

## **Service interministériel des Archives de France**

### **Bulletin sur les ressources archivistiques numériques, n° 43, septembre 2011**

**Compte rendu du 75<sup>e</sup> congrès annuel de la Société des archivistes américains  
(Chicago, 22-27 août 2011)**

Le 75<sup>e</sup> congrès annuel de la Société des archivistes américains (SAA) s'est tenu à Chicago, siège de l'association, les 22-27 août 2011. Selon une formule bien rodée, cette manifestation, qui rassemble chaque année plus de 2 000 participants, a vu alterner réunions de groupes et de commissions spécialisés et séances plénières présentant des projets et réalisations en cours. Cette année, les débats ont été dominés par les évolutions attendues de l'EAD dans le contexte du Web sémantique. On trouvera ci-après le compte rendu des réunions des différentes instances de la SAA puis de quelques sessions plénières.

## **Sommaire**

### **Comités et groupes de travail spécialisés**

#### **Comité des normes**

#### **Sous-comité technique sur l'EAD et Table ronde des utilisateurs de l'EAD**

Site officiel, site de la Table ronde EAD et liste de discussion

Enquête sur les usages de l'EAD

Calendrier, processus de révision, financement

Synthèse des propositions d'évolutions de l'EAD

Premières orientations du TS-EAD et de l'équipe technique de développement

#### **Sous-comité technique sur l'EAC-CPF**

#### **Section sur la description**

#### **Table ronde sur les métadonnées et les objets numériques**

#### **Table ronde de l'OCLC**

#### **Table ronde des utilisateurs des logiciels Archivists'Toolkit et Archon et du projet ArchiveSpace**

### **Sessions plénières**

**Session 304 – Les archives dans le Web de données : vers un flux vertueux de données au Royaume-Uni et au-delà**

**Session 610 – Approches pratiques pour les archives nativement numériques : qu'arrivera-t-il demain ?**

**Session 710 – Le projet « Réseaux sociaux et contexte archivistique (SNAC) : l'EAC-CPF au travail »**

## Comités et groupes de travail spécialisés

### Comité des normes

#### => Sommaire

Entièrement réorganisé en 2010, le Comité des normes de la SAA comprend à présent six sous-comités techniques :

- sous-comité sur les bâtiments d'archives ;
- sous-comité sur la réévaluation et l'aliénation des collections ;
- sous-comité sur la norme DACS (« Décrire les archives : une norme de contenu ») ;
- sous-comité sur l'EAD (TS-EAD) ;
- sous-comité sur l'EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles) (TS-EAC-CPF) ;
- équipe de développement des schémas de la SAA (TS-SDT).

Un portail sur les normes de la SAA a été développé sous Drupal, un système de gestion de contenu libre et open source. Cet outil est consultable à l'adresse suivante : <<http://www2.archivists.org/standards>> Chaque norme fait l'objet d'une notice descriptive avec les informations suivantes : intitulé, brève description (sur la page d'accueil du portail), description détaillée, date d'approbation / de validation par le Conseil de la SAA, liens vers le texte de la norme, groupe(s) chargé(s) de la maintenance intellectuelle de la norme, groupe(s) sponsor(s), partenaire, thème traité (conservation préventive, description, évaluation et sélection, etc.), typologie(norme, guide de bonnes pratiques, etc.), statut (approuvé, validé, en cours de développement ou de révision, etc.). Des liens seront établis vers les mini-sites de certaines instances de la SAA, et inversement (TS-EAD, Section des archives universitaires et de l'enseignement supérieur, etc.). Le portail comprendra également des références, voire des liens, vers les ressources en relation (autres normes, articles, études de cas, etc.), l'objectif étant de fournir un corpus de connaissances sur telle ou telle norme.

Parmi les réalisations ou les travaux en cours du Comité, il convient de noter la publication d'un guide destiné aux archivistes comme aux architectes et relatif à la construction et à la rénovation des bâtiments d'archives. Ce document comprend des règles sur l'évaluation du site, la construction, les systèmes environnementaux, la protection incendie, la sécurité, l'éclairage, les matériaux, les équipements et les différents espaces fonctionnels. Pour en savoir plus, voir : <http://www2.archivists.org/standards/archival-and-special-collections-facilities-guidelines-for-archivists-librarians-architect>

En ce qui concerne la description archivistique, la SAA a lancé le processus de révision de la norme DACS en 2010. L'achèvement des travaux est prévu pour août 2013. Un appel à commentaires a été lancé début 2011 et un forum de discussion était organisé pendant le congrès de Chicago. Les utilisateurs de la norme demandent notamment la mise en ligne gratuite de la norme (diffusée actuellement sous version imprimée payante), l'ajout d'une section correspondant à l'En-tête <eadheader> de l'EAD et comprenant les métadonnées de l'instrument de recherche (auteur et service responsable de la publication, date de création, règles utilisées, révisions effectuées), l'insertion d'un guide sur l'application de DACS aux archives électroniques, l'établissement de règles pour inclure des dates dans les intitulés, une clarification des objectifs de l'élément « Mode de classement » ou encore le caractère obligatoire de l'élément « Histoire administrative/biographie » à tous les niveaux. La deuxième partie sera complètement réécrite pour être mise en conformité avec l'EAC-CPF. La troisième partie, relative à la création de points d'accès normalisés, est dans les faits très peu utilisée ; elle fait référence aux Règles de catalogage anglo-américaines (AACR2) et devrait donc être actualisée pour prendre en compte le code de catalogage remplaçant les AACR2 (« Ressources : Description et Accès » ou RDA).

Voir ci-après pour les comptes rendus des sous-comités techniques sur l'EAD et l'EAC-CPF.

### **Sous-comité technique sur l'EAD et Table ronde des utilisateurs de l'EAD**

#### => Sommaire

Tandis que le sous-comité technique sur l'EAD (TS-EAD) et l'équipe de développement (TS-SDT) sont responsables de la maintenance intellectuelle et technique de l'EAD, la Table ronde EAD de la SAA est plutôt un lieu d'échange et de partage d'expériences entre les différents utilisateurs du format. La Table ronde EAD de la SAA dispose de son propre site (<http://www.archivists.org/saagroups/ead/>), alors que la documentation officielle (Tag Library) et les fichiers de la DTD sont disponibles sur un site hébergé par la

Bibliothèque du Congrès (<http://www.loc.gov/ead/index.html>). Les débats ont été marqués cette année par un retour sur les usages de l'EAD et par les propositions d'évolution du format.

Site officiel, site de la Table ronde EAD et liste de discussion

### => Sommaire

Une enquête a été menée auprès des utilisateurs du site de la Table ronde de l'EAD de l'EAD pour connaître leurs attentes. Les demandes portent sur la mise à jour des listes d'expériences de mise en œuvre de l'EAD, l'ajout de liens pointant sur des tutoriels ou des présentations et la clarification des relations entre le site officiel hébergé par la Bibliothèque du Congrès et celui de la Table ronde EAD.

La liste officielle de discussion de l'EAD (<http://listserv.loc.gov/listarch/ead.html>) est toujours aussi active, puisqu'au 19 août 2011, elle comprenait quelques 1 215 membres représentant 31 pays : les États-Unis (952), le Royaume-Uni (61), le Canada (30), l'Australie (24), les Pays-Bas (21), l'Allemagne (17), la France (15), l'Italie (13), l'Espagne (11) et la Suisse (6) sont les plus représentés. De juillet 2010 à juillet 2011, 11 523 messages ont été postés sur la liste officielle EAD portant sur 173 sujets.

Enquête sur les usages de l'EAD

### => Sommaire

Une enquête sur les pratiques de mise en œuvre de l'EAD a été réalisée par Kathy Wisser (Simmons College, Boston) et Jackie Dean (université de Caroline du Nord, Chapel Hill) à partir de l'envoi à la SAA, fin 2010, parallèlement à l'appel à commentaires, de quelques 1 136 instruments de recherche allant de 30 Ko à 12 033 Ko provenant de 108 institutions (dont une quinzaine d'instruments de recherche provenant de bibliothèques et de services d'archives français). L'objectif de l'enquête était d'identifier les pratiques d'encodage sans évaluer la qualité de l'encodage ou le contenu des instruments de recherche. Il s'agissait simplement de noter la présence ou l'absence d'éléments et d'attributs et la façon dont ces éléments ou attributs sont utilisés dans le cadre d'une instance EAD. Les résultats n'ont pour le moment été communiqués qu'oralement lors de la dernière réunion annuelle de la Table ronde, le 24 août 2011, à Chicago.

D'une manière générale, la balise la plus utilisée est Identification et description <did>, tandis que les moins utilisées sont Abbréviation <abbr>, Groupe de descriptions archivistiques <archdescgrp>, Composant (niveaux 10-12) <c10> - <c12>, Subdivision de texte <div>, Groupe d'instances EAD <eadgrp>, Premier en-tête <head01>, Second en-tête <head02> et Pointeur <ptr>.

Au sein de l'en-tête EAD <eadheader>, outre les éléments obligatoires, l'élément le plus utilisé est Description du profil <profiledesc> (98,1 %), alors que Description des révisions <revisiondesc> ne se retrouve que dans 32,7 % des instruments de recherche. Au sein de <filedesc>, après les éléments obligatoires, viennent Mention de note <notestmt> (9 %), mention d'édition <editionstmt> (4 %) et enfin Mention de série <seriestmt> (0,1 %). Si l'en-tête EAD a été conçue d'abord comme la transposition en XML de la page de titre des instruments de recherche, il est intéressant de noter l'importance des métadonnées liées au contrôle de la description dans la Description du profil <profiledesc>, où les éléments Création <creation> (94,7 %) et Langue utilisée <langusage> (92,2 %) devançant Règles de description <descrules> (42,8%).

Dans Préliminaires <frontmatter>, si l'élément Page de titre <titlepage> est utilisé à une large majorité (92,8 %), Subdivision de texte <div> l'est beaucoup moins (2,2 %). Il est à noter que 5 % des instruments de recherche comportent un élément <frontmatter> vide. Les résultats de l'enquête ne donnent par contre aucune indication sur le pourcentage d'utilisation de <frontmatter> dans l'échantillon d'instances EAD.

Les valeurs attribuées à l'attribut @level de Description archivistique <archdesc> pour qualifier le niveau de l'ensemble documentaire décrit sont, dans l'ordre décroissant : "collection" (90,9 %), "fonds" (4,8%), "recordgrp" (1,4 %), "subgrp" (1 %), "series" (0,6 %), "file" (0,4 %), "subfonds" et "class" (0,3 %), "item" et "otherlevel" (0,1 %). Aucun service n'a utilisé "subseries" à ce niveau.

Les éléments les plus utilisés dans l'élément Identification et description <did> de Description archivistique <archdesc> sont Intitulé de l'unité documentaire <unittitle> (100%), Organisme responsable de l'accès intellectuel <repository> (99,6 %), Description physique <physdesc> (97,2 %), Date de l'unité documentaire

<unitdate> (97 %), Identifiant de l'unité documentaire <unitid> (90 %), Origine <origination> (89 %), Langue des unités documentaires <langmaterial> (89 %) et Résumé <abstract> (86,6 %). Les éléments Unité de conditionnement <container> (0,4 %) et Particularités de certains types de documents <materialspect> (1,6%) sont les moins utilisés à ce niveau de description.

Les autres éléments les plus employés de l'élément <archdesc> sont Présentation du contenu <scopecontent> (93,4 %), Biographie ou histoire <bioghist> (87,3 %), Conditions d'accès <accessrestrict> (86,2 %), Mention conseillée <prefercite> (85 %), Vedettes et accès contrôlés <controlaccess> (85 %), Informations sur les modalités d'entrée <acqinfo> et Conditions d'utilisation <userrestrict> (68 %), Classement <arrangement> (65 %), Historique de la conservation <custodhist> (14%), tandis qu'on ne retrouve Autres données descriptives <odd> que dans 9,7% des instruments de recherche, Index <index> dans 1,2 % et Plan de classement <fileplan> dans 0,6 %.

90 % des instruments de recherche ont un seul élément Description des sous-composants <dsc>, 7,3 % n'en ont pas, et, plus surprenant, 2,4 % ont de multiples <dsc>. Les valeurs les plus fréquentes de l'attribut @type de l'élément <dsc> sont "combined" (66,5 %) et "in-depth" (16,7 %), mais cet attribut n'est pas renseigné pour 8 % des instances fournies.

Seulement 8 % des instruments de recherche utilisent des composants non numérotés. L'utilisation des composants numérotés est donc majoritaire, avec une prépondérance des composants de niveau 2 (<c02>), ce qui est contradictoire avec certains commentaires demandant la suppression des commentaires numérotés (cf. infra).

Les valeurs de l'attribut @level les plus représentées aux niveaux inférieurs de la description sont "series" (77,7 %), "file" (56,9 %), "subseries" (35,3 %), "item" (24,2 %) et "otherlevel" (9,1 %).

Les attributs de l'élément <unitdate> les plus utilisés sont @calendar, @era et @type et la valeur "inclusive" pour les intervalles de dates (la valeur "bulk" pour les dates prépondérantes au sein d'un dossier est beaucoup moins utilisée). On peut toutefois s'interroger sur l'intérêt de la valeur "inclusive" de l'attribut @type avec la suppression de la valeur "single" dans l'EAD 2002. Le modèle de données choisi pour exprimer les datations dans le schéma EAC-CPF est plus satisfaisant et devrait être également implémenté dans le futur schéma EAD.

Enfin, les équivalences dans d'autres formats via l'attribut @relatedencoding sont surtout établies avec les formats MARC (66,5 %) et Dublin Core (32,1 %), les références à la norme ISAD(G) ne représentant que 1,2 %.

Calendrier, processus de révision, financement

## => **Sommaire**

Le TS-EAD a succédé au groupe de travail international EAD en février 2010. Les deux co-présidents sont Michael Rush (université de Yale) et William Stocking (British Library). Le sous-comité comprend 8 membres et 12 membres ex officio (équipe de développement, présidents du Comité des normes et du TS-EAC-CPF, un représentant d'OCLC Recherche). Outre les États-Unis, les pays représentés sont l'Allemagne, le Canada, la France, l'Italie, les Pays-Bas et la Suède. La liste des membres du TS-EAD est disponible sur le site de la SAA ([http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/committees/Roles.html?Action=Show\\_Comm\\_Roles&CommCode=SAA\\*\\*CS-TSEAD&Name=Officers&Status=Active&Time=-712703095](http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/committees/Roles.html?Action=Show_Comm_Roles&CommCode=SAA**CS-TSEAD&Name=Officers&Status=Active&Time=-712703095)).

Le TS-EAD s'est réuni pour la première fois dans sa nouvelle configuration en août 2010. Il s'agissait de définir un calendrier pour la révision et les objectifs initiaux. Un appel à commentaires a été lancé d'octobre 2010 à février 2011. 127 commentaires ont été reçus des États-Unis, du Royaume-Uni, de France, d'Allemagne, d'Espagne, de Suède, reflétant ainsi la nature internationale de l'EAD.

Les commentaires seront examinés et discutés plus en détail par le TS-EAD au cours de deux réunions téléphoniques d'ici février 2012.

Une réunion du TS-EAD est prévue au printemps 2012 à l'université de Yale, grâce à un financement conjoint des Archives nationales des Pays-Bas, de la Fondation Gladys Krieble-Delmas et de la Dotation nationale pour les humanités (National Endowment for Humanities ou NEH), pour l'établissement d'une

feuille de route pour l'équipe de développement du futur schéma.

Une version provisoire du schéma devrait être mise à disposition de la communauté internationale fin 2012, suivie d'un appel à commentaires et d'une réunion consacrée à la documentation du schéma (refonte de la Tag Library). La version définitive du schéma devrait être publiée pour l'été 2013.

### Synthèse des propositions d'évolutions de l'EAD

#### => **Sommaire**

Les nombreux commentaires (127) sont à la fois archivistiques et techniques, d'ordre général ou au contraire très spécifiques.

La plupart des commentaires demandent à ce que l'EAD évolue vers un schéma, qui soit la version de référence : les éléments devront être rendus plus explicites et plus conformes aux règles actuelles d'écriture des schémas XML, les bibliothèques des modèles XML (<http://www.w3.org/XML/1998/namespace>) et Xlink (<http://www.w3.org/TR/xlink/>) devront être utilisées, la Tag Library devra être réactualisée et des scripts de conversion d'instances conformes à la DTD EAD 2002 vers la nouvelle version devront être développés.

L'EAD devrait :

- être alignée avec l'EAC-CPF en ce qui concerne le modèle de relations sous-jacents et les noms et la sémantique des éléments (en particulier Biographie ou Histoire <bioghist>), mais une utilisation de l'EAD seule devra rester possible ;
- inclure les données provenant d'autres espaces de noms (EAC-CPF, MARC21, MARC XML, MODS, TEI, KML, HTML) ;
- permettre des descriptions riches d'objets numériques au sein de <dao> ou y faire référence, y compris les métadonnées techniques et de structure ;
- intégrer les données générées par les utilisateurs ou y faire référence ;
- être plus simple et avoir moins de balises ;
- être plus orientée données : elle devra avoir moins de balises de mise en forme (<head>, <emph> et <table>) et avoir moins d'éléments de contenu mixte (pouvant contenir à la fois du texte et des balises comme par exemple <physdesc>).

L'en-tête EAD <eadheader> devra être plus simple, ne plus servir simplement à la publication d'une page de titre mais être aligné sur l'élément Contrôle de la description <control> du schéma EAC-CPF.

Certains suggèrent de supprimer les composants numérotés et la balise Description des sous-composants <dsc> et de faire de Description archivistique <archdesc> une balise englobante contenant des composants non numérotés <c> à tous les niveaux (les informations de haut niveau seraient elles-mêmes englobées dans un <c>).

Le rôle et les objectifs des éléments Description et identification <did> et Description physique <physdesc> sont à revoir, de même que Groupe d'éléments de description <descgrp> (un commentaire demande la suppression de cet élément).

Dans Nom de personnes <persname>, il faudrait donner la possibilité de baliser séparément les différentes parties de noms dans différents éléments <part> en y incluant les noms de lieux et les sujets. Il faudrait aussi permettre le géo-référencement des lieux dans <geogname>.

Les éléments pouvant être qualifiés par des codes de langues/d'écriture devront être plus rendus nombreux.

Enfin, les fourchettes chronologiques et les datations incertaines et complexes devront pouvoir être encodées et normalisées.

### Premières orientations du TS-EAD et de l'équipe technique de développement

#### => **Sommaire**

Les grands principes de révision définis par le TS-EAD à Chicago sont les suivants :

- développement d'un processus de migration des instruments de recherche encodés en EAD 2002 vers le nouveau format ;
- interopérabilité (mieux prendre en compte les sémantiques des relations et simplifier les

- mécanismes de liens) ;
- échange (simplifier l'EAD, éviter les ambiguïtés et favoriser une utilisation plus prévisible du format).

Des outils devraient être développés pour convertir des instances produites conformément à la DTD EAD 2002 ou au schéma EAD 2002 en instances conformes avec le nouveau schéma, comme cela a été fait pour la migration des instances de la DTD EAD 1.0 vers la DTD EAD 2002 et de la DTD EAD 2002 vers le schéma EAD 2002.

Le schéma devrait être rendu disponible sous une double syntaxe, W3C et RelaxNG, avec les caractéristiques des schémas XML : typage des données, documentation de leur sens et de leur utilisation, gestion des espaces de noms. Une réflexion sera lancée afin de réduire le nombre des éléments de contenu mixte.

L'EAD devra être simplifiée, avec notamment l'abandon des composants numérotés au profit des composants non numérotés.

Les mécanismes d'encodage des dates seront améliorés sur le modèle de ce qui s'est fait dans le schéma EAC-CPF.

L'EAD devra bien s'articuler avec l'EAC-CPF. Il faudrait notamment permettre l'intégration d'instances EAC-CPF ou de fragments d'instances EAC-CPF dans des instances EAD comme il est désormais possible d'intégrer des instances EAD ou des fragments d'instances EAD dans des instances EAC-CPF par le biais des éléments <objectXMLWrap>.

Enfin, il est prévu de développer des profils, c'est-à-dire des sous-ensembles de l'EAD correspondant à certains besoins. Mais pour le moment, rien n'a été décidé sur leur statut.

### ***Sous-comité technique sur l'EAC-CPF***

#### ***=> Sommaire***

Le TS-EAC est composé de 20 membres, dont 12 membres de la communauté des archivistes américains et 8 membres représentant la communauté internationale (France, Suisse, Royaume Uni, Canada, Suède, Grèce, Italie, Australie).

Depuis la publication du schéma EAC-CPF et de son dictionnaire de balises en mars 2010, de nombreuses initiatives ont vu le jour. De nombreuses communications, publications et actions de formation ont eu lieu et le format a déjà été implémenté dans de nombreux projets (pour la France, prise en compte de l'EAC-CPF dans le futur système d'information archivistique, implémentation aux Archives nationales d'outre-mer pour améliorer l'accès aux ressources décrites dans l'application IREL, projet de fichier national d'autorités pour les services extérieurs de l'État). Enfin, dans une perspective d'interopérabilité entre des ressources archivistiques et des ressources bibliographiques, des correspondances ont été établies entre l'EAC-CPF et le modèle FRAD, ainsi qu'entre l'EAC-CPF et le format MARC21 autorités.

La portée internationale du standard a suscité des initiatives de traduction de son dictionnaire de balises. La traduction en français, par le groupe d'experts AFNOR, est achevée et est prête à être publiée. Des traductions en italien et en allemand sont en cours. D'autres traductions vont voir le jour prochainement, notamment en grec, en espagnol et en portugais.

En France le processus de traduction de la documentation a conduit à une relecture approfondie portant non seulement sur les aspects éditoriaux, sur la cohérence intellectuelle et sur la clarté du texte lui-même, mais aussi sur les aspects techniques relatifs au schéma. Le groupe AFNOR des traducteurs a ainsi rédigé au fur et à mesure de la traduction une série de commentaires et les a ensuite adressés au TS-EAC et à l'équipe de développement de schémas de la SAA. Les commentaires français serviront de base pour une révision du dictionnaire des balises en 2011-2012. Si un premier travail de révision (corrections éditoriales, mise en cohérence du vocabulaire, ajout d'exemples, etc.) pourra être facilement réalisé, d'autres actions nécessiteront une collaboration avec l'équipe de développement de schémas de la SAA : révision des concepts de base et de la terminologie utilisée (par exemple, dans le dictionnaire de balises, les deux termes « entité » et « identité » sont utilisés de manière pratiquement interchangeable), de tout ce qui a trait aux identifiants et aux « objets » qu'ils sont censés identifier...



Par ailleurs, le dictionnaire des balises de l'EAC-CPF doit être considéré comme une documentation dynamique, vivante, multilingue et intégrable dans les outils de production, de manière à favoriser l'implémentation du schéma. Pour atteindre cet objectif, il faudra choisir un modèle d'encodage permettant une maintenance simple, efficace et précise (segmenter le dictionnaire des balises en autant de fichiers que d'éléments et d'attributs plutôt que de gérer le tout dans un seul fichier ?) ; établir l'infrastructure technique (système de contrôle des versions) ; établir le flux et les méthodes de travail coordonnées pour faciliter les mises à jour ; enfin, décider du mode de gestion des versions linguistiques (centralisé ou distribué, chaque communauté linguistique gérant sa propre traduction isolément ?).

Les choix adoptés pour la gestion de la documentation technique de l'EAC-CPF devront également s'appliquer à la documentation du futur schéma EAD. Une réflexion d'ensemble devra être menée sur les concepts exprimés par les balises et sur le modèle de gestion de la documentation considérée comme un vocabulaire contrôlé. Dans cette perspective seront exploités les nouveaux modèles de gestion de vocabulaires contrôlés (clarification et précision des concepts, identification de chaque concept par un URI, choix de l'expression formelle en SKOS, en RDF ou en OWL, implémentation dans des schemata, complémentarité avec d'autres vocabulaires, etc.).

Le TS EAC-CPF a enfin convenu de prendre contact avec les initiatives qui se développent autour de l'EAC-CPF. On peut mentionner notamment :

- le travail de Silvia Mazzini de la société regesta.exe, en collaboration avec l'Institut des biens culturels d'Émilie-Romagne, pour une représentation ontologique de l'EAC-CPF (<http://templates.xdams.net/RelationBrowser/genericEAC/RelationBrowserCIA.html>) ;
- la « carte topique »<sup>1</sup> établissant des équivalences entre le modèle FRAD (« Fonctionnalités requises des données d'autorité »), la norme ISAAR(CPF) et le format EAC-CPF (<http://culturalheritage.wordpress.com/2010/09/23/frad-isaarcpf-eac-cpf-topic-maps-mapping/>) ;
- les équivalences entre MARC21 et l'EAC-CPF effectuées dans le cadre des projets annexes d'Europeana par l'Espagne (voir le compte rendu de ce travail à : <http://conference.ifla.org/sites/default/files/files/papers/ifla77/80-agenjo-en.pdf>).

## **Section sur la description**

### **=> Sommaire**

Chaque année, la section sur la description de la SAA présente dans une « exposition » quelques expériences de production et de mise en ligne de descriptions archivistiques. C'est ainsi qu'était présenté EADitor, outil open source utilisé par la Société américaine de numismatique pour produire des instruments de recherche EAD à partir de formulaires Web (<http://code.google.com/p/eaditor/> et <http://eaditor.blogspot.com/>).

L'université Brigham Young a quant à elle développé une nouvelle interface de consultation d'instruments de recherche, permettant de présenter aux chercheurs les informations d'une manière plus intuitive sans perdre le contexte archivistique (<http://findingaid.lib.byu.edu/>). Les informations de gestion sont beaucoup plus réduites et ne sont plus accessibles que depuis l'un des trois onglets apparaissant à l'affichage. La présentation hiérarchisée des niveaux de description est simplifiée, avec un volet plus discret de navigation contextuelle. Le système est également conçu pour faciliter la réutilisation des métadonnées des instruments de recherche. Les composantes des instruments de recherche peuvent être moissonnées en utilisant le protocole OAI-PMH avec la possibilité de filtrer les résultats aux différents niveaux de description. Le système fournit actuellement du Dublin Core en sortie, mais il est prévu d'ajouter une fonctionnalité pour supporter l'EAD natif.

L'exposition a vu également la présentation de l'interface de consultation des collections de la Bibliothèque présidentielle John F. Kennedy, qu'elles soient numérisées ou non, certains instruments de recherche détaillés comportant des liens vers les objets numériques (<http://www.jfklibrary.org/Research/Search-the->

---

<sup>1</sup> Une carte topique (en anglais « Topic Map ») est un outil très général de représentation des connaissances, dont le but est d'agrégier autour d'un point unique d'indexation (appelé topic) toutes les informations disponibles concernant un sujet donné, et de relier ces points par un réseau sémantique de relations appelées associations.

[Digital-Archives.aspx](#)). **Erika Boudreau**, archiviste numérique à la Bibliothèque Kennedy, a montré comment la numérisation de masse des collections au sein de son institution avait entraîné la production d'instruments de recherche EAD.

Bien que la Bibliothèque Kennedy ait une politique de production normalisée d'instruments de recherche antérieure à la refonte de sa plate-forme d'accès aux fonds numérisés (application de la norme DACS, établissement en interne de règles de bonnes pratiques, mise en ligne de la plupart des instruments de recherche aux formats HTML ou PDF, participation au catalogue des Archives nationales américaines ARC : <http://arcweb.archives.gov/arc/action/BasicSearchForm?jScript=true>), la numérisation à grande échelle a modifié les pratiques et nécessité l'établissement de règles :

- mise en œuvre de la description à plusieurs niveaux ;
- établissement de règles de nommage des fichiers qui reflètent leur arborescence physique ;
- utilisation de normes et élaboration de dictionnaires de données pour rendre les métadonnées plus cohérentes.

Le développement d'une nouvelle interface de consultation en 2009-2010 a conduit l'institution à approfondir sa réflexion sur les métadonnées des objets numériques. Toutes les collections ont été dotées de descriptions de haut niveau et des équivalences ont été établies entre les champs du logiciel documentaire Documentum et les éléments EAD 2002 correspondants pour permettre les exports des descriptions archivistiques conformes à la DTD. Les notices descriptives de haut niveau comportent ainsi des éléments d'identification de la collection, une présentation de son contenu, des informations sur son producteur, ses modalités d'entrée, ses conditions d'accès et de reproduction, son plan de classement et des termes d'indexation. Les notices au niveau inférieur contiennent quant à elles des informations sur le classement interne du dossier, une description de son contenu et des liens vers les objets numériques au moyen de l'élément `Objet archivistique numérique <dao>`.

### **Table ronde sur les métadonnées et les objets numériques**

#### **=> Sommaire**

La Table ronde sur les métadonnées et les objets numériques tend à favoriser les échanges, la formation et la collaboration entre les archivistes intéressés par les objets archivistiques numériques et les métadonnées qui en permettent l'accès, la gestion et la conservation. La réunion a vu la présentation des stratégies des universités de Stanford (Californie), de Princeton (New Jersey), du Minnesota ou encore de l'Alabama pour une numérisation de masse à faible coût et une capture rapide des métadonnées.

C'est ainsi que l'archiviste numérique des bibliothèques de l'université de Stanford, **Peter Chan**, a décrit l'utilisation du logiciel Forensic par son institution pour la génération de métadonnées techniques et l'attribution de métadonnées descriptives et administratives aux collections numériques natives (<http://lib.stanford.edu/digital-forensics>). Les outils Jhove et DROID (Digital Record Object Identification)<sup>2</sup> ne peuvent identifier que des formats de fichiers très limités. Est-il possible d'affecter aux objets numériques des métadonnées descriptives sans en connaître le contenu ? Comment ouvrir les nombreux fichiers créés par le passé avec des logiciels obsolètes comme WordPerfect, Lotus 1-2-3, etc. ? L'université de Stanford avait besoin d'un outil intégré pour identifier les formats de fichiers, visualiser les fichiers de format obsolète et enrichir les métadonnées de fichiers individuels ou de groupes de fichiers. Le choix s'est donc porté sur le logiciel Forensic Toolkit, reconnu comme une référence en matière d'investigation informatique (déchiffrement de fichiers, analyse de mails, recherche par mots clés, émission de rapports...). Les métadonnées techniques comprennent l'empreinte, le format, la taille, la date de création, la date de dernière modification et la date de dernière consultation du fichier. Les métadonnées descriptives comprennent le contexte archivistique (série, sous-série, etc.), les vedettes-matières et les médias sources. Les métadonnées administratives concernent principalement les restrictions d'accès.

**Jody DeRidder** a ensuite présenté Acumen, une solution logicielle libre développée par les Bibliothèques de l'université d'Alabama pour mettre à faible coût du contenu numérique en ligne (<http://acumen.lib.ua.edu/project/>). Basé sur PHP, MySQL et Apache, Acumen est un outil modulaire et facile à configurer. Le logiciel fonctionne avec les principaux navigateurs et ne nécessite pas de compétences supplémentaires pour des archivistes. Il est composé de trois modules pour l'indexation des fichiers, la génération automatique de métadonnées MODS (Schéma de Description Objet de Métadonnées) selon des profils définis par l'utilisateur et la consultation des documents numériques à partir des instruments de

---

<sup>2</sup> Outils libres en Java permettant d'identifier, de valider et de caractériser le format d'un objet numérique en vue de son archivage.



recherche EAD les décrivant.

**Erik Moore**, archiviste-adjoint à l'université du Minnesota, département des Sciences de la santé, a quant à lui décrit un modèle de capture rapide des métadonnées dans le cadre de la numérisation de masse de contenus institutionnels. À ce jour, plus de 500 000 pages ont été numérisées et mises en ligne. Le projet a évolué pour devenir une des principales composantes du programme de travail des Archives de l'université du Minnesota.

On notera enfin avec intérêt la mise en place, par la SAA, d'un cursus de formation pour les archivistes numériques (<http://www2.archivists.org/prof-education/das>). Élaboré par des experts de l'archivage numérique, ce parcours diplômant est structuré en différents niveaux :

- cours de base se concentrant sur les compétences essentielles que les archivistes doivent acquérir pour gérer des archives numériques ;
- cours destinés aux archivistes ayant à encadrer des équipes sur les stratégies à mettre en place ;
- cours pratiques mettant l'accent sur les outils et services spécifiques à utiliser pour la capture et la préservation d'archives numériques ;
- cours relatifs aux compétences que les archivistes doivent acquérir pour faire évoluer leur vie professionnelle et transformer leurs institutions en services d'archives numériques.

### **Table ronde de l'OCLC**

#### **=> Sommaire**

Fondé en 1967, l'OCLC (Online Computer Library Center) est une organisation à but non lucratif mondiale au service des bibliothèques, dont le but est d'offrir un meilleur accès public aux informations et d'en réduire le coût. Plus de 72 000 bibliothèques dans le monde utilisent les services de l'OCLC afin de trouver, de cataloguer ou de conserver leurs ouvrages. Cette mise en réseau planétaire a notamment conduit à la création d'un catalogue mondial de recherche documentaire (WorldCat), qui rassemble à ce jour plusieurs centaines de millions de références (livres, articles, CD, DVD), dans plus de 450 langues. Un autre catalogue en ligne, ArchiveGrid (<http://archivegrid.org>), permet de consulter plus d'un million de descriptions de fonds d'archives détenues par des milliers d'établissements, bibliothèques, musées, sociétés d'histoire ou services d'archives, à travers le monde. Les descriptions proviennent des notices bibliographiques de WorldCat et d'instruments de recherche moissonnés à partir des sites Web des contributeurs. Le moteur de recherche et l'interface d'ArchiveGrid permettent d'afficher des résultats comprenant le titre de la collection, le nom de l'institution détentrice, une brève description ainsi qu'un lien vers une description plus détaillée. Pour le moment, l'accès à ArchiveGrid est payant, mais l'OCLC est en train de développer une nouvelle interface qui devrait être librement consultable (voir le prototype à : <http://experimental.worldcat.org/archivegrid/>).

L'OCLC conduit également différents programmes de recherche. L'un d'eux, intitulé « Mobiliser les matériaux uniques », tend à favoriser une plus grande collaboration entre les bibliothèques, les archives et les musées pour partager les données, les services et les compétences. L'objectif est de réaliser des économies d'échelle et d'être plus efficace dans la description, la mise à disposition et la découverte par les chercheurs des documents conservés par les différentes institutions patrimoniales. Le programme comprend plusieurs volets :

- « Capture rapide : numérisation de masse des collections spéciales » :

L'OCLC a mené une enquête auprès des équipes impliquées dans neuf projets de numérisation de masse de documents de différents formats. Le rapport de recherche (disponible à <http://www.oclc.org/research/activities/capture/default.htm>) se concentre sur la capture au sein de la chaîne de numérisation (la préparation des documents à numériser, la création de métadonnées ou la préparation des documents numérisés pour l'accès ne sont pas abordées dans le rapport) ;

- « Recherche unique des collections des bibliothèques, des archives ou des musées » :

De nombreuses institutions souhaitent avoir une interface unique et simple d'accès à toutes leurs collections, mais les collections sont souvent décrites et gérées très différemment. Neuf institutions partagent leur expérience et donnent des recommandations pour faciliter la mise en œuvre d'interfaces permettant d'interroger de manière transverse les descriptions des collections des bibliothèques, des archives ou des musées (<http://www.oclc.org/research/activities/lamsearch/default.htm>) ;

- « Numérisation et diffusion : conduire des projets de numérisation initiés par les utilisateurs dans les collections spéciales et les services d'archives :

De plus en plus visibles sur Internet, les collections spéciales et les archives des universités américaines intéressent de plus en plus le public et sont de plus en plus utilisées par les chercheurs. La découverte de ces collections s'accompagne d'une forte augmentation des demandes de copies numériques. Ce rapport propose des stratégies aux institutions ayant à délivrer aux usagers des copies numériques de documents (<http://www.oclc.org/research/publications/library/2011/2011-05.pdf>). Si les bibliothèques de prêt considèrent la numérisation à la demande comme de la routine, les demandes de reproduction de documents rares et uniques prennent souvent beaucoup plus de temps dans d'autres institutions. Ces processus pourraient être allégés pour s'adapter à la fois aux circonstances de la demande et aux ressources institutionnelles. La numérisation devrait être envisagée comme le prêt entre bibliothèques, c'est-à-dire comme un service qui est fourni aux utilisateurs sur demande (souvent sans aucun coût pour l'utilisateur) et pour lequel un budget et des infrastructures sont prévus.

- « Les collections spéciales nativement numériques » :

Ce projet de recherche se concentre sur l'amélioration de la gestion des documents nativement numériques dans les bibliothèques de recherche. Il est issu d'une enquête de l'OCLC sur la gestion des archives numériques dans les collections spéciales et les services d'archives des universités, qui reste embryonnaire (<http://www.oclc.org/research/publications/library/2010/2010-11.pdf>). Un tiers des institutions ayant répondu à l'enquête ont créé un service spécialisé dans la gestion de leurs collections numériques et seulement un tiers savent combien de documents numériques ont été collectés. Le manque d'expertise reste l'obstacle le plus commun pour la préservation des documents numériques. Le programme de l'OCLC vise donc à identifier les compétences des bibliothécaires et les archivistes des collections spéciales, indispensables pour une gestion efficace des différents types de documents nativement numériques ; à décrire comment ces domaines d'expertise se rapportent ou non aux différents types de documents numériques ; enfin, à définir les étapes élémentaires qu'une bibliothèque peut suivre pour commencer à exercer un contrôle minimal sur les documents numériques.

Par ailleurs, pendant le congrès de Chicago, l'OCLC organisait une réunion sur les « consortiums » ou projets collaboratifs EAD, tels qu'APEnet, le portail européen des archives (<<http://www.apenet.eu/>> et <<http://www.archivesportaleurope.eu/index.action>>) ou encore Arizona Archives Online (<http://azarchivesonline.org>), Archives Florida (<http://palm2.fcla.edu/afll/>), Northwest Digital Archives ou NWDA (<http://nwda.orbiscascade.org/>), dans l'Oregon, OhioLINK (<http://ead.ohiolink.edu>), Online Archive of California ou OAC (<http://www.oac.cdlib.org/>), Texas Archival Resources Online ou TARO (<http://www.lib.utexas.edu/taro>), etc., APEnet est le plus ambitieux de tous ces projets regroupant des institutions par zones géographiques, puisqu'il vise à donner un accès unifié aux archives européennes.

L'objectif de la réunion était de favoriser les échanges entre les différents consortiums représentés pour développer une approche commune, autour de trois interrogations :

1°) Quelle est votre vision idéale pour des collaborations régionales plus larges ? Qu'est-ce qui inciterait ces collaborations ?

Les participants ont identifié trois facteurs pour le développement de collaborations plus importantes :

- fournir aux partenaires des ressources humaines et financières constantes (par exemple, le coût du projet APEnet a été de près de 3 millions d'euros et une équipe de quinze personnes, composés d'archivistes et d'informaticiens des différents États membres de l'Union européenne, a été mobilisée) ;
- créer un réseau d'entraide : manuels et guides de bonnes pratiques, outils de formation, listes de discussion, autres outils (plusieurs consortiums ont mis à disposition des feuilles de style XSLT que les partenaires peuvent facilement adapter à leurs propres besoins) ;
- élaborer un ensemble d'outils de conversion et de plug-ins pour les institutions participantes afin de faciliter l'échange de données et l'interopérabilité.

C'est ainsi que le projet APEnet a défini un schéma EAD APEnet qui est un sous-ensemble du schéma de l'EAD 2002 et qui répond aux exigences de la présentation conjointe des descriptions dans le portail européen des archives. Le schéma EAD APEnet contient uniquement les éléments et attributs de l'EAD nécessaires pour l'affichage commun, les mises en relation et l'identification des ressources archivistiques. Le schéma a été développé à partir de la comparaison des pratiques des directions des archives et des Archives nationales participant actuellement au projet. Un outil de conversion locale a été développé, permettant à chaque partenaire de convertir ses données – déjà structurées en EAD « locale » ou

extraites de bases de données compatibles avec ISAD(G) – dans le format EAD APENet, avant de les intégrer dans le portail. Par ailleurs, toutes les institutions d'archives intéressées par le projet peuvent tester localement leurs données par rapport aux normes archivistiques utilisées dans le portail avec cet outil. L'outil est open source et peut être téléchargé à l'adresse : <http://www.apenet.eu/> (Développements).

La participation au portail européen des archives sera ouverte à tous les services d'archives européens qui peuvent partager leurs descriptions de leurs collections structurées conformément aux normes internationales archivistiques (EAD, EAC-CPF, EAG et le format METS) ou dans un format qui peut être converti par eux-mêmes avec les outils fournis par le projet.

L'équipe APENet a par ailleurs développé un outil logiciel pour convertir les informations sur les objets numériques archivistiques auxquels donnent accès les instruments de recherche APENet EAD vers le modèle ESE de métadonnées d'Europeana basé sur le Dublin Core.

2°) Quels sont les obstacles administratifs, techniques ou structurels à la collaboration ?

Les participants ont identifié les obstacles administratifs, structurels et techniques suivants : petite taille des services, insuffisance des infrastructures institutionnelles, manque de normalisation dans les pratiques descriptives et diversité des usages de l'EAD, préoccupations concernant la qualité ou l'exhaustivité des instruments de recherche qui entraînent les archivistes à réécrire les instruments de recherche existants avant leur encodage en EAD, mauvaise qualité des instruments de recherche qui peut être un obstacle à la mise en œuvre EAD dans certains établissements, utilisation de flux logiciels complexes, serveurs et outils de publication inadaptés, etc.

3°) Quel est votre prochain grand projet ?

L'équipe de projet d'APENet étudie actuellement les modalités de mise en œuvre du schéma EAC-CPF dans le portail. Les Archives nationales des Pays-Bas souhaitent quant à elles améliorer le référencement de leurs instruments de recherche EAD sur Google. Enfin, dans le cadre de la Réforme générale des politiques publiques en France, le conseiller d'État en service extraordinaire Maurice Quénet a été mandaté le 27 octobre 2010 pour explorer « les voies et moyens de moderniser la gestion des archives de l'État et de faciliter l'accès de tous nos concitoyens à ce patrimoine culturel inestimable ». Parmi les onze propositions formulées dans le rapport Quénet, on parle d'entreprendre « le regroupement virtuel des fonds au sein d'un portail interministériel intégré au projet européen APENet ». Un tel portail interministériel permettrait aux utilisateurs non seulement d'avoir accès aux instruments de recherche mais aussi aux documents d'archives numérisés disponibles en ligne.

### ***Table ronde des utilisateurs des logiciels Archivists'Toolkit et Archon et du projet ArchiveSpace***

#### ***=> Sommaire***

Le logiciel Archivists' Toolkit (AT ou « Boîte à outils des archivistes ») est une application libre développée depuis 2007 à l'initiative des universités de San Diego et de New York et des « Cinq collègues » (consortium d'établissements d'enseignement supérieur du Massachusetts), qui permet de rédiger des descriptions conformes aux normes DACS (« Décrire les archives : une norme de contenu », norme de la SAA) et ISAAR(CPF) et de faire des imports et des exports conformes à la DTD EAD 2002. Le site du projet est consultable à : <http://archiviststoolkit.org/>

Cet outil, qui a atteint en août 2011 sa version 2.0-11 (téléchargeable à [http://archiviststoolkit.org/download/release/2\\_0](http://archiviststoolkit.org/download/release/2_0)), est aujourd'hui l'un des plus utilisés par les départements de collections spécialisées des universités américaines en Alabama, en Arizona, en Californie, dans le Connecticut, en Géorgie, dans le Maryland, dans le Massachusetts, dans le Michigan, dans le New Jersey, dans l'Oklahoma, en Pennsylvanie, dans le Texas, etc. Une communauté d'utilisateurs s'est peu à peu constituée, donnant naissance à une Table ronde spécifique au sein de la SAA, très dynamique comme en témoignent les échanges sur la liste de discussion. Certains utilisateurs ont développé des fonctionnalités répondant à leurs besoins immédiats pour ensuite en faire profiter la communauté. Ainsi, l'université de Harvard a enrichi la liste des valeurs disponibles dans le champ « Description physique » pour pouvoir enregistrer des informations sur les différents types de conditionnements et de supports de ses collections archivistiques (voir : <http://archiviststoolkit.org/node/251>). L'université Brigham Young a automatisé le processus de création de liens entre des objets numériques et les descriptions archivistiques (voir : <http://archiviststoolkit.org/node/246>). Mais l'apport le plus important au cours des derniers mois a été celui du

Centre d'archives Rockefeller, qui a développé un module pour gérer l'enregistrement des lecteurs et les demandes de reproduction de documents (voir la présentation du projet à : <https://github.com/RockefellerArchiveCenter/ATReference/wiki>). Les objectifs étaient de consolider les informations semblables/dupliquées saisies, stockées et gérées dans plusieurs systèmes et d'améliorer l'efficacité, la qualité et la sécurité des chercheurs. Cinq priorités ont été identifiées, dont seule la première pour le moment a donné lieu à des développements :

- enregistrement des lecteurs (mai 2011) : extension du module Noms de AT pour rentrer les informations d'enregistrement des lecteurs, ajout d'une fonctionnalité permettant de tracer les visites et les publications des chercheurs en rendant possibles l'interrogation et la manipulation de ces données pour produire des statistiques et des rapports ;
- services de reproduction (décembre 2011) : gestion de tous les types de demandes de reproduction, création de formulaires normalisés en PDF, ajout d'une fonctionnalité permettant de tracer les demandes de reproduction, en rendant possibles l'interrogation et la manipulation de ces données pour produire des statistiques et des rapports ;
- traçabilité des consultations de documents par les lecteurs et fonctionnalité de codage barre des documents (juillet 2012)
- gestion de la salle de lecture et des demandes d'aide à la recherche ;
- interface Web et comptes utilisateurs personnalisés.

Les concepteurs d'AT réfléchissent à la combinaison de certaines des fonctionnalités du logiciel avec un autre outil, développé par l'université de l'Illinois à Urbana-Champaign, Archon (<http://archon.org/>). Le logiciel Archon permet de créer des descriptions archivistiques, d'y associer des objets numériques ou numérisés (images, sons, vidéos) et de gérer des plans de classement complexes. Archon permet aussi l'export des données en EAD ou en MARC. AT et Archon sont donc à présent en train de migrer vers un système unifié, ArchivesSpace (<http://www.archivesspace.org/>). Le projet, initié en juillet 2009 et bénéficiaire d'une subvention de la Fondation Mellon, devrait s'achever en février 2013.

## Sessions plénières

### *Session 304 – Les archives dans le Web de données : vers un flux vertueux de données au Royaume-Uni et au-delà*

#### *=> Sommaire*

Comment le partage ouvert des métadonnées peut-il aider à améliorer la visibilité et l'utilisation des collections ? Comment faire en sorte que nos données « travaillent plus » ? Les intervenants ont retracé les initiatives en cours au Royaume-Uni et États-Unis visant à aider les archives à partager ouvertement leurs métadonnées et à favoriser la réutilisation de ces métadonnées.

C'est ainsi que **Kira Omo** (université de l'État de l'Oregon) a présenté un projet collaboratif de « découverte ouverte » (open discovery) impliquant une trentaine d'institutions du nord-ouest des États-Unis. Accessible en ligne depuis 2005, le portail Northwest Digital Archives ou NWDA (<http://nwda.orbiscascade.org/>), permet d'accéder aux descriptions des collections d'archives conservées dans les bibliothèques des universités des États de l'Idaho, du Montana, de l'Oregon, de l'Alaska et de Washington. Le catalogue NDWA est un programme de l'Alliance Orbis Cascade, consortium regroupant 36 bibliothèques d'universités et d'établissements d'enseignement supérieur des États de l'Oregon et de Washington.

Le NDWA donne accès à plus de 5 800 instruments de recherche encodés en EAD, mais seul un très petit nombre comprend des liens vers les contenus numériques, que les institutions membres hébergent sur leurs propres plates-formes avec des systèmes de gestion très différents : CONTENTdm, logiciel de gestion des collections numérisées d'OCLC, DSpace, logiciel libre permettant la construction d'archives électroniques ouvertes, ou encore le système de gestion de contenu open source Drupal, etc. Les mécanismes d'interrogation permettant la « découverte » de ces ressources sont donc hétérogènes et il n'est pas possible d'effectuer des recherches transverses depuis un guichet unique. Les études utilisateurs soulignent l'insuffisance des métadonnées pour les objets numériques individuels et la difficulté à retrouver ces ressources hébergées dans des systèmes distincts. Or, ce que les utilisateurs veulent avant tout, c'est pouvoir accéder aux contenus numériques avec le plus d'informations contextuelles possibles et avoir à leur disposition des moyens faciles pour communiquer avec les professionnels pour plus d'informations. Par contre, ils ne veulent pas de sites se focalisant sur une thématique donnée.

En 2009, l'Alliance Orbis Cascade a obtenu une subvention de l'Institut des Services de Musées et de Bibliothèques (Institute of Museum and Library Services ou IMLS) pour le développement d'un prototype

pour la découverte des ressources numériques. L'outil ainsi réalisé, XCU, réunit les contenus numérisés de 36 systèmes de 25 institutions, ainsi que les métadonnées détaillées provenant des collections d'archives et de manuscrits de 33 établissements. La plate-forme permet d'accéder aux objets numériques dans le contexte des collections auxquels ils sont liés et d'afficher de manière conviviale les objets numériques et les métadonnées qui leur sont associées.

**Joy Palmer** (université de Manchester, Archives Hub) a ensuite montré comment la « découverte » (discovery) était considérée au Royaume-Uni, dans les domaines de l'enseignement et la recherche, comme une véritable « écologie de métadonnées », afin d'aboutir à un « flux vertueux de métadonnées à travers le Web ».

La découverte est une initiative nationale financée par le JISC (Joint Information Systems Committee ou Comité conjoint pour les systèmes d'information, équivalent du TGE Adonis), les Archives nationales, la British Library et le Collections Trust (organisation indépendante travaillant pour l'amélioration et le développement des collections culturelles). Cette initiative repose sur une vision commune selon laquelle les chercheurs et les étudiants auront un accès flexible et continu au contenu et aux services à travers un cadre collaboratif, agrégé et intégré de découverte et de distribution des ressources. Comment atteindre cet objectif ? Par une infrastructure unique permettant de tout relier ? Par des services partagés et des plates-formes de cloud computing<sup>3</sup> ? La solution semble plutôt passer par les Linked data et par un « écosystème de métadonnées » pour reprendre l'expression employée par Rufus Pollock, l'un des fondateurs de l'Open Knowledge Foundation<sup>4</sup>. De fait, « le modèle de base pour le traitement des données est actuellement une rue à sens unique. Les sources de données, comme par exemple le gouvernement, publient des données..., qui sont traitées par des intermédiaires ... avant d'être finalement consommées par les utilisateurs finaux. C'est une rue à sens unique, car il n'y a pas de retours, pas de partage de données vers les éditeurs et pas de partage entre les [utilisateurs] intermédiaires... Ce que nous devrions avoir est un écosystème... Avec l'introduction de cycles de données, nous avons un véritable écosystème qui n'est pas une rue à sens unique et cet écosystème se développe sur la collaboration, la modularisation et les données ouvertes » (<http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/>).

En 2010, le JISC et RLUK Resource Discovery Taskforce (RDTF) ont donc travaillé avec les bibliothèques, les archives et les musées britanniques pour définir les grands principes d'une « écologie de métadonnées pour la découverte » (<http://discovery.ac.uk/>), visant à rendre les données culturelles accessibles à tous et à s'affranchir des restrictions imposées par les licences payantes ou autres formes de contrôle qui en limitent notamment l'exploitation et la reproduction.

Portail collaboratif fournissant un point d'accès unique aux descriptions de collections d'archive détenues dans plus de 180 établissements universitaires britanniques, Archives Hub est une des briques de cette « vision nationale » pour la découverte des ressources. Le service est financé par le Comité conjoint pour les systèmes d'information (JISC) et est supervisé par Bibliothèques de recherche du Royaume-Uni (Research Libraries UK). Le centre de données MIMAS de l'université de Manchester héberge le service et le suivi et la maintenance de l'application sont assurés par l'équipe de développement de Cheshire à l'université de Liverpool.

Pour **Jane Stevenson** (université de Manchester, Archives Hub), la découverte repose sur trois principes :

- toujours penser à des approches flexibles afin d'ouvrir et d'optimiser le potentiel de ses données
- penser « visuel » mais combiner cette approche avec une solide épine dorsale fournissant du contenu et des informations contextuelles
- les données « en sortie » reposent sur la qualité des données « en entrée ». Il faut donc s'efforcer de maintenir la qualité des données entrant dans le système d'information.

L'objectif des développeurs de Archives Hub est de partager les descriptions de collections d'archives pour les valoriser et montrer combien les archives sont pertinentes dans de nombreuses disciplines et à de nombreuses fins. À bien des égards, le partage des données permet de s'éloigner de l'approche selon laquelle les archives n'intéressent qu'un groupe restreint d'utilisateurs (généalogistes, historiens des familles, universitaires...).

Toutes les descriptions dans Archives Hub ont donc des URIs pérennes, ce qui leur permet d'avoir une identité sur le Web, d'être exposées en Linked Data et d'être mises en relation les unes avec les autres ou

---

<sup>3</sup> Le cloud computing ou « informatique en nuage » consiste à déporter sur des serveurs distants des traitements informatiques traditionnellement localisés sur des serveurs locaux ou sur le poste client de l'utilisateur.

<sup>4</sup> Association à but non lucratif promouvant les données ouvertes. Elle a été fondée en 2004 à Cambridge au Royaume-Uni. La fondation maintient plusieurs projets, tel que CKAN (Comprehensive Knowledge Archive Network), répertoire utilisé par de nombreux gouvernements pour leurs projets de données ouvertes



avec d'autres ressources externes. Les noms, les lieux, les sujets, les institutions de conservation, etc. au sein de Archives Hub ont ainsi leur propre identifiant, ce qui permet par exemple de mettre en relation une personne identifiée dans Archives Hub avec la même personne identifiée dans d'autres ensembles de données.

**Adrian Stevenson** (université de Bath) a présenté un premier projet, LOCAH (Linked Open Copac Archives Hub), consistant à rendre les notices d'Archives Hub et du catalogue collectif COPAC<sup>5</sup> disponibles en tant que Linked Data, et à créer des liens avec d'autres sources de données (DBPedia, BBC, OCLC pour les autorités Noms, fichiers de vedettes matières de la Bibliothèque du Congrès). Un modèle conceptuel a été développé à partir de l'EAD pour créer un réseau de relations entre les données du système. Le site Linked Data est accessible à : <http://data.archiveshub.ac.uk>.

Un deuxième projet, intitulé « Linking Lives », tend à développer une nouvelle interface Web pour Archives Hub, en offrant un accès plus direct aux données contenues dans le portail relatives à des personnes et éventuellement à des organismes. Il s'agira ensuite de confronter ces données avec des ressources externes. Les chercheurs veulent en effet connaître toutes les ressources, archivistiques ou non archivistiques, disponibles sur des personnes ou des organismes. Ils peuvent ne pas avoir pensé à utiliser les archives, mais ils peuvent aussi être très intéressés par les informations figurant dans les instruments de recherche archivistiques et concernant la biographie d'une personne, les événements survenus dans sa vie, les autres personnes avec lesquelles elle était en relation, etc. L'idée est de montrer que différentes sources d'information existent, y compris les archives, et donc d'élargir le public des archives.

Cette approche est une alternative intéressante à la production de notices d'autorité nativement en EAC-CPF. Les développeurs d'Archives Hub sont d'ailleurs très intéressés par un autre projet de recherche visant à constituer un fichier virtuel d'autorités à partir de différentes ressources. Il s'agit du projet SNAC (Social Networks and Archival Context project ou Réseaux sociaux et contexte archivistique), qui implique l'Institut des technologies de pointe en sciences humaines de l'université de Virginie, l'École d'information de l'université de Californie à Berkeley et la Bibliothèque numérique californienne (cf. infra compte rendu de la session 710). Un prototype de données d'autorité provenant de sources diverses (archives, bibliothèques, musées) a été développé pour relier entre elles les descriptions des ressources historiques et fournir un système d'accès à ces ressources (<http://socialarchive.iath.virginia.edu/>). Le standard choisi pour l'encodage des données d'autorité générées est l'EAC-CPF. Les données provenant d'Archives Hub pourraient alimenter les données d'autorité produites dans le cadre de SNAC.

### ***Session 610 – Approches pratiques pour les archives nativement numériques : qu'arrivera-t-il demain ?***

#### ***=> Sommaire***

Archivemata est une plateforme de préservation numérique réunissant une large suite d'outils libres d'extraction de métadonnées, de conversion de formats, etc., et qui soutient le traitement automatisé des documents d'archives électroniques (<http://www.archivemata.org>). Le projet est piloté par la société Artefactual Systems, en collaboration avec le sous-comité sur les technologies du programme Unesco Mémoire du monde, les Archives de la ville de Vancouver, la Bibliothèque de l'université de Colombie britannique et le Centre d'archives Rockefeller. L'outil est déjà en cours de déploiement aux Archives de la ville de Vancouver, au Centre d'archives Rockefeller, à l'université de l'Illinois ou encore aux Archives du Fonds monétaire international.

Le projet présente les caractéristiques suivantes :

- l'outil crée des paquets d'information archivés (AIP) qui peuvent être conservés sur le long terme,. Il doit être interfacé, pour de ces AIP, avec une infrastructure matérielle ;
- Archivemata est conçu de manière évolutive ; l'application peut fonctionner en monoposte dans un petit service ou sur plusieurs serveurs ;
- Archivemata intéressera plus particulièrement les petits services ; cependant, tous ne seront pas en mesure de répondre aux exigences logicielles de la plate-forme ou d'acquérir les connaissances techniques requises ;
- Archivemata demande un certain degré de savoir-faire technique. Ainsi, un environnement réel ou virtuel Linux est nécessaire. Or, la plupart des archivistes ne sont pas familiers avec Linux et doivent être prêts à en apprendre les rudiments. Par ailleurs, le processus d'installation n'est pas aussi simple qu'il pourrait l'être.

<sup>5</sup> Catalogue collectif proposant l'accès direct à l'information contenue dans les principaux catalogues en ligne de bibliothèques universitaires de recherche du Royaume-Uni.



Archivematica demande des développements supplémentaires, mais étant donné que l'outil n'a pas encore atteint le stade de développement bêta, ce besoin n'est pas surprenant ! Les intervenants ont fait part de souhaits d'améliorations qui permettrait aux archivistes de spécifier les formats à employer pour la pérennisation des documents numériques, de réinsérer ou de traiter autrement les fichiers rejetés par Archivematica, et de comprendre pourquoi le processus de réception des paquets d'information échoue parfois. Les participants souhaiteraient également que le traitement des e-mails soit amélioré et que le logiciel devienne un outil de records management.

### **Session 710 – Le projet « Réseaux sociaux et contexte archivistique (SNAC) : l'EAC-CPF au travail »**

#### **=> Sommaire**

Le projet SNAC est un projet de recherche collaboratif entre l'Institut des technologies de pointe en sciences humaines de l'Université de Virginie, l'École d'information de l'Université de Californie à Berkeley et la Bibliothèque numérique californienne de l'Université de Californie. Le financement est assuré en partie par la Dotation nationale pour les sciences humaines (National Endowment for Humanities).

Le projet a débuté en mai 2010 pour une durée de deux ans. Le but de ce projet est de créer un prototype de données d'autorité provenant de sources diverses issues de données d'archives, de bibliothèques et de musées afin de relier entre elles les descriptions des ressources historiques et de fournir un système d'accès à ces ressources. Les sources utilisées sont les instruments de recherche encodés en EAD de trois consortiums d'archives ainsi que ceux de la Bibliothèque du Congrès<sup>6</sup>.

Le projet utilisera également les données d'autorité du fichier d'autorité de la Bibliothèque du Congrès (NAACL/LCNAF ; 3,8 millions de noms de personnes et 900 000 noms de collectivités), du fichier d'autorité virtuel international (VIAF ; <http://www.viaf.org> ; 5 millions de noms de personnes) et de la Liste collective de noms d'artistes (ULAN) de la Fondation Getty (soit 293 000 noms de personnes et de collectivités).

Le standard choisi pour l'encodage des données d'autorité générées est l'EAC-CPF. Les données concernant les personnes, collectivités et familles se trouvant dans les instruments de recherche archivistiques EAD sont dérivées des éléments <origination> et <controlaccess> ainsi que des séries décrivant les correspondances dans lesquelles des noms peuvent être identifiés. Les descriptions relatives aux producteurs d'archives sont enrichies avec les informations qui se trouvent dans les éléments <bioghist> et <occupation> et dans les autres éléments considérés comme appropriés. Ces informations sont confrontées entre elles et fusionnées. Ensuite, elles sont confrontées avec celles contenues dans les fichiers d'autorité NAACL/LCNAF, VIAF et ULAN pour vérification et enrichissement (normalisation des points d'accès, ajout de formes alternatives de noms, de titres, de données historiques). Un ensemble unique de données d'autorité est ainsi créé et est relié aux instruments de recherche archivistiques. L'ensemble des données est exposé en tant que Données liées (Linked Data). Une première version du prototype peut être consultée à l'adresse : <http://socialarchive.iath.virginia.edu/>

Les difficultés rencontrées sont multiples : confusion de la description des producteurs d'archives avec les noms associés aux contenus des documents d'archives, qualité très variable dans le nombre des noms identifiés et encodés, dans la forme des noms (directe, inversée, utilisation des majuscules, ponctuation, etc.) et dans leur catégorisation (personnes, collectivités, familles), présence de nombreux noms qui ne sont pas identifiés en tant que tels... L'extraction porte en priorité sur les noms balisés en tant que tels, les autres noms qui ne sont pas identifiés en tant que tels seront progressivement traités. Par ailleurs, la qualité de description des collectivités, des personnes et des familles est très variable d'un instrument de recherche en EAD à l'autre et si de nombreux noms obéissent au contrôle d'autorité dans les instruments de recherche de la Bibliothèque du Congrès, dans d'autres (Online Of California et NWDA), on trouve moins de noms et le niveau de contrôle est variable. Les procédures d'extraction seront affinées avec l'application d'outils de Traitement Automatisé de la Langue (TAL), pour vérifier le type d'entité (collectivité, personne ou famille), pour normaliser les formes de noms et pour identifier les chaînes de caractères qui sont potentiellement des noms mais qui ne sont pas identifiés en tant que tels.

Les responsables du projet viennent d'obtenir le financement par IMLS d'une deuxième phase prévoyant

<sup>6</sup> Les instruments de recherche utilisés sont ceux des consortia Online Archive of California (environ 15 400 instruments de recherche), Northwest Digital Archive (5 160 instances EAD), Virginia Heritage (8 390 instances EAD), auxquels s'ajoutent 1 159 instruments de recherche EAD de la Bibliothèque du Congrès.

d'étendre le travail vers les instruments de recherche archivistiques d'autres institutions, notamment ceux des plus importantes archives et bibliothèques de recherche des Etats-Unis (NARA, Institution smithsonienne, notices MARC du catalogue WorldCat d'OCLC), afin de créer une infrastructure pour un système national d'autorité pour les archives. Un appel à contribution (fourniture d'instruments de recherche EAD et notices EAC-CPF) est également lancé auprès de la communauté internationale.

---

Merci de faire parvenir toute information ou tout article pour le prochain numéro à : Claire Sibille – de Grimouard, Direction générale des patrimoines, Service interministériel des Archives de France