



*Production
et publication des instruments de recherche archivistiques
au format XML/EAD :
quelques repères, et présentation d'outils*

MCC/DAF/CHAN/SNTIC/F. Clavaud, réunion du groupe Pérennisation des informations numériques, CNES, 4 mai 2004

1



I - Produire des documents XML/EAD

2



A 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 :
 - étudier la structure source et la structure cible ;
 - faire ou faire faire un programme de conversion
 - reprendre le résultat...
- Exemple : la conversion en XML/EAD, au moyen de programmes XSL-T, des notices documentaires de la base ARCHIM du CHAN : cahier des charges ; résultats

3



A partir d'instruments de recherche non structurés

- Beaucoup d'IR 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
- Deux exemples très différents :
 - l'état des fonds d'archives privées du CHAN
 - l'inventaire du fonds Napoléon

4



Conversion en XML/EAD d'instruments de recherche non structurés (suite)

- 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...)
 - conversion automatique (par programme) de ce fichier texte vers XML/EAD
 - relecture attentive et reprise du document XML/EAD...

5



Production directe de documents XML/EAD

- Peut se faire au moyen de divers outils :
 - un SGBD permettant de faire un export des données vers le format XML/EAD (ex. : Arkheïa de la société Anaphore)
 - un traitement de texte configuré pour permettre une conversion vers XML par programme sur la base des styles employés
 - un logiciel éditeur de documents XML

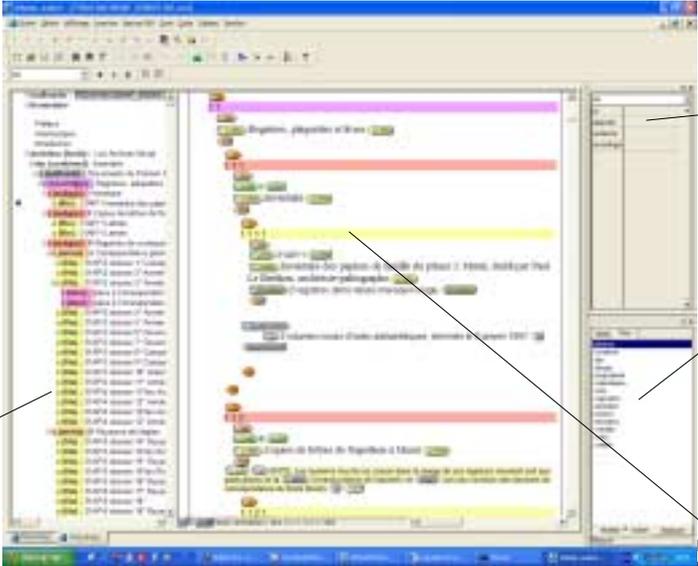
6

Un exemple de logiciel éditeur de documents XML

- XMetaL de Blast Radius (<http://www.blastradius.com>)
 - permet de saisir un document XML et d'en vérifier la conformité par rapport au modèle choisi
 - quatre fenêtres interactives de travail (structure XML, fenêtre de saisie, éléments, attributs)
 - les contenus de ces fenêtres et leur apparence peuvent être adaptés, diverses fonctionnalités mises en œuvre
 - il existe une configuration française de XMetaL pour la DTD EAD, propriété du CHAN et gratuitement diffusée. Sera probablement mise sous licence libre (GPL) bientôt.

7

Un écran de travail XMetaL 4



The screenshot displays the XMetaL 4 software interface. It features a central workspace with XML content, a left-hand pane showing a hierarchical tree structure of the document, and a right-hand pane for editing attributes and elements. The interface is annotated with three text boxes and arrows:

- La fenêtre donnant à voir la structure du document (les niveaux de description)**: Points to the left-hand pane showing the document's hierarchy.
- Fenêtre des attributs, avec noms et valeurs des attributs pour l'élément courant**: Points to the right-hand pane showing attribute details for the selected element.
- Fenêtre des éléments, avec noms des éléments disponibles à l'endroit où l'on se trouve**: Points to the right-hand pane showing a list of available elements for the current location.
- La fenêtre principale pour la saisie de l'IR. Chaque élément est représenté par des balises de couleur**: Points to the central workspace where the XML content is displayed with colored tags.



Quelques remarques techniques

- Un éditeur XML sera indispensable pour vérifier la qualité d'un travail de conversion rétrospective
- Permet de produire tous types de documents XML (pas seulement EAD)
- Attention, un éditeur XML ne gère pas les documents XML produits (stockage, recherche, gestion des accès... sont à confier à un autre logiciel, par exemple à une base native XML)

9



Et surtout, pour produire des IR en XML/EAD, en général bien connaître la DTD EAD ne suffit pas...

- La DTD EAD, comme ISAD (G), n'est qu'un modèle générique :
 - utilisable pour tout type de fonds d'archives ou d'IR
 - peu contraignant
 - qui ne traite guère du contenu des éléments.
- Donc pour mettre en œuvre l'EAD dans un service ou dans le cadre d'un projet, il faut réfléchir et écrire des spécifications (comme lorsqu'on crée une base de données documentaire).

10



Quelques manuels d'application existants

- Manuel d'application de la SAA : concerne la DTD EAD v. 1.0 ; techniquement périmé
- "Guidelines" du Research Libraries Group (http://www.rlg.org/en/page.php?Page_ID=450)
- Manuel de l'AFNOR, pour la mise en œuvre de la DTD EAD en France : recommandations générales (version officielle disponible à : <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/repertoireEAD.html>)
- Exemples de manuels d'application institutionnels ou par projet, en cours de rédaction : CHAN ; projet Archives-Canada-France

11



Quelques mots du manuel du CHAN

- Introduction insistant sur enjeux, continuité, ruptures, concepts de base (notamment celui de l'héritage et de la non-redondance des informations, de la nécessité de connaître le plan de classement)
- Nombreux exemples et annexes (documentation des projets en cours, modèles XMetaL, modèle CINDOC, normes ISO/AFNOR...)

12



Quelques mots du manuel du CHAN (suite)

- Plan de nommage des instruments de recherche et métadonnées (bibliographiques) de l'inventaire
- Réflexion sur les parties liminaires de l'instrument de recherche
- Travail sur le contenu des éléments, par ex. la formulation des informations administratives, des cotes, la normalisation ISO 8601 des dates, les intitulés et présentations de contenu, le formalisme des éléments d'indexation, les modalités de l'indexation

13



II - Publier les instruments de recherche en XML/EAD

14

XML/EAD, un format pivot

- XML/EAD est donc un format pour la structuration des IR électroniques, et un format d'archivage
- A 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 peuvent s'appuyer sur la structure du document XML pour transformer le document en un autre format dédié.
Ex. programmes XSL-T

15

XML/EAD, un format pivot (suite)

- Exemples de fichiers PDF : ceux produits à partir des documents EAD de la base BORA (<http://sdx.archivesdefrance.culture.gouv.fr/ap/>) ; faire une recherche, puis après avoir affiché une notice, cliquer sur le bouton 
- Exemples de fichiers HTML : ceux produits en HTML statique par le CHAN (<http://www.archivesnationales.culture.gouv.fr/chan/> rubrique Quoi de neuf ?)
- Les propriétés inhérentes au document XML placent donc les IR au format XML/EAD au cœur des nouvelles chaînes éditoriales

16



On peut aller beaucoup plus loin...

- Plusieurs documents EAD obéissant au même modèle, avec une probabilité forte de la présence des mêmes éléments de même nom explicite (ISAD(G)), dont certains sont conçus comme des points d'accès, d'autres contiennent l'analyse archivistique...
- Si ces documents respectent des règles communes de mise en œuvre de l'EAD, leur homogénéité est renforcée...
- Donc potentiellement des matériaux à exploiter ensemble
- Il faut un logiciel capable de chercher des données au sein des documents XML et de restituer les résultats à l'écran de façon claire

17



Un exemple d'outil pour construire une application Web de consultation

- **PLEADE** : une application générique (donc configurable et adaptable à ses propres besoins) conçue pour la publication et la recherche sur le Web des IR en XML/EAD
- Initiative privée des sociétés AJLSM et Anaphore
- logiciel libre (*open source*) fonctionnant à l'aide d'autres logiciels libres (plate-forme SDX, voir le site de l'ADNX, Association pour le développement du numérique en XML, <http://www.adnx.org>), dont les sources et la documentation sont disponibles à <http://www.pleade.org/>

18



Un exemple d'outil (suite) : principales fonctionnalités de PLEADE

- *Import* d'un ou plusieurs documents EAD dans PLEADE : processus d'indexation qui se fait selon des paramètres choisis par le responsable des contenus. Quelques paramètres importants : sous-ensembles d'IR, fragmentation en unités documentaires recherchables, descripteurs
- *Recherche* : se fait grâce à un moteur de recherche, en recherche simple (plein texte) ou avancée (exploitant le contenu des index préalablement formés). Formulaire de recherche avancée généraux ou propres à un sous-ensemble

19



Un exemple d'outil (suite) : principales fonctionnalités de PLEADE

- *Consultation* : le document EAD résultat d'une recherche est affiché au format HTML (créé dynamiquement par PLEADE à partir du document EAD), dans un écran divisé en trois fenêtres, selon une présentation choisie lors de l'import.
 - Les éléments de lien internes ou externes sont restitués par des liens hypertexte
 - PLEADE peut être couplée avec Navimages, (<http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages/index.html>), logiciel libre pour l'organisation, le traitement et la visualisation de lots d'images numériques, ce qui permet de consulter à partir du document EAD, les séries d'images numériques des documents d'archives décrits

20



Quelques exemples d'utilisation de PLEADE

- application de test de PLEADE sur le serveur de la société AJLSM : <http://ajlsm-sdx.hopto.org/sdx-21h/pl/>
- applications en cours de construction sur le serveur du Muséum national d'histoire naturelle et de la Bibliothèque nationale suisse (<http://ida.snl.ch>)
- l'application du CHAN, en ligne sur l'intranet du ministère de la Culture (<http://chan1.culture.gouv.fr/sdx/pleadechanv6/>), et dès que possible sur un serveur Internet
- Archives Canada-France (<http://bd.archivescanadafrance.org/acf/>) : une application construite avec Navimages, qui va bientôt utiliser PLEADE

21



Un logiciel en pleine évolution et hautement configurable

- PLEADE en est seulement à une version 1.0rc3, et va continuer d'évoluer en fonction des besoins exprimés et des orientations de la plate-forme SDX
- Adaptation aux besoins institutionnels ou sectoriels :
 - modification de l'apparence du site
 - informations statiques
 - organisation des IR en sous-ensembles
 - définition des formulaires de recherche
 - logiciel open source, donc développements complémentaires éventuellement

22



Une brique pour construire un système d'information

- Une application PLEADE peut :
 - être intégrée à un site Web ou constituer un site Web
 - être mise en relation avec d'autres applications SDX (Navimages en est un exemple)
 - plus tard, servir d'entrepôt pour fournir des contenus à un portail d'accès à divers types de ressources (entrepôt exposant des métadonnées OAI, Open Archives Initiative, <http://www.openarchives.org>)

23



III - Conclusion : défis et perspectives

24



Normaliser pour améliorer le service public, une vraie révolution

- Normaliser la production des IR, se doter des moyens techniques pour créer de nouvelles chaînes éditoriales : un investissement financier et humain important, un changement dans les habitudes, du temps et de la rigueur, mais une garantie, celle de la pérennité des contenus, et des résultats qui changent tout :
 - avec les technologies d'aujourd'hui, on peut dès maintenant faire sortir les IR de leur confidentialité en les diffusant comme jamais ils ne l'ont été ;
 - grâce aux applications de consultation, on peut dès maintenant faciliter considérablement l'accès aux sources secondaires et primaires, renouveler les fondements de la recherche historique

25



La normalisation : partager, y compris 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
- Ces communautés disposent aussi de modèles XML pour leur documentation scientifique, dont les archivistes peuvent avoir besoin
- Construire des portails interdisciplinaires devient possible

26



Impacts sur la pratique et la théorie archivistiques...

- L'IR comme une simple partie vivante d'un corpus, et non comme un tout figé et isolé...
...le défi des systèmes descriptifs, thesaurus et listes d'autorité, métadonnées
- EAD, un modèle unique autorisant toutes les évolutions internes à l'IR...
...un gain pour le public, qui remet aussi en cause la typologie classique des IR
- Les fragments d'IR : une nouvelle unité de lecture contextualisée et munie d'index (hors ou dans le texte)
- Le rapprochement avec les substituts numériques des documents d'archives : changement dans les contenus et dans les modes de production (IR à la fois plus légers et plus précis, plates-formes d'annotation coopérative...)

27



Le besoin de formation...

- Formation initiale et formation continue à la mise en œuvre des nouvelles technologies dans les services d'archives : une clé pour répondre correctement aux défis et réussir la révolution en cours
 - donner à tous les archivistes les repères indispensables, par la théorie et par la pratique
 - former aussi de vrais spécialistes

28



Pour tous renseignements complémentaires

- Florence Clavaud
conservateur du Patrimoine
responsable du service des nouvelles technologies
Centre historique des Archives nationales
60 rue des Francs-Bourgeois
75141 PARIS Cedex 03
tél. : 01 40 27 62 84
courriel : florence.clavaud@culture.gouv.fr