

Sommaire

Journée d'études « En avant avec les autorités », Washington, 9 août 2010

74^e Congrès annuel de la Société des archivistes américains, Washington, 10-15 août 2010

Journée d'études « En avant avec les autorités », Washington, 9 août 2010

Le 9 août 2010, les Archives nationales des Etats-Unis (NARA, acronyme de National Archives and Records Administration) ont accueilli la pré-conférence du 74^e Congrès de la Société des archivistes américains, intitulée « En avant avec les autorités ». La publication, le 5 mars dernier, du schéma EAC-CPF et de son dictionnaire de balises sur le site officiel <http://eac.staatsbibliothek-berlin.de/> a reçu un accueil enthousiaste de la part des archivistes, car elle répond au besoin croissant d'un outil rigoureux et en même temps simple et souple pour la description des informations contextuelles relatives aux collectivités, personnes et familles.

Anila Angjeli (Bibliothèque nationale de France), présidente du groupe d'expert AFNOR « Données d'autorité » et membre du Groupe international EAC, est intervenue au cours de cette journée pour présenter différents projets de mise en œuvre du schéma EAC-CPF en France. Voici son compte rendu.

Introduction à l'EAC-CPF, par Kathy Wisser (présidente du Groupe de travail pour la révision de l'EAC, Simmons University, Boston)

En ouverture de la journée, Kathy Wisser a présenté le schéma EAC-CPF, son architecture et ses caractéristiques. Le schéma résulte du travail de plus de deux ans d'un groupe de travail international composé de 15 archivistes et bibliothécaires. Il est l'héritier de la DTD EAC dont les origines remontent à 2001, avec la réunion à Toronto d'un groupe d'archivistes à Toronto pour définir les principes pour l'élaboration d'un modèle de haut niveau pour la description des producteurs d'archives. Le schéma EAC-CPF 2010 a pris en compte les commentaires des utilisateurs recueillis lors de deux enquêtes internationales et de deux webinaires. Le résultat est prometteur et semble satisfaire les utilisateurs tant effectifs que potentiels.

Projet « Réseaux sociaux et contexte archivistique » (Social Networks and Archival Context Project – SNAC), par Daniel Pitti (Université de Virginie)

Il s'agit d'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. Le projet utilisera également les données d'autorité du fichier d'autorité de la Bibliothèque du Congrès (NACL/LCNAF), du fichier d'autorité virtuel international (VIAF ; <http://www.viaf.org>) et de la Liste collective de noms d'artistes (ULAN) de la Fondation Getty.

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 seront 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 seront enrichies avec les informations qui se trouvent dans les éléments <bioghist> et <occupation> et dans les autres éléments considérés comme appropriés. Ces informations seront confrontées entre elles et fusionnées. Ensuite, elles seront confrontées avec celles contenues dans les fichiers d'autorité NACL/LCNAF, VIAF et ULAN pour vérification et enrichissement. Un ensemble unique de données d'autorité sera ainsi créé et sera relié aux instruments de recherche archivistiques. L'ensemble des données sera exposé en tant que Données liées (Linked Data). Une première version du prototype peut être consultée à l'adresse : <http://socialarchive.iath.virginia.edu/>

Dans une deuxième phase, il est prévu d'étendre le projet vers d'autres instruments de recherche archivistiques, notamment ceux des plus importantes archives et bibliothèques de recherche des Etats-Unis. Cette deuxième phase, déjà baptisée « SoNAAR » (Contexte social et Notices d'autorité archivistiques – Social Context and Archival Authority Record) dépendra du financement obtenu.

APEnet : un projet financé par l'Union européenne pour un portail consacré aux archives en Europe, par Angelika Menne-Haritz (Bundesarchiv)

APEnet (Archives Portal Europe : <http://www.apenet.eu>), initiative visant à créer un portail Internet des archives en Europe, rassemble les efforts et l'expertise des archives aux niveaux national et européen. Ce projet a été défini comme prioritaire et approuvé par le Conseil européen dans sa recommandation du 14 novembre 2005 (JO de l'Union européenne, 22 novembre 2005, 2005/535/EC). Il est mis en œuvre dans le cadre du programme eContent+ de la Commission européenne qui le finance à hauteur de 80%. Il a officiellement été lancé en janvier 2009 ; le pilote doit être mis en ligne en janvier 2011 et la version finale en janvier 2012.

Le projet est complexe en raison du nombre important de pays, de langues, de traditions archivistiques différentes et de la multitude des systèmes archivistiques qu'il faut mettre en commun. Cependant, toutes les institutions participant au projet décrivent leurs archives selon les standards définis par la communauté archivistique internationale : certaines archives ont des bases de données conformes à l'ISAD(G), certaines utilisent l'EAD pour les exports de données, un grand nombre utilisent l'EAD pour l'encodage et l'affichage de leurs données.

Dans ce contexte, pouvoir assurer l'interopérabilité à court et à long terme est une question cruciale pour la réussite du projet. L'EAD a été défini comme le format pivot du projet. La seule contrainte imposée aux partenaires est donc de fournir leurs données sous la forme de fichiers EAD-XML. En effet, au centre d'APEnet, un instrument de recherche global réunira toutes les données afin de proposer un accès commun aux descriptions archivistiques et, le cas échéant, aux documents numérisés eux-mêmes.

Les trois piliers du modèle de cet instrument de recherche global sont :

- l'outil de conversion et de validation des données originales dans le format commun (APEnet-ead). Ce format commun a été défini sur la base d'une analyse fine des pratiques de chaque pays dans la mise en œuvre de l'EAD ; il est adapté selon l'évolution des besoins ou l'entrée de nouveaux partenaires dans le projet.
- l'outil de chargement et d'intégration qui transmet les données à l'hébergement central,
- l'outil d'indexation et d'affichage qui permet de combiner les données pour la navigation et l'affichage.

L'instrument de recherche global est structuré en trois niveaux :

1. Pays/institution. Les différents services d'archives seront créés au sein de leurs pays respectifs, en suivant l'organisation administrative propre, au fur et à mesure de l'intégration des contenus. Ces « descriptions » d'institutions ou services sont en EAD.

2. Guide des fonds. Les guides des fonds d'archives intégrés par chaque partenaire fournisseur de contenus comprennent des liens horizontaux vers les producteurs (décrits en EAC-CPF) et des liens verticaux vers les instruments de recherche détaillés. Si les services d'archives n'ont pas encore de

guide des fonds en EAD, ils pourront en créer semi-automatiquement à partir des instruments de recherche qu'ils auront fournis.

3. Instrument de recherche. Les instruments de recherche en EAD comportent des liens vers les documents numérisés le cas échéant.

La navigation entre tous les fichiers fournis est assurée par des liens rendus transparents en html.

Chacun des niveaux présente les fichiers de données complets avec leur hiérarchie interne (en EAD). Les liens verticaux entre les fichiers sont réalisés en exploitant <otherfindaid>, <p> et <extref>. Il n'y a pas d'utilisation des espaces de nom. Les noms des fichiers font office d'identifiants (documentés dans le guide des fonds ou dans l'instrument de recherche par l'attribut @identifier au sein de l'élément <eadid>) et sont utilisés comme des ancres pour les liens.

Il est à noter que les institutions participantes maîtrisent totalement leurs données : en effet, l'infrastructure technique centralisée leur est largement accessible afin qu'ils puissent maintenir, corriger ou enrichir leurs données originales. Les procédures pour les partenaires fournisseurs de contenus seront au choix automatisées ou non, pour assurer un maximum de transparence.

Les fonctionnalités seront très limitées pour l'internaute en janvier 2011. La première partie du travail s'est en effet concentrée sur les fonctionnalités pour les partenaires fournisseurs de contenus. Deux types de recherche seront toutefois disponibles : navigation « à vue » dans le paysage archivistique européen, en suivant les liens verticaux et horizontaux présents dans l'instrument de recherche global, ou recherche plein texte, avec un affichage des résultats dans l'environnement hiérarchique de l'instrument de recherche. Dans un second temps, le portail proprement dit offrira davantage de fonctionnalités et de contenus.

Enfin, APEnet jouera le rôle d'agrégateur de contenu en matière d'archives pour le portail Europeana. Europeana est d'ailleurs membre du projet APEnet.

APEnet a développé un outil de conversion du format APEnet-EAD vers le format de données Europeana (ESE). La difficulté principale à résoudre résidait dans la radicale différence d'approche des deux portails. Europeana ne s'intéresse en effet qu'aux objets numérisés, décrits individuellement, à plat, en suivant un modèle de données basé sur Dublin Core. Pour retrouver les informations contextuelles nécessaires à la compréhension des documents d'archives, l'outil de conversion incorpore donc dans chacune des descriptions un certain nombre d'éléments pris dans les niveaux de description hiérarchiques supérieurs, notamment les titres de série, fonds etc. Les premiers tests, concluants, ont été mis en ligne sur Europeana grâce aux archives des Pays-Bas. L'évolution annoncée du schéma de données d'Europeana vers un schéma beaucoup plus riche et complexe sera suivie de près par l'équipe du projet APEnet qui adaptera son outil de conversion ; par ailleurs, de façon symétrique, une « dose de hiérarchie » est désormais prise en compte dans les réflexions de l'équipe Europeana afin de mieux restituer les spécificités des archives, ce qui permettra également de mieux répondre à certains besoins des musées ou des bibliothèques.

Premiers pas avec EAC-CPF en France, par Anila Angjeli (Bibliothèque nationale de France)

La publication de la norme ISAAR(CPF), surtout après sa deuxième publication en 2004, a été un événement important pour les archivistes français leur fournissant les règles de rédaction de fichiers d'autorité de collectivités, personnes et familles. Elle a permis que la parution de la DTD EAC bêta en 2004 soit rapidement adoptée par un certain nombre d'archives en France pour la description séparée des collectivités, personnes et familles, comme le complément indispensable de la description des documents d'archives en EAD. Cependant la DTD EAC bêta n'était pas pleinement satisfaisante et les archivistes français attendaient sa révision. C'est ainsi qu'ils n'ont pas hésité à s'impliquer dans la révision de l'EAC, notamment en commentant les versions de travail successives de l'EAC-CPF émises par le Groupe international pour la révision de l'EAC. La présente intervention relate trois expériences françaises en cours de réalisation dans l'application de l'EAC-CPF.

La première concerne l'implémentation de l'EAC-CPF dans le nouveau Système d'information archivistique (SIA) lié au nouveau site des Archives nationales, à Pierrefitte-sur-Seine. Le SIA sera un système d'information commun aux trois sites des Archives nationales de France : Paris,

Fontainebleau et Pierrefitte-sur-Seine. Un fichier d'autorité central faisant partie des applications documentaires servira à faire le lien entre les applications de gestion et l'interface d'interrogation. Ce projet fait l'objet du « Chantier scientifique : producteurs des documents d'archives ». L'EAC-CPF a été choisi comme schéma de métadonnées pour la description des producteurs. Ce projet scientifique concerne tant la conversion des descriptions existantes des collectivités, personnes et familles en EAC-CPF que la réalisation des guides méthodologiques pour la création des nouvelles instances. Deux cas ont permis d'illustrer la problématique de conversion, l'un étant la conversion en EAC-CPF de l'application Etanot (<http://chan.archivesnationales.culture.gouv.fr/sdx-222-chan-etanot/etanot/index.xsp>), le fichier d'autorité des notaires parisiens, réalisé en EAC bêta - l'autre étant la conversion en EAC-CPF des instances de la dite « Nomenclature N1 », créée depuis 1981 et décrivant les services versants d'administration centrale. Les problématiques sont différentes. Pour Etanot, il fallait juste s'assurer que toutes les informations pouvaient être converties sans perte. Des tests de conversion ont été effectués et les résultats sont assez satisfaisants. Pour la Nomenclature N1 l'exercice a obligé à modifier radicalement l'organisation de l'information contenue dans le fichier : mise en conformité avec ISAAR(CPF), redéfinition des éléments permettant l'identification unique de l'entité décrite (importance de la forme autorisée du nom) et redéfinition des relations entre les instances.

La deuxième expérience relatée concerne les tests d'implémentation de l'EAC-CPF pour la description des collectivités, personnes et familles liées aux IREL (Instruments de recherche en ligne) des Archives nationales d'outre-mer (ANOM). L'accent a été mis sur l'importance scientifique des fichiers d'autorité pour la compréhension des documents d'archives relatifs à l'histoire de la présence coloniale française. Les ANOM ont été précurseurs dans l'application de l'ISAAR(CPF) et dans la réalisation de fichiers d'autorité pour les personnes en EAC bêta dans une application interne. Dès la publication de l'EAC-CPF, les ANOM ont procédé à des tests en vue de la modification de l'application existante pour la rendre conforme à EAC-CPF et la relier aux IREL. Les descriptions actuelles de personnes réalisées en EAC bêta seront converties et y seront intégrées ; les descriptions de collectivités, non encodées en EAC en attendant la publication de l'EAC-CPF, seront, elles aussi, traitées directement dans la nouvelle application. Des tests ont déjà été réalisés et un échantillon des résultats peut être consulté en ligne à l'adresse : <http://irel2.tech.fr/eac/cpf/>

Le troisième cas présenté est celui du projet d'implémentation de l'EAC-CPF à la Bibliothèque nationale de France (BnF) pour la description des producteurs des archives institutionnelles de la BnF. La BnF réalise des instruments de recherche en EAD pour la description des fonds d'archives et des manuscrits. Les archives contemporaines sont également en cours de description. Pour les autorités, le choix a été fait d'établir des liens entre les instruments de recherche en EAD et les Fichiers d'autorité de la BnF en MARC. Or, pour la description des producteurs d'archives institutionnelles, un traitement en MARC obéissant aux règles de rédaction des fichiers d'autorité de la BnF n'est pas satisfaisant. C'est plutôt une description conforme à ISAAR(CPF) en utilisant EAC-CPF comme standard de production et de communication qui conviendrait le mieux. Dans ce contexte, le défi au niveau de l'établissement est d'éviter les risques de multiplication de dispositifs d'autorité divergents. Il a donc été décidé d'intégrer la méthodologie de création des instances EAC-CPF pour les producteurs des archives institutionnelles de la BnF dans la stratégie générale de création de fichiers d'autorité de l'établissement. Dans ce but, un mapping entre les éléments de données en MARC et en EAC-CPF va être réalisé pour assurer la cohérence entre les deux dispositifs et éviter les divergences. Pour une même entité, l'identifiant pérenne, dont les notices en MARC sont équipées, va être renseigné dans l'instance correspondante en EAC-CPF. D'autres éléments, comme la description de l'histoire de l'entité, peuvent avoir un traitement différent dans les deux dispositifs et jouer un rôle complémentaire.

People Australia, par Kathy Wisser (en remplacement de Basil Dewhurst de la NLA - Bibliothèque nationale d'Australie)

People Australia est un service en ligne créé par la Bibliothèque nationale d'Australie et dont l'objectif stratégique est de permettre un accès facile non seulement aux informations biographiques et contextuelles sur les personnes et collectivités importantes australiennes, mais aussi aux collections et aux ressources les concernant. L'accès se fait via l'interface « Trove People and Organisations » <<http://trove.nla.gov.au/people>>. Le service sera basé sur le fichier d'autorité des noms australiens (Australian Name Authority File - ANAF), maintenu depuis 1980 par les bibliothèques australiennes

qui alimentent la base bibliographique nationale australienne. ANAF est utilisé dans ce cadre pour fédérer également d'autres informations sur les personnes et les collectivités australiennes provenant d'autres institutions dans d'autres domaines. People Australia favorise le partage des données permettant de développer de nouveaux services en ligne, minimise la duplication du travail de création d'information. Le service est développé dans le respect du droit de la propriété intellectuelle et des contextes respectifs des partenaires dans le projet.

People Australia accepte des données en formats MARC et Dublin Core, mais le schéma EAC-CPF constitue le modèle de données du service. Il a été choisi pour sa richesse, sa capacité d'ingérer les différents types d'informations concernant les personnes et les collectivités australiennes provenant de différentes sources et sa capacité de gérer les relations entre les entités elles mêmes et entre les entités et les ressources les concernant. Les différents types de données peuvent être mappées sans perte dans EAC-CPF, et People Australia assistera les différents contributeurs pour développer des mappings de leurs formats vers EAC-CPF. Le concept d'identité, au cœur du schéma EAC-CPF, a été utilisé par People Australia pour réaliser des « notices d'identité » portant sur des personnes ou des collectivités. Une « notice contenante » d'identité en EAC-CPF sera initialement créée à partir de chaque notice d'autorité d'ANAF. Les notices des autres contributeurs, après conversion en EAC-CPF, seront confrontées avec les notices ainsi créées et les informations seront fédérées dans la « notice contenante ». Bien entendu, pour les noms qui ne sont pas présents dans ANAF, une nouvelle notice contenante sera créée considérant les informations du contributeur comme source primaire pour la personne ou la collectivité en question.

Pour plus d'information sur People Australia, voir la page du projet sur le site de la NLA, à l'adresse : <http://www.nla.gov.au/initiatives/peopleaustralia/>, ainsi que le Wiki consacré à People Australia, à l'adresse : <https://wiki.nla.gov.au/display/peau/Home;jsessionid=C1088B6B2AD285DE6D9EF2BAE0B4EE8C>

NACO et VIAF, par Ana Cristán (Bibliothèque du Congrès)

Tout d'abord, Ana Cristán a dressé le panorama des projets coopératifs menés par la Bibliothèque du Congrès et dont l'histoire remonte au début du XX^e siècle. Ils ont abouti à l'adoption à large échelle des AACR2 (Règles anglo-américaines de catalogage) et du format MARC, actuellement MARC21. La Bibliothèque du Congrès a lancé et mis en place progressivement plusieurs programmes de coopération dont les plus substantiels sont réunis sous le chapeau PCC (« Programme de coopération de catalogage », en anglais Program Cooperative Cataloguing). Pour en savoir plus, consulter les informations sur ces programmes sur le site de la Bibliothèque du Congrès (<http://www.loc.gov/catdir/pcc>). Un autre facteur important dans le développement de ces programmes a été le partenariat de longue durée avec l'OCLC (Online Computer Library Center, 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). Ce dernier fournit les infrastructures et les moyens logiciels pour l'échange, la fourniture et la diffusion des notices dans le réseau des partenaires.

Le plus important des programmes est NACO (« Programme de coopération des autorités de noms », en anglais Name Authorities Cooperative Program), dont le succès a été dû au partage de caractéristiques communes des données fournies par les membres : mêmes règles de catalogage (AACR2), même format (MARC21), utilisation de la même langue de catalogage (l'anglais, étant donné que les bibliothèques membres appartiennent aux pays anglophones : Australie, Canada, Royaume-Uni, Nouvelle-Zélande, Afrique du Sud). De plus, tous les partenaires sont adhérents d'OCLC. Cette situation a en même temps des inconvénients, notamment la difficulté de faire participer dans le programme des bibliothèques qui ne sont pas adhérentes d'OCLC et qui pratiquent d'autres langues que l'anglais et pratiquent d'autres règles que les AACR2. Coûteux à l'origine, ce programme est devenu rentable pour les partenaires grâce aux évolutions technologiques et aux souplesses introduites pour satisfaire les besoins des partenaires.

Le lancement du projet VIAF (<http://www.viaf.org>) en 2003 visait à faciliter le partage des données d'autorité, la réduction des coûts de catalogage, la simplification du contrôle d'autorité (création et maintenance) dans un environnement international et la création d'un dispositif de fourniture de données d'autorités dans des langues, écritures et formes appropriées aux besoins des utilisateurs divers.

Le projet a démarré en 2003 comme partenariat entre la Bibliothèque du Congrès, l'OCLC et la Bibliothèque nationale allemande (Deutsche National Bibliothek). La Bibliothèque nationale de France est entrée dans ce partenariat en 2007. D'autres participants sont entrés au fur et à mesure – actuellement ils sont au nombre de 16 (20 fichiers d'autorité avec environ 13 millions de noms pour environ 10 millions d'identités et 4,5 millions de grappes, en plusieurs langues et plusieurs écritures). Depuis septembre 2009, VIAF est disponible comme Données liées (Linked Data) sur le web (URIs, données en RDF, des liens internes et externes riches). Les scripts sont en Unicode et les données peuvent être soumises en UNIMARC, MARC21 ou MARCXML. En plus des noms de personnes, le projet est en cours d'extension aux noms géographiques.

La puissance de VIAF réside dans l'utilisation non seulement de données provenant des fichiers d'autorité de tous les participants au projet, mais aussi dans l'exploitation (data mining) des bases bibliographiques liées à ces fichiers. Ainsi, pour tout nom de personne utilisé dans les notices bibliographiques, une notice d'autorité est générée comportant des informations sur les documents publiés relatifs à la personne en question. Toutes les notices ainsi générées pour une même personne seront réunies dans une même grappe. Une notice d'autorité enrichie est générée pour une personne. Pour cela, la notice d'autorité d'origine est confrontée à toutes les autres notices d'autorité fournies par les partenaires. La notice d'autorité enrichie comporte à la fois des informations issues des fichiers d'autorité et des informations issues des notices bibliographiques.

Les informations extraites des notices bibliographiques pour être injectées dans la notice d'autorité enrichie sont les suivantes : variantes de noms d'auteurs trouvées aussi bien dans les zones des points d'accès que dans les mentions de responsabilité, indice de classification de la Bibliothèque du Congrès, titre de la manifestation (au sens du modèle FRBR : « Fonctionnalités requises pour les notices bibliographiques »), type de document, noms des éditeurs, lieux et dates de publication, langue, sujets, noms de personnes associées.

VIAF est enrichi avec de multiples fonctionnalités donnant une information multi-facettes sur une personne : information sur la personne, ses identités publiques, variantes de noms (dans des langues et des écritures différentes), publication des œuvres de et sur la personne, vue cartographique sur les territoires et lieux de publication, ligne de temps, liens externes (Wikipedia ; WorldCat Identities, base de données d'autorité en ligne de l'OCLC).

L'information issue des notices bibliographiques est stockée dans des zones Unimarc de données locales 9XX. Il va sans dire qu'à partir de la notice d'autorité enrichie on pointe directement vers les fichiers d'autorité source des bibliothèques partenaires en ligne.

VIAF s'est imposé comme une source inestimable d'informations pour les bibliothèques en matière d'autorités. Par ailleurs, VIAF s'ouvre à d'autres utilisations du fait de son positionnement à la pointe des technologies du web et de l'exposition des données en tant que Linked Data.

Noms de personnes et de collectivités : ULAN, introduction et perspectives, par Patricia Harpring (Fondation Getty)

La présentation de Patricia Harpring peut être consultée à l'adresse suivante : http://www.getty.edu/research/conducting_research/vocabularies/

A l'Institut de recherche Getty à Los Angeles, la gestion de la liste ULAN (« Liste collective des noms d'artiste », en anglais Union List of Artist Names) s'inscrit dans le programme plus large de vocabulaire de la Fondation Getty (Getty Vocabulary Program), qui comprend :

- le thésaurus d'art et d'architecture (AAT) (34 000 notices et 13 1000 termes) ;
- le thésaurus Getty des noms géographiques (TGN) (895 000 notices ; 1 115 000 noms) ;
- la liste collective de noms d'artistes (ULAN) (127 000 notices et 375 000 noms) ;
- les autorités de noms d'objets culturels (CONA) (en cours d'élaboration).

Ces vocabulaires sont centrés sur les domaines de l'architecture et des arts visuels. Ils sont alimentés par toute la communauté de leurs utilisateurs, à savoir les catalogueurs ou indexeurs décrivant des œuvres d'art, des documents d'archives, des substituts visuels pour des œuvres d'art ou des

ressources bibliographiques, mais aussi les chercheurs et les développeurs de systèmes d'information qui les exploitent pour enrichir les possibilités d'accès aux ressources en ligne.

Les vocabulaires de la Getty respectent les standards nationaux et internationaux de construction de thésaurus et les spécialistes de la Getty participent à l'élaboration de ces standards. Un thésaurus se présente comme un réseau de concepts uniques ; il peut être monolingue ou multilingue ; les termes sont mis en relation de type : d'équivalence, hiérarchique et associative. ULAN est également un thésaurus bien que souvent présenté sous forme de liste. Les autorités de la Bibliothèque du Congrès sont également utilisées comme source lorsque c'est approprié. Un mapping de différents formats et schémas de métadonnées standard sert à faire correspondre entre elles les sources utilisées.

Les termes de vocabulaires contrôlés de la Getty sont dérivés des sources publiées et sont représentatives de la recherche et des usages en matière de l'histoire de l'art. Les institutions qui contribuent à l'enrichissement de ces vocabulaires versent les données en masse dans les formats XML prescrits par la Getty ou en utilisant des formulaires en ligne. Les versements en masse vont de quelques milliers à quelques millions d'enregistrements. Les données sont ensuite vérifiées, enrichies, fusionnées avec les données existantes, intégrées dans les hiérarchies et autres types de relations et publiées dans le Getty Vocabulary Program. Les données sur le site web sont mises à jour toutes les deux semaines. Parmi les contributeurs actuels, on compte des musées, des bibliothèques, des archives et d'autres projets bibliographiques et documentaires.

Les vocabulaires contrôlés à la Getty sont gérés par une petite équipe hautement qualifiée. Cette équipe élabore et édite des outils de travail et assure la formation des contributeurs.

ULAN couvre toutes les périodes, de l'Antiquité à nos jours. Il traite des personnes, collectivités, familles impliquées dans la conception ou la réalisation des œuvres visuelles. Il est à noter qu'ULAN traite les familles comme des collectivités, à savoir comme groupes de personnes. Les informations suivantes sont données pour chaque entité : la forme préférée du nom, les variantes de formes, les sources, la biographie, les rôles, la nationalité, le sexe, les dates de naissance et de mort, ainsi que la position hiérarchique pour les collectivités. Une réflexion assez poussée est menée autour de la nature et de la forme des noms en fonction des entités désignées ainsi qu'autour de la nature et de la forme de chacune des informations pouvant caractériser une personne ou une collectivité. Dans ce sens, les règles de la Getty reflètent les besoins des domaines de l'architecture et des arts visuels.

74^e Congrès annuel de la Société des archivistes américains, Washington, 10-15 août 2010

La Société des archivistes américains (SAA) a tenu son 74^e congrès à Washington DC, du 10 au 15 août 2010. Comme les autres années, cette manifestation réunissait trois organisations : la SAA, le COSA (Conseil des archivistes d'État) et la NAGARA (Association nationale des archives gouvernementales et des gestionnaires de documents) autour du thème « Archives et records ». Le congrès a été caractérisé par une grande variété de sujets allant des instruments de recherche structurés et de leur visibilité sur le web jusqu'au métier d'archiviste aujourd'hui. L'EAD a fait l'objet d'une session plénière, elle a également été traitée lors de tables rondes et de réunions spécifiques.

Le programme complet du congrès se trouve sur le site de la SAA, à l'adresse <http://www2.archivists.org/conference/2010/washington>

Les supports de quelques-unes des interventions y sont déjà disponibles, d'autres seront mis en ligne au fur et à mesure.

Les archives twitter se trouvent à l'adresse <http://twrapperkeeper.com/hashtag/saa10>

Différents blogs ont également assuré une couverture sur un certain nombre de sujets traités (<http://larchivista.blogspot.com/>, <http://inkdroid.org/>).

Sessions plénières

Session 102 – « Des données structurées sont essentielles pour l'efficacité de la description et la découverte des ressources archivistiques : vrai ou faux ? »

Trois décennies après l'implémentation des premières données structurées, quelques archivistes continuent à se poser des questions sur leur utilité en argumentant que les systèmes d'information

n'exploitent pas leur potentiel, ce qui ne justifie pas leur coût. Les interventions de cette session s'efforcent d'apporter des arguments en faveur de la nécessité des données structurées en mettant l'accent sur l'émergence d'une nouvelle génération d'outils qui permettent de tirer profit de ces données.

Dans son intervention « Pourquoi fait-on cela à nous-mêmes ? Structurer dans un monde non structuré », **Michael Fox (Société historique du Minnesota)** a mis l'accent sur l'effort des professionnels des archives pour l'élaboration des standards de structuration de données. Cet effort avait pour but de fournir à la communauté des archivistes des outils adaptés à leur description. La découverte et l'identification des ressources deviennent de plus en plus des objectifs stratégiques des institutions de conservation des ressources historiques. Si ces ressources sont complexes, leur description sous forme électronique doit rendre compte de cette complexité. Les données structurées obligent les archivistes à réfléchir sur l'apport intellectuel du métier. Les instruments de recherche contiennent des données caractérisées et non pas simplement du texte. C'est leur organisation structurée qui facilite ensuite la localisation et l'identification des ressources souhaitées et permet aux utilisateurs d'effectuer des recherches précises en exploitant la structure qui respecte la nature et le sens des données et les met en contexte.

A l'heure actuelle, l'accès aux ressources est exponentiel, mais il ne faut pas oublier que les instruments de recherche font partie de l'économie du web et ne sont pas isolés des autres moyens d'accès aux ressources. Dans ce contexte, réfléchir sur le sens des données pour les structurer correctement charge les catalogueurs d'une plus grande responsabilité pour mieux satisfaire les besoins de la recherche. Chercheurs et catalogueurs doivent être amenés à réfléchir sur toutes les utilisations possibles des métadonnées structurées. Il est évident que plus la structuration des données est précise, meilleur sera le taux de satisfaction dans la découverte des ressources. Cependant, les archives se posent rarement la question sur le rapport coût – bénéfice dans la mise en œuvre des standards de métadonnées et il est temps que cette réflexion soit menée.

Dans son intervention « Plus que simplement lier : intégrer MARC et EAD dans une seule interface de découverte à l'Université de Duke de l'État de Caroline », **Noah G. Huffman (Université de Duke)** a présenté un projet sponsorisé par l'Université de Duke portant sur la réalisation d'une unique interface de recherche. Le projet utilise la couche découverte du moteur de recherche Endeca pour offrir une interface de type catalogue de nouvelle génération offrant un accès par facettes, la navigation par onglets et la suggestion de termes de recherche.

Alors que MARC offre des descriptions concises et très structurées, sa dépendance aux OPAC actuels (catalogues collectifs) limite ses utilisations par les chercheurs. EAD est plus souple, offre des descriptions plus complètes et permet même de rechercher et de trouver les instruments de recherche via Google.

L'objectif de l'interface de recherche réalisée est triple : permettre la découverte des contenus encodés en EAD en même temps que des autres contenus de la bibliothèque ; offrir une recherche plein texte et afficher des instances complètes EAD dans le catalogue ; tirer profit des fonctionnalités « nouvelle génération » d'Endeca. Les résultats atteints sont confirmés par les statistiques recueillies avec Google analytics de février 2009 à juillet 2010, qui montrent une augmentation de trafic de l'ordre de 475% de l'Endeca vers les instruments de recherche individuels.

Dans sa communication « Vrai (et faux) : les données structurées sont nécessaires pour la description et la découverte des ressources archivistiques », **Barbara Aikens (Institution smithsonienne de Washington, Archives de l'Art américain)** a proposé une réflexion et une prise de distance sur la valeur ajoutée apportée par les données structurées. Barbara Aikens a débuté son intervention en posant différentes questions. Comment faire pour rendre nos données plus intéressantes et structurées ? Les études d'usage des données font également émerger d'autres interrogations. Les utilisateurs sont-ils perturbés par la structuration des données et par les présentations offertes ? Doit-on permettre le balisage des données par les utilisateurs ?

Pour donner des réponses à ces questions, Barbara Aikens a comparé les instruments de recherche en EAD de la première génération, conçus à l'image des instruments de recherche papier, dont les fonctionnalités de recherche et de mise en contexte étaient limitées, avec ceux de la deuxième génération qui mettent l'accent sur la valeur des contenus améliorés, fiables et d'autorité et leur exploitation. Elle a illustré son propos par l'exemple des Archives de l'art américain qui offrent une

présentation par producteurs d'archives et remarque que la publication de l'EAC-CPF ouvre de nouvelles perspectives dans ce sens. De même, la possibilité de visualiser en même temps les documents numérisés disponibles est un autre pas en avant vers ce qu'elle appelle « redéfinir le but des métadonnées » (repurpose), c'est-à-dire produire du sens et du contexte. Elle a conclu en remarquant que les archivistes sont seulement aux débuts de la découverte du potentiel qu'offrent les capacités d'encodage du XML pour augmenter les possibilités de recherche.

Session 206 – « L'EAD est-elle trop complexe ? Abattre les barrières pour la mise en œuvre de l'EAD »

Une forte communauté de pratiques s'est développée autour de l'EAD accompagnée d'une masse impressionnante de littérature et d'outils. Toutefois, plusieurs institutions affrontent encore les barrières d'ordre tant philosophique que technologique ou financier pour l'implémentation de l'EAD. Les intervenants ont évoqué ces barrières en mettant l'accent sur des solutions pratiques pour contourner les obstacles.

Merrilee Proffitt (OCLC Research) a présenté un rapport publié récemment par OCLC Research, intitulé « Par-dessus, par-dessous, autour et à travers : un rapport d'OCLC Research sur le franchissement des barrières à la mise en œuvre de l'EAD (présentation consultable à l'adresse : <http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/store/PDFs/conf/dc2010/Session206-Proffitt.pdf>). Le rapport lui-même est accessible en ligne à l'adresse : <http://www.oclc.org/research/publications/library/2010/2010-04.pdf>

Le rapport porte sur un certain nombre de difficultés rencontrées par les archivistes dans l'adoption de l'EAD et propose des solutions pour la mise en valeur de l'EAD comme élément clé de systèmes d'information archivistiques et pour franchir les barrières potentielles dans sa mise en œuvre. Le rapport a le mérite de traiter le problème tant du point de vue politique et logistique (comment gagner le soutien des décisionnaires, dépasser le perfectionnisme, trouver les moyens de maximiser les ressources, franchir les premiers obstacles dans la mise en place d'un standard complexe) que technique (comment réaliser des conversions sans perte d'informations, sélectionner un logiciel, en particulier open source, publier et relativiser la complexité du standard).

Le rapport énonce différents arguments pour « vendre » l'EAD : c'est un standard international ; il peut être articulé avec d'autres standards de métadonnées déjà en place (mappings existants) ; il peut être utilisé comme outil de production mais également comme syntaxe de transfert de données ; il facilite l'agrégation des données, de plus en plus d'outils supportant l'implémentation de l'EAD et une forte communauté promouvant la formation et la collaboration autour du standard ; c'est bénéfique pour les chercheurs ; les projets EAD sont vus favorablement par les organismes de financement ; la connaissance acquise lors de la mise en place de projets EAD est réutilisable pour d'autres projets ; l'EAD ouvre la voie vers le Web sémantique et... tout le monde le fait !

Merrilee Proffitt a choisi de mettre l'accent sur certains points du rapport, notamment : le rôle des consortia, les possibilités de contourner la complexité de l'EAD entre autres en réutilisant des outils existants, l'utilisation des méthodes itératives et la gamme des outils, etc.

La complexité de l'EAD peut être contournée en faisant des choix d'application et d'élaboration du projet par étapes, par exemple en réalisant des descriptions au niveau de la collection ou des descriptions minimales et ensuite, en les complétant, en fonction des moyens. L'utilisation des modèles préétablis, des feuilles de style, etc. sont autant de moyens de facilitation de la mise en place d'un projet.

L'intérêt du rapport réside également dans les nombreux appendices très riches donnant de l'information concrète et adaptée aux besoins pour les projets de mise en place de l'EAD.

Dans son intervention intitulée « Réduire les obstacles de mise en œuvre de l'EAD : l'approche consortium » (présentation consultable à : <http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/store/PDFs/conf/dc2010/Session206-Gaidmore.pdf>)

Jay Gaidmore (Université de la Caroline du Nord à Chapel Hill) est revenu sur l'intérêt des consortia dans la mise en place de projets d'implémentation de l'EAD. Les consortia mobilisent des expériences, des talents et des ressources collectives et en font bénéficier un projet. Ils peuvent fournir une infrastructure logistique, des moyens techniques et peuvent offrir des possibilités de financement. Les problèmes de ressources humaines peuvent également être résolus par des

moyens souvent impossibles à mobiliser par une seule institution, par exemple en engageant un chef de projet (payé par subventions, employé par l'institution leader du projet, travaillant pour l'ensemble du consortium et tirant des bénéfices directs pour l'institution qui met en place le projet), en mobilisant des assistants étudiants, ou en tirant profit des membres du consortium reconnus pour leurs compétences dans tel ou tels aspect du projet.

Dans son intervention intitulée « Défis et opportunités dans la mise en œuvre de l'EAD dans les Archives de L'État de New York » (présentation consultable à : <http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/store/PDFs/conf/dc2010/Session206-Backman.pdf>), **Prudence Backman (Archives de l'État de New York)** a commencé par rappeler que les Archives de l'État de New York gèrent une large quantité d'enregistrements. elles jouent le rôle d'un consortium fédérant la communauté des « Archives historiques de New York » et réunissent dans une sorte de catalogue collectif divers instruments de recherche (<http://iarchives.nysed.gov/xtf/search>). Les défis à relever sont multiples, notamment collecter et générer des données archivistiques des collections, retravailler les données pour les adapter à des outils différents, publier dans une variété de formats tout en assurant la robustesse des données. les outils open source Archivists Toolkit et Archon (voir ci-après) sont utilisés, mais ils sont de type PHP et présentent des problèmes d'indexation. Du point de vue de l'administration des données, la structure EAD doit être appliquée à une variété d'instruments de recherche (Archives de l'État de New York, documents d'archives historiques dans l'État de New York), des outils faciles à utiliser par les non professionnels doivent être réalisés et le workflow doit être ajusté et étendu à d'autres composants de gestion. Pour l'interface publique, des images numérisées et des traductions sont à intégrer mais aussi des descriptions des producteurs d'archives en EAC. Du point de vue technique se pose la question de l'articulation avec les enregistrements en MARC avec tous les problèmes d'accroissement des données et leur intégration incrémentielle. Les listes d'unités de conditionnement posent la question de niveaux de détail de traitement pour répondre à différents besoins : pour les professionnels ou pour le public. Dans ce contexte, des priorités sont à fixer en fonction des questions suivantes. Quel niveau de perfection pour les instruments de recherche ? Quels utilisateurs viser : les avertis ou penser à d'autres utilisateurs potentiels ?

Session 502 – « Pas sur Google ? Cela n'existe pas : l'optimisation des moteurs de recherche pour les archives »

La recherche commence souvent par des moteurs comme Google ou Bing. Pour apparaître dans les résultats de recherche, les données des inventaires et des catalogues doivent être optimisées dans leur affichage et mises à disposition pour l'indexation externe. La conception des sites web devrait s'attacher aussi bien à l'accès au contenu à partir des moteurs de recherche qu'à l'accès des visiteurs aux pages d'accueil de nos sites.

Dans sa communication intitulée « Construire des sites web archivistiques que Google aimera », **Jeanne Kramer-Smyth** a commencé par présenter les concepts généraux de « l'optimisation pour les moteurs de recherche » (en anglais, Search engine optimization ou SEO). Il s'agit d'un ensemble de techniques visant à favoriser la compréhension de la thématique et du contenu d'une ou de l'ensemble des pages d'un site Web par les moteurs de recherche. Ces techniques visent donc à apporter un maximum d'informations concernant le contenu d'une page web aux robots d'indexation des moteurs de recherche.

L'objectif de ce procédé est d'orienter le positionnement d'une page Web dans les résultats de recherche des moteurs sur des mots-clés correspondant aux thèmes principaux du site. On considère généralement que le positionnement d'un site est bon lorsqu'il est positionné (classé) dans l'une des dix premières réponses d'une recherche sur des mots-clés correspondant précisément à sa thématique.

La deuxième partie de la présentation donnait quelques conseils pratiques pour bien construire des sites web :

1. Un seul titre par page, de moins de 65 caractères.
2. Ce titre doit avoir un mot clé cible compréhensible.

3. Une URL par page.
4. Une méta description unique, de moins de 164 caractères.
5. Un texte riche en contenu - pas une grande image avec quelques mots, mais une grande (ou une petite) image avec un paragraphe d'explication.

Quelques éléments d'analyse :

1. Où les utilisateurs arrivent-ils ?
2. A quel moment et pourquoi les utilisateurs quittent-ils un site web ?
4. Où suis-je quand je fais une recherche Google ?

Le blog SEOmoz Daily (<http://www.seomoz.org/blog>) donne quelques astuces et recommandations pour optimiser les moteurs de recherche.

Les deux autres intervenants, **Matt Herbison (Musée du port de l'Indépendance)** et **Mark Matienzo (Université de Yale)**, ont discuté plus en détail des questions liées à l'architecture du site web, aux URLs et aux liens internes à un site web. Ainsi, le protocole Sitemaps permet à un webmestre d'indiquer aux moteurs de recherche quelles adresses d'un site web sont disponibles pour l'indexation automatique. Il s'agit d'un plan de site (« sitemap ») compréhensible par les robots d'indexation, rédigé sous forme d'un fichier XML ou texte qui répertorie les URL d'un site permettant ainsi d'inclure des informations complémentaires sur chaque adresse, comme sa date de dernière modification, la fréquence de mise à jour et son importance par rapport aux autres adresses du site.

Session 604 – « Le contrôle bibliographique des documents d'archives : l'impact des normes de bibliothèques sur les archives »

Les intervenants ont examiné comment les changements dans les normes des bibliothèques affectaient la description archivistique. Les évolutions récentes, notamment les « Ressources : description et accès » (RDA) et l'application par les RDA des « Fonctionnalités requises pour les notices bibliographiques » (FRBR), sont en train de modifier la description bibliographique des documents d'archives. Ces changements ont des impacts sur les systèmes d'information et auront des conséquences substantielles sur les catalogues de bibliothèques et la découverte des ressources. Ce sont autant de défis que d'opportunités pour les bases de données archivistiques.

La session a commencé par une intervention de **Cory Nimer (Université de Brigham Young)** intitulée « En testant les RDA : appliquer les normes de contenu du domaine des bibliothèques aux documents d'archives » (présentation consultable à l'adresse : <http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/store/PDFs/conf/dc2010/Session604-Nimer.ppt>).

Cory Nimer a d'abord présenté les fondements des RDA, à savoir la référence forte aux FRBR et à son pendant autorité, les Fonctionnalités requises des données d'autorité (FRAD), qui par ailleurs constituent les fondements des nouveaux Principes internationaux de catalogage.

Les RDA organisent les règles de description des ressources par rapport aux entités FRBR (œuvre, expression, manifestation, item, personne, collectivité, concept, objet, événement, lieu, ainsi que l'entité famille ajoutée de FRAD). Chacune de ces entités est caractérisée par ses attributs. Les RDA mettent l'accent sur l'exploitation des relations entre ces entités.

Il est à noter que la théorie et les pratiques de description archivistiques ont nourri la réflexion pour l'élaboration des RDA. On peut mentionner les normes suivantes qui ont servi à cet effet : Contexte Archivistique Encodé (EAC) beta, ISAD(G), ISAAR(CPF), « Décrire les archives : une norme de contenu » (DACS) et les Règles canadiennes pour la description des documents d'archives (RDDA).

Bien qu'une organisation des règles selon les FRBR puisse ne pas convenir pour la description archivistique et que les FRBR soient d'une application limitée pour les archives, des éléments RDA sont transposables dans la pratique de la description archivistique, notamment :

- pour la description de la Manifestation, les éléments titre propre, date de production, identifiant, type de contenant, étendue de la ressource ; pour l'œuvre : les éléments identifiant, titre préféré ; le lien

entre la manifestation et l'œuvre ; le lien avec des personnes, familles et collectivités associées ;

- pour la description des Personnes : nom, titres, dates, autres désignations, professions et occupations, domaine d'activité, identifiant ;
- pour les familles : nom, type de la famille, dates, identifiant ;
- pour les collectivités : nom, dates, autres désignations, identifiant ;
- pour les congrès : localisation, numéro.

D'autres principes adoptés par les RDA sont également applicables à la description archivistique: limitation des abréviations, importance des descriptions, règles indépendantes des formats, normalisation accrue des termes, définition d'éléments de base et adoption du principe des choix alternatifs et des options selon les besoins de l'agence de catalogage.

Cory Nimer a présenté plus en détail les changements apportés par les RDA dans le format MARC21 pour l'encodage des descriptions archivistiques.

Les RDA ont été publiées en juin dernier sur une application dédiée – RDA Toolkit (<http://www.rdatoolkit.org>). L'accès a été gratuit jusqu'à fin juillet pour que la communauté des utilisateurs puisse tester. Il sera accessible en ligne sur abonnement.

En conclusion, si les RDA portent l'influence des concepts, des pratiques et des règles archivistiques, c'est une raison supplémentaire pour que les archivistes s'y intéressent et participent à leur évolution.

Dans sa présentation intitulée « Les documents d'archives et Worldcat : mélange de métaphores dans un catalogue de type nouvelle génération », l'intervenant suivant, **John W. Chapman (OCLC)** a commencé par rappeler qu'OCLC avait participé activement au processus d'élaboration des RDA et qu'il était représenté dans les groupes de travail sur les différents sujets liés à l'application des RDA. OCLC a publié une déclaration sur la politique de l'application des RDA dans le catalogue Worldcat (voir les pages consacrées aux RDA sur le site de OCLC : <http://www.oclc.org/rda/about.htm>).

Pour l'OCLC, la mission de description et d'accès aux ressources s'inscrit dans une démarche d'innovation incessante. Les RDA ont révolutionné la conception de l'organisation de l'information de l'ancien modèle où l'organisation des ressources dans le catalogue se faisait par grands types (ressources numériques, imprimées, etc.) vers un nouveau modèle orienté vers les relations entre les objets, où les identifiants jouent un rôle prééminent et où les métadonnées sont considérées comme des déclarations à exploiter. Le but de ce nouveau type de catalogue est de donner le contexte dans lequel les ressources évoluent pour que les institutions puissent s'appuyer de manière fiable sur les descriptions fournies. Le nouveau paradigme organisationnel est de type « sujet -> prédicat -> objet » où l'on s'efforce d'identifier les sujets et les objets informationnels et de les relier entre eux en précisant la nature de l'information. Les FRBR sont le fondement d'une telle conception. On s'appuie notamment sur les relations définies dans les FRBR et sur les « Tâches des utilisateurs » permettant aux utilisateurs d'effectuer différentes opérations de recherche. Les déclarations portent : sur les objets (contenants), sur les contenus, sur la différence entre le contenant et le contenu, sur la portée de la ressource (coverage), sur la propriété de la ressource et sur l'unicité de la ressource.

Concernant les ressources archivistiques, le catalogue Worldcat s'efforce de répondre aux questions suivantes. Quels sont les documents conservés ? Qui les conserve ? OCLC traite les données pour proposer différents services. Un de ces services proposés est Worldcat Identities qui a été développé en s'appuyant sur un algorithme centré sur l'œuvre (au sens FRBR du terme). Toujours en exploitant les relations entre les ressources, ce service permet de découvrir l'univers des œuvres, publications, archives, sujets, etc., associé à une « identité, le cas échéant une personne.

Dans sa démarche, OCLC vise à mélanger les métaphores et à développer des produits personnalisés. Par exemple, pour les documents soumis à des droits, un objectif serait de les grouper autour de la notion du possesseur. Mais à côté, les documents libres de droits ou les documents uniques sont également traités pour assurer et accroître la diversité. Qui ? Par qui ? Quoi ? Quand ? Où ? Comment ? Ce sont les questions qui nourrissent le paradigme de la diversité et de la discontinuité cultivées par OCLC. Les métadonnées dépendent de la communauté qui les produit. Il appartient à OCLC de mobiliser la matière grise pour élaborer des algorithmes visant à réaliser des services adaptés aux besoins. La solution pour cela est la gestion à plusieurs couches en tenant compte à chaque fois du contexte, de la communauté qui a produit les données et de celle qui les

cherche. Pour cela, les approches sont combinées en vue d'une plus grande efficacité.

Pour **Robin McElheny (Archives de l'Université de Harvard)** (« Les instruments de recherche dans le nouvel OPAC de l'Université : MARC, EAD, recherche intégrée et satisfaction des utilisateurs »), la différence entre les documents d'archives et les documents traditionnellement conservés dans les bibliothèques est un vaste sujet de réflexion. Les différences entre les types de documents ont conduit à des normes métier de description distinctes. Mais les frontières entre les deux métiers ne sont pas étanches et les archivistes s'inspirent des normes et des pratiques des bibliothécaires tant en matière de rigueur de traitement qu'en matière d'outils réalisés. Par exemple le domaine des vocabulaires contrôlés est un des champs d'investigation où les deux métiers se rencontrent. Les normes de description des bibliothèques ont également exercé un impact important sur les normes de description archivistique. On voit clairement cet impact dans des normes comme « Décrire les archives : une norme de contenu » (DACS), en ce qui concerne l'aspect traitement bibliographique de l'information. Cet impact se traduit également dans la conception des outils destinés au traitement archivistique comme Archivists' Toolkit et Archon (voir ci-dessous). Il est important de noter que les normes des deux métiers visent à satisfaire les utilisateurs.

Comme dans d'autres institutions, l'Université de Harvard a fait évoluer les outils de recherche mis à disposition des utilisateurs. Autrefois, les recherches ne pouvaient s'effectuer que directement sur place, à l'aide des index et des inventaires. La fin des années 1990 a marqué une évolution importante ; les utilisateurs se voient offrir des OPAC, des sites web contenant des informations pour la recherche et des inventaires en EAD. Mais actuellement, l'heure est à l'accroissement des possibilités de recherche tout en ne dispersant pas l'utilisateur. L'augmentation du « trafic » pousse à intégrer les instruments de recherche dans les OPAC. D'où l'importance de la convergence entre les normes de bibliothèques et les normes archivistiques pour une augmentation des possibilités de recherche et d'accès aux ressources. Le défi reste comment tenir compte de la diversité des standards - diversité qui est justifiée par la diversité des contenus traités, comme par exemple la norme VRA Core (norme de description des ressources visuelles ; <http://www.vraweb.org/projects/vracore4/>) ou le Cataloguing Cultural Objects (guide pour décrire les objets culturels ; <http://www.vrafoundation.org/ccoweb/index.htm>) – tout en facilitant l'accès aux ressources pour l'utilisateur.

Les nouveaux standards comme RDA et EAC-CPF relèvent justement ce défi en aidant à « déverrouiller » les collections et les notices des catalogues des bibliothèques.

Commissions et groupes de travail spécialisés

Table ronde des utilisateurs du logiciel « Boîte à outils des archivistes » (Archivists' Toolkit™ ou AT)

Cette Table ronde est un lieu de discussion et d'échange des utilisateurs du logiciel « Boîte à outils des archivistes » (Archivists' Toolkit™ ou AT), développé 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). En juillet dernier, il a été proposé de modifier la mission de cette table ronde afin qu'elle puisse servir de forum commun à la fois pour la communauté des utilisateurs du logiciel AT et celle des utilisateurs du logiciel Archon, développé par l'université de l'Illinois. Par ailleurs, un projet, financé par la Fondation Andrew W. Mellon, vise à développer à partir de 2011 un nouvel outil, ArchivesSpace, destiné à remplacer les logiciels AT et Archon. Cette fusion de communautés d'utilisateurs est supposée accroître l'implication des utilisateurs dans le développement des outils, promouvoir le partage de l'expertise entre eux et faciliter la collaboration entre les utilisateurs d'AT, ceux d'Archon et l'équipe de développement de ArchivesSpace.

La proposition de modification de mission est publiée sur le site de la SAA : <http://www2.archivists.org/groups/archivists-toolkit-roundtable/proposal-to-revise-the-roundtables-mission-goals-and-name-draft>

ArchivesSpace sera orienté vers les services. Une collaboration entre ArchivesSpace et le logiciel ICA-AtoM en cours de développement pour le Conseil international des archives est souhaitée.

Jackie Dooley de l'OCLC a ensuite annoncé la publication prochaine des résultats d'une enquête sur les départements des collections spéciales des bibliothèques universitaires et de recherche. Quelques

conclusions intéressantes :

- sur les 275 établissements interrogés, environ le tiers utilise AT, 11% utilisent Archon ;
- 70% ont utilisé l'EAD pour leurs instruments de recherche ;
- environ 75% utilisent des logiciels de traitement de texte pour leurs instruments de recherche ;
- moins de 50% des instruments de recherche de ces institutions sont en ligne.

La séance a enfin vu des présentations d'utilisateurs du logiciel, notamment celle de Nancy Enneking du Getty, relative à la production, à partir d'AT, de statistiques sur les acquisitions et les opérations de traitement des archives, ou encore celle de Corey Nimer de la Bibliothèque universitaire de New York, sur l'utilisation de plug-in complémentaires de AT. Marissa Hudspeth de la Fondation Rockefeller et Sibyl Shaefer de l'Université du Vermont ont quant à elles présenté un module permettant l'enregistrement des lecteurs, le suivi de leurs demandes de communication et de reproduction, la gestion des comptes personnels d'utilisateurs, etc. Bien qu'il y ait beaucoup de débats dans la communauté des archives pour savoir si cela correspond à une bonne utilisation de AT (le logiciel a été initialement conçu pour la description et la gestion de contenu des archives), ce module devrait être disponible à l'automne 2010. Les développeurs d'AT feront prochainement une annonce officielle sur la liste de discussion du logiciel.

Table ronde EAD

Les présentations lors de cette table ronde ont porté sur l'ergonomie des instruments de recherche en EAD sur le web, sur le développement d'applications pour la gestion de volumes importants d'instruments de recherche et leur exposition sur le web, sur la recherche d'une nouvelle génération d'outils facilitant la production des instruments de recherche en EAD, etc.

Dans son intervention « Filtrer les listes de contenus », **Linda Sellars (Bibliothèque de l'Université d'État de la Caroline du Nord)** a mis l'accent sur la nécessité de la réalisation d'outils d'aide à la navigation dans les instruments de recherche qui donnent de l'information sous forme de longues listes de boîtes d'archives. Assurer une bonne ergonomie de navigation et une compréhension de la part de l'utilisateur de la logique de l'instrument de recherche passe par des questions comme : faut-il montrer les séries sans un résultat de recherche concret à la clé ? Comment rendre transparent pour l'utilisateur qu'un changement dynamique se fait lors du lancement de la recherche ? Faut-il laisser des fenêtres de recherche multiples ? Ces questions ont été illustrées à travers l'exemple de la navigation sans l'instrument de recherche des papiers de James Webb (http://webdev.lib.ncsu.edu/jnronall/ead_filter/mc00102.html).

Dans son intervention « Guides des collections de la bibliothèque Joiner », **Mark Custer (Université de la Caroline de l'Est)** a montré la publication d'instruments de recherche en ligne permettant de mettre en avant la visualisation des objets numériques associés aux instruments de recherche. L'application offre des possibilités d'affichage par facettes (<http://digital.lib.ecu.edu/special/ead/>).

L'intervention conjointe de **Mary Lacy (Bibliothèque du Congrès, Département des manuscrits)** et de **Michael Ferrando (Bibliothèque du Congrès, Systèmes des technologies de l'information)** portait sur une nouvelle application de découverte pour les instruments de recherche de la Bibliothèque du Congrès ». Cette nouvelle application, actuellement en beta test, permet une présentation plus claire et plus simple de l'information sur les collections. Construite en englobant des instances EAD dans METS, l'application stocke, indexe et affiche les instruments de recherche à partir d'une plateforme de stockage de données en XML natif. Il est à noter que la Bibliothèque du Congrès rend disponible en ligne plus de 1035 instruments de recherche, totalisant 31 millions de documents d'archives.

Les instances EAD sont transformées en schéma EAD XML au moyen d'un workflow Java et XSL et fusionnées avec les métadonnées descriptives en MARCXML dérivées des éléments EAD et des notices associées de niveau collection. Des identifiants uniques structMap sont assignés aux différents niveaux de séries et de sous-séries des instruments de recherche pour faciliter la subdivision logique de ces documents XML, souvent de taille très importante, pour la recherche et

l'affichage.

Les recherches par mot clé et par balayage sont construites en utilisant XQuery. Des listes de réponses proposent un accès transversal aux noms, titres, sujets, dates et bases de données. Sur les pages des instruments de recherche individuels, ces termes contrôlés sont reliés aux entrées parallèles dans « Tous les instruments de recherche », de même qu'aux vedettes du catalogue en ligne de la Bibliothèque du Congrès, ce sont les éléments <controlaccess> qui permettent le lien entre les instruments de recherche et le Catalogue de la Bibliothèque du Congrès en ligne.

Grâce à l'extraction des métadonnées des notices bibliographiques de collections et de celles de données locales, les utilisateurs sont alertés sur les restrictions d'accès et les conditions de préservation hors site. Les versions PDF des instruments de recherche de la Bibliothèque du Congrès générées au moyen de XSL FO continuent d'être moissonnées par Google et Yahoo et permettent la connexion directe au système de recherche des instruments de recherche par les identifiants handle et les permaliens LCCN (type d'URL conçu pour référer un élément d'information et pour rester inchangé de façon permanente, ou du moins, pour une certaine période de temps).

Ethan Gruber (Scholar's Lab, Université de Virginie) a présenté « EADitor : Xforms pour l'EAD » (<http://code.google.com/p/eaditor/>), une application pour la production, la vérification et la publication des collections d'instruments de recherche. Pour que les instruments de recherche puissent devenir de véritables outils de recherche utiles pour les usagers des archives, les collections archivistiques doivent être décrites au moyen de métadonnées robustes et bien structurées. Or, l'EAD est complexe et il n'est pas facile d'exiger de tous les archivistes de maîtriser l'encodage en XML. L'utilisation de XForms, un standard W3C, pour produire de l'XML via un formulaire web nouvelle génération, fait tomber les barrières et permet de produire des guides EAD tout en réduisant le taux d'erreurs humaines dans l'utilisation sémantique de l'XML.

Même si le formulaire web constitue l'aspect le plus important de l'application car intégrable dans les systèmes de gestion de diffusion de contenus, EADitor comprend également une interface publique, facilement personnalisable, pour la recherche des collections d'instruments de recherche. Cela crée la possibilité d'utiliser une même application pour la production et la publication des contenus. L'interface utilise Solr (serveur de recherche Open Source basé sur Lucene) pour la recherche par facettes. EADitor permet la gestion des vocabulaires contrôlés. Prochainement, il est prévu d'intégrer un module pour la gestion de l'EAC-CPF.

Randall Miles (Kheel Center for Labor management Documentation and Archives, Université de Cornell) a présenté EAD McTaggart, outil sous forme de base de données programmée avec Microsoft Access utilisant Virtual basic pour Applications (VBA) permettant d'insérer automatiquement les balises EAD et leurs attributs. Un fichier Excel contenant une liste d'environ 7000 lignes, où chaque ligne représente une série, une sous-série ou un titre de dossier, est converti en une liste de composants balisés en quelques minutes. EAD McTaggart peut traiter jusqu'à six niveaux hiérarchiques mais il y a la possibilité de le modifier pour en gérer plus. L'outil minimise ainsi le travail et le temps de conversion. La réalisation a fait l'objet d'un article consultable en ligne (<http://journal.code4lib.org/articles/2025>).

Enfin, **Elizabeth Perkes (Archives de l'État d'Utah)** a présenté un logiciel développé par les Archives de l'État d'Utah, Axaem, qui est sous licence GPL et que tout service d'archives peut donc utiliser. Cette application a été développée à partir d'un outil propriétaire, le logiciel APPX, nécessaire pour exécuter Axaem. Le logiciel Axaem permet de gérer des archives intermédiaires et définitives, de décrire des documents d'archives en utilisant la norme américaine DACS (Décrire les archives : une norme de contenu), de générer des calendriers de conservation (généraux et spécifiques), d'identifier les producteurs et les sujets des documents à travers les fichiers d'autorité; d'assurer le suivi des questions des utilisateurs, d'exporter des notices MARC, de produire des exports EAD et EAC. L'outil peut s'interfacer avec le site d'une institution.

Section sur la description

La réunion de la section sur la description a vu les comptes rendus d'activités de différentes instances et commissions spécialisées.

Parmi ces initiatives, on peut retenir celle d'un groupe de travail créé en 2007 au sein de la SAA, suite au rapport de deux archivistes, Mark Greene et Dennis Meissner. Intitulée « Plus de résultats, moins de procédures », cette étude donnait des recommandations pour réduire les arriérés de traitement des fonds dans les services d'archives. Lors du congrès annuel de la SAA en 2008, le groupe de travail avait fait le compte rendu d'une première enquête auprès de différentes institutions. Il en ressortait que les archivistes voulaient des conseils pratiques de mises en œuvre des recommandations de Greene et Meissner. Le souci de l'impact sur les services publics ne faisait que commencer à émerger. Après un an de discussions sur les différentes formes que pourraient prendre de telles orientations, le groupe a été reconstitué et ses objectifs ont été redéfinis l'hiver dernier.

Le groupe représente désormais les intérêts et les perspectives de différentes instances de la SAA : les sections « Institutions de conservation de manuscrits », « Description » et « Archives des collègues et universités », Description. Il a pour mission de faire la synthèse des échanges et des débats sur ce que doit être un traitement minimal des fonds d'archives, et de proposer des modèles pour mettre en œuvre ce traitement minimal et évaluer son impact. Il doit effectuer des enquêtes de suivi sur le traitement des fonds et les pratiques de référence tout en s'appuyant sur les informations déjà recueillies par Greene, Meissner et d'autres. Enfin, il doit rendre compte de ses travaux via une plateforme interactive permettant des commentaires et des ajouts par des archivistes et d'autres professionnels intéressés par le sujet.

Le groupe de travail a créé différentes pages de ressources répertoriant des articles et d'autres publications en ligne, des blogs et des présentations pour des conférences ; des liens vers les textes complets de ces ressources sont inclus lorsqu'ils sont disponibles. En outre, les manuels de classement et de description élaborés par les institutions représentées dans le groupe ont fait l'objet d'une compilation et d'annotations complémentaires. Les mises à jour concernant le déroulement et le contenu de ce travail sont disponibles à : <http://raonews.blogspot.com/>, dans les bulletins de la section sur la description, et dans des listes de discussion.

Le groupe étudie actuellement la faisabilité d'applications permettant de poster ses contenus en ligne et de recueillir des commentaires. Il espère finaliser ce processus au cours de l'automne et rendre disponibles d'ici la fin de l'année les informations ainsi recueillies et créées. Ce contenu constituera l'ossature à partir de laquelle d'autres ressources seront élaborées pour aider les archivistes à prendre en connaissance de cause des décisions sur la mise en œuvre des recommandations de Greene et Meissner. D'autres outils d'enquête seront créés et déployés début 2011 pour une restitution des résultats lors du 75^e congrès annuel de la SAA.

Un autre groupe, intitulé « Catalogage descriptif des documents rares (manuscrits) » (DCRM) a été constitué à la demande de la SAA par le Comité des normes bibliographiques de la Section des livres rares et des manuscrits de l'Association des bibliothèques d'enseignement supérieur et de recherche et de l'Association des bibliothèques américaines (ALA). En 2004, la SAA adoptait la norme « Décrire les archives : une norme de contenu » (DACS), qui se concentre sur la description des fonds et collections. Alors que DACS était mis en œuvre par les archivistes américains, le besoin d'une norme pour la description des manuscrits isolés s'est fait sentir. Après discussion, la SAA a demandé au Comité des normes bibliographiques de développer cette nouvelle norme, tout en reconnaissant que la description des manuscrits au niveau de la pièce établirait un pont entre la norme ISBD (Description bibliographique internationale normalisée), cadre de référence de la future norme de catalogage des manuscrits, et ISAD(G), qui a inspiré DACS. Le groupe DCRM est donc une équipe éditoriale chargée d'élaborer une suite de manuels, dont on espère qu'ils seront les « normes d'accompagnement » envisagées par DACS.

Pour en savoir plus, voir :

http://www.rbms.info/committees/bibliographic_standards/dcrm/dcrmmss/dcrmmss.html

Sous-comité technique EAD

Comme annoncé dans le précédent numéro, les groupes de travail internationaux hébergés par le Comité des normes de la SAA viennent d'être réorganisés, avec la mise en place d'un sous-comité technique pour l'EAD (TS-EAD), co-présidé par Michael Rush (Université de Yale) et William Stocking (British Library), et la constitution d'une équipe restreinte, chargée du développement et de la maintenance technique des formats d'échange de la SAA. Le descriptif du mandat et la liste des membres de ces différentes instances sont consultables sur le site de la SAA :

<http://www2.archivists.org/> puis Groups / Subcommittees.

La première réunion du TS-EAD dans sa nouvelle configuration s'est donc tenue le 15 août 2010. Il a été décidé de donner à la communauté des utilisateurs la possibilité de contribuer à la révision de l'EAD par des commentaires. Un calendrier prévisionnel a été ainsi établi :

- 30 septembre 2010 : lancement de l'appel officiel à commentaires (en s'inspirant de ce qui a été fait pour l'EAC-CPF) ; les membres « internationaux » du TS-EAD seront chargés de diffuser largement l'annonce ; pour éviter que des commentaires de toute nature arrivent au même niveau, un formulaire sera proposé, permettant d'identifier l'auteur des commentaires (individu ou personne morale) et son affiliation, un résumé de ses attentes et différentes rubriques concernant l'architecture de l'EAD, sa structure, des éléments ou des attributs spécifiques...
- 28 février 2011 : date limite pour l'envoi des commentaires ;
- 1^{er} mars – 15 août 2011 : traitement des commentaires et développement par l'équipe restreinte d'une première version du schéma XML ;
- printemps 2012 : réunion de travail du TS-EAD (sous réserve de financement)
- décembre 2012 : publication d'un premier projet pouvant être testé par la communauté des utilisateurs ;
- 1^{er} semestre 2013 : rédaction de la Tag Library à partir de la Tag Library de la DTD EAD 2002 ; une version imprimée sera publiée mais la publication en ligne restera la référence ;
- août 2013 : publication finale du schéma et de sa documentation.

Groupe EAC-CPF

La réunion du groupe EAC-CPF s'est tenue après celle du TS-EAD. Les deux groupes se réuniront ensemble lors des congrès annuels de la SAA. Toutefois, les deux structures resteront distinctes pour assurer une large représentativité.

La traduction en italien du Dictionnaire des balises est presque achevée. Elle a permis de relever des manques de clarté et d'erreurs dans la version d'origine, erreurs qui seront corrigées. La version italienne sera donc, de fait, un test pour les procédures de traduction telles que les envisage le groupe EAD.

Les travaux pour la traduction en français seront lancés assez rapidement par le groupe d'experts AFNOR. Une traduction en allemand devrait également être disponible prochainement.

Enfin, la bibliographie sur l'EAD sera enrichie avec des références d'articles et de publications sur l'EAC-CPF. Des liens seront faits entre les sites officiels de l'EAD et de l'EAC-CPF.

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