

## KOST-Newsletter Quartal 4, 2017

### PDF/A-2u

Im Dezember 2017 hat die Preservation-Planning-Expertengruppe PPEG der KOST die Version 5.1 des [Katalogs archivischer Dateiformate KaD](#) publiziert. Darin empfiehlt sie, bei der Konvertierung in PDF/A nach Möglichkeit die Version 2 und das Konformitätslevel U zu wählen, also PDF/A-2u-Dokumente zu erzeugen. Hier erklären wir, auf welchen Überlegungen diese Empfehlung beruht.

#### **Warum Version 2 und nicht Version 1?**

Jede Version von PDF/A beruht auf einer Version von PDF. Der PDF/A-Standard macht nichts anderes als festzuhalten, welche Elemente der zugrundeliegenden PDF-Spezifikation verboten oder zwingend sind. Umgekehrt heisst das aber auch: Was in der zugrundeliegenden PDF-Spezifikation nicht möglich ist, ist in PDF/A ebenfalls ausgeschlossen. PDF/A-1 beruht auf PDF 1.4 von 2001, also auf einer sechzehnjährigen Spezifikation. Einiges von dem, was heute in PDF gang und gäbe ist, fehlt darin – zum Beispiel die Möglichkeit, mit Transparenz umzugehen, und die Verwendung der JPEG2000-Komprimierung für Bilder. Bei der Migration neuerer PDF-Dokumente in PDF/A-1 besteht also das Risiko, dass gewisse Elemente nicht wunschgemäss konvertiert werden können. Aus diesem Grund empfehlen wir, direkt Version 2 zu wählen.

#### **Warum Version 2 und nicht Version 3?**

Das ist schnell erklärt: PDF/A-3 erlaubt das Einbetten von Dateien in sämtlichen möglichen Formaten in das PDF/A-Portfolio, also genau das, was wir mit einem Formatkatalog und der Beschränkung auf archivtaugliche Formate zu verhindern suchen. PDF/A-3 ist für die dauerhafte digitale Archivierung nicht geeignet.

#### **Warum Konformitätslevel U und nicht B?**

Ab PDF/A-2 existiert zusätzlich zu den Konformitätslevels A und B auch das Level U. Dieses entspricht dem weniger strikten Minimallevel B, erfordert aber zusätzlich die Abbildung des gesamten Texts in Unicode. Dies garantiert die Indexierbarkeit des Textes und die Volltextsuche. In ihrer Praxis ist die KOST auf Dokumente in PDF/A-1b gestossen, in denen fehlerhafte Schriften eingebettet sind, welche die Volltextsuche verunmöglichen. Deshalb sollte für neue Migrationen das Konformitätslevel U gewählt werden.

#### **Warum Konformitätslevel U und nicht A?**

Das Konformitätslevel A entspricht der vollständigen Übereinstimmung mit dem Standard. Dies umfasst insbesondere sämtliche Massnahmen, um das Dokument barrierefrei zu machen und eine Ausgabe z.B. auf Screenreader zu gewährleisten (Tagged PDF). Diese Konformität ist schwierig zu erreichen. Sie sollte gewählt werden, wenn es möglich ist, kann aber kein Killerkriterium für die digitale Archivierung sein.

#### **Was sollen wir mit unseren alten PDF/A-1b tun?**

Nichts. PDF/A-1b bleibt ein praktisch ideales archivtaugliches Format, zumal der Standard nicht abgelöst wurde durch spätere Versionen. Archivalien in PDF/A-1b brauchen nicht migriert zu werden. Es ist nicht abzusehen, dass sie mittelfristig unlesbar würden.

### KaD v5.1

Die neue Version 5.1 des Katalogs archivischer Dateiformate KaD enthält neben unzähligen Fehlerkorrekturen und aufdatierten Weblinks zwei grössere Änderungen:

- Die einzelnen Formate werden im KaD gemäss einem Katalog von Kriterien analysiert, wobei der Erfüllungsgrad jedes Kriteriums mit einer Note von 1 (Minimum) bis 4



## Newsletter CECO du 4e trimestre 2017

### PDF/A-2u

En décembre 2017, le groupe d'experts Preservation planning du CECO a publié la version 5.1 du [Catalogue des formats de données d'archivage \(Cfa\)](#). Il y recommande, en cas de conversion en PDF/A, de choisir, dans la mesure du possible, la version 2 et le niveau de conformité U, donc de générer des documents PDF/A-2u. Nous vous expliquons ci-après quelles réflexions motivent cette recommandation.

#### ***Pourquoi la version 2 et pas la version 1 ?***

Chaque version PDF/A est basée sur une version PDF. La norme PDF/A ne fait rien d'autre que fixer quels éléments de la spécification PDF qui est à la base sont interdits ou obligatoires. A l'inverse, cela signifie cependant également que ce qui n'est pas possible dans la spécification qui est à la base, est également exclu dans le PDF/A. PDF/A-1 est basé sur la version PDF 1.4 de 2001, donc vieille de seize ans. Une partie de ce qui est aujourd'hui courant dans le PDF manque dans cette version, par exemple la possibilité de travailler dans la transparence et d'utiliser la compression JPEG2000 pour la compression des données graphiques. Il existe donc un risque de ne pas pouvoir convertir certains éléments comme on le souhaite lors de la migration de documents PDF plus récents en PDF/A-1. C'est pour cette raison que nous recommandons de choisir directement la version 2.

#### ***Pourquoi la version 2 et pas la version 3 ?***

L'explication est simple : le PDF/A-2 autorise l'intégration de fichiers de tous les formats possibles dans le portfolio PDF/A, c'est-à-dire exactement ce que nous cherchons à éviter en publiant un catalogue des formats et en nous restreignant à des formats adaptés pour l'archivage. Le PDF/A-3 ne convient pas pour un archivage numérique durable.

#### ***Pourquoi le niveau de conformité U et pas le B ?***

En plus des niveaux de conformité A et B, il existe depuis la version PDF/A-2 le niveau U. Ce dernier correspond au niveau minimal B, moins strict, mais avec l'exigence supplémentaire de représenter l'ensemble du texte en Unicode. L'indexation du texte et la recherche en plein texte sont ainsi garanties. Dans sa pratique, le CECO a rencontré des documents en PDF/A-1b dans lesquels sont intégrées des polices d'écriture incorrectes qui rendent la recherche en plein texte impossible. C'est pour cette raison qu'il faut choisir le niveau de conformité U pour de nouvelles migrations.

#### ***Pourquoi le niveau de conformité U et pas le A ?***

Le niveau de conformité A correspond à une conformité totale avec la norme. Cela comporte en particulier toutes les mesures rendant le document totalement accessible et garantissant une édition par exemple sur lecteur d'écran (tagged PDF). Cette conformité est difficile à atteindre. Il y a lieu de privilégier dans la mesure du possible, mais elle ne devrait pas représenter un critère d'exclusion pour l'archivage numérique.

#### ***Que faire avec notre ancien PDF/A-1b ?***

Rien. En pratique, le PDF/A-1b reste un format adapté pour l'archivage idéal d'autant que la norme n'a pas été remplacée par des versions ultérieures. Il n'est pas nécessaire de migrer les documents archivés en PDF/A-1b. Il est peu probable qu'à moyen terme ils ne soient plus lisibles.

### Cfa v5.1

Outre d'innombrables corrections d'erreurs et des liens mis à jour, la nouvelle version 5.1 du Catalogue des formats de données d'archivage (Cfa) contient deux modifications plus importantes :



## **Veranstaltungshinweise / Calendrier des événements**

Nachfolgend Hinweise auf Veranstaltungen von Archiven, Organisationen und Firmen, die für die KOST-Trägerarchive relevant sind und in der Schweiz stattfinden.

*Ci-après, le calendrier des événements organisés en Suisse par des archives, organisations et entreprises sur des thèmes importants pour les archives membres du CECO.*

- 29.01.18** Fedora-Austausch  
*Echange Fedora*  
Bern, Schweizerisches Bundesarchiv / *Archives fédérales suisses*
- 20.03.18** Kolloquium „AIS-Modell“  
*Colloque « Modèle pour logiciels de gestion d’archives »*  
Bern, Schweizerisches Bundesarchiv / *Archives fédérales suisses*
- 11.04.18** KOST.Forum 07  
KOST-Projekt 14-016 Amtliche Vermessung  
*CECO.Forum*  
*Projet du CECO 14-016 Mensuration officielle*  
Zürich, Walcheturm (Walchetor 2)

Wenn Sie einen Veranstaltungshinweis im KOST-Newsletter publizieren wollen, kontaktieren Sie uns bitte unter [info@kost-ceco.ch](mailto:info@kost-ceco.ch).

*Si vous souhaitez publier un événement dans le calendrier de la Newsletter du CECO, veuillez s.v.pl. nous contacter à l'adresse [info@kost-ceco.ch](mailto:info@kost-ceco.ch).*