

KOST-Newsletter Quartal 2, 2016

KaD, Version 5.0

Der [Katalog archivischer Dateiformate KaD](#) ist in der Schweiz die Referenz für Informationen und Bewertungen von Formaten im Hinblick auf ihre Archivtauglichkeit. Seit seiner Publikation 2007 ist der KaD mehrfach aufdatiert und ergänzt worden. Die neue, Anfang Juli 2016 publizierte Version 5.0 umfasst 37 Dateiformate. Neu gibt der KaD fundiert Auskunft über die Kategorie Hypertext und ermöglicht einen umfassenderen Überblick über mögliche Videoformate.

Hypertext

Hypertext bezeichnet eine Menge von Texten, die über Verknüpfungen (Links) miteinander verbunden sind. Die Gestaltung der Texte wie auch die Verknüpfungen erfolgen mit einer Auszeichnungssprache (*markup language*). Die bekannteste Hypertext-Auszeichnungssprache ist HTML. Damit ist das *World Wide Web* der grösste Hypertext. Dieses gliedert sich wiederum in thematisch-inhaltliche Unterbereiche, die wir Website oder Internetauftritt nennen und die durch eine gemeinsame Domain ausgezeichnet sind, und weiter bis hin zur einzelnen Webseite oder Webpage, die genau genommen auch wieder ein Hypertext ist, weil sie aus einer HTML-Seite und einer Menge referenzierter Webressourcen besteht.

Hypertextformate

Im KaD 5.0 analysiert die KOST die sechs Formate HTML, HTML5, MHTML, ARC, WARC und PDF/A-2 und bewertet ihre Eignung für die Archivierung von Webinhalten. Die Empfehlung: Obwohl WARC das eingeführte Format der Webarchivierung ist, kann es als Langzeitformat nicht empfohlen werden. Die Vielzahl möglicher eingebetteter Formate und die Schwierigkeit, diese zu migrieren, sprechen dagegen. Für die Archivierung eines gesamten Webauftritts empfiehlt sich längerfristig eindeutig eine Umwandlung in PDF/A.

Studie zur Webarchivierung

Die Archivierung von Webinhalten ist freilich mehr als ein Formatproblem und darf nicht auf ein solches reduziert werden. Um die Formatdiskussion in den richtigen Kontext zu stellen, hat die KOST parallel zum neuen KaD-Kapitel auch eine Studie zur Webarchivierung erarbeitet. Diese untersucht die grundlegenden Konzepte wie Hypertext, Auszeichnungssprachen und Schlagworte wie Web 2.0. Sie setzt sich mit dem Objekt der Archivierung und dessen signifikanten Eigenschaften auseinander und fokussiert danach intensiv auf konkrete Methoden der Bewertung und Archivierung von Webinhalten. Die Studie ist auf der KOST-Website zugänglich.



Arbeitsgruppe Hypertext

Die neue KaD-Kategorie und die Studie zur Webarchivierung wurden von einer Arbeitsgruppe entwickelt, welcher neben der KOST-Geschäftsstelle die folgenden Kolleginnen und Kollegen angehörten: Natasa De Maio (StadtAZH), Emmanuel Ducry (AEG), Nicole Martini (BAR), Tobias Mérinat (BAR), René Quillet (StABL), Brigitte Sacker (StadtABE), Pascal Schneiter (OAEN). Die Preservation-Planning-Expertengruppe PPEG der KOST war für das Review verantwortlich. Wir danken allen Beteiligten für die gründliche Arbeit.

Video

Die regelmässigen Anfragen an die KOST-Geschäftsstelle zeigen, dass das Thema Video auf die Agenda der KOST-Träger drängt. Version 5.0 des KaD trägt dieser Entwicklung Rechnung, indem sie zwei weitere Videoformate detailliert beschreibt und bewertet (FFV1 und ProRes). Zusammen mit der Erweiterung aus Version 4.0 analysiert und vergleicht der KaD nun 7 Videoformate.

Im Bereich Video stösst ein Formatkatalog allerdings an die Grenzen dessen, was er leisten kann. Die Auswahl eines geeigneten Archivformats für Videomaterial kann nicht nur auf einer reinen Kriterienbewertung beruhen. Das Problem der Dateigrössen und deshalb des Speicherplatzbedarfs führt zur Bewertungsfrage, welche Komprimierung in Kauf genommen werden soll. Die Verluste beim Umcodieren von komprimiertem Material sollten zu einer zurückhaltenden Migrationspolitik führen. Solche und andere Fragen diskutiert das erneuerte Videokapitel im KaD. Ergänzt und illustriert wird es durch konkrete Szenarien, die in der neuen Rubrik FAQ auf der KOST-Website dargestellt werden.

Arbeitsgruppe Video

Für die Überarbeitung und Erweiterung der Informationen zu Videoformaten im KaD konnte die KOST-Geschäftsstelle auf die Hilfe einer breiten, interdisziplinären Arbeitsgruppe zählen: Bruna Casagrande (Tanzarchiv), Andrea Ehrat (Tanzarchiv), Peter Fornaro (Uni Basel), Rudolf Gschwind (Fachlabor Gubler AG), David Gubler (Fachlabor Gubler AG), Tabea Lurk (FHNW), Emilie Magnin (Tanzarchiv), Christoph Manasse (StABS), Katrin Oettli (Tanzarchiv), Romano Padeste (StAZH), David Pfluger (Memoriav), Moritz Tewes (Fachlabor Gubler AG). Wiederum übernahm die PPEG das Review der neuen Inhalte. Wir danken allen Beteiligten für ihre Zeit und die Zusammenarbeit.

Weitere Aktivitäten der KOST

eCH-0165, SIARD-Format, Version 2.0

Der Expertenausschuss von eCH hat am 1. Juni die Version 2.0 von eCH-0165, SIARD-Format, verabschiedet und zur Publikation freigegeben. Diese Version ist in enger Zusammenarbeit mit dem EU-Projekt E-ARK entstanden und damit mit mehreren europäischen Archivinstitutionen, welche SIARD ebenfalls nutzen. Neben kleinen Korrekturen und Vereinheitlichungen umfasst die neue Version zwei hauptsächliche Änderungen: die externe Speicherung von BLOBS und LOBS sowie die vollständige Realisierung der SQL:1999- bzw. SQL:2008-Kompatibilität durch die Übernahme von User Defined Types und des Datentyps ARRAY.

Der neue Standard ist auf der Website von eCH zugänglich: <http://www.ech.ch/vech-web/page?p=dossier&documentNumber=eCH-0165&documentVersion=2.0>. Die KOST-Tools für den Umgang mit SIARD (KOST-Val, csv2siard) werden gegenwärtig angepasst.

Newsletter CECO du 2e trimestre 2016

Cfa Version 5.0

Le [Catalogue des formats de données d'archivage Cfa](#) est la référence en Suisse pour les informations et évaluations de formats au regard de leur aptitude pour l'archivage. Depuis sa publication en 2007, le Cfa a été mis à jour plusieurs fois et complété. La nouvelle version 5.0 publiée en juillet 2016 comprend 37 formats de données. Le Cfa donne dorénavant des renseignements solides sur la catégorie de l'hypertexte et offre un aperçu plus détaillé des formats vidéo possibles.

Hypertexte

Le terme hypertexte désigne un ensemble de textes reliés entre eux par des liens (*links*). La présentation des textes et celle des liens s'effectuent avec un langage de balisage (*markup language*). Le langage de balisage hypertexte le plus connu est le HTML. Le Web constitue donc le plus grand hypertexte. Celui-ci se subdivise en sous-domaines thématiques appelés sites web ou sites internet et qui se caractérisent par un nom de domaine commun, jusqu'à en arriver à une page internet qui à proprement parler est également un hypertexte, parce qu'elle est constituée d'une page HTML et d'un ensemble de ressources web référencées.

Formats hypertextes

Dans le Cfa, le CECO analyse les six formats HTML, HTML5, MHTML, ARC, WARC et PDF/A-2 et évalue leur aptitude pour l'archivage de contenus web. Recommandation : bien que WARC soit le format instauré pour l'archivage web, il ne peut être recommandé pour l'archivage à long terme en raison de la diversité potentielle de formats intégrés et de la difficulté de les migrer. Pour archiver un site internet entier, il est clairement recommandé à long terme de convertir en PDF/A.

Étude sur l'archivage web

L'archivage de contenus web ne peut cependant pas se réduire uniquement à un problème de format. Afin de resituer la discussion sur les formats dans son contexte, le CECO a également élaboré une étude sur l'archivage web parallèlement au nouveau chapitre du Cfa. Cette étude examine des concepts élémentaires tels que l'hypertexte, les langages de balisage et des notions comme le web 2.0. Elle se penche sur l'objet de l'archivage et les propriétés significatives de celui-ci et se concentre ensuite de manière intensive sur des méthodes d'évaluation et d'archivage de contenus web. L'étude est disponible sur le site du CECO (en allemand).



Groupe de travail hypertexte

Un groupe de travail a élaboré la nouvelle catégorie du Cfa et l'étude sur l'archivage web. En plus de l'équipe du bureau du CECO, les membres du groupe sont : Natasa De Maio (StadtAZH), Emmanuel Ducry (AEG), Nicole Martini (AFS), Tobias Mérinat (AFS), René Quillet (StABL), Brigitte Sacker (StadtABE), Pascal Schneiter (OAEN). Le groupe d'experts Preservation Planning du CECO était responsable de la révision. Nous remercions tous les participants et toutes les participantes pour leur travail minutieux.

Vidéo

Les demandes régulières au bureau du CECO montrent que le thème de la vidéo se bouscule à l'agenda des membres du CECO. La version 5.0 du Cfa prend en compte cette évolution en décrivant et en évaluant deux autres formats vidéo (FFV1 et ProRes). En comptant l'enrichissement de la version 4.0, le Cfa analyse et compare maintenant sept formats vidéo.

Dans le domaine de la vidéo, un catalogue de formats se heurte toutefois aux limites de ce qu'il peut offrir. Le choix d'un format d'archivage adéquat pour du matériel vidéo ne peut reposer uniquement sur une simple évaluation de critères. Le problème des tailles de fichiers et donc du besoin d'espace de mémoire engage à évaluer quelle compression on est prêt à risquer. Les pertes engendrées lorsqu'on modifie le codage de matériel compressé doivent inciter à une politique de migration mesurée. Le chapitre vidéo remanié du Cfa discute de ce genre de questions et d'autres encore. Il est complété et illustré par des scénarios concrets qui sont présentés dans la rubrique FAQ sur le site du CECO.

Groupe de travail vidéo

Pour remanier et enrichir les informations sur les formats vidéo du Cfa, le CECO a pu compter sur l'aide d'un vaste groupe de travail interdisciplinaire composé de : Bruna Casagrande (Collection suisse de la danse), Andrea Ehrat (Collection suisse de la danse), Peter Fornaro (Uni Basel), Rudolf Gschwind (Fachlabor Gubler AG), David Gubler (Fachlabor Gubler AG), Tabea Lurk (FHNW), Emilie Magnin (Collection suisse de la danse), Christoph Manasse (StABS), Katrin Oettli (Collection suisse de la danse), Romano Padeste (StAZH), David Pfluger (Memoriav), Moritz Tewes (Fachlabor Gubler AG). Le groupe d'experts Preservation Planning a une fois encore assuré la révision des nouveaux contenus. Nous remercions tous les participants et toutes les participantes pour le temps consacré et leur collaboration.

Avancement des travaux dans les autres projets du CECO

eCH-0165, format SIARD, version 2.0

Le comité d'experts d'eCH a adopté le 1^{er} juin la version 2.0 d'eCH-0165 format SIARD et l'a publiée. Cette version a été élaborée en étroite collaboration avec le projet E-ARK de l'UE et donc avec plusieurs institutions d'archives européennes qui utilisent également SIARD. Outre quelques petites corrections et uniformisations, la nouvelle version comporte principalement deux changements : les BLOBS et les LOBS peuvent être enregistrés en externe et la compatibilité SQL:1999 respectivement SQL:2008 est intégralement concrétisée par la prise en charge de User Defined Types et du type de données ARRAY.

La nouvelle norme est accessible sur le site d'eCH : <http://www.ech.ch/vech-web/page?p=dossier&documentNumber=eCH-0165&documentVersion=2.0>. Le CECO adapte actuellement ses outils permettant de travailler avec SIARD (KOST-Val, csv2siard).

Potentielle KOST-Projekte / Projets potentiels du CECO

Die Liste der potentiellen KOST-Projekte listet Projektvorschläge auf, die für eine Betreuung als KOST-Projekt in Frage kommen, und für die weitere Partner gesucht werden.
La liste des projets potentiels du CECO répertorie les propositions de projets des archives membres qui entrent en ligne de compte pour devenir un projet CECO et pour lesquels des partenaires additionnels sont recherchés.

Kompass3 StAZH	Archivierung der Daten der Berufsbildungsämter <i>Archivage des données des services de la formation professionnelle</i>
JURIS offen/à compléter	Archivierung von Unterlagen der Rechtspflege <i>Archivage de documents de justice</i>

Archive, die an der Mitarbeit an einem dieser Projekte interessiert sind, werden gebeten, dies der Geschäftsstelle mitzuteilen (info@kost-ceco.ch).
Les archives souhaitant collaborer à l'un de ces projets sont priées de le communiquer au Bureau (info@kost-ceco.ch).

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.

06.09.16 KOST-Workshop „Kosten der digitalen Archivierung“
Atelier du CECO « Coûts de l'archivage numérique »
Bern, Schweizerisches Bundesarchiv
Berne, Archives fédérales suisses

12.12.16 KOST.Forum 06, „KOST-Projekte 2016“ / « Projets du CECO en 2016 »
Zürich, Staatsarchiv
Zurich, Archives de l'Etat

Wenn Sie einen Veranstaltungshinweis im KOST-Newsletter publizieren wollen, kontaktieren Sie uns bitte unter 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.