

## Avantages et inconvénients d'un encodage en EAD

Claire Sibille

---

### Citer ce document / Cite this document :

Sibille Claire. Avantages et inconvénients d'un encodage en EAD. In: La Gazette des archives, n°220, 2010. Les instruments de recherche : évolutions, publics et stratégies. pp. 179-194;

[http://www.persee.fr/doc/gazar\\_0016-5522\\_2010\\_num\\_220\\_4\\_4757](http://www.persee.fr/doc/gazar_0016-5522_2010_num_220_4_4757)

---

Document généré le 15/03/2017

# Avantages et inconvénients d'un encodage en EAD

---

Claire SIBILLE-DE GRIMOÛARD

L'EAD a été développée à la fin des années 1990 par la Société des archivistes américains<sup>1</sup>. Au départ destinée à des utilisateurs américains, elle a connu un très vif succès dans de nombreux pays et elle est aujourd'hui le meilleur outil pour la publication électronique des instruments de recherche archivistiques. Elle est le format pivot de nombreux projets collaboratifs (portail archivistique européen). En France, l'intérêt pour l'EAD s'est manifesté dès 1999, avec les premiers tests d'encodage financés par le ministère de la Culture et de la Communication. Aujourd'hui, une quarantaine d'institutions (archives, bibliothèques, organismes de recherche, associations, etc.) donnent accès sur leurs sites Internet à des descriptions structurées en EAD. Depuis deux ans, les deux tiers des instruments de recherche soumis au visa des Archives de France se présentent sous forme de fichiers XML-EAD. Enfin, les instruments de recherche font de plus en plus souvent l'objet d'une conversion rétrospective, avec comme objectif la récupération de documents électroniques structurés et donc interrogeables par ordinateur, généralement au format XML suivant la DTD-EAD.

L'EAD fait ce pour quoi elle a été développée, à savoir permettre la mise en ligne de descriptions archivistiques structurées. Elle permet de produire des instruments de recherche pérennes et en principe échangeables, mais également d'atteindre de nouveaux publics pour de nouveaux usages. Pourtant, une mise en œuvre de l'EAD au sein d'un service n'est pas neutre et ne peut se faire que progressivement car elle implique des changements d'habitudes de travail et de vocabulaire, la coexistence de plusieurs systèmes, etc. Les critiques les plus vives envers l'EAD ont porté sur la complexité de sa structure, sur les

---

<sup>1</sup> La documentation officielle (ou dictionnaire des balises) et les fichiers de la DTD sont accessibles sur un site hébergé par la Bibliothèque du Congrès : <<http://www.loc.gov/ead/>> (tous les sites mentionnés dans cet article ont été consultés en janvier 2010).

coûts de balisage et sur les difficultés d'encodage par du personnel non spécialisé. L'application de l'EAD contraint également l'archiviste à adapter ses instruments de recherche : elle l'oblige, par exemple, à éclater les introductions traditionnelles entre plusieurs éléments de description de haut niveau afin d'obtenir des informations lisibles sur écran.

Quels sont les avantages et les inconvénients d'un encodage en EAD ? L'investissement en vaut-il la peine ? Quelles sont les contraintes ? Pour quels résultats ? Est-il possible d'encoder tous les instruments de recherche d'un service en EAD ?

Seront envisagés successivement les avantages et inconvénients de l'EAD en tant que telle, et les avantages et inconvénients des outils et des stratégies de mise en œuvre de l'EAD.

## **Avantages et inconvénients de l'EAD**

### *Avantages techniques*

Certaines caractéristiques techniques de l'EAD sont héritées du langage XML. Le langage XML permet d'utiliser et de manipuler le contenu des documents produits comme des données, d'échanger des données standardisées, de constituer de vastes réservoirs d'informations (portails archivistiques), de faciliter la recherche du public grâce à cette interopérabilité (les contenus sont structurés et peuvent donc être interrogés intelligemment), d'assurer longévité et migration aux données. En d'autres termes, l'EAD permet de passer d'une production « manuelle » et d'une interprétation des instruments de recherche par des humains à des modes de production et d'exploitation par des programmes informatiques dont la puissance peut être mise à profit.

L'EAD facilite également l'échange et la portabilité des instruments de recherche : il est donc possible d'avoir des présentations adaptées aux contextes d'édition et d'utiliser le même document encodé pour des restitutions multiples, en fonction des besoins et des publics visés. À partir d'un document XML (donc d'un fichier électronique structuré) on peut produire divers documents destinés à la consultation, donc présentés de façon pertinente pour un usage et un type de media donné : fichier PDF pour l'impression, fichier HTML pour la consultation au moyen d'un navigateur

Web. En effet, des programmes informatiques (par exemple, les programmes XSLT) peuvent s'appuyer sur la structure du document XML pour transformer le document en un autre format dédié. Les propriétés inhérentes au document XML placent donc les instruments de recherche au format XML-EAD au cœur des nouvelles chaînes éditoriales.

L'EAD permet la construction d'entrepôts de métadonnées OAI, première étape pour l'établissement de portails documentaires permettant d'accéder de manière homogène à des sources hétérogènes de données<sup>1</sup>. Les instruments de recherche structurés en XML-EAD accessibles sur le site d'une institution peuvent être « moissonnés », c'est-à-dire interrogeables par d'autres interfaces de recherche. Des équivalences doivent au préalable avoir été établies dans les fichiers EAD avec les métadonnées du format Dublin Core<sup>2</sup>, afin de prévoir l'identification future des documents EAD par une application utilisant le protocole d'interopérabilité OAI, au sein d'un portail documentaire externe. Par exemple, le portail Culture.fr/Collections<sup>3</sup> respecte les outils informatiques préexistants de stockage et de production de données comme la Base d'orientation et de recherche dans les archives (BORA)<sup>4</sup> ou la plateforme Instruments de recherche en ligne (IREL)<sup>5</sup>; le moteur de recherche sémantique utilisé produit des index en format XML et une passerelle permet le moissonnage des ressources *via* le protocole OAI-PMH. De même, le portail européen des archives s'appuiera sur l'EAD pour permettre de rendre accessibles des instruments de recherche et des documents numérisés conservés par des services d'archives de divers pays d'Europe<sup>6</sup>. Les ressources ainsi disponibles seraient également consultables par le portail Bibliothèque numérique européenne, dont le portail des archives pourrait constituer un maillon intermédiaire.

---

<sup>1</sup> L'*Open Archives Initiative* (initiative pour des archives ouvertes) ou OAI est un projet visant à faciliter l'échange et la valorisation d'archives numériques. Il permet à des fournisseurs de services de « moissonner » des métadonnées sur les sites de fournisseurs de données. Il est ainsi possible d'utiliser un protocole OAI pour créer un outil de recherche simultanée dans plusieurs applications informatiques. L'implémentation technique de ce concept est définie dans l'*Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH).

<sup>2</sup> Schéma de métadonnées générique permettant de décrire des ressources numériques ou physiques et d'établir des relations avec d'autres ressources.

<sup>3</sup> <[www.culture.fr/fr/sections/collections/accueil](http://www.culture.fr/fr/sections/collections/accueil)>

<sup>4</sup> La base BORA du Service interministériel des Archives de France comprend les descriptions des fonds d'origine privée et des fonds photographiques conservés dans les services d'archives publics ([www.archivesdefrance.culture.gouv.fr/ressources/bases-de-donnees/](http://www.archivesdefrance.culture.gouv.fr/ressources/bases-de-donnees/)).

<sup>5</sup> <[www.archivesnationales.culture.gouv.fr/caom/fr/index\\_irel.html](http://www.archivesnationales.culture.gouv.fr/caom/fr/index_irel.html)>

<sup>6</sup> Pour en savoir plus, voir le site officiel du projet : <[www.apenet.eu/](http://www.apenet.eu/)>

*Avantages « métier »*

L'EAD est conçue pour être utilisée avec des normes de description, en particulier ISAD(G). Développée ultérieurement, au départ sans lien avec le groupe d'experts du Conseil international des Archives, puis en collaboration avec lui, l'EAD est, dans sa version actuelle, totalement compatible avec ISAD(G). L'EAD est fondée sur le même principe de description par niveaux et à chacun des éléments définis dans ISAD(G) correspond un élément ou un attribut dans l'EAD. L'EAD, compatible avec l'ISAD(G), explicite celle-ci, la précise et, en quelque sorte, lui offre une application concrète. Avec l'EAD, il est possible de déplacer aisément des niveaux, d'en ajouter, d'affiner à volonté la description du fonds à la pièce, etc.

Pour autant, l'EAD ne remet absolument pas en cause les normes de description. La norme ISAD(G) reste un préalable indispensable. L'EAD a été conçue dès le départ comme un outil de publication et de consultation Web des instruments de recherche alors qu'avec la norme ISAD(G), il est possible de faire des instruments de recherche papier ou électroniques. L'informatique n'est qu'un outil, cet outil ne pouvant être mis en œuvre que sur des instruments de recherche rédigés de manière homogène et structurée. Un instrument de recherche ou un fichier manuscrit établi avec beaucoup de rigueur pourra être rapidement encodé en EAD et sans trop de changements. Inversement, un fichier informatique établi sans principe directeur stable ne pourra pas être structuré en EAD.

L'encodage nous contraint à une grande rigueur dans la structuration des descriptions, et l'encodage d'instruments de recherche déjà saisis sur traitement de texte révèle sans aucune concession la moindre faille logique. Par exemple, les informations contenues dans le guide des sources de l'esclavage ont été structurées conformément aux éléments de la norme ISAD(G). Toutefois, l'élément « Présentation du contenu » a été renseigné de manière très différente. Certaines notices présentaient, en effet, le plan de classement d'un fonds, d'autres montraient l'intérêt de séries organiques ou de groupes d'articles sous forme de texte rédigé, d'autres enfin mentionnaient quelques exemples de documents représentatifs de la thématique du guide. Un encodage simple a donc été retenu<sup>1</sup>, ce qui a conduit à utiliser des éléments de mise en forme (listes) pour différencier les différents niveaux existant au sein de la présentation du contenu.

---

<sup>1</sup> Utilisation de la balise <scopecontent> équivalant à l'élément « Présentation du contenu » d'ISAD(G)) et non pas de l'élément Description des sous-composants <dsc> et de la hiérarchie des composants.

L'EAD permet de créer de nouveaux instruments de recherche mais elle est également très bien adaptée à des opérations de conversion rétrospective, notamment par la souplesse de l'ordre de ses éléments. Toutefois, cela suppose que les instruments de recherche soient parfaitement conformes aux règles de la description à plusieurs niveaux. L'EAD a été conçue de manière à ce qu'elle puisse fonctionner avec toutes les formes d'instruments de recherche existants, même s'il est souvent nécessaire de retravailler les instruments de recherche avant leur encodage. L'EAD rend néanmoins possible la publication sur le Web d'instruments de recherche anciens avec toutes leurs faiblesses et leurs défauts.

L'EAD permet d'intégrer aux instruments de recherche des liens vers des représentations numériques des documents décrits. Mais les différentes expériences en cours aux Archives nationales ou en Archives territoriales pour la publication sur Internet de corpus organiques complets de documents d'archives numérisés ont montré que la mise en relation informatique des instruments de recherche avec les documents qu'ils décrivaient n'était pas sans conséquence pour le contenu des descriptions documentaires. Il est probable qu'en préparant cette mise en relation, l'archiviste ressent la nécessité d'affiner l'instrument de recherche pour rendre plus commodes la visualisation ou le feuilletage des images numériques.

Il est possible de choisir la finesse de l'encodage en fonction des besoins et des moyens, l'EAD comportant 146 éléments dont très peu sont obligatoires (entête EAD, description du fonds, niveaux hiérarchiques). Il est par exemple inutile de renseigner tous les attributs LEVEL (niveaux de description) d'un fichier EAD si ces données ne sont pas utilisées au moment de la publication, par exemple pour fragmenter l'instrument de recherche en unités documentaires interrogeables.

### *Inconvénients techniques*

L'EAD est une DTD. Le schéma développé en 2007 par la Société des archivistes américains n'est en fait qu'un clone de la DTD de 2002 et n'est pas abouti : il présente de nombreuses imperfections du point de vue des éléments et attributs Xlink<sup>1</sup> et n'est donc pas utilisé pour le moment. La DTD de 2002 reste donc la version de référence. Comme toutes les DTD, l'EAD est assez permissive. Un schéma est plus contraignant car il permet d'homogénéiser les contenus. Par exemple, il est utile de normaliser l'écriture des dates pour

---

<sup>1</sup> Technologie permettant de créer des liens entre fichiers XML ou portions de fichiers XML.

permettre l'interrogation de descriptions EAD avec des critères de dates. Avec une DTD, on doit s'obliger à entrer une forme normalisée. Avec un schéma XML, il est possible de définir que l'écriture des dates sera conforme à un modèle donné, la norme ISO 8601 par exemple<sup>1</sup>, et la saisie des dates normalisées se fera automatiquement. La transformation de l'EAD en schéma devrait donc permettre une saisie plus cohérente des données. Elle devrait également donner la possibilité aux utilisateurs du format d'inclure le schéma EAD dans d'autres schémas ou d'inclure d'autres DTD ou d'autres schémas dans l'EAD, et enfin de gérer la documentation de l'EAD et les règles de gestion dans un document unique, de définir des profils d'utilisateurs de l'EAD (choisir par exemple d'utiliser tel élément plutôt que tel autre).

L'EAD n'est pas conforme aux règles d'écriture des schémas XML, ce qui explique pourquoi elle n'a pas été intégrée en tant que telle dans le standard d'échanges de données pour l'archivage (SEDA) développé par les Archives de France et la Direction générale de la modernisation de l'état (DGME). À la suite d'un premier appel à commentaires, fin 2005, l'élaboration du SEDA a en effet été réorientée suivant les règles définies au niveau international par l'UN-CEFACT<sup>2</sup>, organisme dépendant des Nations unies spécialisé dans les échanges de données. Or, les règles de l'UN-CEFACT ne permettent pas de supporter le format EAD (utilisation de modèles de données communes remplaçant certaines balises EAD, règles de nommage des balises incompatibles, les abréviations ne sont pas permises, besoin d'éléments spécifiques du numérique non présents dans l'EAD). En revanche, il est tout à fait possible, après transfert au service d'archives, d'exporter les métadonnées et de les convertir en EAD.

L'EAD est centrée sur le document et devrait être plus cadrée sur les données. L'EAD mélange les éléments structurant l'information et les éléments de mise en forme (titres de rubriques, paragraphes, listes, tableaux, etc.). L'EAD a été conçue comme une transposition des habitudes issues du papier, enfermant ainsi les instruments de recherche dans une seule vision documentaire. Elle se concentre sur un mode de représentation de l'information, le document structuré, alors que le schéma EAC-CPF, en cours de développement par la Société des archivistes américains pour l'encodage en XML des notices

---

<sup>1</sup> La norme internationale ISO 8601 spécifie la représentation numérique de la date et de l'heure. Cette notation est particulièrement destinée à éviter tout risque de confusion dans les communications internationales dû au grand nombre de notations nationales différentes. De plus, cette notation a de nombreux avantages pour une utilisation informatique par rapport aux autres notations.

<sup>2</sup> En anglais : *United Nations Centre for Trade Facilitation and Electronic Business*.

d'autorité de producteurs, a été conçu comme un ensemble structuré de métadonnées. Certains professionnels, inspirés par la flexibilité du Web des données, commencent à ne plus voir les instruments de recherche EAD comme des documents mais plutôt comme des collections de données structurées. Lors d'un dernier congrès de la Société des archivistes américains à San Francisco en 2008, Daniel Pitti<sup>1</sup> signalait ainsi l'existence, à l'université de Virginie, d'un séminaire sur la conception de systèmes de description archivistiques. Les participants au séminaire tentent « d'explorer en détail les technologies de balisage et de bases de données, les efforts entrepris pour intégrer les deux technologies et l'opportunité offerte par cette intégration pour transformer la description archivistique ». Daniel Pitti s'est pris à imaginer le jour « où les bases de données et les technologies de balisage seraient utilisées de manière complémentaire pour former un système de description archivistique complet et flexible qui créerait un réseau d'interrelations entre les descriptions des archives, les descriptions des producteurs et les descriptions des fonctions et activités »<sup>2</sup>. La structure hiérarchisée des fonds serait affichée à l'écran mais la structure de base de données permettrait l'établissement de liens multiples au sein des descriptions EAD, faisant apparaître les relations au sein d'un fonds qu'une représentation hiérarchique ne révèle pas nécessairement.

Si les travaux de Daniel Pitti sont annonciateurs des évolutions des modalités d'accès aux descriptions archivistiques, faut-il pour autant abandonner l'EAD ? Les archivistes devraient plutôt s'inspirer des problèmes rencontrés par les bibliothécaires pour la conversion de leurs fichiers bibliographiques en catalogues accessibles en ligne. C'est l'homogénéité des données dans les fichiers qui a fait la différence en matière de coût, d'efficacité et de succès des projets de conversion dans les bibliothèques. Pour ce qui est de l'EAD, nous devons faire porter tous nos efforts sur une homogénéité de l'encodage des instruments de recherche archivistiques. Les éléments EAD que nous encodons aujourd'hui de manière homogène seront plus facilement repris par une nouvelle technologie plus centrée sur les données.

#### *Inconvénients « métier »*

L'EAD est un format générique par nature ; comme la norme ISAD(G), elle est utilisable pour tout type de fonds d'archives ou d'instrument de recherche, elle ne traite guère du contenu des éléments, elle est peu contraignante et elle est très souple. L'encodage, même encadré, laisse donc une certaine liberté

---

<sup>1</sup> Directeur adjoint, *Institute for Advanced Technology in the Humanities*, université de Virginie.

<sup>2</sup> Intervention lors réunion de la table ronde de l'EAD.

dans la précision de la description ou dans le nombre d'informations, souvent facultatives, que l'on va ou non renseigner. Il est possible, pour un même instrument de recherche, avec le même niveau de précision, de réaliser plusieurs encodages différents. Autre point à préciser : celui de la définition des niveaux de description, définition rendue plus difficile en France par nos cadres de classement qui ont une existence parallèle aux notions de fonds. Par ailleurs, la puissance de l'encodage permet un nombre théoriquement infini de niveaux de descriptions. Pour faciliter la compréhension de l'utilisateur, il faut exercer une certaine retenue dans la multiplication de ces sous-divisions et s'assurer qu'à la création d'un niveau intermédiaire correspond la réalité d'une activité, d'une fonction ou d'un service administratif. On doit éviter de créer des niveaux artificiels ne correspondant à aucune réalité dans les activités du producteur ou pour lesquels aucun document d'archives ne sera conservé.

Pour mettre en œuvre l'EAD dans un service ou dans le cadre d'un projet, il faut donc réfléchir et écrire des spécifications (comme lorsque l'on crée une base de données documentaire). C'est ainsi que parallèlement à la conversion rétrospective du Catalogue général des manuscrits des bibliothèques publiques de France et à la mise en ligne de différents catalogues collectifs (Catalogue collectif de France, ou CCFr, et Calames), un guide des bonnes pratiques de l'EAD en bibliothèques est en cours d'élaboration<sup>1</sup>. Le guide n'est pas axé sur la conversion rétrospective et se base sur les recommandations d'un groupe d'experts de l'AFNOR pour la description des manuscrits modernes et contemporains. Ce guide est destiné à aider les bibliothèques pour la réalisation des spécifications de mise en œuvre de l'EAD. De même, les Archives nationales ont rédigé un manuel pour les utilisateurs de l'outil SOSIE, avec des recommandations sur l'utilisation des différents éléments et attributs de l'EAD.

Plus généralement, il convient de souligner que le respect formel des normes et des formats d'échange ne garantit pas à lui seul la qualité des descriptions archivistiques. Certes, la norme ISAD(G) et sa déclinaison sous la forme de DTD offrent à la fois un cadre bien conçu pour la saisie des informations et des possibilités d'interrogation élargie et d'échanges de notices descriptives,

---

<sup>1</sup> Le Catalogue collectif de France intègre, depuis 2008, une partie du Catalogue général des manuscrits des bibliothèques publiques de France, converti rétrospectivement en EAD ; il s'agit de la partie consacrée aux bibliothèques municipales qui a été versée dans le CCFr : <[www.ccf.fr/bnf.fr/](http://www.ccf.fr/bnf.fr/)>.

Calames est le catalogue collectif des archives et des manuscrits des bibliothèques universitaires françaises mais également de grands établissements nationaux (Institut de France, Muséum d'histoire naturelle, etc.) et de plusieurs bibliothèques de recherche (Bibliothèque littéraire Jacques-Doucet) : <[www.calames.abes.fr/pub/](http://www.calames.abes.fr/pub/)>

mais la description, opération intellectuelle, ne peut se réduire à une simple technique. Ce sont bien les informations que l'archiviste va mettre derrière chaque niveau et chaque élément de description qui importent, et cela relève d'une démarche scientifique que ne garantit pas la normalisation.

## **Avantages et inconvénients des stratégies d'encodage et de mise en ligne de descriptions EAD**

Les avantages et les inconvénients de l'EAD sont également très dépendants de la méthode et des choix d'encodage et de mise en ligne.

### *Encodage courant ou conversion rétrospective ?*

Un service ayant un arriéré dans le traitement et la description des fonds qu'il conserve se mettra peut-être plus facilement à l'EAD qu'un service ayant un corpus hétérogène d'instruments de recherche produits à différentes époques et pas toujours conformes aux normes de description. Il est parfois plus facile de commencer un projet EAD par l'encodage courant plutôt que par des opérations de conversion rétrospective. Toute conversion doit en effet être précédée d'une étude approfondie de l'instrument de recherche à convertir, voire de sa restructuration préalable. Bien souvent, on a tendance à privilégier l'intérêt du fonds au lieu de commencer par évaluer l'instrument de recherche lui-même et se demander s'il est un bon candidat pour l'encodage. En cas de conversion rétrospective, le dialogue avec le prestataire est essentiel. C'est ainsi que les formations EAD organisées aux Archives de France, à l'Association des archivistes français, etc. ont vu apparaître un nouveau public : celui des agents chargés de la rédaction de cahiers des charges et du suivi des opérations d'encodage par un prestataire extérieur. Ces agents doivent être en mesure de relire de manière critique les fichiers XML-EAD pour pouvoir déceler les anomalies d'encodage. Toute prestation doit en effet être validée au moyen de tests techniques (validité, tests de publication avec une interface de production ou de publication) et de conformité archivistique (maîtrise des normes archivistiques, qualité de la structuration et de l'encodage).

Dans tous les cas, il faut se dire que tous les problèmes ne peuvent pas être résolus au moment de la conversion : le résultat n'est jamais parfait, il faut

souvent faire des reprises manuelles ; ce n'est pas l'EAD qui remédiera aux imperfections de l'instrument de recherche initial (description à plusieurs niveaux non mise en œuvre, hétérogénéité des notes de bas de page, etc.).

### *Que faut-il encoder ?*

Chaque service doit faire ses choix selon l'état initial des instruments de recherche et selon les moyens dont il dispose. Certains services offrent un accès homogène à l'ensemble de leurs instruments de recherche, tous structurés en XML-EAD. Leurs sites permettent d'accéder à un état des fonds XML-EAD permettant un premier repérage et à des instruments de recherche détaillés, également encodés en XML-EAD, accessibles directement ou à partir de la présentation générale du fonds et interrogeables grâce à un moteur de recherche. Les outils utilisés par ces services pour la publication et la consultation de leurs descriptions EAD génèrent des pages dynamiques qui, bien plus que le format HTML statique, permettent de rendre compte de la richesse du format EAD concernant la structuration et la contextualisation de la description des fonds d'archives. L'internaute peut ainsi explorer les descriptions archivistiques de manière approfondie, et accéder, lorsque les fonds ont été numérisés, aux documents eux-mêmes.

Toutefois, que faire lorsqu'on a un corpus d'instruments de recherche hétérogènes, produits à des époques très différentes et pas toujours parfaitement conformes à la description à plusieurs niveaux ? Dans ce cas de figure, l'encodage en EAD devra être réservé en priorité aux descriptions de haut niveau (état des fonds) et les instruments de recherche détaillés ne pouvant pas être encodés à moins de reprendre le classement des fonds, pourront simplement être transformés en fichiers pdf et rattachés à l'état des fonds. On pourra décider de s'arrêter à cette étape, l'avantage étant bien évidemment la facilité et la rapidité de l'opération, ainsi que la facilité pour l'internaute d'imprimer l'inventaire sur un support papier. Toutefois, le mode de recherche ne sera pas structuré : il sera impossible de faire une recherche sur un champ précis et de restituer le résultat dans son contexte de description.

Comment faire de l'EAD à partir de fichiers électroniques structurés ? Les bases de données documentaires contiennent des notices structurées : il est donc possible d'opérer une conversion vers le format XML. Les étapes seront les suivantes : étudier la structure source et la structure cible, faire ou faire faire un programme de conversion, reprendre le résultat, etc.

Beaucoup d'instruments de recherche n'existent que sous la forme d'un imprimé, d'un dactylogramme ou d'un fichier issu d'un traitement de texte. Dans de tels cas, l'information n'est pas structurée ; sa présentation peut toutefois donner des indications. La conversion d'instruments de recherche non structurés comporte en général plusieurs phases : éventuellement, numérisation et reconnaissance optique des caractères, étude de l'existant, identification des éléments de données et de leurs caractéristiques visuelles, des niveaux de description (y compris implicites), des données posant problème (notes de bas de page, index, introductions, etc.), conversion automatique (par programme) de ce fichier texte vers XML-EAD, et enfin relecture attentive et reprise du document XML-EAD.

### *Quelles stratégies ?*

Avant de produire de l'EAD, il faut toujours se poser certaines questions pour connaître les besoins du service (et choisir les réponses appropriées en fonction des avantages et des inconvénients de telle ou telle solution) :

- quel est l'environnement d'informatique documentaire (GED, systèmes de gestion de bases de données, bureautique, sécurité du réseau, schémas directeurs) ?
- qui va produire : combien de personnes, quelle formation ?
- quels fonds vont être traités : fonds sériel, fermé, ouvert, manuscrit ?
- quel type d'instruments de recherche va être produit : répertoire synthétique ou analytique, catalogue ?
- quel va être l'usage du document EAD : publication en ligne (intranet ou Internet), publication papier ?

En fonction des réponses apportées, on peut être amené à choisir de confier la production à une société extérieure, si l'établissement n'a pas assez de ressources pour investir (matériel, formation) et si de grandes quantités sont à produire rapidement (conversion rétrospective), ou de s'orienter vers une production interne. Toutefois, il convient de noter que les circonstances propres à chaque service peuvent conduire à prendre des décisions mêlant le recours à une société extérieure et une production en interne. Par exemple, les Archives nationales d'outre-mer ont réalisé en interne l'encodage de leur état des fonds, ce qui a impliqué la formation de tous les agents chargés du traitement des fonds à l'utilisation d'un éditeur XML. En revanche, l'encodage

de certains inventaires détaillés décrivant des fonds sériels a été externalisé. Inversement, en 2003-2004, dans le cadre d'un programme de numérisation, les Archives départementales des Pyrénées-Atlantiques ont confié l'encodage d'une quarantaine d'instruments de recherche papier à un prestataire privé, tandis qu'un autre projet visait à récupérer sous un logiciel et à exporter en EAD plus de 500 instruments de recherche papier. Quant aux Archives départementales du Val-de-Marne, elles ont travaillé avec la société produisant leur progiciel de gestion des archives à la réalisation d'un import et d'un export de leurs instruments de recherche en EAD<sup>1</sup>.

Les choix faits lors de la production auront des répercussions sur la publication. Ainsi, il ne sert à rien d'encoder finement si l'outil de publication et de consultation n'exploite pas la granularité de l'encodage. Quand on réfléchit à l'affichage des instruments de recherche, il est nécessaire de faire des essais de publication avec un outil gratuit.

Une attention particulière doit également être portée à l'affichage des résultats : il faut veiller à ce qu'ils soient toujours contextualisés. Tout résultat doit être affiché dans une hiérarchie de manière à situer l'unité documentaire par rapport aux niveaux de description supérieurs ou inférieurs.

### *Quels outils ?*

Quels sont les avantages et les inconvénients des outils permettant une production d'instruments de recherche EAD en interne ?

Les éditeurs XML sont les outils les plus appropriés pour produire des inventaires rédigés de fonds complexes (structure complexe, description très détaillée, etc.) et ils peuvent être paramétrés en fonction des besoins particuliers (macros, habillages, etc.). La contrainte de structuration rigoureuse de son instrument de recherche, au fur et à mesure de sa réalisation, est sans doute le point principal si on compare au travail sur traitement de texte. La possibilité d'enrichissements est infinie, qu'il s'agisse des balises de structure, essentielles, définissant les différentes zones de l'instrument de recherche et leurs relations hiérarchiques, ou les balises permettant d'apporter des enrichissements typographiques (comme le gras, l'italique) ou encore des précisions sur la nature d'une partie de la zone (préciser qu'il s'agit d'une citation, d'une référence bibliographique, d'un sigle, d'un nom de personne, etc.).

---

<sup>1</sup> La base d'orientation des Archives départementales du Val-de-Marne est accessible en ligne : [<archives.cg94.fr/>](http://archives.cg94.fr/)

Toutefois, le gros reproche que l'on peut faire aux éditeurs XML est qu'ils sont d'une prise en main plus ou moins ergonomique et qu'il faut se les approprier par une pratique régulière. Par exemple, celui utilisé aux Archives nationales (Paris) dès 2000-2001 dispose d'une configuration française des fichiers de la DTD-EAD, mais la lourdeur de son maniement (insertion manuelle des balises et des attributs) a été un frein à l'implémentation de l'EAD dans cette institution. Par ailleurs, les éditeurs XML ne gèrent pas les documents XML produits (stockage, recherche, gestion des accès, qui sont alors à faire porter par un autre logiciel, par exemple par une base native XML).

Les logiciels documentaires avec des fonctions d'export en EAD représentent une alternative intéressante. Ce sont les outils les plus appropriés pour décrire des documents dont la nature se prête à une saisie dans une base de données structurée ; ils présentent l'avantage d'être très ergonomiques et de pouvoir être pris en main rapidement. Les agents chargés de rédiger des instruments de recherche avec ces logiciels utilisent l'EAD de manière transparente, sans avoir à connaître les 146 éléments de la norme, le seul pré-requis étant de connaître la norme ISAD(G). Le temps consacré à l'encodage sera quasi nul et l'homogénéité des instruments de recherche sera assurée, pour autant que l'éditeur ait procédé aux « bons choix » parmi les options possibles. Si tous ces outils se réfèrent de façon plus ou moins complète à ISAD(G), ils offrent en effet des possibilités plus ou moins souples d'adaptation (dépendantes du prestataire et de l'évolution du produit) et les exports EAD doivent souvent être retravaillés (enrichissements possibles). Des instruments de recherche déjà saisis ou importés dans le logiciel seront automatiquement encodables *a posteriori* et pratiquement sans frais. Quelques compléments d'information de haut niveau seulement seront parfois nécessaires. Néanmoins, rares sont les logiciels d'archives permettant pour le moment d'importer des documents XML-EAD enrichis.

Une dernière solution consiste à utiliser les produits bureautiques classiques : saisie directe en XML, outils de conversion de bases de données dont certains sont disponibles sur le site officiel EAD, etc. Les Archives nationales ont fait développer par un prestataire un outil de conversion d'instruments de recherche en XML-EAD à partir d'OpenOffice. Ce projet est le prolongement des travaux du groupe EAD aux Archives nationales qui avait essayé de définir des règles de bonnes pratiques de l'EAD en préconisant ou en déconseillant l'utilisation d'éléments et attributs. SOSIE constitue une solution provisoire en attendant le développement du système d'information du centre de Pierrefitte-sur-Seine. SOSIE sera utilisé aussi bien en amont par les missions dans les ministères qu'aux Archives nationales à Paris et à Fontainebleau.

Dans tous les cas, il faut qu'il y ait adéquation entre les besoins de la production et l'outil retenu. Dans un service d'archives utilisant un logiciel de gestion, il peut suffire d'utiliser les fonctions d'export du logiciel en spécialisant une ou deux personnes (équipées d'un éditeur de XML) dans le contrôle et l'enrichissement des documents EAD automatiquement produits. Un service peut aussi envisager une solution mixte en fonction de ses ressources humaines et financières.

Avant d'offrir au public ses instruments de recherche, il faut là aussi se poser certaines questions et choisir les réponses appropriées. Les questions de publication et de recherche sont, en effet, indissociables des choix qui sont faits lors de la production des instruments de recherche. Souhaite-t-on diffuser les instruments de recherche en intranet ou sur Internet ? Souhaite-t-on publier les instruments de recherche sous plusieurs formes (inventaire imprimé et inventaire en ligne) ? Souhaite-t-on se contenter d'un affichage statique type HTML ? Souhaite-t-on donner la possibilité de manipuler les informations pour la recherche (Web dynamique, conversion des données vers d'autres formats pour permettre le moissonnage par d'autres applications) ? En fonction des réponses apportées, on peut être amené à choisir des solutions très différentes. À partir d'un fichier XML-EAD, on peut en effet produire différents documents présentés de façon pertinente pour un usage et un type de media donné (fichier PDF pour l'impression, fichier HTML pour la consultation au moyen d'un navigateur Web). Des programmes informatiques peuvent s'appuyer sur la structure du document XML-EAD pour transformer ce document en un autre format dédié. L'utilisation d'un outil de gestion et de publication de documents XML permettant la production de pages dynamiques est la solution la plus élaborée. Les avantages sont multiples : facilité de mise à jour des informations, ajustement beaucoup plus fin de l'information fournie à l'internaute en fonction de sa question, richesse de la recherche et pertinence forte des résultats, utilisation d'un format normalisé qui pourra faciliter l'intégration dans des portails à venir, pérennité sur le long terme des fichiers XML-EAD. Toutefois, les outils de gestion et de publication de documents XML sont encore très peu nombreux, qu'ils soient propriétaires ou libres.

Dans tous les cas, il n'existe pas de solution ne demandant pas un investissement de la part de l'établissement : lors de la rédaction d'un cahier des charges complet et cohérent (pour acquérir un logiciel propriétaire ou pour faire paramétrer un logiciel libre), ou pour s'approprier les logiciels (« libres » ou non), ou encore pour garantir l'interopérabilité entre les systèmes.

## **Défis et perspectives**

### *Faire évoluer l'EAD pour la rendre plus conforme aux évolutions technologiques*

La diffusion du schéma EAC-CPF, plus conforme aux règles d'écriture du XML<sup>1</sup>, rend inéluctable une prochaine révision de l'EAD. Il faudrait ainsi rendre les éléments plus explicites et plus conformes aux règles actuelles d'écriture des schémas XML et utiliser les bibliothèques de modèles XML, l'usage d'éléments et d'attributs génériques normalisés étant un gage d'interopérabilité.

Toutefois, s'il est nécessaire de faire évoluer l'EAD vers un schéma XML qui soit la version de référence, il semble préférable de repartir de la DTD de 2002, base conceptuelle de l'EAD et dont le schéma développé en 2007 n'est qu'une formalisation.

Pour autant, en ce qui concerne la nécessité de faire évoluer l'EAD par rapport aux demandes des utilisateurs, il faut faire preuve de prudence. Ce n'est pas parce que l'on rencontre des difficultés dans la mise en œuvre de l'EAD qu'il faut nécessairement apporter des changements : les pratiques peuvent être mauvaises ou les outils inadaptés.

### *Normaliser pour partager avec d'autres communautés*

Les archivistes ont la chance de disposer d'un arsenal de normes informatiques internationales et vivantes, une occasion unique de partager savoir-faire, ressources et outils dans le domaine informatique. D'autres communautés patrimoniales ont besoin des normes, savoir-faire et outils des archivistes. Du fait des limites des formats MARC, l'EAD a très vite suscité l'intérêt des bibliothèques pour décrire de manière satisfaisante leurs fonds d'archives ou de manuscrits. Désormais, il n'est pas rare que des bibliothèques proposent des inventaires en EAD à côté de leur catalogue en MARC, voire un « catalogue » en EAD avec des recherches transverses sur les différents inventaires publiés. On peut citer comme exemple le catalogue Calames<sup>2</sup>, mis au point par l'Agence bibliographique de l'enseignement supérieur (ABES), ou BnF archives et manuscrits, regroupant les inventaires en EAD de la Bibliothèque nationale de France.

---

<sup>1</sup> Éléments et attributs plus explicites, approche plus « sémantique », possibilité d'inclure des segments XML conformes à d'autres DTD ou schémas XML dans une instance EAC-CPF, homogénéisation des mécanismes d'établissement de liens avec d'autres entités, avec des ressources archivistiques ou avec des fonctions.

<sup>2</sup> Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur.

L'utilisation de l'EAD par les bibliothèques les a par ailleurs conduites à prendre connaissance de la norme à laquelle l'EAD fait référence, l'ISAD(G), à revoir les règles de description des manuscrits et fonds d'archives à la lumière des principes et éléments de l'ISAD(G) mais également à se demander en quoi l'EAC-CPF définit un traitement des données d'autorité différent de celui exercé traditionnellement dans les bibliothèques.

*Homogénéiser le rendu des descriptions archivistiques*

En diffusant sur Internet leurs instruments de recherche, les services d'archives permettent à un public plus large d'accéder à la description de leurs collections. Or, la plupart des internautes sont peu familiarisés avec la terminologie archivistique et la structure des instruments de recherche. Les professionnels se sont intéressés jusqu'à présent à la structuration des contenus mais pas à l'homogénéisation de leur affichage sur Internet, d'où une grande disparité de présentation des descriptions archivistiques sur les quelques sites d'Archives nationales ou territoriales françaises utilisant l'EAD. Les études utilisateurs montrent que l'utilisation d'un vocabulaire « technique » est à reconsidérer sur Internet et que les modalités de navigation au sein de la hiérarchie des niveaux de description ne sont pas toujours bien comprises des utilisateurs. La stagnation voire la diminution de la fréquentation des salles de lecture et la spectaculaire progression des consultations en ligne des archives rendent vitale la publication en ligne des instruments de recherche. Une plus grande attention portée aux utilisateurs, avec des facilités offertes pour la consultation et l'interrogation des instruments de recherche, devra donc être au cœur des préoccupations des prochaines expériences de mise en œuvre des normes et formats d'encodage.

Claire SIBILLE – DE GRIMOÛARD  
Service interministériel des Archives de France  
claire.sibille@culture.gouv.fr