Devoir : encodage d’un instrument de recherche en XML/EAD 2002 et d’une notice de producteur en XML/EAC-CPF

Cours sur la description archivistique (ENC2408)

# Objectifs du devoir

Il s’agit de **convertir manuellement un instrument de recherche produit au traitement de texte en fichier XML conforme à la DTD EAD 20002 et d’enrichir la notice de producteur correspondante, qui existe déjà au format XML/EAC-CPF mais qui est vide de tout contenu.**

Vous mettrez ainsi en œuvre concrètement vos connaissances tout en travaillant sur un cas réel. En effet l’instrument de recherche à convertir concerne une partie des archives de la Conservation régionale des monuments historiques d’île-de-France, versées en 2011 aux Archives nationales et décrites par le service versant. Les fichiers produits dans le cadre de ce devoir seront, après correction et relecture, intégrés au système d’information archivistique des Archives nationales et publiés dans la [salle des inventaires virtuelle](https://www.siv.archives-nationales.culture.gouv.fr/).

# Fichiers à convertir et ressources complémentaires

Ces fichiers peuvent être téléchargés depuis le dépôt GitHub du cours ENC2408 (voir plus précisément le dossier <https://github.com/florenceclavaud/cours-ENC2408/tree/master/devoir>).

En voici la liste :

* instrument de recherche archivistique à convertir en EAD (fichier OpenOffice.org, nommé 20110282\_CRMHIDF.odt) ;
* DTD utilisée dans le SIA des Archives nationales (fichier nommé ead\_sia.dtd) ;
* notice de producteur à enrichir (fichier nommé FRAN\_NP\_005412.xml) et schéma EAC-CPF (fichier cpf.xsd) ;
* référentiels à utiliser pour l’indexation (référentiel FRAN\_RI\_005.xml des lieux hors Paris, référentiel FRAN\_RI\_026.xml pour les édifices parisiens, référentiel FRAN\_RI\_104.xml pour les statuts juridiques des collectivités, référentiel FRAN\_RI\_011.xml pour les fonctions des collectivités)

# Recommandations générales

**En ce qui concerne le fichier EAD** :

* **vous utiliserez**, non pas directement la DTD EAD 2002, mais **la DTD dite “EAD\_SIA”** (fichier nommé ead\_sia.dtd) **qui est utilisée au quotidien pour produire les IR dans le SIA des Archives nationales** et qui restreint les possibilités offertes par la DTD EAD 2002 ; pour rappel, dans le dépôt GitHub, **le dossier exercices/encodage-EAD/corriges contient un exemple d’utilisation de cette DTD**, pour l’encodage de la notice descriptive du fonds Léon Blum - fichier nommé Notice-Fonds-Blum-v2.xml) ;
* utilisez la chaîne de caractères ‘20110282\_CRMHIDF’ comme identifiant provisoire de l’instrument de recherche EAD (cet identifiant sera modifié automatiquement lors de l’import du fichier dans le SI des AN) ;
* l’objectif est d’**encoder toute la substance de l’instrument de recherche** (IR), à l’exception du contenu de la page 5 (le plan de classement) ;
* vous veillerez par ailleurs à **relier le fichier EAD avec la notice EAC/CPF décrivant le producteur**, via les éléments ad hoc ;
* vous veillerez également à **encoder de la manière la plus précise et la plus utile possible** (donc si possible en créant des liens vers les ressources mentionnées) l**es informations fournies à la fin de la page 3 et au début de la page 4** sur d’autres versements d’archives de la CRMH d’Ile-de-France ;
* **en ce qui concerne la description des sous-composants**, concertez-vous avant de commencer le travail ; **mettez en œuvre strictement le principe de la description par niveaux en privilégiant l’organisation intellectuelle** (par contenu) **des dossiers par rapport à l’organisation en unités de conditionnement** (les articles, dotés de cotes); les cotes peuvent être répétées d’un sous-composant à un autre lorsque c’est nécessaire ;
* vous procéderez à **une indexation des lieux mentionnés dans l’IR,** c’est-à-dire au minimum : les noms des départements, les noms des communes ; si possible, vous commencerez à indexer les édifices mentionnés. Pour cela vous utiliserez les référentiels des AN, qui sont disponibles au format RDF dans le dépôt <https://github.com/ArchivesNationalesFR/Referentiels>, et dont nous vous fournissons dans le dossier du devoir la version XML source pour les types de lieux concernés (fichier FRAN\_RI\_005.xml pour tous les lieux hors département de Paris ; fichier FRAN\_RI\_026.xml pour les édifices parisiens). Si vous ne parvenez pas à finir l’indexation, veillez à la terminer pour quelques cotes au moins.

**En ce qui concerne le fichier EAC-CPF** :

* il s’agira de **compléter voire de corriger la notice existante**, en vous inspirant des corrigés de l’exercice que vous avez fait pendant le cours ;
* les corrections pourront concerner **la forme autorisée du nom de l’organisme, son statut juridique** (c’est un service déconcentré de l’Etat ; voir le référentiel des statuts juridiques fourni dans le dossier, FRAN\_RI\_104.xml), **ses dates d’existence;**
* en ce qui concerne les enrichissements, il faudrait surtout **élaborer un historique rédigé**, en utilisant comme source le contenu de l’IR (vous êtes libres de vous en écarter et de ne pas tout reprendre) et des informations tirées principalement des pages web suivantes, dans l’ordre de priorité décroissant, et en mentionnant les sources dans les éléments EAC ad hoc :
  + <https://www.culture.gouv.fr/Regions/Drac-Ile-de-France/Patrimoines-Architecture/Monuments-Historiques>
  + <https://aaf.ica-atom.org/region-conservation-regionale-des-monuments-historiques>
* vous tâcherez de procéder à l**’indexation du siège de la CRMH** (indexation lieu) **et de ses fonctions**, en utilisant les référentiels des AN idéalement (voir plus haut) ;
* vous créerez dans la notice l**e lien vers l’IR** que vous encodez par ailleurs.

# Organisation du travail

Vous composerez **7 groupes de 3 étudiants, chaque groupe prenant en charge une partie des travaux, selon la répartition suivante** :

|  |  |  |
| --- | --- | --- |
| Groupe | Tâche | Composition du groupe |
| Groupe 1 | Fichier EAD : production de l’en-tête du fichier (élément <eadheader>) ; contrôle de cohérence et de qualité du fichier EAD ; rédaction d’un fichier externe de documentation explicitant les choix faits | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 2 | Fichier EAD : encodage des informations de description globale du versement (pages 2-4) ; indexation du fichier | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 3 | Fichier EAD : encodage de la description des dossiers conservés sous les cotes 20110282/1 à 20110282/21 inclus (p. 6-12) | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 4 | Fichier EAD : encodage de la description des dossiers conservés sous les cotes 20110282/22 à 20110828/40 inclus (p.12-18) | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 5 | Fichier EAD : encodage de la description des dossiers conservés sous les cotes 20110282/41 à 20110282/59 inclus(p. 18-26) | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 6 | Fichier EAD : encodage de la description des dossiers conservés sous les cotes 20110282/60 à 20110282/74inclus (p. 26-33) | * Nom, Prénom * Nom, Prénom * Nom, Prénom |
| Groupe 7 | Fichier EAC-CPF : enrichissement de la notice du producteur FRAN\_NP\_005412.xml | * Nom, Prénom * Nom, Prénom * Nom, Prénom |

# Documents à rendre et date de restitution

Nous attendons de vous que vous nous adressiez, **avant le mardi 7 juin 2022 à 18 h**, soit par messagerie (à l’adresse [le-lab.archives-nationales@culture.gouv.fr](mailto:le-lab.archives-nationales@culture.guov.fr)), soit via un dépôt public que vous aurez créé sur GitHub pour la gestion du projet (et dans ce cas en nous prévenant bien sûr), un dossier contenant :

* **le fichier EAD que vous aurez produit** et associé à la DTD fournie ;
* **le fichier EAC-CPF que vous aurez produit**, associé au schéma EAC-CPF ;
* **un fichier de documentation produit au traitement de texte**, dans lequel devront figurer au moins :
  + le tableau de répartition des personnes par groupe ;
  + une présentation générale du travail réalisé, qui expliquera les choix que vous aurez faits en ce qui concerne les points mentionnés ci-dessus dans la section *Recommandations générales*, et précisera le cas échéant sur quels aspects précis vous n’avez pas atteint les objectifs fixés et pourquoi.

# Méthode et critères d’évaluation du devoir

Les fichiers XML produits seront évalués principalement sur la base des critères suivants :

* **conformité à XML**
* **validité par rapport à la DTD fournie et au schéma EAC-CPF**
* **exhaustivité** (tout doit être encodé, sauf la page 5 de l’instrument de recherche comme indiqué plus haut)
* **exactitude de la représentation arborescente des unités de description**
* **pertinence du choix des éléments et attributs**
* **homogénéité** (la même structure XML et les mêmes attributs doivent être appliqués systématiquement aux mêmes catégories d’informations)
* **existence des liens entre les deux fichiers**
* **pertinence de l’indexation** (un nom de lieu ou d’édifice doit être lié à la bonne entrée dans le référentiel)
* **clarté de la documentation**

La même note sur 20 sera attribuée aux 3 membres de chaque groupe.

Un corrigé sera fourni, accompagné d’une documentation.

# Pour toute précision ou question

N’hésitez pas à nous contacter si quelque chose n’est pas clair, si vous jugez nécessaire d’avoir des précisions ou si un problème bloquant survient. Désignez dans ce cas une personne dans la promotion pour échanger avec nous (logiquement, une personne du groupe 1).

Si besoin, une réunion en visio-conférence pourra être organisée.

Bon courage !

Florence Clavaud et Pauline Charbonnier (Lab des Archives nationales), 31 mars 2022