

## KOST-Newsletter Quartal 1, 2020

### In eigener Sache: COVID-19

*Die Mitarbeitenden der KOST-Geschäftsstelle arbeiten bis auf Weiteres im Home Office. Dringende Sitzungen werden als Telefonkonferenz oder Web-Meeting abgehalten, weniger dringende sowie auch alle Veranstaltungen werden verschoben. Die KOST-Mitarbeitenden sind telefonisch und per E-Mail wie gewohnt erreichbar. Sie wollen den Austausch mit den Trägern weiterhin so gut als möglich pflegen.*


*Die aussergewöhnliche Lage regt zu Innovation bei den Mitteln an: Die KOST hat als Ersatz für ausgefallene Referate ihr erstes Schulungsvideo produziert (siehe unten zum Thema «PDF mit mangelhaftem Font»), weitere können wenn gewünscht folgen. Wir danken Ihnen für Ihr Feedback dazu an [claire.roethlisberger@kost.admin.ch](mailto:claire.roethlisberger@kost.admin.ch).*


### PPEG-Report: PDF mit mangelhaftem Font


Die Preservation-Planning-Expertengruppe (PPEG) der KOST befasst sich seit längerem mit PDF/A-Dokumenten, welche nicht durchsuch- und extrahierbar sind. Solche Dokumente fallen auf, weil sichtbare Textpassagen nicht gefunden und nicht korrekt extrahiert werden können.

Die KOST konnte 2019 das neue Zusatzmodul zum «3-Heights™ PDF Validator» von PDF Tools AG testen und in KOST-Val integrieren. Ein erstes Assessment mit Testdokumenten aus der Geschäftsstelle und dem Bundesarchiv zeigte auf, dass das Problem durchaus signifikant ist. Gleichzeitig hat die KOST aber auch festgestellt, dass längst nicht alle fehlerhaften PDF/A-Dateien auch wirklich problematisch sind. Eine in KOST-Val integrierte Nachauswertung erlaubt es, die unkritischen Fälle zu ignorieren und auf die wirklich kritischen zu fokussieren. Diese Anpassungen sind ab der KOST-Val Version 1.9.1 enthalten. Es wird empfohlen, die aktuelle Version 1.9.8 von KOST-Val zu verwenden.

Für das abschliessende Assessment wurden neben den Beispieldateien der KOST-Geschäftsstelle (Sammlung aus unterschiedlichen Quellen und Archiven) auch archivierte PDF/A-Dateien von den Staatsarchiven SG, BS und UR sowie dem Schweizerischen Bundesarchiv verwendet. Dabei wurde festgestellt, dass alle PDF-Dateien mit eingebetteten Schriften diesen Mangel enthalten können, und dass der Fehler nicht auf einzelne Schriftarten oder Herstellungsjahre eingeschränkt werden kann. Ohne Korrekturmassnahmen müssen die Archive davon ausgehen, dass rund 1% aller PDF/A-Dateien der Konformität B (in erster Linie PDF/A-1b) bei einer Volltextsuche nicht gefunden würden.

 [KOST\\_PPEG\\_PDF-Font\\_Report\\_v1.pdf](#) , 889.7K, 23.03.20  
Report (März 2020)

 [KOST\\_PPEG\\_PDF-Font\\_Praesentationsfolien\\_v1.pptx](#) , 320.2K, 19.03.20  
Präsentation zum PDF-Font-Report (März 2020)

 [KOST\\_PPEG\\_PDF-Font\\_Praesentation\\_v1\\_pub.mp4](#) , 17.2M, 26.03.20  
Video der Präsentation zum Herunterladen (März 2020)

Für die PPEG ist es wichtig, nicht nur den Mangel und die Ursachen aufzuzeigen, sondern auch darzulegen, wie dieser Mangel vermieden und behoben werden kann.

[Alle Untersuchungen und Erkenntnisse der PPEG flossen in den PPEG-Report ein, welcher am 31.03.2020 publiziert wurde.](#) Als Ersatz für die wegen der Coronakrise abgesagten Präsentationen (am Arbeitskreis «Archivierung von Unterlagen aus digitalen Systemen» Ende März in Köln sowie an einer anschliessenden Schulungsveranstaltung in der Schweiz) stellt die KOST auf ihrer Website eine Video-Präsentation zum Herunterladen von Claire Röthlisberger-Jourdan zur Verfügung.

### **KOST-Val Version 1.9.8 veröffentlicht**

Die aktuelle Version 1.9.8 von KOST-Val wurde vor wenigen Tagen veröffentlicht. Sie erlaubt insbesondere die oben beschriebene Schriftvalidierung. KOST-Val kann von GitHub heruntergeladen werden unter <https://github.com/KOST-CECO/KOST-Val>.

### ***Descriptive Information nach ISAD(G) als Ontology (xlsadgR)***

Im letzten Newsletter haben wir das KOST-Projekt [17-034 xlsadg-Ontologie](#) vorgestellt. Nach der Abschlussveranstaltung vom 28.01.2020 liegt jetzt der von Claire Röthlisberger-Jourdan verfasste [Abschlussbericht](#) vor. Das Projekt ist damit abgeschlossen; als nächstes sollen praktische Erfahrungen mit der xlsadg-Ontologie gesammelt werden.

## Newsletter CECO du 1<sup>er</sup> trimestre 2020

### Information concernant le bureau du CECO : COVID-19

*Les collaboratrices et collaborateurs du bureau du CECO télétravaillent jusqu'à nouvel avis. Les séances urgentes se tiendront sous forme de conférences téléphoniques ou en ligne ; les moins urgentes ainsi que tous les événements sont reportés. Les membres du bureau du CECO sont comme d'habitude joignables par téléphone et courriel. Ils entendent poursuivre au mieux les échanges avec les membres du CECO.*


*Cette situation extraordinaire donne des idées d'innovation en matière de ressources : le CECO a produit son premier tutoriel vidéo en remplacement de conférences annulées (voir ci-dessous sur le thème « PDF contenant des polices de caractères défectueuses », d'autres suivront si souhaité). Nous vous remercions de nous faire parvenir votre avis à ce sujet à [claire.roethlisberger@kost.admin.ch](mailto:claire.roethlisberger@kost.admin.ch).*


### Rapport du groupe d'experts Preservation planning : PDF contenant des polices de caractères défectueuses


Le groupe d'experts Preservation planning du CECO s'occupe depuis longtemps des documents en PDF/A qui ne présentent pas de possibilité de recherche ou d'extraction. On repère ces documents au fait qu'on ne peut pas rechercher ni extraire correctement des passages de textes pourtant visibles.

En 2019, le CECO a pu tester le nouveau module additionnel au «3-Heights™ PDF Validator» de PDF Tools SA et l'intégrer à KOST-Val. Une première évaluation effectuée avec des documents tests du bureau et des Archives fédérales suisses a montré que le problème était vraiment considérable. En même temps, le CECO a cependant aussi relevé que ce ne sont de loin pas tous les fichiers PDF/A défectueux qui posent réellement problème. Une nouvelle évaluation intégrée dans KOST-Val permet d'ignorer les cas qui ne sont pas critiques pour se concentrer sur ceux qui le sont vraiment. KOST-Val à partir de la version 1.9.1 contient ces adaptations. Nous recommandons d'utiliser la version actuelle 1.9.8 de KOST-Val.

Pour l'évaluation finale, nous avons utilisé, en plus des échantillons de fichiers du bureau du CECO (collection en provenance de différentes sources et archives), également des fichiers PDF/A archivés par les Archives de l'État de SG, BS et UR ainsi que par les Archives fédérales. Nous avons constaté à cette occasion que tous les fichiers PDF avec polices intégrées pouvaient présenter ces défauts et que l'erreur ne se limitait pas à certains types de police ou certaines années de production. En l'absence de mesures pour remédier à ce problème, les Archives doivent s'attendre à ce qu'environ 1% de tous les fichiers PDF/A du niveau de conformité B (avant tout PDF/A-1b) ne peuvent faire l'objet d'une recherche en plein texte.

 [KOST\\_PPEG\\_PDF-Font\\_Report\\_v1.pdf](#) , 889.7K, 23.03.20  
Rapport (mars 2020, en cours de traduction)

 [KOST\\_PPEG\\_PDF-Font\\_Praesentationsfolien\\_v1.pptx](#) , 320.2K, 19.03.20  
Présentation du rapport sur les polices de caractères dans le PDF (mars 2020, en allemand uniquement)

 [KOST\\_PPEG\\_PDF-Font\\_Praesentation\\_v1\\_pub.mp4](#) , 17.2M, 26.03.20  
Vidéo de la présentation à télécharger (mars 2020, en allemand uniquement)

Le groupe d'experts estime important de ne pas seulement mettre en évidence le défaut et ses causes, mais également d'expliquer comment l'éviter et y remédier.

[Toutes les analyses et enseignements du groupe d'expert figurent dans son rapport qui a été publié le 31 mars 2020.](#) Le CECO met à disposition sur son site internet une présentation vidéo de Claire Röthlisberger-Jourdan que vous pouvez télécharger en lieu et place des présentations annulées en raison de la crise du coronavirus (lors de la réunion du groupe de travail «Archivierung von Unterlagen aus digitalen Systemen» de la fin mars à Cologne de même que lors d'une journée de formation ultérieure prévue en Suisse).

### **La version 1.9.8 de KOST-Val publiée**

La nouvelle version 1.9.8 de KOST-Val a été publiée il y a quelques jours. Elle permet en particulier d'effectuer la validation de polices de caractères décrite ci-dessus. Vous pouvez télécharger KOST-Val sur GitHub <https://github.com/KOST-CECO/KOST-Val>.

### **Information de description selon ISAD(G) sous forme d'ontologie (xlsadgR)**

Nous vous avons présenté le projet du CECO [17-034 ontologie xlsadg](#) dans la dernière newsletter. Après la présentation du projet lors de l'événement de clôture du 28 janvier 2020, nous disposons maintenant du [rapport final](#) rédigé par Claire Röthlisberger-Jourdan. Le projet est ainsi terminé ; la prochaine étape consistera à recueillir les expériences pratiques accumulées avec l'ontologie xlsadg.

## **Veranstaltungshinweise / Calendrier des événements**

Bis auf weiteres sind keine KOST-Veranstaltungen geplant. Gerne informieren wir Sie per Mail über die nächsten Anlässe der KOST, sobald die allgemeine Lage diese wieder zulässt.

*Le CECO ne prévoit aucune manifestation jusqu'à nouvel avis. Nous vous tiendrons au courant des prochains événements par courriel dès que la situation générale le permettra à nouveau.*