

## 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. Dies 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

(Maximum) bewertet wird. Für die Kriterien Offenheit, Lizenzfreiheit, Implementierung und Verifizierbarkeit hat die PPEG die Erfüllungsgrade neu präzise definiert. Dies erleichtert die Nachvollziehbarkeit der Bewertung und erlaubt Vergleiche über die Kategoriengrenzen hinweg. Bei den anderen Kriterien ist eine präzise Definition des Erfüllungsgrades nicht sinnvoll.

- Der KaD analysiert Formate und macht Aussagen für die sieben Formatkategorien Text, Bild, Audio, Video, Tabellenkalkulation, Datenbanken und Hypertext. Weitere Kategorien sind für Archive zu wenig relevant, um eine ausführliche Behandlung im KaD zu rechtfertigen. Die PPEG hat jedoch im Rahmen der langfristigen Planung verschiedene andere Formatkategorien kurz und knapp untersucht. Diese Untersuchungen werden im KaD v5.1 zu Dokumentationszwecken zugänglich gemacht unter Weitere Kategorien. Diese Dokumente geben den aktuellen Wissensstand der PPEG wieder. Eine weitergehende Analyse ist nicht vorgesehen.

Katalog archivischer Dateiformate (KaD), Version 5.1.0, Dezember 2017																																					
Kriterienkatalog zur Bewertung der Archivtauglichkeit von Dateiformaten																																					
Kriterien	Erklärung	Formate																																			
		TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	OXML	TIFF	JPEG	JPEG2000	PNG	DNG	MP3	WAV	Unkomprimiert	Digital Video	FFV1	MPEG-2	MPEG-4	Microsoft Office Open XML																
Offenheit	<i>Die Spezifikation des Formats ist vollständig und öffentlich zugänglich</i>	1	4	4	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4																
Lizenzfreiheit	<i>Der Gebrauch des Formats ist durch keine Lizenzen eingeschränkt oder einschränkbar</i>	1	3	3	3	1	3	3	3	4	3	3	3	4	3	1	4	1	1	3	3																
Verbreitung	<i>Das Format ist weit verbreitet; deshalb können lange Migrationszyklen erwartet werden</i>	1	4	4	4	4	1	2	4	4	4	3	3	2	4	3	4	1	4	4	4																
Funktionalität	<i>Das Format kann die Eigenschaften seiner Kategorie repräsentieren</i>	1	1	3	2	3	3	4	4	4	2	4	3	4	2	4	2	3	2	3	2																
Implementierung	<i>Für das Format existieren mehrere Implementierungen</i>	0.5	4	4	4	4	3	4	3	4	4	4	4	4	4	3	4	2	4	4	3																
Speicherdichte	<i>Das Format ermöglicht eine hohe Speicherökonomie</i>	0.5	4	3	2	3	2	3	3	2	4	3	3	2	3	1	4	2	3	2	3																
Verifizierbarkeit	<i>Es existieren Methoden und Werkzeuge zur Erkennung und Validierung</i>	1	2	4	4	4	2	3	2	4	4	3	3	3	4	3	1	2	3	2	3																
Komplexität	<i>Das Format ist möglichst wenig komplex.</i>	0.5																																			
Selbstdokumentation	<i>Das Format kann Metadaten einholen und ermöglicht deren Verwendung</i>	0.5														1	1	3	3	3	3																
Best Practice	<i>Das Format wird in Archiven bereits verwendet</i>	0.5	4	2	4	3	1	1	1	3	2	3	1	2	3	4	2	2	1	2	2																
Perspektive	<i>Das Format ist im archivischen Sinn zukunftsfähig und hat noch weiteres Potenzial</i>	0.5	1	1	3	4	1	2	2	2	1	4	2	2	2	1	4	1	3	3	2																
Formatklasse	<i>Einteilung in einer der drei Formattypen: A) selbsterklärende Formate      D) provisorische Formate B) aktueller Formate      E) veraltete Formate C) zukünftige Formate      F) keine Aussage</i>	A	B	B	B	Ø	C	B	A	A	B	B	C	B	A	B	B	A	D	Ø	Ø																
		TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	OXML	TIFF	JPEG	JPEG2000	PNG	DNG	MP3	WAV	Unkomprimiert	Voice	Digital Video	FFV1	MPEG-2	MPEG-4	Microsoft Office Open XML															
Gewichtete Summe / 4		0.73	0.82	0.84	0.89	0.52	0.75	0.77	0.84	0.84	0.89	0.75	0.71	0.84	0.84	0.64	0.61	0.69	0.63	0.67	0.66	0.63															
Logarithmisch gewichtete Summe		0.87	1.38	1.57	2.35	0.42	0.95	1.05	1.20	1.20	1.79	1.73	0.61	1.20	1.07	0.94	1.00	0.91	1.21	0.97	1.27	1.08	1.21	0.69	0.78	1.13	1.64	0.78	1.07	1.65	0.80	1.04	1.00	0.68	0.68	1.51	1.80

## xlsadg 2.0

Die Publikation von Version 2.0 von xlsadg, der Schnittstelle zwischen Ingest und Data Management, haben wir Ihnen bereits im Sommer-Newsletter angezeigt. Das Stadtarchiv Zürich sowie die Staatsarchive AR, BE, BS, LU, SG, TG und ZH, die im Rahmen des KOST-Projektes 17-034 diese Standardversion erarbeitet haben, legen nun als Ergänzung das Mapping von eCH-0160 v1.1 zu xlsadg vor. Dieses wird bei eCH als Addendum zu eCH-0160 publiziert und verbindet so die beiden Standards in nachvollziehbarer Art miteinander.

Auf der KOST-Website finden Sie unter [https://kost-ceco.ch/cms/index.php?xisadg\\_de](https://kost-ceco.ch/cms/index.php?xisadg_de) eine Vorabpublikation des Mappings ([eCH-0160 xlsadg&EAD v2.0.xlsx](#)) sowie eine Referenzimplementierung (<https://kost-ceco.ch/isadg/app/input.php>).

## Fedora-Austausch

Von verschiedenen Archiven ist der Wunsch an die KOST-Geschäftsstelle herangetragen worden, mehr über *Fedora Repository* zu wissen. *Fedora* ist eine Softwarelösung, die bereits in etlichen Archiven im Einsatz ist. Da es sich um eine Open-Source-Lösung handelt, sind die Benutzer stärker zu eigenem Engagement verpflichtet, als das bei einer kommerziellen Lösung der Fall wäre. Darum laden wir alle Interessierten am Montag, 29. Januar 2018, nachmittags, nach Bern zu einem ersten Meinungsaustausch über das Thema ein.

Ziel des Meinungsaustausches ist es zu ermitteln, welche Kenntnisse bereits da sind, welche Bedürfnisse bestehen und wie wir das Thema weiter angehen sollen – ob in einer Weiterbildungsveranstaltung, einer Erfa-Gruppe oder einem Workshop.

## 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 la 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 :

- Chaque format est analysé en fonction d'un catalogue de critères en évaluant le degré de satisfaction pour chaque critère à l'aide d'une note de 1 (minimum) à 4 (maximum). Le groupe d'experts a redéfini précisément les degrés de satisfaction pour les critères ouverture du format, licence libre, implémentation et vérifiabilité. Cela facilite la compréhension de l'évaluation et permet de comparer les différentes catégories. Il n'est pas judicieux de définir précisément le degré de satisfaction des autres critères.

Kriterien	Erklärung	Formatkatalog zur Bewertung der Archivauglichkeit von Dateiformaten																												
		Text		Bild			Audio			Video			Protokoll			Tabellenkalkulation			Datenbanken			hypertext								
		Schreib	TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	DOCML	TIF	JPEG	AVI	MPEG	PNG	DNG	MP3	FFM	XLS	ODF	DOCML	HTML	CSV	XML	JSON	WARC	PDF/A-2				
Offenheit	Die Standardisierung des Formats ist vollständig definiert.	1	4	4	4	4	4	4	4	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4					
Lizenzfreiheit	Der Gebrauch des Formats ist durch keine Lizenzien eingeschränkt oder einschränkbar.	1	3	3	3	3	1	3	3	3	4	3	3	3	3	3	4	1	1	1	3	3	3	3	3					
Verbreitung	Das Format ist weit verbreitet und kann lange Migrationsszyklen erneut verarbeiten.	1	4	4	4	4	1	2	4	4	4	3	2	4	3	4	2	3	2	4	4	4	2	3	3					
Funktionalität	Das Format kann die spezifizierten Anforderungen für Archivierung erfüllen.	1	3	2	3	3	4	4	4	2	4	3	4	3	4	2	3	2	3	2	3	4	3	3	3					
Implementierung	Für das Format existieren mehrere Implementierungen.	0.5	4	4	4	4	3	4	3	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4					
Spezifität	Das Format ist mit einer sehr hohen Spezifität.	0.5	4	3	2	3	2	3	3	2	4	4	3	3	4	3	2	3	2	3	3	1	2	1	2					
Verifizierbarkeit	Es existieren fachliche und Werkzeuge zur Überprüfung und Verifizierung.	1	2	4	4	4	2	3	2	4	4	4	3	3	4	3	2	2	3	3	1	2	2	3	4					
Komplexität	Das Format ist möglichst wenig komplex.	0.5																												
Selbstdokumentation	Das Format kann Metadaten einschließen und ermöglicht deren Verwendung.	0.5																												
Best Practice	Das Format wird in Archiven bereits verwendet.	0.5	4	2	4	3	1	1	1	3	2	3	1	2	3	4	2	2	2	2	1	1	3	3	3					
Perspektive	Das Format ist im archivischen Sinn interessant und hat noch weitere Potenzial.	0.5	1	3	4	1	2	2	2	1	4	2	2	3	2	2	2	1	4	1	3	3	2	1	2					
Formatklasse	A: allgemein; B: Dokument; C: multimediales Format; D: audiovisueller Format; E: benutzerdefinierte	A	B	B	B	O	C	B	A	A	B	B	C	D	A	B	D	A	B	O	A	B	D	O	E					
		TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	DOCML	TIF	JPEG	AVI	MPEG	PNG	DNG	MP3	FFM	XLS	ODF	DOCML	HTML	CSV	XML	JSON	WARC	PDF/A-2					
Gesuchte Summe / #		0.73	0.97	0.84	0.89	0.57	0.75	0.77	0.84	0.84	0.89	0.75	0.71	0.94	0.84	0.62	0.62	0.62	0.69	0.63	0.57	0.60	0.69	0.71	0.80	0.63	0.57			
Logarithmisch gesuchte Summe		0.67	1.30	1.57	1.75	0.42	0.95	1.05	1.20	1.20	1.78	0.72	0.81	1.20	1.09	0.94	1.00	0.91	1.21	0.97	1.12	1.04	1.13	1.64	0.75	1.03	1.00	0.69	1.51	1.60

## xlsadg 2.0

Dans la newsletter de l'été dernier, nous avons déjà annoncé la publication de la version 2.0 de xlsadg, l'interface entre les entités *entrée* et *gestion des données*. Les Archives de la Ville de Zurich ainsi que les Archives de l'Etat de AR, BE, BS, LU, SG, TG et ZH qui ont élaboré cette version standard dans le cadre du projet du CECO 17-034 présentent maintenant en complément le mappage de eCH-0160 v1.1 vers xlsadg. Il sera publié chez eCH en tant qu'addendum à eCH-0160 et interconnecte ainsi les deux normes de manière compréhensible.

Vous trouvez sur le site du CECO à l'adresse [https://kost-ceco.ch/cms/index.php?x-sadg\\_de](https://kost-ceco.ch/cms/index.php?x-sadg_de) une prépublication du mappage ([eCH-0160.xlsadg&EAD\\_v2.0.xlsx](https://kost-ceco.ch/eCH-0160.xlsadg&EAD_v2.0.xlsx)) ainsi qu'une implémentation de référence (<https://kost-ceco.ch/isadg/app/input.php>).

## Echange Fedora

Différentes archives ont fait part au CECO de leur souhait d'en savoir plus sur *Fedora Repository*. *Fedora* est une solution logicielle que bon nombre d'archives utilisent déjà. Comme il s'agit d'une solution open source, les utilisateurs doivent fournir un investissement personnel plus important que si c'était une solution commerciale. C'est pour cette raison que nous vous invitons à un premier échange de vues sur le sujet qui aura lieu le 29 janvier 2018 à Berne.

L'objectif de cet échange est d'établir l'état actuel des connaissances, les besoins et comment continuer à traiter ce sujet, à savoir organiser une formation continue, un groupe d'échange d'expériences ou un atelier.

## **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 (Walcheturm 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.p. nous contacter à l'adresse [info@kost-ceco.ch](mailto:info@kost-ceco.ch).*