

## Sommaire

**L'EAC-CPF, un nouveau schéma XML pour la description des collectivités, personnes, familles**

**Vers un modèle conceptuel commun aux archives, aux bibliothèques et aux musées ?**

**Journées d'études aux Archives nationales d'outre-mer, 2-4 juin 2010**

**Lu pour vous**

---

## **L'EAC-CPF, un nouveau schéma XML pour la description des collectivités, personnes, familles**

La Société des archivistes américains (SAA) vient de publier la version 2010 du schéma XML EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles), ainsi que la documentation correspondante (dictionnaire des éléments et des attributs). Le tout est disponible sur un site hébergé par la Bibliothèque d'État de Berlin : <http://eac.staatsbibliothek-berlin.de/>

Décrire les archives implique aussi de fournir des informations sur les circonstances entourant leur production et leur utilisation. Connaître le contexte de production des archives est essentiel pour comprendre l'organisation et le contenu informationnel de ces documents. L'historique de la conservation des documents d'archives, leur mode de classement et les circonstances entourant leur production et leur utilisation contribue à établir leur authenticité.

Le schéma EAC-CPF permet donc l'encodage de notices d'autorité établies conformément à la norme ISAAR(CPF) et décrivant des collectivités, des personnes et des familles intervenant d'une manière ou d'une autre dans le contexte de production et de gestion des archives. Toutefois, d'autres spécialisations de l'EAC pourraient être élaborées pour aller plus loin dans la description de ce contexte (fonctions exercées par ces collectivités, personnes et familles notamment).

L'EAC-CPF permet de produire et de gérer des descriptions séparées des producteurs d'archives mais liées aux descriptions des documents d'archives qu'ils ont produits. L'EAC-CPF permet donc de mettre en relation les notices d'autorité décrivant des entités (collectivités, personnes et familles) avec les descriptions archivistiques en EAD.

Toutefois, ces entités peuvent aussi être simplement les sujets des documents d'archives ou entretenir d'autres types de relations avec ces documents. Le schéma EAC-CPF s'efforce donc d'explicitier la nature de ces relations. L'EAC-CPF crée ainsi les conditions pour le partage et la réutilisation de l'information sur les collectivités, les personnes et les familles. L'EAC-CPF a vocation à être utilisé non seulement par les archives mais aussi par d'autres métiers, les bibliothèques notamment, dans le cadre de projets fédérant des ressources de provenances diverses relatives à des collectivités, personnes et familles. Dans un souci d'interopérabilité avec d'autres schémas de métadonnées exprimés en XML, il s'appuie sur l'état de l'art en la matière (codes de balises explicites, réutilisation d'éléments de bibliothèques XML).

Le schéma EAC-CPF est le résultat de 30 mois de travaux d'un groupe de 15 experts représentant 9 pays. Le projet a bénéficié du soutien de la Société des archivistes américains, de la Bibliothèque d'État de Berlin, de l'Institut pour les biens artistiques, culturels et naturels de la Région Émilie-Romagne, et de subventions de la Delmas Foundation. Le groupe de travail a également bénéficié de la contribution de la communauté internationale à travers le processus d'appel à commentaires et de tests du projet de schéma fin 2009. Enfin, il convient de saluer la contribution très importante du Groupe d'experts AFNOR CG 46/CN 357/GE4 Données d'autorité dans l'aboutissement de cette entreprise.

Le schéma est téléchargeable sous trois syntaxes différentes : XML Schéma, RelaxNG et RelaxNG compact. Il s'accompagne d'un dictionnaire des éléments (Tag Library) complet avec des exemples d'encodage, également librement téléchargeable. La documentation en ligne continuera à évoluer pour répondre aux besoins des différentes communautés. Le projet EAC-CPF ayant été conçu comme un projet international et multilingue, le site est censé recevoir des versions linguistiques différentes. La version anglaise du dictionnaire des balises est dès à présent intégrée dans les outils de production (définition de chaque élément et attribut), mais une traduction en français sera bientôt lancée, contribuant ainsi à offrir à la communauté des utilisateurs francophones de l'EAC-CPF un environnement de travail en français.

---

## Vers un modèle conceptuel commun aux archives, aux bibliothèques et aux musées ?

*Les 26-28 janvier 2010, le Musée national de Finlande accueillait à Helsinki les experts du groupe de normalisation documentaire du Conseil international des Musées chargé d'assurer la maintenance du modèle sémantique de référence CIDOC-CRM, qui est une ontologie propre à s'appliquer au patrimoine culturel. L'objectif de ce séminaire était de définir les principes et les objectifs d'un modèle conceptuel commun aux archives, aux bibliothèques et aux musées, pour faciliter l'intégration et les échanges des informations provenant de sources diverses.*

### 1. Historique du modèle CIDOC-CRM

Depuis 1996, le Comité international pour la documentation (CIDOC), qui est lui-même une émanation du Conseil international des Musées (ICOM), travaille sur un **modèle conceptuel de référence** (CRM) du domaine du patrimoine culturel. Un groupe de travail spécifique, organisé par le CIDOC mais ouvert également aux personnes extérieures au CIDOC, s'est mis en place pour gérer les mises à jour du modèle : le Groupe d'intérêt spécial CRM (CRM-SIG). Depuis 2002, les grandes lignes du modèle sont stabilisées, et il ne s'agit plus que d'assurer des modifications de détail, notamment en ce qui concerne les définitions et les exemples. En 2006, le modèle CIDOC-CRM a été accepté par l'Organisation internationale de normalisation comme norme internationale, *ISO 21127:2006 Information et documentation -- Une ontologie de référence pour l'échange d'informations du patrimoine culturel*. La traduction en français de la version 3.4 du modèle CRM a été assurée par la Commission CN 357 modélisation, production et accès aux documents de l'AFNOR. La version la plus récente du modèle, dite 5.0.1 (datée de novembre 2009) est disponible sur le site officiel du CIDOC-CRM en .doc et .pdf : [http://cidoc.ics.forth.gr/official\\_release\\_cidoc.html](http://cidoc.ics.forth.gr/official_release_cidoc.html)

### 2. Principes du modèle CIDOC-CRM

Le modèle CRM vise fondamentalement à **fournir un langage commun à des gisements d'information hétérogènes** et à

permettre leur intégration, par delà leurs éventuelles incompatibilités tant sémantiques que structurelles. Il s'agit donc de **faciliter l'échange et la recherche d'informations dans le domaine du patrimoine culturel** et de permettre aux musées de rendre compatibles leurs documentations sans rien perdre de leurs spécificités ni du niveau de précision de leurs données actuelles.

Plutôt que de concevoir un modèle de données unique et énorme, destiné à répondre a priori à tous les types de besoins et à tous les cas de figure, le CIDOC a élaboré un modèle de concepts pouvant servir de point de référence. Les experts du CIDOC ne sont pas partis de rien mais se sont appuyés sur un ensemble d'éléments d'information prédéfinis dans un document normatif publié en 1995, les *Recommandations internationales pour l'information concernant les collections des musées : les catégories d'information du CIDOC*. Cet ensemble comportait 74 éléments informatifs, répartis en 22 grands groupes tels que « Acquisition », « Localisation », « Marques et inscriptions », « Matériau et technique », « Sujet représenté », etc.

Le CIDOC CRM est tout à la fois un modèle orienté objet, un modèle sémantique et une ontologie de domaine.

Le CIDOC CRM est un **modèle orienté objet** qui permet de donner une vue d'ensemble sans s'encombrer de détails superflus, qui est plus facilement gérable que le formalisme entité-relation et qui ne préjuge d'aucun mode d'implémentation concrète.

C'est un **modèle sémantique**, car il permet de faire émerger la signification réelle de tout ce qui est considéré comme « implicite » et « évident » dans une structuration d'informations.

Il peut enfin être considéré comme une **ontologie de domaine**, une analyse formelle des entités et des relations qui sont fondamentales à un champ d'activité, en l'occurrence le patrimoine culturel. Cette structure fournit le cadre pour la description des relations complexes qui existent entre les objets, les acteurs, les événements, les lieux, et les concepts qui constituent le domaine du patrimoine culturel. Il s'agit donc d'établir des équivalences entre un format donné et le CRM, pour qu'ensuite toutes les informations créées dans ce format puissent être interprétées par une machine en référence aux relations sémantiques définies par le CRM.

**Le CRM est extensible** : il peut être rendu très compact, ou étendu pour une plus grande précision et pour couvrir de nouveaux champs. Ce caractère extensible est dû, en grande partie, à l'adoption du schéma orienté objet. Cette approche utilise en effet des entités (ou « classes ») telles que personne, lieu, objet et événement et des relations (ou « propriétés ») organisées dans une arborescence hiérarchique. Les propriétés sont définies (« déclarées ») comme étant applicables entre un domaine et une cible, elles relient une instance de la classe qui fait office de domaine (ou « source du lien ») à une instance de la classe qui sert de « cible » (vers laquelle pointe le lien). Quand une classe est déclarée comme sous-classe d'une autre classe, alors chaque instance de la sous-classe est aussi une instance valide de la super-classe, et hérite des propriétés de cette super-classe. Enfin, les entités ou classes peuvent avoir des caractéristiques ou attributs.

Toutes les classes du CIDOC CRM ont à la fois un nom et un identifiant. L'identifiant d'une classe est composé de la lettre E suivie d'un numéro séquentiel. Les propriétés ont également un nom et un identifiant, qui consiste en la lettre P suivie d'un numéro séquentiel,

Le modèle CIDOC CRM contient ainsi 86 classes et 137 propriétés. Ceci peut paraître beaucoup, mais le noyau au cœur du CRM – un modèle centré sur les événements – est très simple. Il peut être exprimé par les propositions suivantes :

- les choses arrivent (expositions, prêts, etc.) : ce sont des Événements (E5) ;
- lorsqu'un événement a lieu, il est souvent possible d'en préciser le Lieu (E53) et la Durée (E52) ;
- les événements changent le monde de différentes façons, c'est-à-dire que certaines Choses (E70) matérielles ou intangibles sont affectées ;
- des événements peuvent impliquer des agents en tant que témoins, victimes, ou instigateurs ; ces agents sont des Acteurs (E39) ;
- tous ces concepts – Événements, Lieux, Durées, Choses et Acteurs – peuvent exister sous différentes formes, qui sont des Types (E55).

**Au cœur même du CRM, on trouve donc la notion d'événement**, de phénomène temporel. Avant de décrire un objet matériel tel qu'il se présente, il faut avant tout replacer cet objet dans un contexte historique. Ce qui importe, c'est à la fois l'environnement de cet objet au cours du temps (lors de sa création, et lors de chacune de ses éventuelles modifications) et l'énoncé de tout ce qui a pu lui arriver. Le fait de placer systématiquement une Entité Temporelle au milieu d'une nébuleuse constituée d'un objet matériel ou conceptuel, d'un agent, d'un lieu et d'une tranche chronologique peut sembler très contraignant. En fait, le modèle CRM prévoit dans beaucoup de cas la possibilité de relier directement un objet au résultat d'un événement.

### 3. Intérêt et utilisations possibles

#### a. Outil de dialogue entre spécialistes d'horizons différents

En fournissant une analyse riche et détaillée du domaine du patrimoine culturel, **le CRM peut faciliter le dialogue entre experts du patrimoine et informaticiens**. Les classes et les propriétés qui le composent sont toutes clairement définies, grâce à des descriptions textuelles, des notes d'application, des exemples, des références croisées et l'organisation hiérarchique des classes et des propriétés. Cette présentation est censée être accessible aussi bien pour les techniciens que pour les experts culturels. Les professionnels du patrimoine peuvent le considérer comme une représentation formelle de concepts familiers tandis que les informaticiens peuvent le considérer comme le plan général d'un système d'information.

### **b. Outil d'aide à la conception de systèmes**

Une autre utilisation possible du CRM est la spécification technique lors de la conception de nouveaux systèmes d'information. **Le CRM n'est pas conçu pour être implémenté tel quel.** Le modèle doit être adapté aux besoins spécifiques, simplifié ou développé si nécessaire. Le CRM a été conçu afin de rendre ce processus d'adaptation aussi simple que possible. Des points d'ancrage permettent des extensions et une méthodologie explique comment il est possible de développer la structure du CRM tout en restant compatible avec la structure générale du modèle.

### **c. Outil de comparaison, d'échange et de stockage de données**

**Le CRM peut aussi servir de référence technique** lors de la comparaison et de l'évaluation de systèmes d'information, de schémas de données, etc. Une comparaison d'un système existant ou projeté avec le CRM permet de mettre en évidence des divergences qui pourront alors être examinées de plus près.

La valeur du CRM en tant que référence technique devient particulièrement évidente quand il est utilisé pour guider le transfert de données entre systèmes incompatibles. Le CRM peut ainsi fournir l'épine dorsale sémantique d'un format commun de données qui peut être partagé par plusieurs systèmes. Le CRM facilite non seulement les échanges de données entre institutions mais aussi la migration entre systèmes.

### **d. Outil de dialogue entre réservoirs de données hétérogènes**

L'application la plus ambitieuse du CRM est le développement d'outils d'interrogation fédérés par l'intermédiaire de systèmes de médiation. Actuellement, les informations stockées dans les systèmes d'information de bibliothèques, des archives et des musées sont effectivement cloisonnées. Plusieurs ressources doivent être interrogées séparément et les liens entre systèmes sont rares. Le but du CIDOC-CRM est de **permettre la combinaison et l'intégration d'informations de provenance hétérogène** afin de faciliter la recherche et la qualité des services offerts aux utilisateurs améliorée.

Pour des raisons techniques et organisationnelles, l'intégration physique des

données dans un seul système est souvent impossible. Le but des systèmes de médiation est de permettre la fédération des sources d'information, rendant possible la distribution des interrogations sans la nécessité d'une base de données unique. Les systèmes de médiation offrent une interface unique pour les utilisateurs. Ils acceptent et interprètent les interrogations et les distribuent aux systèmes participants. Ces systèmes envoient leurs réponses au médiateur qui peut ensuite consolider les résultats et les présenter à l'utilisateur.

Afin que la médiation puisse fonctionner correctement, la communication doit être rendue possible entre le médiateur et chaque base de données spécifique. L'utilisation du CRM pour la conception du schéma de médiation peut faciliter la conception d'un système distribué. Il permet d'interpréter et d'intégrer des données en provenance de systèmes qui seraient autrement incompatibles. **Le CRM a été conçu pour permettre l'intégration des informations en provenance de systèmes qui sont relativement pauvres ou riches**, sans pertes d'informations.

### **e. Outil d'intégration au Web sémantique**

Enfin, le CRM, en tant qu'ontologie, jouera certainement à terme un rôle majeur dans tous les secteurs du Web sémantique concernés par l'information sur le patrimoine culturel.

C'est ainsi que des travaux sont en cours au sein du JISC (Comité conjoint pour les systèmes d'information, organisme gouvernemental britannique regroupant des organismes de financement de l'enseignement supérieur et de la formation continue), pour **établir des correspondances entre différents vocabulaires de schémas de métadonnées**, dont le modèle CIDOC-CRM.

La finalité de ce projet, intitulé « Vocabulary mapping framework » (VMF) est de développer un outil libre permettant l'interopérabilité entre les différentes communautés productrices de ces schémas. Un premier travail d'interopérabilité réalisé entre RDA (« Ressources : Description et accès », le nouveau code de catalogage à vocation internationale) et ONIX (norme internationale pour la diffusion de métadonnées relatives aux livres et d'autres documents utilisés par les bibliothèques et les éditeurs) a été étendu à d'autres vocabulaires : outre le CRM, il convient de mentionner Dublin Core, DDEX

(ensemble de normes pour encadrer et mesurer les ventes de musique en ligne), le modèle FRBR (Spécifications fonctionnelles des notices bibliographiques ; cf. *infra*), LOM (schéma de description de ressources d'enseignement et d'apprentissage) et MARC21. Chaque terme d'un vocabulaire source a une correspondance avec une ontologie générique ou « matrice », exploitable pour établir des passerelles entre les schémas sources. La matrice n'est pas conçue pour être un autre vocabulaire qui donnerait un sens définitif aux termes qui le constituent. On trouvera une description complète du projet à l'adresse :

<http://cdlr.strath.ac.uk/VMF/index.htm>

Des essais de « mapping » avec l'EAD ont également été tentés par le CRM-SIG, en utilisant le langage informatique XPath pour localiser une portion d'un document XML/EAD. Par exemple, l'intitulé d'une unité documentaire est inclus dans l'élément <unittitle> qui est inclus dans <did>, lui-même compris dans <archdesc>, lui-même emboîté dans l'élément racine <ead>. L'équivalence dans CIDOC-CRM pourrait être la suivante : un Document (E31) ou un Objet linguistique (E33) sont composés de/ont partie (P106) d'un Document (E31) ou d'un Objet linguistique (E33) qui a pour titre (P102) un Titre (E35). Les résultats de ces travaux sont disponibles sur le site du CIDOC-CRM, mais l'EAD est un exemple d'implémentation des principes de la description archivistique, et une approche plus conceptuelle, dont le point de départ serait les normes élaborées par le Conseil international des archives, semble préférable (cf. *infra*).

#### 4. Quelques exemples de mise en œuvre

Quelques exemples d'utilisation du CIDOC-CRM sont décrits sur le site officiel du modèle conceptuel.

Ainsi, le projet Musinfo s'est appuyé sur le modèle CIDOC-CRM pour la conception et l'implémentation d'une base de données multidisciplinaire pour les **musées de la Ville de Genève** (<http://www.ville-ge.ch/musinfo/>).

Le **musée national allemand à Nuremberg** a utilisé le CRM pour la conception d'une série de DTDs XML, celles-ci formant le noyau d'un système de saisie et de stockage des documents.

Aux États-Unis, le **Research Libraries Group** (Groupe de bibliothèques de recherche) a

utilisé le CRM pour la conception de sa base de données Cultural Materials.

**English Heritage** (Patrimoine de l'Angleterre), organisme gouvernemental du Royaume-Uni qui dispose d'une très importante latitude pour surveiller le patrimoine historique de l'Angleterre, utilise le CRM dans ses projets de construction de passerelles entre de multiples bases de données.

Le CRM a également été utilisé par le **Musée national de Finlande** lorsque celui-ci a voulu changer de base de données : le modèle a servi de point de référence dans la conception de la nouvelle base Muusa et des équivalences ont été faites entre la structure de données de l'ancienne base et les sections pertinentes du CRM. Les règles de catalogage des documents ont en outre été révisées pour prendre en compte le CRM afin d'assurer une meilleure cohérence interne de ces règles.

Le modèle CIDOC-CRM est également l'un des onze outils d'échange retenus pour la structuration des métadonnées descriptives dans le cadre du projet de **Bibliothèque numérique de Finlande** (<http://www.kdk2011.fi>). Financé à hauteur de 16 millions d'euros et piloté par le Ministère de l'Éducation, ce projet fait partie des priorités définies par le Plan d'action gouvernemental pour la société de l'information pour la période 2007-2011. L'objectif est d'améliorer les conditions d'accès général à l'information aussi bien à des fins éducatives que de recherche scientifique, de favoriser la numérisation des objets patrimoniaux les plus importants conservés dans les bibliothèques, les archives et les musées finnois, et d'assurer la préservation à long terme des ressources numériques, natives ou numérisées. Le projet de Bibliothèque numérique de Finlande vise donc à créer une interface publique commune donnant accès aux ressources numériques issues des bibliothèques, archives et musées finnois, ainsi qu'à des services en ligne. **Le projet sera mené en collaboration étroite avec Europeana**, afin de garantir la mise à disposition dans la Bibliothèque numérique européenne des matériaux les plus essentiels de la mémoire finnoise. L'interface publique sera accessible en 2011. Sur les quelques 700 institutions culturelles et patrimoniales que compte la Finlande, 35 participent directement au projet, dont le Musée national via l'application Muusa. **Les Archives nationales de Finlande sont quant à elles chargées d'élaborer une stratégie de préservation à long terme des ressources numériques,**

d'ici l'été 2010. Les principales difficultés rencontrées sont les interprétations très diverses des normes et formats de description, la présence de métadonnées locales qui peuvent avoir leur utilité dans d'autres systèmes d'information, l'absence de contrôle d'autorité rigoureux, ou encore l'hétérogénéité des descriptions qui peuvent parfois avoir peu en commun.

Le modèle CRM a également été utilisé dans **CultureSampo** (Kulttuurisampo en finnois ; <http://www.kulttuurisampo.fi/>), portail développé à partir des technologies du Web 2.0 et des concepts du Web sémantique. CultureSampo est une plate-forme destinée à donner accès à des contenus culturels hétérogènes. Le portail comprend au total quelques 128 714 objets culturels provenant d'une vingtaine d'institutions patrimoniales finnoises, par exemple des objets de musées, des photographies, des cartes, des peintures, des poèmes, des livres, des chansons folkloriques et des vidéos, des histoires (l'épopée nationale de Kalevala), ainsi que 276 681 ressources de référence sur des concepts, des lieux, des événements historiques, etc. Des passerelles ont été établies entre les schémas de métadonnées utilisés pour la représentation des données et les ontologies de domaine disponibles sur le serveur d'ontologies partagées ONKI (<http://www.yso.fi/>), afin de rendre les contenus interopérables. Le modèle CIDOC-CRM a été utilisé pour harmoniser les valeurs des attributs de ces différentes ontologies de domaines. CultureSampo a également été relié à de vastes réservoirs de données via des services web 2.0 comme Wikipedia ou Panoramio, un outil de partage de photographies géopositionnées, pour permettre l'enrichissement sémantique et l'annotation collaborative des contenus. Cet enrichissement sémantique a conduit à de nouvelles approches d'accès à l'information. CultureSampo donne ainsi accès aux contenus à partir de différents « points de vue », chaque point de vue étant représenté par un mode particulier de consultation et d'interrogation. Il est notamment possible de naviguer d'un lieu à un autre dans des cartes, de lister les événements historiques et les ressources s'y rattachant, de consulter des biographies enrichies sémantiquement, de parcourir les chants de Kalevala à partir d'un sommaire arborescent, etc.

## 5. Le modèle CIDOC CRM et les bibliothèques

Parallèlement aux travaux du CIDOC, dans les années 1992-1998, la communauté des bibliothèques, à travers la section de catalogage de l'IFLA (Fédération internationale des associations de bibliothécaires et d'institutions), concevait les FRBR (Spécifications fonctionnelles des notices bibliographiques).

**Le modèle FRBR est un modèle entités-relations qui identifie trois grands groupes d'entités intervenant dans la description d'une ressource bibliographique :** un premier groupe de quatre entités correspondant aux produits d'une activité intellectuelle ou artistique (Œuvre, Expression, Manifestation, Item), un deuxième groupe réunissant les deux entités (Personne et Collectivité) correspondant aux agents ayant exercé une responsabilité vis-à-vis de chacune des entités du premier groupe, et enfin un troisième groupe composé des entités ayant une relation de sujet avec l'entité Œuvre ; ce dernier groupe est composé des entités des deux groupes précédents auxquelles s'ajoutent quatre entités supplémentaires (Concept, Objet, Événement, Lieu). Le modèle définit ensuite les principales caractéristiques de chaque entité (ses attributs), ainsi que les relations qui existent entre les différentes entités.

L'idée qu'une harmonisation des deux modèles pouvait être profitable aux musées et aux bibliothèques a été exprimée pour la première fois en 2000, durant un séminaire organisé à Paris par l'ELAG (European Library Automation Group), groupe européen qui rassemble des personnes intéressées par l'informatique documentaire dans les bibliothèques. Cependant il a fallu attendre 2003, suite à la création du Groupe de révision des FRBR, pour qu'un groupe commun soit officiellement constitué dans le but d'harmoniser les deux modèles de conceptualisation. Le Groupe international d'harmonisation FRBR/CIDOC CRM est à la fois un groupe de travail du Groupe de révision des FRBR de l'IFLA et un sous-groupe du CIDOC CRM SIG. Il est co-dirigé par Martin Doerr (ICS-FORTH, Grèce), président du CIDOC CRM SIG et principal auteur du CIDOC CRM, et Patrick Le Bœuf (Bibliothèque nationale de France), le premier président du Groupe de révision des FRBR. Une première version du **modèle FRBRoo (modèle FRBR « orienté objet »)** a été publiée en mai 2009 ([http://cidoc.ics.forth.gr/docs/frbr\\_oo/frbr\\_docs/FRBRoo\\_V1.0\\_draft\\_2009\\_may.pdf](http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_draft_2009_may.pdf)). Cette version correspond aujourd'hui au seul modèle

FRBR, mais a vocation à prendre en compte les modèles FRAD (Modèle conceptuel pour les données d'autorité) et FR SAR (Modèle conceptuel pour les données d'autorité sujet) au fur et à mesure de leur publication pour fournir un modèle global de l'information gérée par les catalogues de bibliothèques.

Les FRBR modélisent les résultats (œuvre, expression) des processus (comme la création, la réalisation, la planification) mais ne traitent pas des processus en eux-mêmes. FRBROO qui s'appuie sur la démarche du CRM, met l'accent sur les processus. Cette approche permet de réfléchir sur les circonstances dans lesquelles, par exemple, des instances d'œuvres ont été conçues ou réalisées. La complexité de FRBROO n'est pas le fruit de l'imagination de ses créateurs pour décrire le monde bibliothéconomique mais résulte de la modélisation de tous les concepts implicites qui étaient nécessaires pour justifier l'existence de certains attributs dans les FRBR(ER).

Parallèlement, le groupe a travaillé à un « FRBROO Core », schéma de métadonnées reprenant les principales structures de FRBROO, afin de montrer aux professionnels une façon simple de traduire les concepts clés dans une application.

## 6. Le modèle CIDOC-CRM et les archives

A présent, le CRM-SIG étudie la possibilité d'initier un processus similaire avec les archives via le Comité des normes et bonnes pratiques du Conseil international des archives.

Depuis les années 1990, quatre normes internationales de description ont été élaborées par le Conseil international des archives, permettant ainsi de décrire les documents d'archives (*Norme générale internationale de description archivistique* ou norme ISAD(G)), leurs producteurs/détenteurs (*Norme internationale pour les notices d'autorité archivistiques Collectivités, personnes, familles* ou ISAAR(CPF) et *Norme internationale pour la description des institutions de conservation des archives* ou ISDIAH) et les fonctions et activités exercées par ces producteurs (*Norme internationale pour la description des fonctions* ou ISDF).

L'intérêt de cette réflexion normative est d'avoir débuté au moment où les progrès de l'informatique permettaient de concevoir de nouveaux outils de travail, plus simples et plus

efficaces. Les normes de description ont offert aux archivistes l'instrument nécessaire à la conception de ces nouveaux outils informatiques. C'est ainsi que sont apparus les formats EAD (Description archivistique encodée) et EAC-CPF (Contexte archivistique encodé- Collectivités, personnes, familles), qui permettent d'insérer dans des fichiers informatiques des codes ou « balises » permettant d'identifier la nature des informations et de développer à partir de là des outils d'interrogation et de recherche transverse.

Les normes et formats de description sont complémentaires les uns des autres dans la mesure où les informations sur les archives, leurs producteurs/détenteurs et les fonctions des producteurs peuvent être recueillies et mises à jour dans des descriptions séparées mais liées entre elles. ISAD(G) constitue ainsi le premier niveau de normalisation des descriptions archivistiques. Un usage conjoint d'ISAD(G) et d'ISAAR(CPF) est un deuxième niveau, mais l'élément Histoire administrative / Biographie d'ISAD(G) ne doit plus être utilisé pour éviter les répétitions de mêmes informations entre descriptions archivistiques proprement dites et notices d'autorité décrivant des producteurs. De même, une utilisation conjointe d'ISAD(G), d'ISAAR(CPF) et d'ISDF rendra caduques certains éléments d'ISAAR(CPF) qui ne devront donc plus être employés pour éviter les redondances.

Toutefois, il ne suffit pas de recenser et de définir les éléments de données. Organiser les informations pour construire un réseau de relations entre les éléments de données constitue l'étape suivante, afin de proposer un modèle générique permettant l'interopérabilité. Les archivistes australiens ont commencé à réfléchir à l'élaboration de modèles entités - relations, distinguant les entités (documents d'archives, producteurs ou détenteurs des documents et fonctions) et les relations entre ces entités. Dans ces modèles, les « séries » d'entités décrites et leurs relations ont des propriétés ou attributs. Suite à l'appel à commentaires pour les projets de normes sur les fonctions et les institutions de conservation, en 2007, les Australiens ont adressé au Conseil international des archives des propositions d'évolution des normes de description allant dans le sens d'un modèle unique de référence pour les normes de description, identifiant trois entités correspondant aux documents d'archives (ISAD), aux autorités (ISAAR) et aux fonctions (ISDF). Les attributs communs aux trois entités

(identifiant, type, intitulé/forme autorisée de nom, dates d'existence, etc.) serait regroupés dans une entité de niveau supérieur. Les relations entre ces trois entités pourraient elles-mêmes être qualifiées par un jeu d'attributs.

Par ailleurs, le système australien des « séries » a servi de cadre conceptuel pour un projet de recherche dirigé par Sue McKemmish à l'université de Monash visant à développer un schéma de métadonnées pour l'archivage (RKMS – Recordkeeping Metadata Schema). Le RKMS utilise différentes façons d'aborder la gestion des documents en vue de rendre explicites les relations entre les fonctions et activités, les agents (individus ou collectivités) qui exercent des fonctions et activités et les documents qui sont produits dans le cadre de ces fonctions et activités.

Un rapprochement des archives avec les bibliothèques et les musées en vue de l'élaboration d'un modèle conceptuel commun semble donc essentiel. Ce projet commun représente une grande opportunité de coopération avec l'IFLA et l'ICOM et les archivistes pourront bénéficier de l'expertise des musées et des bibliothèques en matière de modélisation conceptuelle. Ce rapprochement permettrait aussi une meilleure articulation des normes de description de l'ICA, problème majeur pour ceux qui veulent mettre en œuvre ces outils.

Le séminaire d'Helsinki a permis d'identifier plusieurs objectifs.

Le premier objectif est de développer un modèle conceptuel archivistique à partir des

quatre normes de l'ICA, ISAD(G), ISAAR(CPF), ISDF, ISDIAH, mais en se référant à d'autres projets (modèle basé sur les entités du logiciel ICA-AtoM, travaux en cours en Espagne pour développer un modèle conceptuel archivistique, schéma de métadonnées australien pour l'archivage, norme ISO 23081 sur les métadonnées de gestion des documents, etc.) et en collectant des exemples significatifs de normes nationales.

Il ne s'agit pas de développer un format d'encodage mais d'envisager les relations conceptuelles entre les différentes « entités » archivistiques (archives, producteurs, fonctions des producteurs) et de clarifier les concepts-clés (par exemple, comment définir un fonds d'archives ?). Il ne s'agit pas non plus d'atteindre une complexité technique telle que le résultat de ce travail ne soit compris que par des informaticiens.

Il s'agira dans un deuxième temps de comparer le modèle conceptuel archivistique ainsi élaboré avec les modèles élaborés pour les musées et les bibliothèques, afin d'harmoniser les processus métiers des musées, des archives et des bibliothèques à travers la compréhension des concepts communs qui se cachent derrière la documentation scientifique des différents métiers patrimoniaux.

Un groupe de travail conjoint (ICA-ICOM-IFLA) pourrait donc être créé officiellement dès la prochaine réunion du Comité des normes et bonnes pratiques du Conseil international des archives, à Paris, les 25-28 mai prochains.



## **Journées d'études sur la production et la publication électronique des instruments de recherche**

à l'occasion du lancement de la nouvelle version de l'application IREL (Instruments de recherche en ligne)

**Archives nationales d'outre-mer  
29 chemin du Moulin Detesta 13090 Aix-en-Provence**

*2-4 juin 2010*

En juin 2004, les Archives nationales d'outre-mer rendaient accessible sur Internet une application, IREL (Instruments de recherche en ligne), permettant l'interrogation et la consultation de l'état des fonds d'archives conservés dans leurs locaux, ainsi que des instruments de recherche propres à certaines séries (archives notariales, archives du ministère des affaires algériennes). Six ans après, c'est une nouvelle version d'IREL qui est mise à disposition des internautes, avec de nombreux instruments de recherche détaillés encodés en XML et des outils communs d'orientation (listes et notices d'autorité) permettant de guider plus efficacement l'utilisateur.

A l'occasion du lancement de cette nouvelle version, le Service interministériel des archives de France a souhaité organiser des journées d'études sur la production et la publication électronique de descriptions archivistiques structurées en XML/EAD (Description archivistique encodée). Développée à la fin des années 1990 par la Société des archivistes américains (SAA), l'EAD a connu un très vif succès dans de nombreux pays et elle est aujourd'hui le meilleur outil pour la mise en œuvre informatique des normes de description archivistiques. Elle est le format pivot de nombreux projets collaboratifs, parmi lesquels on peut citer le portail archivistique européen, et en France, près de quarante institutions (archives, bibliothèques, organismes de recherche, associations, etc.) donnent accès sur leurs sites Internet à des descriptions structurées en EAD.

Parallèlement, un autre format, l'EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles), vient d'être publié par la SAA. Parfaitement compatible avec la norme internationale sur les notices d'autorité (collectivités, personnes, familles) ISAAR(CPF), complémentaire du format UNIMARC/Autorités, l'EAC-CPF est une combinaison des notices d'autorité bibliographiques et des notices d'autorité archivistiques, qui donnent des informations sur le producteur et le contexte de production des documents.

Les participants présenteront les résultats de leurs recherches et leurs idées autour de trois grands axes :

- la normalisation des descriptions, un préalable indispensable à leur structuration en XML,
- les stratégies de production et de publication,
- l'EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles), un format pour la description contextuelle.

Ces journées se dérouleront aux Archives nationales d'outre-mer, 29 chemin du Moulin Detesta, 13090 Aix-en-Provence

**de 14h à 18h le mercredi 2 juin, de 9h à 17h30 le jeudi 3 juin et de 10h à 13h le vendredi 4 juin 2010**

L'inscription est gratuite. Les frais de déplacement et de restauration sont à la charge des participants.

L'inscription est cependant obligatoire en raison du nombre de places limité de la salle de conférences des Archives nationales d'outre-mer.

Si les demandes excédaient la capacité de la salle, nous demanderions aux services qui ont inscrit plusieurs personnes, d'annuler certaines inscriptions.

L'inscription doit être faite avant le 1er mai par courrier électronique auprès de Madame Isabelle Andréoli, Archives nationales d'outre-mer, en donnant nom, prénom, qualité, institution, toutes coordonnées et en particulier adresses électroniques.

Des attestations de présence pourront être retirées les 2 et 3 juin auprès de Madame Andréoli à l'accueil des Archives nationales d'outre-mer.

Courriel : [isabelle.andreoli@culture.gouv.fr](mailto:isabelle.andreoli@culture.gouv.fr)

Merci de diffuser cette annonce sur les listes que vous connaissez.

### Programme prévisionnel

#### **Mercredi 2 juin**

14h Accueil des participants

14h30 Allocution de bienvenue

14h45 Ouverture des travaux

*Première séance : la normalisation des descriptions, un préalable indispensable à leur structuration en XML*

15h Les normes de description et les formats apparentés

15h30 Élaborer des modèles de description pour produire de l'EAD : de Sosie au futur Système d'information archivistique des Archives nationales

16h Pause

16h15 La formation, facteur d'homogénéité de la production EAD

16h45 ICA-AtoM, un logiciel permettant de produire des descriptions conformes aux normes de l'ICA et des exports/imports en EAD

17h15 Questions

18h Visite des Archives nationales d'outre-mer

#### **Judi 3 juin matin**

*Deuxième séance : stratégies de production et de publication*

9h Production directe en EAD, conversion rétrospective en interne à partir d'instruments de recherche bureautiques et intégration d'inventaires en XML produits par externalisation : l'exemple des Archives départementales du Doubs

9h30 Entre exigence normative et pragmatisme : la reprise des bases de données documentaires aux Archives départementales du Var

10h L'EAD et ses relations possibles avec d'autres formats : utilisation et exploitation conjointe de l'EAD et de la TEI dans le cadre d'un projet sur les archives médiévales de l'abbaye de Saint-Denis

10h30 Pause

- 10h45 Calames : un réseau de catalogage en EAD pour l'Enseignement Supérieur
- 11h15 Quels modes de recherche et de consultation de descriptions EAD offrir aux usagers ? L'exemple du portail documentaire des Archives départementales du Val-de-Marne
- 11h45 EAD et interopérabilité : comment permettre le moissonnage de métadonnées normalisées par des portails (APEnet, Collections)
- 12h15 Questions
- 12h30 Buffet offert par les Archives nationales d'outre-mer

### **Jeudi 3 juin après-midi**

*Troisième séance : l'EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles), un format pour la description contextuelle*

- 14h30 Le schéma EAC-CPF : origines, structure, objectifs, impacts sur l'EAD, relations avec la norme sur les fonctions des producteurs
- 15h Structurer en XML des notices d'autorité de producteurs : exemple du Guide des sources de la Deuxième Guerre mondiale des Archives départementales du Morbihan
- 15h30 Comment passer de la DTD EAC au schéma EAC-CPF ? Le chantier scientifique « Producteurs » des Archives nationales
- 16h Pause
- 16h30 Produire des notices d'autorité et des référentiels pour guider plus efficacement les utilisateurs dans leurs recherches : la version 2 de l'application IREL
- 17h15 Questions
- 17h30 Conclusion

### **Vendredi 4 juin**

- 10h-13h Lancement de la nouvelle version de l'application IREL des Archives nationales d'outre-mer

---

### **Lu pour vous**

*Surmonter les obstacles de mise en œuvre de l'EAD (<http://www.oclc.org/research/activities/eadtools/>)*

*Le département Programmes et Recherche d'OCLC-RLG vient de mettre en ligne une étude identifiant les obstacles de mise en œuvre de l'EAD et proposant des solutions pour les surmonter. Les auteurs de ce document sont Michele Combs (université de Syracuse), Mark Matienzo (université de Yale), Merrilee Proffitt (OCLC Recherche) et Lisa Spiro (Université de Rice). L'objectif est d'aider les professionnels à mieux communiquer sur les avantages d'une implémentation du format. Malgré une utilisation de l'EAD très largement répandue, des réticences, d'ordre politique, logistique ou technique, existent encore.*

*Quels arguments utiliser pour convaincre les sceptiques du bien-fondé de l'utilisation de l'EAD ? Que répondre à des questionnements*

*du style « Qu'est-ce que l'EAD et pourquoi mon institution doit-elle s'investir dans sa mise en œuvre ? »*

L'EAD est un format d'encodage des instruments de recherche utilisé internationalement (Amérique du Nord, Europe, Australie, Nouvelle-Zélande, Asie) et développé pour répondre aux besoins à la fois des utilisateurs finaux et des archivistes. L'EAD est compatible avec les normes de description comme ISAD(G), élaborée par le Conseil international des archives, et DACS (« Décrire les archives : une norme de contenu ») élaborée par la Société des archivistes américains.

L'EAD est portée sur le langage XML, un format de données neutre qui assure la longévité des données en cas de migration d'un environnement logiciel à un autre. L'EAD assure la viabilité à long terme des données en s'attachant à la structure plutôt qu'à la mise en forme, contrairement au HTML.

On peut produire de l'EAD à partir d'autres formats comme le MARC ou le Dublin Core. Par ailleurs, on peut très bien faire de l'EAD à partir des données saisies dans une base de données relationnelle, via un filtre d'export, l'environnement de production n'est pas nécessairement un environnement EAD.

Un encodage en EAD facilite l'agrégation des données. Il serait difficile, voire impossible, de créer des portails thématiques (comme Archives Canada France) ou territoriaux (comme le portail archivistique européen) sans la cohérence imposée par l'EAD.

Il existe désormais de nombreux outils permettant d'implémenter l'EAD, de l'encodage à la publication.

Il existe également de nombreuses possibilités pour se former, s'informer et échanger (listes de discussions, formations, table ronde EAD, manuels de bonnes pratiques en ligne, etc.).

L'EAD facilite l'interrogation et la consultation des instruments de recherche par les utilisateurs.

Un encodage en EAD permet d'offrir de multiples points d'accès aux instruments de recherche.

Des chercheurs inexpérimentés peuvent utiliser des instruments de recherche beaucoup plus facilement. La cohérence des contenus et de la présentation facilite l'utilisation des descriptions par des utilisateurs novices. Par ailleurs, si un mode d'affichage est peu clair pour les usagers, il est possible de le modifier assez rapidement et assez facilement sans repartir à zéro.

Les chercheurs peuvent filtrer et affiner leurs recherches. La structuration des informations dans des balises permet par exemple de limiter les recherches aux intitulés/analyses des unités documentaires.

L'affichage des descriptions peut être adapté aux besoins des utilisateurs. Un fichier EAD peut avoir de multiples restitutions (version en ligne, version imprimée, etc.) en fonction des différents publics.

Une connaissance de l'EAD permet de comprendre les principes du XML et donc de travailler avec d'autres formats comme MARCXML, MODS ou METS.

Il n'y aura pas de place pour des données non structurées pour le Web du futur, le futur étant le « Web sémantique » ou « Web des données ». Mettre en œuvre l'EAD aide à positionner son institution par rapport à de futures applications Internet.

*« Qui fera le travail et quand ? » ou comment trouver des solutions de mise en œuvre peu coûteuses en temps et en argent ?*

Aux Etats-Unis, de nombreuses institutions se sont regroupées en consortium, pour mutualiser les ressources. Ces groupements offrent généralement des services comme des guides de bonnes pratiques, des feuilles de style, des fichiers et d'autres outils, des formations, l'hébergement de données. Il n'est pas nécessaire d'appartenir à un de ces groupements, qui mettent à disposition ces ressources tout à fait librement.

Avec des notices de niveau « fonds » au format MARC, il faut envisager la création de notices simples EAD par export (ce qui peut être facilement fait en utilisant un outil comme MARCEdit, outil libre permettant de convertir du MARC en XML ou au format Dublin Core : <http://people.oregonstate.edu/~reaset/marcedit/html/index.php>). Les fichiers issus de cette conversion permettront de contribuer à un consortium EAD. On peut décider ensuite d'enrichir ces notices minimales avec les niveaux inférieurs de la description ou se contenter de ce niveau minimal d'informations, suffisant pour faire connaître les ressources dont on dispose.

Il convient d'avoir une approche itérative. L'encodage EAD n'est pas fait une fois pour toutes, et ce n'est pas parce qu'un fonds n'est pas classé que l'on ne peut pas produire dans un premier temps une notice de haut niveau décrivant l'ensemble du fonds, que l'on complètera par la suite au fur et à mesure de

l'avancement du traitement des documents.

« *C'est si compliqué* ».

Il est possible de faire à l'avance, avec les éditeurs XML du marché, des documents EAD pré-remplis, avec par exemple les informations qui ne changeront jamais (nom de l'institution, adresse, etc.) et des commentaires guidant l'utilisateur dans le choix des balises.

Utiliser le schéma XML plutôt que la DTD permet de contraindre les contenus, par exemple les valeurs de l'attribut @normal, en réduisant les chances d'erreur.

L'encodage d'un instrument de recherche peut se faire via un formulaire HTML, de sorte que l'encodeur ne verra jamais de balises EAD.

Des feuilles de style peuvent aider à la saisie. Si on utilise oXygen, XMetaL ou d'autres éditeurs XML, des feuilles de style peuvent être automatiquement générées la première fois que l'on ouvre un document EAD. Ces feuilles de style peuvent être modifiées, complétées, etc. pour aider les encodeurs dans leur saisie.

« *Je ne sais pas par où commencer* ».

La solution est de prendre le projet étape par étape. Que est l'existant ? Quels sont les objectifs à atteindre. Les étapes sont généralement les suivantes : harmoniser les pratiques descriptives et réfléchir à l'application des normes de description, choisir une méthode d'encodage (éditeur XML, export depuis une base de données), saisir les données et les mettre en ligne. D'autres décisions concerneront la capacité du moteur de recherche, la mise à disposition de formats alternatifs (comme des versions imprimées), les liens à établir entre les descriptions et les images numériques, etc.

« *Je veux que mes données soient stockées dans un format flexible qui me permette d'évoluer* »

Tout processus de conversion s'accompagne de la nécessité de maintenir une certaine flexibilité et de la volonté de ne pas perdre d'information. La conversion des instruments de recherche papier au format électronique permet de réutiliser plus facilement les données. L'EAD offre plus de flexibilité puisque des équivalences peuvent être établies avec d'autres formats comme le MARCXML ou le Dublin Core. L'étude propose

trois solutions possibles.

1°) Convertir des fichiers Word en EAD. Une prestation externe est la solution la plus facile et la plus rapide, mais ce n'est pas nécessairement la plus économique ; une plus grande conformité des descriptions avec les normes de description permet d'économiser du temps et des efforts et des spécifications précises d'encodage devront être établies pour le prestataire. Une conversion en interne permet un contrôle direct sur les processus et les résultats, mais cela implique l'acquisition d'éditeurs XML et la manipulation de documents textuels en utilisant des scripts.

2°) Faire des exports en EAD à partir d'un système de gestion de contenu (CMS) : des logiciels comme Archivists' Toolkit (<http://www.archiviststoolkit.org/>) ou Archon (<http://www.archon.org>) permettent de produire de l'EAD. Si un système de gestion de contenu ne permet pas d'exporter directement de l'EAD, il faut se poser deux questions : les éléments EAD requis correspondent-ils à des champs distincts dans la base et la base de données exporte-t-elle du XML ?

3°) Migrer les données vers une base de données ou un système de gestion de contenu capable de produire de l'EAD pour un stockage permanent des données.

*Un obstacle majeur à l'adoption de l'EAD est l'impossibilité pour certaines institutions de mettre en ligne leurs descriptions encodées.*

Plusieurs solutions sont possibles : participer à un catalogue collectif, restituer du XML avec les navigateurs les plus récents (Internet Explorer 5, Firefox 9) en incluant une instruction de traitement dans le document XML pointant vers une feuille de style XSLT ou CSS, convertir les fichiers XML/EAD en fichiers HTML ou PDF, utiliser des outils bon marché permettant l'interrogation de fichiers HTML et XML (par exemple le moteur d'indexation et de recherche Swish-e), utiliser un système de gestion permettant aussi la publication Web (par exemple ICA-AtoM, le logiciel développé pour le Conseil international des archives) ou une plate-forme de publication XML comme XTF (<http://sourceforge.net/projects/xtf/>), Mark Logic (<http://www.marklogic.com/>), PLEADE (<http://pleade.com>), DLXS (<http://www.dlxs.org/>) ou Cheshire 3 (<http://www.cheshire3.org/>).

En annexe, on trouvera des liens vers les

ressources utiles (exemples de fichiers XML/EAD pré-remplis, formulaires Web pour la saisie, feuilles de style pour des éditeurs XML, feuilles de style XSLT, références d'éditeurs

XML, de systèmes de gestion de contenu, de plates-formes de publication XML, outils de migration ou de conversion, etc.).

---

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