

KOST-Newsletter Quartal 3, 2019

Standardisierte Ablagestruktur im Repository

Das digitale oder digitalisierte Archivgut liegt in der Regel in Form von Dateien in einem Dateisystem (wenn es nicht in einem Cloud-basierten *Object Store* gespeichert ist). Das ist grundsätzlich eine gute Sache, denn das POSIX¹-spezifizierte Dateisystem gehört zu den bis anhin langlebigsten Standards der Informatik und ist seit mehr als vierzig Jahren stabil. Die Organisation der Dateien im Dateisystem (Verzeichnisstruktur/Dateinamen, d.i. Pfad, PID, UID, GUID etc.) hingegen wird durch das *Digital Repository Layer* bestimmt. *Digital Repositories* gibt es kaum länger als 15 Jahre, und die ersten Lösungen sind schon wieder vom Markt verschwunden. Migrationen in diesem Bereich sind aber besonders aufwändig und heikel, weil hier Terabytes, wenn nicht gar Petabytes bewegt oder umbenannt werden müssen.

Inzwischen liegt ein Vorschlag vor, die Speicherung im Dateisystem derart standardisiert vorzunehmen, dass mit einer *Repository*-Migration oder sonstigen Weiterentwicklung der Archivsoftware keine Änderungen an der Ablagestruktur notwendig sind. Der Vorschlag heisst *Oxford Common File Layout* (OCFL, <https://ocfl.io/>) und spezifiziert nur die auf der Ebene Dateiablage minimal notwendigen Themen: *Naming Convention*, *Fixity*, *Versioning*, *Forward Delta*

- *Naming Convention*: *Pairstree*-Konvention
- *Fixity*: SHA512-Checksumme und Legacy-Hashwerte
- *Versioning*: Jede Veränderung wird in einer Version festgehalten
- *Forward Delta*: Nur veränderte Dateien werden gespeichert

OCFL behandelt gemeinhin Objekte, das sind Primär- und Metadateien, die in einem Ordner zusammengefasst sind und ein *Archival Object* bilden. Ein Objekt kann ein Buch, ein Dokument oder z.B. ein Bild sein. Eine OCFL *Inventory*-Datei im JSON-Format (die damit auch menschenlesbar ist) hält *Naming*, *Fixity* und *Versioning* in diesem Objekt fest.



Weil sich OCFL in keiner Art um den Inhalt der Metadaten oder das Format der Primärdaten kümmert, muss bei zukünftigen Erweiterungen und Migrationen bereits gespeicherte Information nicht mehr migriert werden, sondern das Objekt kann fortlaufend ergänzt werden. Neue *Repository*-Applikationen müssen nur den OCFL-Standard verstehen und erfüllen – eine Datenmigration ist nicht mehr nötig.

¹ Portable Operating System Interface: ISO/IEC/IEEE 9945, siehe auch https://de.wikipedia.org/wiki/Portable_Operating_System_Interface

Die Initiative zur Erarbeitung des Oxford Common File Layout wurde 2017 ergriffen von Gedächtnisinstitutionen (vornehmlich Bibliotheken aus dem angelsächsischen Raum mit grossen Datensammlungen) und Stiftungen, die sich intensiv mit dem Thema Repository-Software beschäftigen: Bodleian Library, Oxford; Hesburgh Libraries, University of Notre Dame; Emory Libraries, Emory University; Stanford University Libraries; Cornell University Library; Fedora Commons; DuraSpace und Lyris. Im Moment liegt OCFL in einer Beta-Version vor. Die KOST-Geschäftsstelle beobachtet dieses Projekt und wird ihre Träger bei der Umsetzung unterstützen, falls sich der Standard tatsächlich durchsetzt.

KOST-Projekt 14-017 TAXAR

Die Staatsarchive von Zürich, Uri, Schwyz, Bern und Basel-Stadt wollen im [TAXAR-Projekt](#) eine schweizweit einsetzbare, produktunabhängige Empfehlung für die Archivierung der kantonalen Steuerakten (Steuerdossiers, Registerdaten, Reports, Statistiken) erarbeiten. Der Entwurf dazu befindet sich aktuell in der öffentlichen Vernehmlassung bei eCH unter <http://www.ech.ch/standards/52953>. Die Projektgruppe hofft, die Empfehlung im Februar 2020 als eCH-Best Practice veröffentlichen zu können. Parallel dazu wird in verschiedenen der beteiligten Kantone an der Implementierung gearbeitet.

In eigener Sache: Neues KOST-Mitglied

Mit der Überarbeitung der Verwaltungsvereinbarung hat sich die KOST über den Kreis der staatlichen Träger hinaus für weitere Institutionen geöffnet, die ein öffentlich zugängliches Archiv betreiben. Wir freuen uns, als erste solche Institution ein Universitätsarchiv im Kreis der KOST begrüßen zu dürfen, und zwar das [UZH Archiv](#).

Newsletter CECO du 3^e trimestre 2019

Normalisation de la structure de classement dans un dépôt numérique (*repository*)

Le document d'archives numérique ou numérisé se présente en général sous la forme de fichiers rangés dans un système de fichiers (quand il n'est pas stocké dans un dépôt ou référentiel d'objets reposant sur une solution en nuage). C'est en principe une bonne chose étant donné que le système de fichiers spécifié selon la norme POSIX² fait partie des normes informatiques les plus durables jusqu'à ce jour et qu'il est stable depuis plus de quarante ans. L'organisation des fichiers dans le système de fichiers (structure des répertoires/noms des fichiers, c'est-à-dire chemin, PID, UID, GUID, etc.) est en revanche déterminée par la couche dépôt numérique (*digital repository layer*). Il existe des dépôts numériques depuis à peine 15 ans et les premières solutions ont déjà disparu du marché. Les migrations dans ce domaine sont particulièrement coûteuses et délicates parce que les fichiers à déplacer ou à renommer se mesurent en téraoctets voire en pétaoctets.

Entre-temps est apparue une solution permettant le stockage dans le système de fichiers sous une forme standardisée ne nécessitant plus de modifier la structure de classement en cas de migration de dépôt ou d'autre évolution des logiciels d'archivage. Cette initiative a pour nom *Oxford Common File Layout* (OCFL, <https://ocfl.io/>) et elle ne spécifie que le minimum de sujets nécessaires au niveau du classement des fichiers : *convention de nommage, immuabilité, gestion des versions, sauvegarde delta*

- Convention de nommage (*naming convention*) : convention *Pairtree*
- Immuabilité (*fixity*) : somme de contrôle SHA512 et valeurs de hachage existantes
- Gestion des versions (*versioning*) : chaque modification est consignée dans une version
- Sauvegarde delta (*forward delta*) : seuls les fichiers modifiés sont stockés

OCFL traite communément des objets, des fichiers primaires et des métafichiers, qui sont rassemblés dans un classeur et constituent un *objet archivistique*. Cet objet peut être un livre, un document ou par exemple une photo. Un fichier d'inventaire OFCL en format JSON (le rendant lisible pour l'humain) consigne les données de nommage, d'immuabilité et de gestion des versions dans cet objet.



Comme OCFL ne s'occupe absolument pas du contenu des métadonnées ou du format des données primaires, l'information déjà stockée ne doit plus faire l'objet d'une migration lors d'enrichissements et migrations futures, mais l'objet peut être complété en continu. Les nouvelles applications de dépôt doivent uniquement comprendre et respecter la norme OCFL, il n'y a plus besoin de migrer les données.

² Portable Operating System Interface: ISO/IEC/IEEE 9945, voir aussi <https://fr.wikipedia.org/wiki/POSIX>

Des institutions patrimoniales (parmi celles-ci surtout des bibliothèques anglo-saxonnes détentrices d'importantes collections de données) et des fondations qui s'occupent activement du thème des logiciels de dépôt ont pris l'initiative d'élaborer le Oxford Common File Layout en 2017 : Bodleian Library, Oxford; Hesburgh Libraries, University of Notre Dame; Emory Libraries, Emory University; Stanford University Libraries; Cornell University Library; Fedora Commons; DuraSpace et Lyris. OCFL est actuellement disponible en version beta. Le bureau suit le développement de ce projet et apportera son soutien aux membres du CECO lors de la mise en œuvre si la norme devait effectivement s'imposer.

Projet du CECO 14-017 TAXAR

Avec le projet [TAXAR](#), Les Archives de l'État de Zurich, d'Uri, de Schwyz, de Berne et de Bâle-Ville entendent élaborer une recommandation utilisable dans toute la Suisse indépendamment d'un produit pour archiver les dossiers fiscaux cantonaux (dossiers fiscaux, registre fiscal, rapports, statistiques). Le projet fait actuellement l'objet d'une consultation publique pilotée par eCH disponible sous <http://www.ech.ch/fr/standards/52953>. Le groupe de projet prévoit de publier la recommandation en février 2020 sous la forme de bonnes pratiques eCH. Il travaille en outre en parallèle à une implémentation dans différents cantons participants.

Développement du CECO : nouveau membre

Le remaniement de l'accord administratif a ouvert les portes du CECO, jusque-là réservé à des membres strictement étatiques, à un cercle plus large d'institutions pour autant qu'elles exploitent des archives accessibles au public. Nous sommes heureux d'accueillir les Archives de l'Université de Zurich ([UZH Archiv](#)) qui est la première représentante de ces institutions à rejoindre le CECO.