

KOST-Newsletter Quartal 2, 2013

Kolloquium „Wer archiviert das Web, und wie?“

Die Archivierung von Webinhalten hatte für die meisten Trägerarchive der KOST bislang geringe Priorität. Neben Bedenken über die technische Komplexität der Aufgabe wurde vielfach eine gewisse Unklarheit über den Umfang der Aufgabe als Grund für die Zurückhaltung genannt. Ziel des diesjährigen KOST-Kolloquiums war es, hier im Austausch zwischen Bibliotheken und Archiven aus dem In- und Ausland mehr Klarheit zu schaffen.

In einem ersten Teil zeigten verschiedene Institutionen auf, unter welchem Blickwinkel sie das Thema in Angriff nehmen. Barbara Signori (Schweizerische Nationalbibliothek NB) berichtete einleitend über das Webarchiv Schweiz. Dieser führenden Initiative zur Webarchivierung in der Schweiz stellten Kerstin Brunner (StABS), Felix Stadler (StASG), Martin Akeret (StASH) und Krystyna Ohnesorge (BAR) in vier Kurzstatements den Stand der Überlegungen und Aktivitäten in einzelnen KOST-Trägerarchiven gegenüber. Ein Bericht aus dem Landesarchiv Baden-Württemberg (LA-BW, Johannes Renz) ergänzte diese im Umfang noch limitierten Vorhaben durch die Präsentation eines funktionellen Webarchivs in einem grossen staatlichen Archiv. Im zweiten Teil stellte Sara Aubry von der Bibliothèque nationale de France (BnF), einer der weltweiten Pionierinstitutionen, die im Rahmen des IIPC entwickelten Tools zum Web-Harvesting vor. Andere mögliche technische Vorgehensweisen präsentierte abschliessend Simon Scheurer (Qumram AG, Zürich).

Aus den vielen Erkenntnissen aus Präsentationen und Diskussion greifen wir drei heraus:

1. Die Archivierung von Webinhalten umgibt vielfach noch die Aura des Speziellen, vom archivischen Tagesgeschäft Losgelösten. Es muss jedoch unbedingt Ziel der Archive sein, archivwürdige Webinhalte in den gleichen Prozessen und mindestens teilweise mit den gleichen Werkzeugen zu behandeln wie alle digitalen Archivalien. Beispiele dafür sind die Bemühungen des BAR, im Rahmen seiner Machbarkeitsstudie zusammen mit dem EDA die relevanten Websites in den etablierten BAR-Prozessen zu archivieren, oder die Verortung der Webressourcen in der Archivtektonik und ihre Erschliessung im AIS durch das LA-BW.
2. Die Bibliotheken gehen bei der Webarchivierung einem Sammlungsauftrag nach, der je nach gesetzlichem Umfeld grösser oder kleiner sein kann: In Frankreich autorisiert das *dépôt légal* die BnF zur Archivierung der gesamten .fr-Domain sowie weiterer Websites aus Frankreich; die NB archiviert Websites aus der Schweiz gemäss einer kuratierten Auswahl. Während einzelne Archive wie das LA-BW teilweise ebenfalls Sammlungen anlegen, ist das eigentlich Archivische bei der Webarchivierung schwieriger zu fassen. Gemäss Punkt 1 oben sind auf Websites des Internets und Intranets der Verwaltung die gleichen Bewertungskriterien anzuwenden wie für das übrige Verwaltungsschriftgut. Was dies konkret bedeutet, und wo die Grenze zu dokumentarischer Sammlung genau liegt, ist bislang erst im Einzelfall entschieden worden.
3. Der vom IIPC propagierte und von vielen Gedächtnisinstitutionen angewendete Ansatz ist ein Harvesting der Webressourcen und eine Archivierung der Dateien in ihren Originalformaten. Für das offene und standardisierte Format HTML ergeben sich damit mittelfristig kaum Probleme, wohl aber für eingebettete Dateien, aktive und multimediale Inhalte. Ein Funktionalitätsverzicht durch Archivierung im PDF-Format ist deshalb unter Umständen angezeigt. Web-2.0-Ressourcen und interaktive Websites legen, wie von Qumram beschrieben, keine inhalts-, sondern eine transaktionsbasierte Archivierung nahe.

[Die Präsentationen des Kolloquiums sind auf der KOST-Website verfügbar.](#)

Weitere Aktivitäten der KOST

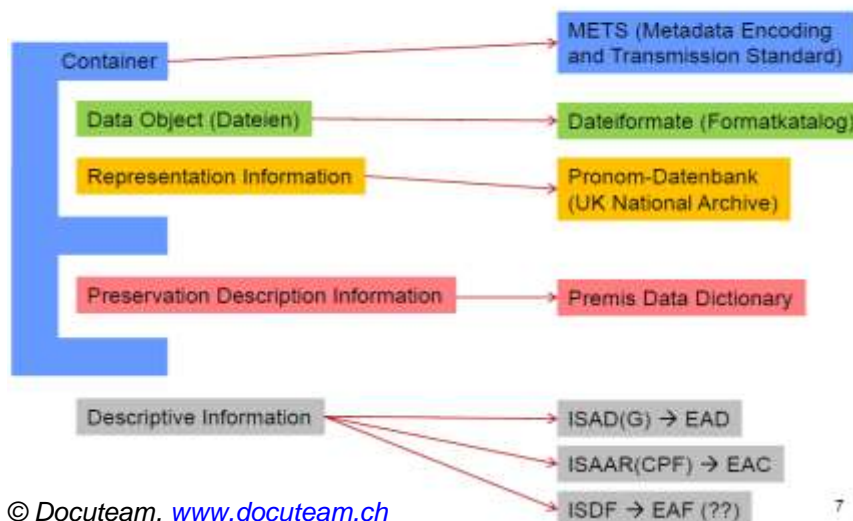
KOST.Forum no4: Das Matterhorn METS Profil

Der *Metadata Encoding and Transfer Standards* METS, ein sehr offen definiertes Containerformat für die digitale Archivierung, wird mittels sogenannter Profile für konkrete Anwendungen konfiguriert. Das Staatsarchiv Wallis hat zusammen mit Docuteam zur Verwendung als AIP das Matterhorn METS Profil entwickelt, das sich durch die Einbettung von EAD für die Verzeichnungsinformationen sowie von PREMIS für die Archivierungs-Metadaten auszeichnet. Am 18.04. präsentierten Alain Dubois (AEVS) und Tobias Wildi (Docuteam) diese Entwicklung in einem KOST.Forum und stellten sie in den Kontext der digitalen Archivierung beim Staatsarchiv Wallis.

[Die Veranstaltung ist auf der KOST-Website dokumentiert.](#)



Das Matterhorn METS Profil: Abbildung in archivische Metadatenstandards



TIFF

Das *Tagged Image File Format* TIFF, ein Urgestein der digitalen Archivierung und in den meisten Archiven seit Jahren vorhanden, hat einen gewichtigen Nachteil: seine Komplexität. TIFF erlaubt eine Vielzahl von Kompressionen, Farbräumen und weiteren Charakteristiken; die Kombination dieser Merkmale stellt die Anzeigeprogramme vor Probleme. Bestimmte TIFF-Dateien können nur von wenigen oder gar keinen Viewern korrekt dargestellt werden. In einem Workshop am 17.06. diskutierten die KOST-Träger dieses Thema und trugen erste Elemente einer Bestandsanalyse zusammen. Eine KOST-Studie fasst die Diskussion zusammen und empfiehlt die Formulierung von Anforderungen an ein archivtaugliches TIFF. Auf dieser Basis stellt die KOST einen Validator für TIFF-Dateien zur Verfügung, der auf JHOVE beruht und die Implementierung dieser Anforderungen erlaubt.

AudioVault

Das Liechtensteinische Landesarchiv, die Staatsarchive von Bern, Uri und Jura sowie die Stadtarchive von Bern und St. Gallen arbeiten im Projekt AudioVault zusammen an Fragen der Archivierung von Tondokumenten. Angestrebt werden eine zusammenfassende Übersicht über *Best Practice* in der Audioarchivierung sowie die Erarbeitung neuer Erkenntnisse zu Fragen, bei denen noch Klärungsbedarf besteht, beispielsweise bei der Feinerschließung und dem Zugang zu Audioarchivalien. Das Projekt läuft bis im Sommer 2014 und wird danach seine Resultate der KOST-Gemeinschaft präsentieren.

Newsletter CECO du 2e trimestre 2013

Colloque «Qui archive le web et comment?».

Jusqu'ici, l'archivage de contenus web n'était pas la grande priorité de la plupart des archives membres. Outre les réserves émises sur la complexité technique de la tâche, la retenue était souvent motivée par un certain flou autour de son ampleur. Le but du Colloque du CECO de cette année était d'apporter des éclaircissements sur le sujet au travers d'échanges entre bibliothèques et archives suisses et étrangères.

Dans une première partie, différentes institutions ont montré sous quel angle elles appréhendaient la problématique. En introduction, Barbara Signori (Bibliothèque nationale suisse BN) a présenté Archives Web Suisse. Kerstin Brunner (AEBS), Felix Stadler (AESG), Martin Akeret (AESH) et Krystyna Ohnesorge (AFS) ont mis en perspective, au travers de quatre brefs témoignages, l'état des réflexions et des activités dans certaines archives membres du CECO avec cette initiative majeure d'archivage du Web en Suisse. Un rapport des Archives du land de Bade-Wurtemberg (LA-BW, Johannes Renz) est venu compléter ces approches, limitées dans leur ampleur, par la présentation d'archives web déjà en service dans de grandes archives d'État. Durant la deuxième partie, Sara Aubry de la Bibliothèque nationale de France (BnF), une des institutions pionnières au niveau mondial, a présenté les outils de moissonnage du Web (*web harvesting*) développés dans le cadre de l'IIPC. Pour terminer, Simon Scheurer (Qumram AG, Zurich) a présenté d'autres procédés techniques possibles.

Voici trois des nombreux enseignements tirés des présentations et de la discussion :

1. L'archivage de contenus web est encore souvent paré d'une aura extraordinaire, comme si cette tâche ne relevait pas des affaires courantes des archives. Pourtant, celles-ci se doivent d'avoir pour objectif de traiter les contenus web ayant une valeur archivistique dans les mêmes processus et, du moins partiellement, avec les mêmes outils que toutes les archives numériques. Mentionnons, à ce titre, les efforts des AFS, dans le cadre de leur étude de faisabilité en collaboration avec le DFAE, pour archiver les sites web importants dans les processus établis par les AFS ainsi que la localisation des ressources web dans la structure hiérarchique des archives et leur description dans l' AIS par les LA-BW.
2. Lors de l'archivage web, les bibliothèques suivent un mandat de collection d'envergure plus ou moins importante selon le contexte législatif. En France, le dépôt légal autorise la BnF à archiver la totalité du domaine .fr ainsi que d'autres sites de France. La BN archive des sites de Suisse en fonction d'une sélection basée sur des critères définis. Alors que certaines archives comme les LA-BW constituent parfois aussi des collections, la nature archivistique effective de l'archivage web est plus difficilement saisissable. En vertu du point 1 ci-dessus, il faut utiliser les mêmes critères d'évaluation pour les sites internet et intranet que pour les documents ordinaires de l'Administration. Jusqu'ici, il n'a été décidé qu'au cas par cas ce que cela signifie concrètement et où se situe exactement la limite de la collection documentaire.
3. L'approche diffusée par l'IIPC et utilisée par de nombreuses institutions patrimoniales est le moissonnage de ressources web et l'archivage des fichiers dans leurs formats d'origine. Si, à moyen terme, le format ouvert et standardisé HTML ne pose que très peu de problèmes, il en existe toutefois pour les fichiers insérés, les contenus actifs et multimédias. C'est pourquoi, dans certaines circonstances, il est indiqué de renoncer à certaines fonctionnalités par un archivage en format PDF. Pour les ressources du Web 2.0 et les sites interactifs, il convient de procéder à un archivage basé non pas sur les contenus, mais sur les transactions, comme le décrit l'entreprise Qumram.

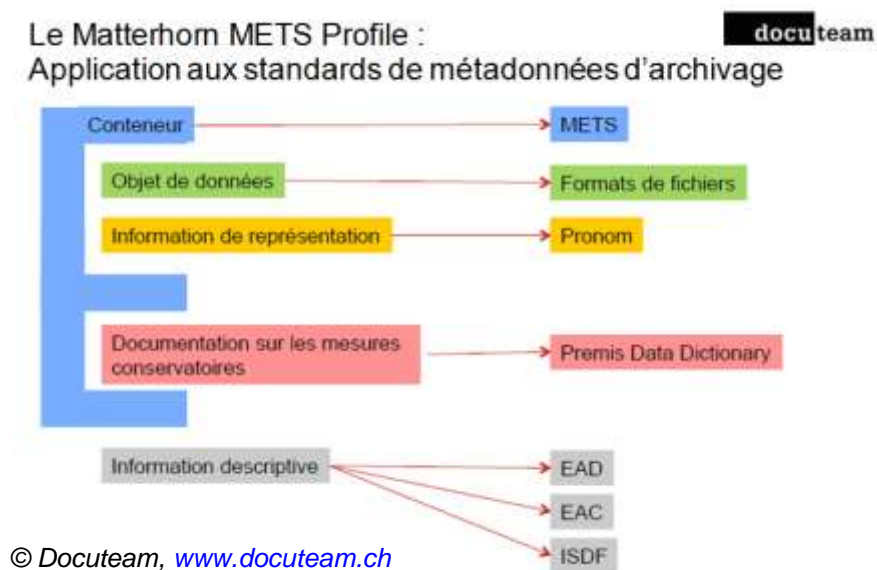
[Vous trouvez les présentations du Colloque sur le site du CECO.](#)

Autres activités du CECO

CECO.Forum no 4: Le Matterhorn METS Profile

Le *Metadata Encoding and Transfer Standards* METS, un format conteneur très ouvert pour l'archivage numérique, est configuré pour les applications concrètes au moyen dudit profil. Les Archives d'État du Valais ont développé le Matterhorn METS Profile en collaboration avec Docuteam pour l'utiliser comme AIP. Il se caractérise par l'intégration d'EAD pour les informations descriptives et de PREMIS pour les métadonnées d'archivage. Le 18 avril, Alain Dubois (AEVS) et Tobias Wildi (Docuteam) ont présenté ce développement durant un CECO.Forum en le plaçant dans le contexte de l'archivage numérique aux AEVS.

[Documentation sur cette manifestation disponible sur le site du CECO](#)



TIFF

Le *Tagged Image File Format* TIFF, un ancien format d'archivage numérique éprouvé et disponible depuis des années dans la plupart des archives a un inconvénient de poids: sa complexité. TIFF permet de nombreuses compressions, une large palette de couleurs et d'autres caractéristiques. Or la combinaison de ces caractéristiques pose problème aux programmes d'affichage. Certains fichiers TIFF ne peuvent être affichés correctement que par un petit nombre d'outils de visualisation, voire par aucun. Les membres du CECO ont discuté de ce sujet lors d'un atelier le 17 juin et ont rassemblé les premiers éléments en vue d'une analyse de l'état des lieux. Une étude du CECO résume la discussion et recommande de formuler des exigences pour un format TIFF adapté pour l'archivage. Sur cette base, le CECO met à disposition un validateur pour fichiers TIFF qui repose sur JHOVE et permet l'implémentation de ces exigences.

AudioVault

Les Archives nationales du Liechtenstein, les Archives d'État de Berne, Uri et du Jura ainsi que les Archives municipales de Berne et St-Gall collaborent pour les questions d'archivage de documents sonores dans le cadre du projet AudioVault. Le projet vise à obtenir un récapitulatif des bonnes pratiques dans l'archivage audio ainsi qu'à acquérir de nouvelles connaissances sur des questions nécessitant encore des éclaircissements, notamment en matière de catalogage fin et d'accès aux archives audio. Le projet se déroulera jusqu'en été 2014 et ses résultats seront ensuite présentés à la communauté du CECO.