

Service interministériel des Archives de France

Bulletin sur les ressources archivistiques numériques, n° 42, avril-juin 2011

Sommaire

- [Réunion annuelle du Comité des normes et bonnes pratiques du Conseil international des archives, Bucarest, 23-26 mai 2011](#)
- [Journée d'étude Référentiels et données d'autorité à l'heure du Web sémantique, BnF, 27 mai 2011](#)
- [Le projet LOCAH \(Linked Open Copac Archives Hub\)](#)
- [La politique de normalisation en Espagne](#)

Réunion annuelle du Comité des normes et bonnes pratiques du Conseil international des archives, Bucarest, 23-26 mai 2011

Le Comité des normes et bonnes pratiques du Conseil international des archives (ICA/CBPS) s'est réuni à Bucarest, les 23-26 mai derniers.

L'essentiel des discussions a porté sur les objectifs du recueil des normes, visant à expliciter les relations entre entités archivistiques (archives, collectivités, personnes ou familles producteurs/détenteurs d'archives, fonctions exercées par les collectivités, personnes ou familles) et à poser les bases d'un modèle conceptuel pour les archives. Ce recueil est-il destiné à remplacer les quatre normes actuelles comme un nouvel outil normatif ? Ou s'agit-il simplement d'une nouvelle présentation éditoriale des normes avec quelques ajouts (zone commune du contrôle de la description, fusion des chapitres 6 d'ISAAR et d'ISDF sur les relations entre archives, collectivités, personnes ou familles, et fonctions) ? Le problème est que le degré d'utilisation des normes, élaborées à des dates différentes, est inégal. ISAD(G) est la plus connue, ISAAR(CPF) l'est beaucoup moins, les deux dernières normes, encore moins.

Le Comité a choisi de se limiter pour le moment à l'élaboration d'un chapitre commun sur les relations entre les différents types d'entités archivistiques et de préparer un rapport d'étape pour le Congrès international de Brisbane (Australie) en 2012. Ce rapport se concentrera sur les incohérences du contenu des normes dans leur version actuelle et tentera de définir les grandes orientations de leur future révision, comprenant le développement d'un modèle conceptuel archivistique.

Il a été décidé d'établir une distinction entre les relations entre entités archivistiques du même type (par exemple, un fonds et ses composantes, une collectivité, personne ou famille en relation avec une autre collectivité, personne ou famille, et une fonction reliée à une autre fonction) et les relations entre différents types d'entités archivistiques (par exemple entre une série organique et la fonction dont cette série est issue). A cela s'ajoutent les relations entre des notices d'autorité archivistiques, des descriptions de fonctions et des descriptions archivistiques et des descriptions d'autres types de ressources (bibliographiques, muséales, etc.).

Le Comité a identifié trois éléments de description pour établir des relations entre plusieurs entités archivistiques ou entre des entités archivistiques et d'autres ressources d'information : « Identifiant des entités et ressources en relation », « Nature de la relation » et « Dates de la relation ». Deux diagrammes ont été élaborés pour offrir une représentation graphique des relations entre les normes de description de l'ICA et les entités archivistiques représentées par leurs descriptions. Ces diagrammes ne contiennent pas toutes les relations possibles mais seulement celles pertinentes pour le recueil de normes. Le premier diagramme représente les relations entre les quatre normes de l'ICA. Le deuxième diagramme représente les relations entre entités archivistiques (il a été suggéré de regrouper les institutions de conservation avec les collectivités, personnes et familles). Les règles et objectifs de ces éléments seront illustrés par des exemples de relations entre archives, collectivités, personnes ou familles et fonctions.

Journée d'étude Référentiels et données d'autorité à l'heure du Web sémantique, BnF, 27 mai 2011

[=>Sommaire](#)

Le rôle des référentiels et vocabulaires contrôlés est de plus en plus déterminant dans la gestion comme dans la découverte des ressources ou dans la création des réseaux sémantiques. Piliers de l'intelligence du Web sémantique, ils sont exposés en Linked Data et utilisés comme pivots dans des projets et applications visant à faciliter la recherche et la mise en relation des ressources.

Il y a trois ans, la journée d'étude AFNOR CG46/BnF « Référentiels, données d'autorité, thésaurus, ontologies, taxonomies... Pour en savoir plus ! » incitait déjà à réfléchir au rôle des référentiels dans l'interopérabilité des systèmes d'information et dans le Web sémantique. Depuis, de nombreuses initiatives ont vu le jour, qui témoignent d'un environnement en perpétuelle évolution et qui offrent des perspectives enthousiasmantes.

Coordonnée par **Anila Angjeli** (Bibliothèque nationale de France ; présidente du groupe AFNOR « Données d'autorité » de la CN 357), cette journée s'est efforcée de porter un regard sur les travaux de normalisation, les réalisations et les usages se rapportant à ces systèmes d'organisation de connaissances et à l'ouverture des données.

La matinée a vu la présentation de trois tutoriels sur le Web sémantique.

La mise en place de l'approche *Linked Data* donne aux systèmes d'organisation des connaissances une visibilité inédite sur le Web. Il s'agit de mettre des données sur le Web de manière à permettre leur traitement direct par les machines. L'intérêt pour ces dispositifs, qui avait déjà été ravivé par les travaux autour du Web sémantique, se renouvelle, alors que l'avènement des moteurs de recherche avait passablement remis en cause leur pertinence. En introduction à cette journée, le premier intervenant, **Bruno Menon** (Laboratoire Paragraphe, Université Paris 8), est revenu sur la nature, la structure et la fonction des systèmes d'organisation des connaissances, ainsi que sur leur rôle dans divers univers informationnels.

Que recouvre exactement la notion d'organisation des connaissances ? Vocabulaires contrôlés, classifications et autres référentiels ou « schéma de concepts », à la suite d'autres outils plus modestes ou d'autres projets plus ambitieux, ont tous vocation à instrumenter la saisie de la connaissance. Par la négociation du sens, la normalisation de la forme et l'adressabilité des symboles, ils furent longtemps un remède plus ou moins utopique à la confusion de Babel. Le XVIIe siècle vit notamment apparaître des projets de langues universelles ou « langues philosophiques a priori » fondées sur la critique du langage. Les référentiels représentent aujourd'hui un gage d'interopérabilité des systèmes en assurant la médiation entre l'ordre foisonnant et trop humain de la langue et l'ordre machinique des algorithmes pour l'accès aux informations. Enfin, ils sont un élément de réponse au besoin de structure et de sens exprimé par nombre d'utilisateurs et de concepteurs du Web depuis sa création.

L'intervention d'**Emmanuelle Bermès** (Centre Pompidou) a ensuite présenté les principes de l'utilisation des technologies du Web sémantique pour représenter des référentiels et les publier, sur le Web de données. Le Web sémantique est une extension du Web permettant de relier pas seulement les documents mais les données elles-mêmes. Il repose sur un ensemble de normes et de standards développés par le W3C pour construire un langage permettant aux machines de comprendre et d'exploiter les données.

Dans la perspective d'un Web sémantique, les ressources doivent être clairement identifiées, sans quoi il n'est pas possible de partager l'information. La notion de personne fournit un bon exemple : il existe par exemple des dizaines de François Martin en France ; comment décrire l'un d'eux sur le Web, sans risquer de le confondre avec d'autres ? La solution proposée par le web sémantique est d'associer un identifiant unique à chaque ressource : il s'agit de l'URI (Uniform Resource Identifiers). L'URL peut être choisie comme URI. Toutefois, pour être pérenne, elle ne doit pas dépendre de la localisation du fichier sur un serveur.

Le Web sémantique est sous-tendu par un modèle conceptuel, RDF, permettant de décrire des choses, simplement et sans ambiguïté. RDF définit une ressource (URI) sous la forme d'un triplet : ressource, propriété, valeur. Formulé encore autrement, on peut dire qu'une chose est décrite sous la forme d'une phrase : sujet, verbe, complément. Ce modèle conceptuel permet de représenter un nombre considérable de ressources.

« Tout est référentiel et rien n'est référentiel dans le Web sémantique ». Les référentiels sur lesquels peut s'appuyer le Web sémantique sont des jeux de données (Dbpedia, Geonames), des jeux de métadonnées (Dublin Core, RDA, FOAF) ou des référentiels de valeurs (LCSH ou fichiers d'autorité matières de la Bibliothèque du Congrès, fichier international virtuel d'autorité VIAF, listes Marc Code). (Voir le référencement des jeux de données en relation avec les bibliothèques à : <http://ckan.net/group/ld>).

SKOS est un vocabulaire RDF permettant de décrire des systèmes d'organisation de la connaissance : thesauri, vocabulaires contrôlés, classifications, ethno-classifications, cartographies conceptuelles, etc. L'intérêt d'exprimer les référentiels en SKOS est de ne plus les avoir isolés dans une base locale mais de pouvoir aligner différents thésaurus ou systèmes de connaissances.

En continuité avec la présentation d'Emmanuelle Bermès, **Gautier Poupeau** (Antidot) est revenu sur les nouvelles perspectives ouvertes par l'exposition des référentiels, selon les règles du Web de données. Il s'est efforcé de montrer le rôle primordial que joue l'utilisation de technologies standard comme celle du Web sémantique pour améliorer l'interopérabilité, simplifier la réutilisation et la mise en relation de données qui, jusqu'à maintenant, ne l'étaient pas. Partager une sémantique pour décrire les référentiels comme SKOS permet de les relier et de les utiliser dans une application unique. Les URI jouent un rôle de désambiguïsation d'une chaîne de caractères, elles peuvent permettre par exemple de faire le lien entre Boris Vian et Vernon Sullivan. A partir d'exemples concrets, l'intervenant a détaillé les différentes techniques existantes pour exploiter des référentiels en vue de l'enrichissement d'informations ou pour améliorer des plates-formes de recherche : proposition de concepts liés aux ressources du corpus, à la frappe de l'utilisateur ; suggestion de concepts pour désambiguïser les termes de la requête ou proposer un élargissement de recherche ; recherche par facette ; regroupement des résultats en fonction des différents concepts du référentiel ; création automatique d'un nuage par annotation automatique des métadonnées avec Rameau et Pactols¹ ; navigation par concepts ; prise de connaissance d'un corpus par l'analyse de l'utilisation d'un concept ; classement selon leurs fréquences des concepts présents dans les mêmes documents que le concept recherché...

L'après-midi était consacré à des retours d'expériences.

Le projet VIAF (*Virtual International Authority File* ou Fichier d'autorité international virtuel) constitue une brique importante pour le Web sémantique. **Vincent Boulet** (Bibliothèque nationale de France) a présenté ce projet international dont l'objectif premier est la création de notices d'autorité par la mise en relation de différents fichiers d'autorité. Il regroupe aujourd'hui 19 établissements et 22 fichiers d'autorité personnes physiques, collectivités et noms géographiques.

En reliant entre elles des données d'autorité de provenances diverses, en les confrontant et en exposant librement et gratuitement ces données en RDF, VIAF est devenu une brique essentielle du Linked Data y compris en dehors du monde des bibliothèques. Le cheminement de ces données, depuis leur création dans des formats d'origine de type MARC, jusqu'à leur transformation en RDF, leur appariement et leur exposition sous forme réutilisable illustrent les multiples potentialités ouvertes par VIAF. En effet, les notices VIAF sont elles-mêmes dotées d'une URI et rendues en RDF pour faciliter leur réutilisation.

Il est à noter que la documentation du modèle sous-jacent est donnée sous forme d'une ontologie. Divers projets utilisent ou vont utiliser les données de VIAF en RDF, tels que data.bnf.fr² ou encore le

¹ Les thésaurus PACTOLS (Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux et Sujets), sont des micro-thésaurus spécialisés sur les sciences de l'Antiquité et l'archéologie de la préhistoire à l'ère industrielle, gérés par le programme FRANTIQU (programme regroupant les Centres de recherches CNRS, des Universités et du Ministère de la culture).

² Le projet data.bnf.fr vise à garantir l'interopérabilité entre les bases internes et avec l'extérieur et à permettre d'utiliser des données structurées de façon renouvelée. Les premières réflexions de la Bibliothèque nationale de France ont porté sur les évolutions des modèles bibliographiques pour rassembler les données décrivant des entités (œuvres, auteurs), notamment le modèle FRBR. Elles ont été complétées par un travail de modélisation visant à exprimer ces données en RDF en vue de leur exposition dans le Web de données.

projet SNAC (The social network and archival context project)³.

L'intervention suivante, celle de **François-Xavier Nuttal** (CISAC), traitait de l'ISNI (International Standard Name Identifier), le dernier né des identifiants de la famille ISO SC9 : ISBN (International Standard Book Number), ISSN (Serial Number), mais aussi ISMN (Music Number), ISRC (Recording Code)... L'International Standard Name Identifier propose, comme son nom l'indique, un identifiant unique, pérenne et normalisé pour les personnes (physiques, morales ou institutions publiques) jouant un rôle de création, gestion et distribution des contenus. En phase d'adoption par les bibliothèques, les producteurs, les sociétés de gestion collectives ou encore les éditeurs, ISNI permettra l'échange d'information entre les différents secteurs d'activités aux fins de référencement ou de gestion des droits. Composé de 16 chiffres et accompagné de ses métadonnées (type, nom, rôle, lieu, date et identifiants associés), l'ISNI permettra ainsi de renforcer l'interopérabilité entre les différentes bases de données bibliographiques et culturelles notamment.

Valoriser les autorités du Sudoc, telle est la vocation de la nouvelle application mise en production par l'ABES : www.idref.fr. **Yann Nicolas** (ABES) a rappelé que depuis sa création, le SUDOC s'appuyait sur des référentiels pour structurer l'activité de son réseau de catalogage partagé : référentiels de personnes et de collectivités, vocabulaires Rameau et Mesh⁴, etc. Depuis 2006, l'ABES a le souci d'étendre cette logique au-delà du seul catalogue SUDOC, pour ses propres applications, mais également d'autres applications documentaires de l'enseignement supérieur et de la recherche. Cette stratégie de service se décline sous différentes formes : lancement en 2010 d'une application Web dédiée (IdRef), ouverture des référentiels d'IdRef sur le Web de données (2010), lancement en 2011 des Web Services d'IdRef, projet de recherche avec ADONIS et le LIRMM⁵ (2010-2011).

Claire Sibille – de Grimoüard (Service interministériel des Archives de France) y a ensuite présenté les projets en cours de constitution de fichiers d'autorité par les Archives nationales et par les services territoriaux d'archives, visant non plus seulement à mettre en place un cadre d'interopérabilité entre les données archivistiques, mais bien à l'envisager avec les autres données culturelles et patrimoniales (bibliothèques, musées, archéologie...).

Pour **Hélène Rabaut** (Lexis Nexis) et **Hélène Zysman** (Smile - Open Source Solutions), la réalisation de la norme *ISO 25964 Thésaurus et interopérabilité avec d'autres vocabulaires* répond à la longue attente des professionnels construisant et gérant des systèmes d'organisation de connaissances. Faciliter la recherche de l'information dans un contexte donné, mais aussi dialoguer avec d'autres outils dans d'autres contextes, tel est l'approche suivie dans la réalisation de cette norme qui vient remplacer deux anciennes normes, l'une sur les thésaurus monolingues (ISO 2788:1986) et l'autre sur les thésaurus multilingues (ISO 5964:1986). Elle s'inspire fortement des standards du Web sémantique relatifs aux systèmes d'organisation des connaissances et de leurs modèles sous-jacents, notamment des ontologies et du standard SKOS. Inversement, les travaux de cette norme ont également nourri la réflexion sur les standards du Web sémantique.

Enfin, **Jamal Rezzouk** (Société ViaSema) a montré comment le Web sémantique pouvait répondre aux besoins et aux contraintes de l'univers de l'accès à l'information juridique, à la fois classique dans ses besoins fonctionnels, mais spécifique dans les contraintes qu'il impose. L'univers juridique est un univers très organisé (le Web sémantique est intéressant en ce qu'il propose un cadre formel permettant de prendre en compte et d'exploiter toutes ces informations) et en perpétuelle évolution (dans le Web sémantique, le nom, la signification et les relations d'un concept peuvent changer). L'intervenant a présenté la cartographie d'un référentiel « juridique » typique et de tous les outils qu'il met en œuvre (de la gestion des ressources lexicales en passant par les listes d'autorité, thésaurus, ontologie documentaire, ontologie de domaine, ontologie de publication et taxonomies) et a décrit les

³ Projet de recherche collaboratif (Institut des technologies de pointe en sciences humaines de l'Université de Virginie, École d'information de l'Université de Californie à Berkeley, Bibliothèque numérique californienne), visant à créer un prototype de données d'autorité provenant de sources diverses (archives, bibliothèques, musée pour relier entre elles les descriptions des ressources historiques et fournir un système d'accès à ces ressources (<http://socialarchive.iath.virginia.edu/>).

⁴ Le MeSH (Medical Subject Headings), thésaurus biomédical de référence, est un outil d'indexation, de catalogage et d'interrogation des bases de données de la NLM (National Library of Medicine, Bethesda, États-Unis), notamment MEDLINE/PubMed. L'Inserm, partenaire français de la NLM depuis 1969, a traduit le MeSH à l'intention des utilisateurs francophones en 1986, le met à jour chaque année et présente la version bilingue en ligne (<http://mesh.inserm.fr/mesh/>).

⁵ Le TGE Adonis est un très grand équipement à destination de l'ensemble des sciences humaines et sociales, piloté par une unité de service du CNRS. Le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) est une Unité Mixte de Recherche de l'Université Montpellier 2 (UM2) et du Centre National de la Recherche Scientifique (CNRS), Institut des sciences informatiques et de leurs interactions (INS2I).

points durs que l'univers juridique nous impose (qualité, traçabilité, prise en compte explicite du temps et cycle de vie au quotidien, réutilisabilité/interopérabilité avec des ressources existantes).

Les textes des interventions sont disponibles à :

http://www.bnf.fr/fr/professionnels/autres_journees_professionnelles/a.referentiel_donnees_autorites_110527.html

Le projet LOCAH (Linked Open Copac Archives Hub)

[=>Sommaire](#)

Le projet LOCAH (Linked Open Copac Archives Hub) est un projet de recherche du JISC (Comité pour la mutualisation des systèmes d'information de l'enseignement supérieur britannique, équivalent du TGE-Adonis) visant à rendre disponibles en tant que Linked Data les notices descriptives de deux catalogues collectifs, Archives Hub et COPAC.

Archives Hub est un Portail national britannique recensant les archives conservées dans plus de 180 établissements d'enseignement (universités et « collèges ») du Royaume-Uni.

COPAC est un catalogue collectif britannique (connu aussi sous le nom de British Union Catalog) regroupant les catalogues en ligne de la plupart des bibliothèques universitaires du Royaume-Uni et d'Irlande, rassemblées en un consortium des bibliothèques de recherche, le Research Libraries UK (RLUK).

Le projet LOCAH est parti du constat que les données contenues dans les systèmes d'information archivistiques étaient incomplètes et souvent méconnues. La mise à disposition de ces données en tant que Linked Data devrait permettre de mettre en relation des archives, souvent très dispersées, relatives aux mêmes personnes, organismes, lieux et sujets. Elle devrait aussi permettre aux utilisateurs de découvrir d'autres ressources liées de manière contextuelle à des ressources archivistiques, donnant un éclairage nouveau sur l'existence et les activités d'un individu, d'un organisme ou d'une famille.

Le projet LOCAH étudie donc la création de liens entre Archives Hub et d'autres sources de données, tels que le COPAC, DBPedia⁶, la BBC⁷, Geonames⁸, le projet VIAF⁹ et les fichiers d'autorité de la Bibliothèque du Congrès (pour les mots-matières). Les objectifs du projet sont décrits plus en détail sur le blog du projet (<http://blogs.ukoln.ac.uk/locah/2010/07/23/locah-project-aims-objectives-and-final-outputs/>).

Les différentes étapes du projet LOCAH sont les suivantes :

- 1) analyse et modélisation des données actuelles et sélection (ou définition) des vocabulaires RDF adéquats ;
- 2) développement d'un modèle pour la construction d'URI claires, pérennes et faciles à utiliser, à partir des règles suivies pour la mise en œuvre du portail data.gov.uk¹⁰ ;
- 3) développement de processus de conversion des formats de données existants en RDF ;
- 4) mise en relation des données de ArchivesHub et de COPAC avec les Linked Data mises à disposition par d'autres partenaires (DBpedia, Geonames, etc.) et utilisation de web services comme Open Calais¹¹ ;
- 5) développement d'interfaces de consultation et d'exploitation des données de Archives Hub et du Copac, confrontées aux jeux de données du Linked Data.

⁶ Projet d'extraction de données de [Wikipédia](#) pour en proposer une version web sémantique structurée ; ce projet est mené par l'Université de Leipzig, l'Université libre de Berlin et l'entreprise OpenLink Software.

⁷ Disposant de multiples sites web consacrés aux différentes thématiques qu'elle traite au quotidien, la BBC a rapidement constaté que ses sites étaient insuffisamment liés entre eux. Par ailleurs, ne pouvant pas facilement recontextualiser un contenu mis en forme une première fois pour s'en servir sous une autre forme, elle a travaillé sur l'intégration des données liées à chaque élément composant un article, une vidéo, un reportage, etc.

⁸ Base de données géographiques gratuite et accessible par Internet sous une licence Creative Commons.

⁹ VIAF (Virtual International Authority File), est un projet d'OCLC Research qui vise à l'origine à aligner des listes d'autorités (notamment sur les noms propres) en vue de constituer une base de référence internationale.

¹⁰ Après les Etats-Unis en 2009, le Royaume-Uni s'est engagé dans l'ouverture de ses données, avec le lancement du portail data.gov.uk en janvier 2010. La mise en ligne des données publiques répond à un triple objectif : encourager l'innovation par la réutilisation des données, participer à la transparence de l'État et offrir une nouvelle forme de services publics aux citoyens. Plus d'un an après, le site a permis d'ouvrir plus de 6 000 jeux de données ainsi que la création de plusieurs applications.

¹¹ Open Calais (<http://www.opencalais.com/>) de Thomson Reuters est une application en ligne qui s'appuie sur des règles linguistiques et des dictionnaires/référentiels issus, entre autres, du Linked Data, et qui permet de repérer des « entités nommées » (lieu, entreprise, personne...) dans des données non structurées.

La politique de normalisation en Espagne

[=>Sommaire](#)

Du 29 mars au 29 avril 2011, la direction générale des patrimoines (département de la formation scientifique et professionnelle) organisait un stage technique international ouvert aux professionnels de la communauté archivistique. Nuria Nieto, archiviste du Centre d'Information Documentaire des Archives (CIDA) du ministère de la Culture espagnol, y a présenté les applications en ligne développées par le ministère de la Culture espagnol ainsi que les travaux de normalisation en cours dans les archives espagnoles.

Créé en 1977, le Centre d'Information Documentaire des Archives (CIDA) dépend de la Sous-direction Générale des Archives de l'État. Il est au service de toute la communauté professionnelle archivistique espagnole, et pas seulement des archives d'État. Sa fonction essentielle est la communication du patrimoine documentaire espagnol à travers les bases de données en ligne. C'est ainsi qu'un Catalogue Collectif du réseau des Bibliothèques des Archives d'État et de la Bibliothèque du CIDA (CCBAE) réunit les notices bibliographiques du réseau de bibliothèques des services d'archives publics relevant de l'État central et de la Bibliothèque du CIDA (<http://www.mcu.es/ccbae/es/inicio/inicio.cmd>). Il contient également la collection numérique des cartes, plans et dessins des Archives Générales de Simancas. Le Guide d'archives d'Espagne et d'Amérique latine ou « Censo Guia » (<http://censoarchivos.mcu.es/CensoGuia/portada.htm>) donne quant à lui accès aux descriptions des archives, tant publiques que privées, relatives à l'histoire de l'Espagne et de l'Amérique latine. Il contient des informations sur les institutions de conservation et une brève description des fonds qu'elles détiennent, incluant leur date limite et leur importance matérielle. Cet outil a un double caractère fonctionnel : c'est un instrument de contrôle chargé de la protection du patrimoine documentaire espagnol, c'est aussi un instrument de diffusion de base pour la connaissance des archives par l'administration, les citoyens et les utilisateurs. Ce projet est pionnier dans l'élaboration d'un modèle de données, l'EAG (Guide archivistique encodé) et pour l'utilisation du langage XML, ce qui lui permet d'être une passerelle dans l'échange avec d'autres projets internationaux comme APEnet, le Portail européen archivistique, et aussi le Portail Espagnol d'Archives du ministère de la Culture ou PARES (<http://pares.mcu.es/>). Enfin, le CIDA est en train de développer un portail relatif au processus d'indépendance des pays d'Amérique latine (http://www.mcu.es/archivos/Novedades/novedades_CIDA_PortalBicentenarios.html). Le portail comprendra trois blocs : un guide des sources avec une sélection de documents issus de PARES, du Censo Guia et du Guide des Sources de l'histoire de l'Amérique (HAME), une exposition virtuelle et des ressources pédagogiques.

Une autre ligne d'action du CIDA est la normalisation de la description archivistique. En mai 2007, le ministère de la Culture a créé une Commission des normes espagnoles de description archivistique (CNEDA) dont le premier projet a consisté à développer un modèle conceptuel de description archivistique (MCDA) identifiant les types d'entités archivistiques, c'est-à-dire les classes fondamentales des « objets » à décrire dans les systèmes d'information archivistiques, leurs relations et leurs attributions, ainsi que les éléments de données essentiels pour la description archivistique. Le modèle de données identifie ainsi six types principaux d'entités :

- documents d'archives : par exemple, les fonds « Conseil de l'Inquisition » et « Gouvernement de Guadalajara » sont des entités « documents d'archives »
- agent : la « Députation de la province de Lugo », la « famille Tellez-Giron (ducs d'Ossuna) » et « Clara Campoamor (1888-1972) » sont des entités « agent »
- fonction : les fonctions « Relations internationales » et « Gestion de la recherche » sont des entités « fonction et subdivisions de fonction »
- texte juridique : la « Charte de Logroño » et l'actuelle Constitution espagnole sont des entités « texte juridique »
- concept, objet ou événement : les « droits de douane » et le « réchauffement climatique » sont des entités « concept », « objet » ou « événement »
- lieu : la « Vice-Royauté de Nouvelle-Espagne » et la « Bolivie » sont des entités « lieu ».

Ces différentes entités sont représentées par des descriptions séparées mais reliées les unes aux autres au sein du système de description archivistique, pour éviter la répétition d'informations dans le système, pour améliorer la recherche et la récupération de l'information, etc.

Le modèle s'attache également à décrire les relations que les entités entretiennent les unes avec les autres : ainsi, les « agents » collectivités, personnes ou familles, « créent, produisent et gèrent » les documents d'archives. Ils peuvent être les responsables du contenu intellectuel des documents d'archives (auteurs) ; les responsables de la production des documents (producteurs) ; les responsables de la constitution de collections (collectionneurs) ; les responsables de la gestion des documents d'archives (gestionnaires) ; etc.

Les documents d'archives sont issus de l'exercice des fonctions par les agents. Les textes juridiques sont les dispositions qui « régissent » les agents, les fonctions exercées par les agents, et les documents d'archives eux-mêmes. Enfin, les agents, les fonctions, les textes juridiques et les lieux peuvent être les sujets traités par les documents d'archives.

Les types d'entités « documents d'archives » « agent » et « fonction » sont eux-mêmes composés de sous-types d'entités, ce qui n'est pas le cas des trois autres types.

Ainsi, le type d'entité « documents d'archives » est composé de dix sous-types reliés hiérarchiquement entre eux :

- « fonds »
- « groupe de documents » (équivalent de la valeur « recordgrp » de l'EAD)
- « sous-fonds » ou « sous-groupe de documents »
- « série organique »
- « sous-série organique »
- « partie de série ou de sous-série organique »
- « unité documentaire »
- « collection »
- « partie de collection »
- « composant documentaire ».

Seuls les sous-types « fonds », « collection » et « série organique » sont obligatoires, les autres sous-types sont facultatifs, ce qui implique que dans tout système de description archivistique, il y ait au moins une description de fonds, de sous-fonds, de collection ou de série organique. Un fonds peut comprendre un ou plusieurs sous-fonds ; un sous-fonds peut regrouper un ou plusieurs sous-fonds ; un sous-fonds peut être formé d'une ou de plusieurs sous-séries ; une sous-série peut inclure une ou plusieurs sous-séries ; une sous-série peut regrouper une ou plusieurs unités documentaires, etc.

Le type « agent » inclut trois sous-types d'entités :

- « collectivité » : par exemple, la « Députation de la province de Lugo » est une entité du type « agent » et appartient au sous-type « collectivité »
- « famille » : la « Famille Tellez-Giron (ducs d'Ossuna) » est une entité du type « agent » et du sous-type « famille »
- « personne » : « Clara Campoamor (1888-1972) » est une entité du type « agent » et du sous-type « personne ».

Si une série organique est le niveau le plus élevé d'un système de description archivistique, cette série doit être reliée à un ou plusieurs agents « producteurs ».

Le type « fonction » inclut quatre sous-types d'entités :

- fonction
- sous-fonction
- activité/processus
- action.

Ces quatre sous-types sont reliés hiérarchiquement les uns aux autres. Ainsi, une fonction peut être composée de une ou de plusieurs sous-fonctions ; une fonction peut regrouper un(e) ou plusieurs activités/processus ; une sous-fonction peut comprendre une ou plusieurs sous-fonctions ; une sous-fonction peut comprendre une ou plusieurs activités (ou un ou plusieurs processus) ; une activité ou un processus peuvent inclure une ou plusieurs actions.

Actuellement, la CNEDA travaille sur l'avant-projet du MCDA, lequel inclura la partie 1 revue (Types d'entités) et la partie 2 (Relations).

Par ailleurs, une Norme pour l'élaboration de points d'accès réglementés des collectivités, familles, lieux et matières dans le système de description archivistique des archives d'État a été élaborée par un Groupe de travail d'indexation des Archives d'État et a été publiée en août 2010 (http://www.mcu.es/archivos/docs/Novedades/Norma_puntos_acceso2010.pdf). Cette norme fournit les lignes directrices pour la préparation de notices d'autorité et points d'accès normalisés des collectivités, des personnes et familles, comme des lieux et des matières, désignés comme « Forme(s) autorisée(s) du nom (pour les collectivités, familles, lieux et matières) et « Terme(s) préférentiel(s) » (pour les lieux et matières).

Il est prévu de fournir des règles et/ou des recommandations pour l'application d'autres éléments ou composants structurels des notices d'autorité archivistiques (collectivités, personnes ou familles) et thésaurus (lieux et matières) qui pourront être développés a posteriori.

Pour le choix du nom et de sa formalisation, une règle générale est établie, avec les normes communes pour la rédaction de chaque type. Cette règle générale se structure en composants et langues. On y détermine les composants que comptera le point d'accès normalisé ou la notice d'autorité, soit le nom de l'entité (collectivités, personnes ou familles) et des qualificatifs, essentiels pour identifier l'entité ainsi désignée et permettre de distinguer sans ambiguïté différentes entités qui ont le même nom ou des noms très proches.

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

claire.sibille@culture.gouv.fr