

# **PDF/A-3 ist nicht per se ein Archivformat**

**Statement zum Einsatz von PDF/A-3 als  
Langzeitarchivformat für die Fachzeitschrift  
„BIT“**

Dr. Ulrich Kampffmeyer



**Hamburg, 2013**

# PDF/A-3 ist nicht per se ein Archivformat

Statement zum Einsatz von PDF/A-3 als Langzeitarchivformat für die Fachzeitschrift „BIT“



## PDF/A-3 ist nicht per se ein Archivformat

PDF/A-3 ist der dritte Teil der ISO-Norm 19005 für das PDF-Archive-Format. Er erschien kurz nach der Veröffentlichung von PDF/A-2. Während PDF/A-2 eine konsequente Weiterentwicklung von PDF/A-1 ist und einige „Unschönheiten“ des ursprünglichen Archivstandards ausgleicht, folgt PDF/A-3 meines Erachtens einem ganz anderen Ansatz. PDF/A-3 ist kein Archivformat per se sondern ein Container für beliebige digitale Inhalte.

PDF/A-1 und PDF/A-2 waren trotz unterschiedlicher Ausprägungen (a, b ...) echte Archivformate. Dynamische Komponenten wie Links, Zeichensätze etc. wurden direkt eingebunden und eingefroren. PDF/A in den ursprünglichen Ausprägungen ist ein „Plattklopfen“ und Normieren von weniger stabilen Formaten. Dabei hat der Anwender die Möglichkeit im Header im XML-Format identifizierende und beschreibende Metadaten beizufügen. Hierdurch kann ein selbstbeschreibendes Informationsobjekt generiert werden, das auch den Langzeitanforderungen z.B. nach ISO 14721:2012, AIP Archive Information Package, genügt. Das Objekt ist mit einem beliebigen PDF-Viewer reproduzierbar, d.h. verlustfrei anzeigbar und in dieser Form druckbar. Hauptsächliches Ziel ist es, ein stabiles, von Menschen lesbares, bildhaftes Format zu gewinnen. PDF/A ersetzt hier zum Beispiel das TIF-Format.

Anders nun bei PDF/A-3. PDF/A-3 ist ein Container auf Basis von PDF/A-1, der beliebige Inhalte aufnehmen kann. Die Komponenten werden nicht kontrolliert. So können nicht ohne die erzeugende Anwendung anzeigbare Spezialdateiformate, Dokumente mit externen Links, sogar Viren und Malware in ein PDF/A-3 eingepackt werden. Erst wenn man ein solches Objekt aufruft, offenbaren sich die Probleme des Inhalts. Der Container wird zwar geöffnet, aber ob der Anwender damit etwas anfangen kann, bleibt offen. Besonders wenn man im Rahmen der Langzeitarchivierung, Digital Preservation, in Jahrzehnten und noch längeren Zeiträumen denkt. Auch eine spätere verlustfreie Migration des Formates in ein Folgeformat ist auf Grund des nicht kontrollierten Inhalts kaum sicherzustellen.

Natürlich gibt es Argumente für PDF/A-3, zum Beispiel von der amerikanischen NARA, die schon länger auf der Suche nach einem geeigneten Containerformat ist. Ziel ist es, in einem Objekt zusammenhängend verschiedene Formate eines Objektes, z.B. Renditionen, gemeinsam zu verwalten und zu speichern. Wo heute noch in der Regel über die Indexdatenbank mit dem gleichen Indexsatz auf zwei verschiedene separat gespeicherte Renditionen eines Objektes, z.B. Original E-Mail und daraus erzeugendes PDF/A-1b, verlinkt wird, tritt nun ein einheitliches Objekt. Auch Initiativen wie ZUGFeRD mit der XML-basierten elektronischen Rechnung favorisieren PDF/A-3 um neben dem visualisierbaren Rechnungsdokument den automatisch auswertbaren ZUGFeRD-XML-Rechnungssatz

## PDF/A-3 ist nicht per se ein Archivformat

Statement zum Einsatz von PDF/A-3 als Langzeitarchivformat für die Fachzeitschrift „BIT“



mitzuspeichern. Dabei wird aber immer unterschätzt, dass es auf der Empfängerseite eines solchen Objektes auch die passenden Anwendungen gibt, die mit den mehreren Komponenten automatisiert auch etwas anfangen können – und dies bei der Archivierung über lange Zeiträume. So stellt sich die Frage, ob PDF/A-3 zusätzlich zur Archivierung auch noch ein geeignetes Transportformat sein kann. Und man kann PDF/A.-3 gegebenenfalls auch noch als Zwischenspeicher im Rahmen von Prozessen sehen, um zusammengehörige Informationen auch als ein transportables Objekt bereitzustellen. Es gibt also trotz der impliziten Unzulänglichkeiten durchaus sinnvolle Anwendungsgebiete – aber außerhalb der Digital Preservation.

Die beschriebenen Probleme für die Langzeitarchivierung lassen sich nicht technisch beim Generieren eines PDF/A-3 abfangen. Um ein echtes Archivobjekt zu sein, müsste nämlich mindestens eine der Komponenten ein PDF/A-2 des gesamten Inhalts sein oder zumindest die für Menschen lesbare Version des Inhalts darstellen können. Ohne mindestens eine Langzeitarchivierungskomponente zu enthalten ist PDF/A-3 nicht als Langzeitarchivformat geeignet. Es wäre meines Erachtens sinnvoller gewesen, im Rahmen der großen PDF-Standard-Familie – man erinnere sich, es gibt auch PDF selbst als ISO-Norm, PDF für die Druckvorstufe, für technische Zeichnungen usw. – eine neue Klasse aufzumachen: PDF/C für PDF/Container. Dies würde einer Verwechslung mit den „echten“ Archivformaten PDF/A-1 und PDF/A-2 entgegenwirken. Nun sind alle Verbände, Hersteller, Berater und IT-Bereiche aufgefordert, allen die PDF/A-3 als „modernsten PDF/A-Standard“ einsetzen wollen, zu erklären, welche Implikationen PDF/A-3 mit sich bringt. Ohne organisatorische Maßnahmen, ohne Verantwortungsbewusstsein der solche Objekte generierenden Anwender, ohne Klarheit über die Restriktionen der freien „Befüllbarkeit“ ist PDF/A-3 nicht standardmäßig als Standard für die Langzeitarchivierung verwendbar.

# PDF/A-3 ist nicht per se ein Archivformat

Statement zum Einsatz von PDF/A-3 als Langzeitarchivformat für die Fachzeitschrift „BIT“



## Über den Autor

Dr. Ulrich Kampffmeyer ist seit über 30 Jahren im Thema Dokumentenmanagement zu Hause. Er gründete und leitete entsprechende Fachverbände, arbeitete bei Standards mit, ist von Anfang an einer der internationalen Verfechter der ECM-Vision, und hat mit zahlreichen Publikationen und Vorträgen den ECM-Markt befruchtet. Er gilt als Mentor der Branche in Europa und wird auch der deutsche „ECM-Papst“ genannt. Seit 1992 ist er als Unternehmensberater für Information Management unterwegs und leitet das Beratungsunternehmen PROJECT CONSULT (<http://www.PROJECT-CONSULT.com>).



Dr. Kampffmeyer ist anerkannter Kongressleiter, Referent und Moderator zu Themen wie elektronische Archivierung, Records Management, Dokumentenmanagement, Workflow, Rechtsfragen oder Wissensmanagement. Auf zahlreichen nationalen und internationalen Kongressen und Konferenzen wirkte er als Keynote-Sprecher mit.

Von der ComputerWoche wurde er 2003 und 2011 unter die 100 wichtigsten IT Macher Deutschlands gewählt. Sein Curriculum findet sich auf Wikipedia [http://bit.ly/WP\\_DrUKff](http://bit.ly/WP_DrUKff).