirements are only one part of what makes a solution appropriate for an organization.



The Implementation Process

This section will help you to identify the key steps required to implement an information management solution. Note that this section does not focus on the implementation tasks that would typically be done by IT, an integrator, etc. Rather, these are the steps that information professionals would be expected to participate in.

The Implementation Process

It is critical to involve users in the design process. No matter how smart and experienced the project team, the business analyst, the manager, or the consultant, the simple fact is that no-one is closer to the work process than the users doing the work. They understand the difference between the work process, as designed in a pretty flow chart, and how it actually gets done. Moreover, they will be the ones doing the work in the redesigned, to-be process or the new process and they may catch omissions that the experts miss because they don't do the work all the time.

Involving the users will also produce two other benefits. First, it will give the users a better sense of ownership of the process. It's hard to argue that the redesigned process cannot succeed if you were one of the people that designed it. This can even be used to bring reluctant users into the fold.

Second, users who are involved in the design and the give and take that is part of the process will have a much better understanding of the system than can be transmitted through the typical training program. Those users who participate will be in a good position not only on go-live but to assist in the pilot, implementation, and rollout within their local area.

Design Work Processes

One of the key steps in the implementation is to incorporate information management into the organization's work processes. There are two things to address:

- Design the new processes required to manage information more effectively. For an organization that has always worked in a more manual, more paper-based way, this means designing processes around scanning, capture, capturing metadata, etc.
- Update existing work processes to incorporate information management practices. In many ways, when we talk about "digitalizing" processes, this is what we mean. It's not *just* scanning and capturing information digitally, but how we can rethink processes in a way to leverage digital capabilities like digital signatures, workflow and process automation, and analytics.

Every organization has a number of work processes in place already, of course; one of the best ways to ensure content is managed appropriately is to incorporate information management processes into existing work processes. In this way, users come to accept information management as just another step in the process, rather than considering it as something "extra" that is added to the end of the process. This also helps to ensure that information is captured in a timely fashion, rather than when the user gets around to it.

Create Procedures and Job Aids

Once the processes have been designed, it is necessary to develop procedures and job aids that instruct users on how to accomplish the tasks that make up the processes.

Sometimes these procedures will be detailed instructions; for example, when an aircraft takes off, the pilot and copilot follow a very detailed checklist to ensure that nothing is missed.

In other cases, the procedures will be more in the way of flexible guidelines that allow users significant leeway in determining how to comply. This might not work as well in landing an aircraft but could be useful in outlining a procedure for delivering training courses.

Often there are existing procedures in place. Part of the business assessment and the process design steps involves reviewing existing policies and procedures to determine whether they remain relevant. If this is the case, leave them alone – there is no sense doing work just to do it. On the other hand, if they are out of date, or if the processes change, the procedures will need to be updated.

It is also helpful to develop or update job aids that support the procedures and processes. For example, a process requires users to classify business documents into the classification scheme. The accompanying procedure specifies that certain metadata fields must be completed as part of the process, including a job code. Users might find a listing of job codes and what they mean very useful as they complete the process.

© AIIM – 2019

Technology System Design

The technology system design is used to design the system that meets the requirements stated in the requirements document. These requirements will include:

- Information management requirements required to capture, use, manage, retain, and disposition information over the entire lifecycle
- Other business and functional requirements that may be required to keep the system running, allow users appropriate access
- And any domain requirements, such as which database the system will interface with or what file formats must be supported

One of the significant system design considerations revolves around how to address multinational organizations with different information management and compliance regulatory regimes in different jurisdictions. This can be addressed effectively through the use of a decentralized approach, where key governance and compliance requirements are addressed initially at the enterprise level and then tailored to those unique requirements.

It is important to ensure that the system design meets all mandatory requirements before the system is built. It is much easier to fix a defect in the design phase than post-implementation.

Interface Design Tasks

Within the technology design process there will be a number of areas to be designed. We look first at the interface design tasks because these will significantly impact how information is presented and ultimately the usability of the information management system.

Requirements for user interfaces will have been identified – whether users will interact with the system from a desktop computer, a browser-based client, or even a smart phone or tablet. In the design phase the team will make decisions about what the client will look like; what browsers will be supported and in what fashion; and whether plug-ins will be required. In the case of mobile devices, a similar determination will need to be made as to how the users will access the system – through a browser or through a client application – and what level of functionality will be provided through those clients.

Another aspect of the user interface is the data entry interface. This involves designing forms and templates that can be used to extract text or enter metadata. In the case of metadata, this might also involve designing controlled vocabularies and exposing them through drop-down lists, or creating business logic that validates the data being entered so that a date field, for example, only accepts valid dates.

As the interface design progresses it is important to take into account good usability practices. For example, functions that are used frequently should be in a top-level menu, not ten layers deep. Routine tasks should not require 27 mouse clicks to execute. Fonts have to be consistent and large enough to read. And so forth.

Design Interoperability

The next design task is to consider interoperability. In many implementations there is a need to integrate the information management system with other systems in the organization. There are a number of ways to do this that should be considered as part of the technology design process.

The first is through system integration. This approach almost always requires custom programming that will integrate some or all of one system's technical functionality into another system. This will be done by IT or the vendor or integrator, but information professionals may need to be involved to ensure the resulting capabilities still meet business needs.

Workflows and business process management solutions provide another approach to integrate disparate systems in a fairly loose yet robust fashion. We address workflow elsewhere in this course.

Finally, and most recently, we come to federation. The New York State Office for Technology defines federation as follows: "Federated Architecture(s) define common or shared architecture standards across autonomous program areas, enabling state government entities to maintain diversity and uniqueness, while providing interoperability." In the context of information management, federated systems are used to provide centralized access to many disparate systems and silos of information. They can also be used to index information in those systems and silos and to work with and through the various security models to provide a comprehensive control framework. Federation can be a valuable part of the overall interoperability design but as with the other approaches will require some thinking through of what is to be done and how it needs to be done.

Design for Data Protection

We address privacy and data protection in much more detail elsewhere in this ebook, but it's better to address privacy and data protection during the design phase rather than as an afterthought.

Information security and access controls are generally developed and implemented by IT, information security, or some similar role within the organization, but they need to know what types of roles and access levels need to be created. Those roles and access levels should be designed by business users to ensure they are compatible with the needs of a particular business area and the organization as a whole. This is particularly important when using federated approaches to information management as outlined earlier, as security can significantly affect findability in a federated search environment.

Go Live

Depending on the size of your organization and the initiative, you may want to roll it out enterprise-wide. This works for smaller, less-complex organizations and initiatives but can be challenging for larger ones.

Alternatively, you may wish to roll out on a business area by business area basis, or a staggered rollout. This may be a lower risk strategy, but also results in potentially complex issues with some parts of the organization managing information more effectively while others aren't.

Whichever approach you choose, your IT supplier or systems integrator may well want to run final testing of the system on the live environment, while floorwalkers may provide a confirmatory check that all users within a business area have managed to access and use the environment. In addition, the systems administrators should be able to provide reporting information from the system to the floorwalkers, so that they can identify individuals who do not seem to be using the system, for whatever reason.

Post Implementation

Once the system has gone live, there are a number of other tasks to be completed. The first is to conduct a post-implementation review. This will help the project team understand the lessons to be learned from the implementation: was there sufficient training, did the pilot meet the needs of the project, how well did the communications work, etc.

On an ongoing basis the team should conduct reviews with the end users to make sure the system continues to meet their needs. They should also connect with the technical support staff to see whether there are trends that should be addressed.

The solution, and program, will need to be rolled out to new groups, processes, departments, and locations. And it will need to be updated periodically to take into account new developments, whether in the organization's strategy and direction, the information management solution, new industry developments such as regulatory requirements, etc.

© AIIM – 2019

Information Management in the Cloud

So what is cloud computing?

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Key Cloud Characteristics

The NIST definition goes on to state that "This cloud model is composed of five essential characteristics…" These characteristics include:

- **On-demand self-service.** A consumer can provision computing capabilities, such as server time and network storage, as needed without requiring human interaction with each service provider.
- **Broad network access.** Capabilities are available over the Internet and accessed through standard mechanisms that promote use by mobile phones, tablets, laptops, and workstations.
- **Resource pooling.** The application provider's computing resources are pooled to serve multiple customers, with different physical and virtual resources dynamically assigned and reassigned according to demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify a location at the country, state, or data center level. Examples of resources include storage, computational processing, memory, and network bandwidth.
- **Rapid elasticity.** Computer resources are rapidly scaled up or down as needed.
- **Measured service.** Cloud systems automatically control and optimize resource usage by leveraging a metering capability at some level appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts).

More than anything, these characteristics define cloud computing and differentiate it from simply putting a business application on a server connected to the Internet.

Cloud Deployment Models

Finally, the NIST definition includes four cloud deployment models.

1. Public cloud. The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them.

2. Private cloud. The cloud infrastructure is provisioned for exclusive use by a single organization. It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist either on- or off-premises. For multinational organizations, this approach would reduce the risks associated with using public clouds in multiple jurisdictions.

A private cloud allows a company host its applications and/or data and content in a private cloud environment that emulates a public cloud, in terms of access from the Internet. Companies that have a private cloud typically host the application in order to maintain highly secure company content.

3. Hybrid cloud. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability.

In a hybrid cloud, a company may establish a private cloud to maintain security for its data but some of that data may be in the form of a report that is publicly available. In this case the private cloud maintains the confidential data but the publicly available reports may be placed in a public cloud.

4. Community cloud. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on- or off-premises.

What Does it Mean?

From the perspective of anyone not architecting cloud solutions, we can boil down "the cloud" to some common characteristics.

Cloud solutions are generally consumed, and paid for, using a utility model: you buy what you need, scaling up or down as needed, and you pay for it, just like you would for gasoline, electricity, or water. This is a significant benefit for organizations with significant spikes in demand – traditionally you would either have to maintain additional resources sufficient for those spikes, or run into bottlenecks during peak periods.

Cloud solutions are paid for on an ongoing basis, rather than the traditional software model of paying at the time of acquisition and then purchasing upgrades or annual maintenance packages. From your finance department's perspective, this means changing the cost of the solution from a capital expense, that is then depreciated over time, to an operating expense. This may also mean that getting started with a cloud solution is cheaper than a traditional solution – and it's almost certainly faster. This does not, however, necessarily mean that a cloud solution is cheaper than its on-premises counterpart.

That said, there can be significant hard and soft cost savings associated with the cloud because maintenance, including patches, new releases, hot fixes, etc. is all done by the cloud provider rather than the organization's IT team. This helps to free up IT resources to do things that provide greater value to the organization rather than simply maintaining legacy applications. In other words, cloud solutions can be significantly easier to deploy and maintain over time.

One way to think of the cloud is using the analogy of a house. If you buy a house, you have absolute control over what appliances to have, what colors to paint it, how often to clean it or replace appliances, etc. You also have absolute responsibility for those things. If you rent a house or apartment, you have less control over those things, but if anything breaks, you call someone and they have the responsibility to fix it. The analogy isn't perfect, but it may be a useful way to consider the cloud.

One more key point. Cloud platforms and solutions are designed with a certain amount of decentralization. This is often for performance reasons, but it can also help mitigate the risks associated with disaster recovery and business continuity. Any number of things could happen to shut down your data center such as power failures, natural disasters, and the like. The decentralized nature of cloud providers means that even if one data center goes down or loses data, there should be other data centers in other geographic locations that will not be impacted and from which the service and data can be recovered.

Security

There is perhaps no greater concern for organizations contemplating the cloud than security. No organization wants to join the ever-growing list of data breaches.

However, with very few exceptions, an argument can be made that cloud-based solutions are more secure than most organizations' on-premises solutions.

First, some attacks are executed using brute-force hacks. The cloud providers, at least the established ones, know that they are high-profile targets for everyone from state actors to criminals to random bored teenagers. They spend significant resources not only on hardening their information security defenses, but on rigorously testing them.

Many other breaches are committed by insiders – disgruntled employees or former employees who, for whatever reason, still have access to those systems. Cloud providers can add an extra layer of protection insofar as a) their employees don't work at your business, so they are a relatively disinterested third party, and b) data can be segregated and stored in such a way as to minimize the results of a breach, for example by encrypting data while it as at rest and any time it is being accessed.

Cloud provider data centers often include robust physical security measures as well as identify and access control mechanisms which compare favorably to those used by organizations.

But the bottom line is that if you use a known, mature cloud service or provider, both the physical and information security measures in place will be at least as secure as those available to most organizations, and significantly better than many organizations are able to provide internally.

© AIIM – 2019

Data Sovereignty

The idea of data sovereignty is that different jurisdictions, especially countries, have different laws around data storage, privacy, and data protection. A related topic is that of data residency – that is, the requirement, typically for government data, to reside exclusively within its country of origin. Sometimes these laws are very broad, such as the General Data Protection Regulation in the European Union. In other cases, they are more targeted to specific types of information or specific industry sectors such as healthcare or financial services.

Historically countries have generally only been able to enforce those laws with respect to entities over which they have legal jurisdiction. So if a company is based in the United States, it wouldn't generally be bound by laws in the European Union or Canada – though any subsidiaries in those countries probably would.

Enter the cloud (and, really, the internet). The distributed nature of cloud-based applications means that an organization's data could be stored anywhere in the world and accessed from anywhere in the world. Data could be stored in – and backed up to – multiple data centers in different geographic locations. And many cloud providers rely on third parties for some of their offerings – partners, hosting services, subcontractors, etc. – who may similarly be located anywhere in the world.

The first step to address this is to know whether or how you are using, or planning to use, the cloud. What services, what approaches, what kind of information? This in turn will help you to identify and assess your potential exposure to data sovereignty issues as well as to any data residency requirements. Next, read the fine print of the agreement and ask the tough questions: Where are the data centers? Are there partners or contractors involved? The fines involved can be significant, so if the vendor is unable to meet these needs or answer these questions you should consider a different provider.

Uptime and Availability

Many organizations have concerns that their cloud-based systems could go down, rendering the information they contain inaccessible for some period of time. At the same time, it's possible that the vendor might go out of business, or change business models, such that access to data is permanently removed.

The first thing to note is that the larger, more mature cloud providers are, frankly, often more robust and available than organizations' on-premises systems. Downtime is generally measured in minutes to hours of downtime per year – compare that with the downtime, planned and unplanned, for your key systems last year. And when they go down, it's noted instantly and fixed very quickly because they have dedicated teams of highly skilled professionals to do that. Many organizations don't have that expertise on staff, and certainly not for every one of their systems.

Similarly, the redundancy and distributed nature of most cloud solutions means that even if there is some kind of adverse event, it may only affect a portion of the service or for a very short period of time.

It's important to review the service level agreements and contracts, to ensure that the target service is robust and reliable, but this is considered generally to be one of the key benefits of using the cloud compared to on-premises solutions.

Vendor Lock-in

Even if the vendor doesn't go out of business, at some point the organization may wish to move its data to another application or provider for any number of reasons. There are a couple of things to consider here.

First, ensure that the vendor does not put your data into proprietary structures and formats with no way to get them out. Many solutions will store the data in whatever format it's uploaded in, but due diligence is still required here.

Next, get agreement on what a process would look like for exporting data from the solution and, potentially, whether there is assistance to move it to another one. This includes things like the data itself, metadata, folder, or filing structures, etc.

And get agreement as to exporting and disposition of any copies of the organization's data upon termination of the agreement.

All of this should be included in the contracts and any related agreements so that there are no surprises when the day comes.

Change Management

A common definition of change management is one that can guide our activities over the course of an implementation.

"Change management is the application of the set of tools, processes, skills and principles for managing the people side of change to achieve the required outcomes of a change project or initiative."

Think back to the goals, objectives and issues that were addressed in the business and technical assessment phases of our implementation so far. The deployment is in fact an initiative intended to achieve change. This change may be to improve customer responsiveness, collaborate faster across geographies, better meet external compliance mandates, or automate key business processes.

Implementing new software, however, is only one part of the possible solution. Changing the habits, priorities and day-to-day activities of information workers is also part of meeting deployment objectives. A program and dedicated plan to manage the change in employee behaviors to parallel the changes in tools or procedures is essential to success.

Change Management Is About People

Your users are the agents for successful transformation of the organization. Their adoption and successful use of the new information management application will be a significant factor in determining the overall success of the initiative.

The key to successful management of organizational change lies in the people. Provide the appropriate level of attention to the human side of projects and you will eliminate one of the greatest causes of project failure.

The Change Readiness Assessment

Regardless of the kind of change, whether technological, cultural, procedural, role-based, or any other, it must first be decided if an organization is ready to face the change and adjust to it. Change may be coming whether it's welcome or not. Determining readiness is a big factor in the potential success of your information management project.

Organizational change is always going to appear threatening to people as it is often linked to job security. Some enterprises freely disseminate information regarding strategy changes. Other firms are very secretive and feel that this is for senior management only.

Practitioners should be as open and honest with staff about change as they possibly can. Typically, people will more readily embrace the change process if clear information is available.

Assess your enterprise's readiness for change. The readiness of both the management to support the change and affected workers to accept and adapt to the change are the most crucial factors in the success, or failure, of your project. Management may be far more ready to change than the potentially-affected workers, particularly if the idea for the proposed change is coming from management – as it typically is. However, just because you have meetings with middle or senior management who are very enthusiastic about this new project, doesn't mean that the organization as a whole is ready to change.

Change Management Strategy

As the part of the implementation team, you must be aware of the extent to which your projects may introduce change, and its impact on the organization. You must deal with this issue.

The change impact should be discussed as early as the project initiation stage. Ask yourself the questions:

- What is changing: technology? process? retention periods? version control procedures?
- How will it affect the employee activities?
- What are the selling points of this new system, procedure, or practice?
- What would make someone WANT to make this change?

It is inevitable that change will lead to a productivity dip as shown here. This includes at a minimum the impact to the organization of pulling people into meetings, conducting assessments and interviews, doing reviews, and the like.

© AIIM – 2019

As the organization works its way through the four stages of change, there can be a negative impact on productivity – the "Productivity Dip." The sooner the change management strategy is implemented, the sooner the productivity dip can be addressed, and the organization can move towards the desired state and see the benefits of the change.

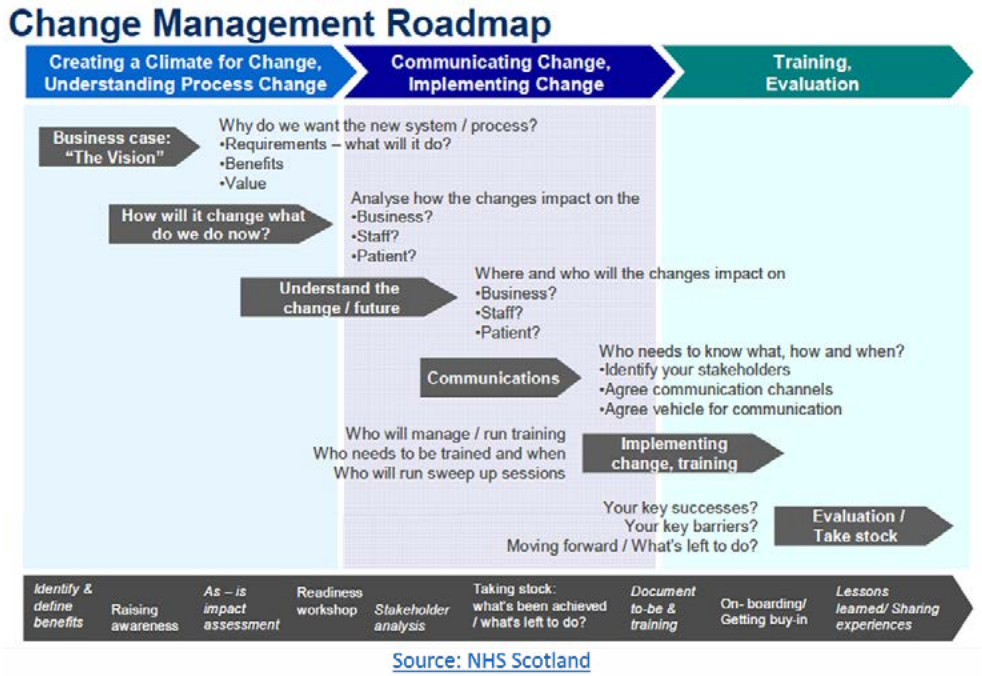
Change isn't easy. With a good strategy, you can make it less painful on the deployment team, sponsors, and the end users that need support.

A change management strategy is much more than a communications plan.

Change management means identifying the possible impacts of the change and coming up with plans to minimize the disruption to the working life of the end users, and even getting them to accept, embrace, and ultimately champion and promote the change themselves to their colleagues.

Change Management Roadmap

This diagram is an example of a change management roadmap, or high-level plan, used by a public sector health care agency. This graphical representation can provide a useful summary of the elements required in a typical change management plan across a content management deployment. As you develop a change management plan for your own deployment, consider creating a single slide or one-page summary to help in workshops, meetings and other venues where communication of the change management tasks and priorities will be required.



Source: NHS Scotland

Accountabilities for Change

The roles and responsibilities for change can be relatively easy to identify and document, but ensuring that accountability is accepted and responsibility taken can be a more significant cultural issue.

IT management must ensure they communicate clearly and proactively with the IM team and business sponsors on the costs and timelines for solution delivery. Has IT agreed to align the design, development and delivery priorities to the business requirements and user pain points?

Senior management must be accountable for ensuring their employees are equipped with the vision, coaching, new tools and encouragement to adopt the new system and put effort into learning new processes and software systems. Management should encourage and lead change management initiatives by example, not merely by issuing directives.

Human resource teams should guide the managers in adopting new reward or recognition systems to align with the new desired behaviors. HR and internal communication teams should also take a close look at any required change to appropriate use policies or social media guidelines if the new IM deployment introduces new collaboration, mobile, cloud, or social media tools or communication expectations on employees.

Trainers, coaches, and workshop facilitators (whether internal or external) should be accountable for delivering the right level of coaching to align with employee skill sets and timing of system adoption.

User champions must be accountable to provide open, constructive feedback through the early assessment phases. They need to be prepared to accept an increased workload during the planning and rollout phases, as they participate in the feedback cycle along with their day-to-day tasks. They should also be prepared - and empowered by management - to help coach their peers during the pilot or production rollout phases.

The project management team is accountable for overall coordination and documentation of the responsibilities that each role must accept, and be prepared to escalate any roadblocks to the executive stakeholders for action.

The Change Management Plan – Development

Let's step through the key phases in the development of a change management plan. Even before we dive into the details of the business and technical assessment phases of our deployment, we should already anticipate the need for change management. The details of the plan can be completed as the implement progresses. Be ready, however, to document and plan for:

First, the governance structure of the team responsible for change management. What roles, titles and business units are represented, and which ones are responsible for decision-making?

Second, be prepared to document the current baseline – the business process and content-handling activities that reflects today's situation. Begin identifying which of these are most in need of change as part of the requirements gathering process. Highlight these as top candidates for change management activities.

Third, document the desired new processes and procedures, the "to-be" or "desired future state." The delta, or difference between the current state and future state, should form the core of the change management focus. These can be documented in greater detail as the assessment phases are completed and requirements are finalized.

Fourth, assess the impact of expected change on people and the processes they run as part of their work activities. Understand their level of technical expertise, process expertise, and ability to adapt to new tools and allocate coaching and training time as necessary.

Fifth, identify potential barriers. For example, if none of the customer service team has ever used a tablet device in their consumer computing activities, and devices will be issued as part of the new case management system, plan for how to surmount this barrier. Or if the sales team is deep into closing a fiscal year end, ensure coaching for this team is not scheduled at this mission critical time.

Sixth, ensure good communication habits start right from day one. Brief all members of the project team – IT management, compliance officers, business stakeholders, and so on – and ensure there is agreement and buy-in to support the change management plan.

The Change Management Plan – Updates

The change management plan must also be updated as the planning and design phases of the deployment are begun. These activities should include:

First, mapping the critical change management activities into the overall project plan so that nothing is missed, dependencies are recognized, and resources are properly allocated. This might mean scheduling time with internal communication staff, pushing notifications through intranet or other in-house channels, preparing or booking classrooms, and scheduling time on employee calendars for training and orientation.

Second, the communication plan should be finalized, identifying the channels of communication that can be used. These may include newsletters, all-staff meetings, management team calls, email, or company collaboration sites. Finalize any variations in the communication message, as there may be different priorities and timelines for management, remote workers, customer service reps and IT support desks.

Third, develop the training plans for key groups, including content, materials, delivery method, timing, and follow up coaching requirements.

The Change Management Plan – Execution

As the deployment moves into pilot development and deployment stages, the execution of the plan can now be completed.

First, schedule the planned training, coaching, and communication activities that we outlined in the previous two sections.

© AIIM – 2019

Second, document the progress of training and communication activities so that there is a clear understanding of what has been completed, what is outstanding, and how users are responding to the activities.

This document is important for the third aspect of change management during the development and deployment stage: the collection of feedback and analysis of areas of improvement. Information management systems that are deployed using an agile or iterative approach should constantly evaluate user reaction and adoption in order to tune and improve the next set of training or communication activities.

Finally, confirm if all of the barriers that were identified in the assessment phase have, in fact, been addressed. Ensure that any technical, cultural, or process changes highlighted as essential in the planning stages have been dealt with and resolved.

The Communications Plan

The purpose of the communications plan is to provide the foundation for project communications. The communications plan should be developed early in the project as communications should start early as well and continue through the entire project. It should include the following:

- **Who?** Identify the intended audiences: stakeholders, project team members, end users, etc.
- **What?** Identify the types of information to be communicated to each audience.
- **When?** Identify the timing and frequency of communications for each audience.
- **How?** Identify the delivery methods and channels that will be used to communicate to each audience.

The Training Plan

The training plan should be developed early in the project. It should address several key considerations:

- **What** types and levels of training will be required? For example, process-focused training will be quite different from application-oriented training (and both may be needed). Similarly, refresher training may be different – shorter, online, even on-demand – compared to a full- or multi-day instructor-led workshop.
- **Who** are the audiences for training? End users have different training needs than system administrators; different end users may have different training needs as well.
- **When** (and where) will the training be delivered – instructor-led vs. online; formal training sessions before a pilot vs. just-in-time training available anytime?
- **How** will the training be acquired? Larger organizations may have a training department to address training in process changes, while smaller organizations may elect to purchase training from an outside training company or consulting firm. Along the same lines, will the organization engage with an outside training company to deliver all of the training, or will it do a “train the trainer” type model and bring the rest of it in-house?

Keys to Success

The keys to change management success include:

- Allocate sufficient time and resources to ensure that change can happen – and then add some more as most organizations consistently underestimate the amount of resistance that will be encountered.
- Communicate value and expectations – to everyone.
- Coordinate across teams and stakeholder communities to ensure that the message is clear and consistent.
- Make user adoption a priority!
- Define a change roadmap for all types of change and in all phases of an initiative.
- Select quick wins to promote success.
- Communicate and train, early and often, and keeping mind different audiences need different messages.
- Sustain the change over time.



Test Your Knowledge

Question 1:

What activities take place after the implementation is complete? (select 2)

a. Requirements definition
b. Lessons learned
c. User acceptance testing
d. Extension of capabilities to new groups

Question 2:

What are the key benefits associated with a cloud-based deployment model? (select 2)

a. It is easier to implement and support over time
b. It provides faster access to information via search
c. It ensures personal data is protected from unauthorized access
d. It scales up and down dynamically to meet demand

Question 3:

When should change management activities for an information management project begin?

a. As soon as the project begins
b. Once the solution has been implemented
c. When senior management determines it is appropriate
d. After HR approves the communications plan

© AIIM – 2019



Test Your Knowledge - Answers:

Answer to Question 1:

What activities take place after the implementation is complete? (select 2)

a. Requirements definition
b. Lessons learned
c. User acceptance testing
d. Extension of capabilities to new groups

The best answers here are B, lessons learned, and D, extension of capabilities to new groups. Requirements definition happens before the implementation (and should shape the solution selection and implementation process). User acceptance testing is one of the last steps required to complete the implementation.

Answer to Question 2:

What are the key benefits associated with a cloud-based deployment model? (select 2)

a. It is easier to implement and support over time
b. It provides faster access to information via search
c. It ensures personal data is protected from unauthorized access
d. It scales up and down dynamically to meet demand

The best answers here are A, it is easier to implement and support over time, and D, it scales up and down dynamically to meet demand. A is correct because the cloud provider is responsible for upgrades, hot fixes, backups, new releases, etc. D is inherent to cloud-based solutions. For B, a cloud-based architecture wouldn't necessarily have any impact on search and retrieval speeds. For C, cloud solutions can help to safeguard personal data, especially from external threats, but ultimately it cannot ensure that personal data is protected from internal threats by people who have access to the system.

Answer to Question 3:

When should change management activities for an information management project begin?

a. As soon as the project begins
b. Once the solution has been implemented
c. When senior management determines it is appropriate
d. After HR approves the communications plan

The best answer is A, as soon as the project begins. Once the solution has been implemented is far too late. Senior management and HR should not be bottlenecks in the way of effective change management activities.

Thank You!

This concludes our study guide on the Foundations of Intelligent Information Management.



Here at AIIM, we believe that information is your most important asset and we want to teach you the skills to manage it. We've felt this way since 1943, back when this community was founded.

Sure, the technology has come a long way since then and the variety of information we're managing has changed a lot, but one tenet has remained constant. We've always focused on the intersection of people, processes, and information. We help organizations put information to work.

AIIM is a non-profit organization that provides independent research, training, and certification for information professionals.

Visit us at www.aiim.org

© 2019

AIIM
8403 Colesville Road, Suite 1100
Silver Spring, MD 20910, USA
+1 301 587 8202
www.aiim.org

AIIM Europe
Office 1, Broomhall Business Centre,
Worcester, WR5 2NT, UK
+44 (0)1905 727600
www.aiim.org