



# **Un projet et une application Web utilisant EAD... et d'autres formats XML : les chartes médiévales de l'abbaye de Saint-Denis**

Florence Clavaud

Conservateur du patrimoine, déléguée à la formation informatique

École nationale des chartes (Paris, France)

[florence.clavaud@enc.sorbonne.fr](mailto:florence.clavaud@enc.sorbonne.fr)

**Journées d'études sur la production et la publication électronique des instruments de recherche  
ANOM, Aix-en-Provence, 2-3 juin 2010**

# Plan de la présentation

- Dans quel cadre et pourquoi ce projet ?
- Principales étapes jusqu'à aujourd'hui
- Comment ?
  - les choix de modélisation
  - l'application
- Perspectives

## Le contexte : un projet de recherche de l'ENC

- Un axe de recherche sur les structures du pouvoir : écrits, hommes, autorité (XI<sup>e</sup>-XIX<sup>e</sup> siècles), et un sous-axe sur les Territoires et pouvoirs en Île-de-France et dans le nord de la France

Ces orientations prolongent, pour le programme quadriennal 2010-2013, des actions entreprises en 2006-2009

- Une collection d'éditions électroniques, et plus globalement une expertise en technologies numériques appliquées aux SHS

A l'origine, les premiers projets d'humanités numériques ont été portés par quelques convaincus, des médiévistes

> travaux sur les cartulaires franciliens, et premiers travaux sur le *Cartulaire blanc* de l'abbaye de Saint-Denis (Archives nationales, LL 1157 et LL 1158) dès 2005-2006 (d'où une première application Web).

Le *Cartulaire blanc*, un objet documentaire à étudier en tant que tel, et aussi un des témoins les plus riches du contenu d'un chartrier aujourd'hui dispersé

Responsable scientifique : Olivier Guyotjeannin

## Une nouvelle stratégie pour ce projet

À partir de 2007, une deuxième époque pour les éditions électroniques à l'ENC:

- évaluation de l'existant ;
- les principes fondamentaux maintenus ;
- une ère plus industrielle avec plus de projets, et aussi des besoins plus complexes.

Dans ce contexte, réorientation du projet Saint-Denis :

- pour **ne pas dépendre du seul calendrier d'avancement de l'édition du *Cartulaire blanc***;
- pour **intégrer à l'édition électronique les images numériques**

Fin 2006, décision de procéder à la numérisation de l'*Inventaire général* manuscrit du chartrier, dressé entre 1680 et 1728 (Archives nationales, LL1189, LL1190 et LL1191).

L'*Inventaire général* décrit 3 307 chartes datant des origines à 1302 ; il est rédigé en français. Le *Cartulaire blanc* donne à lire la copie d'environ 2 600 actes pour la même période, bien sûr pour la plupart en latin, certains en français ancien.

> Présentation au ministère de la Culture d'une demande de financement par les Archives nationales et l'ENC, dans le cadre de l'appel à projets de numérisation 2006 de la MRT.

## Principales étapes jusqu'à aujourd'hui

- **2007** : finalisation de l'édition critique au traitement de texte de deux nouveaux chapitres du *Cartulaire blanc* (Pierrefitte et Beaurain) ; appel d'offres pour la numérisation et l'encodage en EAD de l'*Inventaire général* (IG), remporté par Dataland (un peu par défaut...) ;
- **fin 2007-printemps 2008** : réception et contrôle du travail de Dataland ;
- **automne 2007-printemps 2010** : relecture intégrale et enrichissement des fichiers XML/EAD de l'IG ;
- **2007-2008** : discussions et note de cadrage sur les fonctionnalités de la future application, premiers développements
- **2008-2010** : encodage en XML/TEI de l'édition critique des chapitres Beaurain et Pierrefitte dans le *Cartulaire blanc*, reprise et modifications des fichiers TEI des deux chapitres précédemment publiés (Tremblay et Rueil), définition d'un modèle XML ;
- **fin 2009-2010** : rédaction des principaux textes de présentation ; principaux travaux de développement ; mise en production et tests de l'application
- juin 2010 : ouverture officielle d'une version 1 de l'application

## Informatisation de l'*Inventaire général* : principes

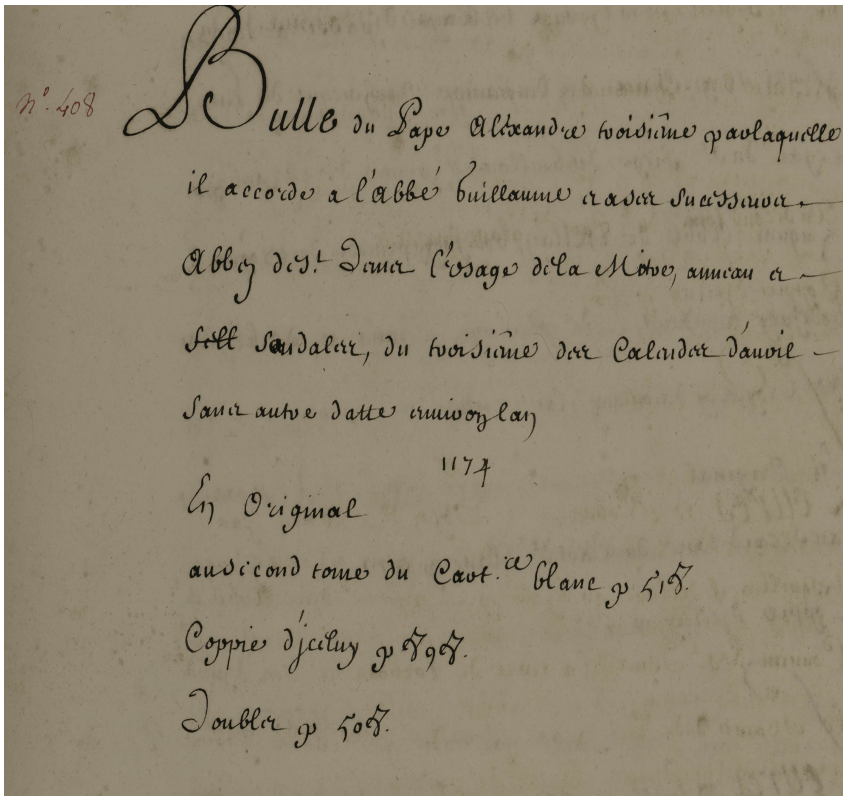
- numérisé en mode image et en couleur, une vue par page, sauf les pages totalement blanches. Plan de nommage séquentiel. L'organisation (et non le nommage) des images est calée sur la structure physique des registres (unités physiques : tome et page).
- pour l'informatisation du texte : choix du modèle XML EAD 2002
  - < inventaire pré-archivistique donnant à lire pour chaque pièce des métadonnées descriptives ; l'unité de description est la pièce, le contenu essentiellement un regeste
    - < principal besoin pour le projet : justement, disposer de ces métadonnées comme point d'accès premier aux actes
      - < EAD est le modèle XML de référence pour le détenteur des documents primaires et principal partenaire, les Archives nationales
- cependant, l'IG est aussi un document ancien qu'il faut éditer pour le rendre utilisable, c'est-à-dire ici, au moins :
  - repérer les scories (d'où des notes)
  - préciser autant que possible la date des actes
  - comprendre les références bibliographiques et sources mentionnées, les enrichir
  - indexer le type d'auteur d'acte

## Informatisation de l'*Inventaire général* : processus

- transcription et encodage XML/EAD simultanés par la société Dataland, sur la base d'un cahier des charges précis ;
  - puis relecture et enrichissement manuels : correction des erreurs de transcription, datation, indexation, ajout de références bibliographiques
  - enfin, contrôles divers et homogénéisation notamment de l'indexation, pose de liens entre notices, encodage des liens entre le texte structuré de l'inventaire et les images reproduisant ce manuscrit, considérés essentiels.  
Ajout de ces liens par programme en récupérant les informations utiles depuis le fichier METS
- > Une **mise en œuvre d'EAD particulière, et un travail en deux grandes étapes**, la deuxième étant la plus lourde, réalisée à trois à l'ENC, avec les images pour base de référence, par paquets de registes.  
L'occasion de travailler vraiment ensemble, directement dans le code XML.

# Le regeste n° 408 du tome 1 de l'IG (Arch. nat., LL 1189 page 436)

URL dans l'application : <http://saint-denis.enc.sorbonne.fr/inventaire.html?tome=1&notice=408>



```
<c id="notice408" level="item">
  <did>
    <unitid>n° 408</unitid>
    <unitdate era="ce" calendar="gregorian">1174</unitdate>
    <unitdate altrender="enc" certainty="estimation indices externes"
      normal="1179-03-30/1179-03-31" era="ce" calendar="gregorian"/>
    <physdesc>original</physdesc>
  </did>
  <scopecontent>
    <p>Bulle du pape Alexandre troisieme par laquelle il accorde à l'abbé Guillaume et à ses
      successeurs abbez de Saint-Denis l'usage de la mitre, anneau et <note altrender="enc">
        <p>raturé sur la ligne: « sell »</p>
      </note>sandales. Du troisieme des calendes d'avril, sans autre datte, environ l'an
      1174.</p>
  </scopecontent>
  <odd altrender="enc" type="liens-images-IG">
    <p>Archives nationales, LL 1189 page <extref href="inventaire/tome1/inventairejpg_0477.jpg"
      title="AN LL 1189 page 436" role="pleade:series" altrender="enc-lien-image-IG"
      >436</extref>
    </p>
  </odd>
  <controlaccess altrender="enc">
    <subject source="liste-auteurs-actes">pape</subject>
  </controlaccess>
  <bibliography>
    <bibref>En original</bibref>
    <bibref>Au second tome du cartulaire blanc p. 518<!--était: 418--></bibref>
    <bibref>Coppie d'iceluy p. 898</bibref>
    <bibref>Doublet p. 508<!--était: 408--></bibref>
    <bibref altrender="enc">Édition : Grosse, n° 72</bibref>
  </bibliography>
  <odd>
    <p>1174</p>
    <p>nombre d'or 16, Pâques<!--était: paquets--> 24 mars, dominicale F</p>
  </odd>
</c>
```



## Informatisation du *Cartulaire blanc*

- Les images numériques couleur des registres originaux existaient déjà, les Archives nationales ont fourni une copie. Là aussi, l'organisation (et non le nommage) des images est calée sur la structure physique du texte (unités de travail : tome et page).
- L'édition critique, d'abord faite au traitement de texte, est une **édition scientifique diplomatique collaborative**, à laquelle contribuent les étudiants archivistes-paléographes de l'ENC. Elle respecte les règles d'édition des documents anciens appliquées en France, tout en se concentrant sur le seul *Cartulaire blanc* comme témoin. Son organisation est calée sur la structure logique des textes (unités de travail : chapitre et acte).
- Choix du modèle d'encodage : XML TEI, plus précisément **TEI P5** (depuis novembre 2007), une ontologie du texte dans laquelle on peut puiser ce dont on a besoin pour dénommer les segments à considérer dans le texte édité.
- Travail de modélisation concomitant avec une réflexion plus large à l'ENC sur la mise en œuvre de la nouvelle version de TEI.

## Informatisation du *Cartulaire blanc*

### - Principes :

- définition d'un sous-ensemble de la TEI comme modèle ;
- un fichier XML par chapitre, intégration au fichier de l'introduction de chaque chapitre ;
- politique d'identifiants définie à l'avance ;
- distinction claire, en zones séparées, entre bloc de transcription et ce qui l'entoure et l'explique : métadonnées, tableau de la tradition, apparat critique et notes historiques ;

### - Particularités (par rapport à d'autres projets de l'ENC):

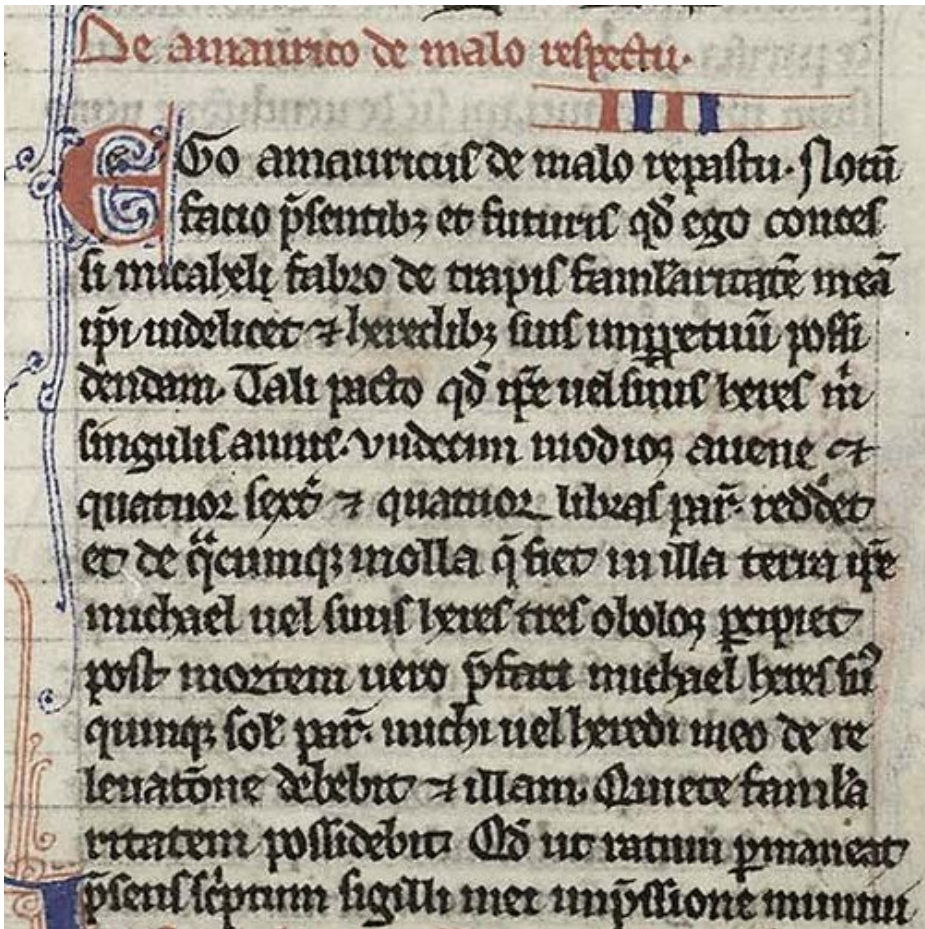
- deux registes, dont un "bref" (intitulé de l'acte)
- tableau de la tradition incluant l'édition de certains registes dans deux inventaires anciens, accompagnée de son apparat critique ;
- indexation de l'auteur de l'acte;
- éléments de liaison avec les images numériques du cartulaire, jugés essentiels ici (ajoutés par programme en récupérant les informations utiles depuis le fichier METS)

- **Travail d'encodage réalisé manuellement pour l'essentiel**, en relation étroite avec l'éditeur scientifique, par deux personnes, les deux éditions anciennes étant en même temps rendues conformes au modèle choisi

- Bientôt, en téléchargement sous licence Creative Commons, les sources XML et leur schéma TEI documenté.

## Chapitre de Beaurain, acte n° 4 (Arch. nat., LL 1157 page 539)

URL dans l'application : <http://saint-denis.enc.sorbonne.fr/cartulaire.html?chapitre=beaurain&acte=4>



### Bloc de la transcription

```
<div type="transcription" xml:lang="lat">
  <p>Ego Amauricus de Malo Repastu<ref type="note" n="1"
target="#beaurain-acte4-n-1"/>
    notum facio presentibus et futuris quod ego concessi <app n="c">
      <lem>Michaeli</lem>
      <rdg wit="#beaurain-acte4-B">Micaheli</rdg>
    </app> fabro de Trapis famularitatem meam, ipsi videlicet et
heredibus suis in perpetuum possidendam, tali pacto quod ipse vel suus
heres mihi singulis annis undecim modios avene et quatuor sextarios et
quatuor libras parisiensium redderet, et de quacumque molla <ref
type="note" n="3" target="#beaurain-acte4-n-3"/> que fiet in illa terra<ref
type="note" n="4" target="#beaurain-acte4-n-4"/> ipse Michael vel suus
heres tres obolos percipiet. Post mortem vero prefati <app
n="d"><lem>Michaelis</lem><rdg wit="#beaurain-acte4-B">Michael</rdg>
</app> heres suus quinque solidos parisiensium michi vel heredi meo de
revelatione debebit et illam quiete famularitatem possidebit. Quod ut ratum
permaneant, presens scriptum sigilli mei impressione munivi.</p>
</div>
```

## Les relations entre les sous-ensembles du corpus

- Un **corpus numérique** formé de quatre sous-ensembles :
  - deux collections d'images ;
  - trois fichiers XML/EAD ;
  - quatre fichiers XML/TEI (et plus bientôt).
- Pour satisfaire un des besoins majeurs du projet (**permettre la navigation entre ces sous-ensembles**), nécessité d'exprimer les relations entre :
  - les images numériques, les pages, et les unités intellectuelles, correspondant à chacun des deux documents primaires ;
  - les registes et les actes :
    - pour un acte du *Cartulaire* : “acte analysé dans l'*Inventaire général*, tome N, regeste n°, page X”
      - > informations données par l'éditeur scientifique, pour les actes des chapitres édités
    - pour un regeste de l'*Inventaire* : “concerne acte du *Cartulaire blanc*, tome N, page X [chapitre Y, acte Z]”
      - > informations directement fournies par l'*Inventaire général* dans chacune des notices
- Volonté de ne pas concrètement lier directement les fichiers EAD et TEI entre eux, afin de pouvoir les maintenir et les réutiliser, à l'ENC ou ailleurs, indépendamment les uns des autres
- > Liens à déclarer au sein de ressources externes

## Les relations entre les sous-ensembles du corpus

- Utilisation de fichiers XML/**METS** (format XML permettant de décrire des paquets de fichiers formant un ensemble documentaire cohérent et de consigner des métadonnées globales, en particulier de préservation en prévision de l'archivage, sur ces fichiers)
- Saisie des fichiers METS de l'*Inventaire général* par la société Dataland au moment de la numérisation
- Dataland a aussi pris en charge la production du fichier METS du *Cartulaire blanc*, pour la section de la structure physique, à partir de la série d'images ; pour la section de la structure logique, saisie manuelle des informations pour les actes déjà édités
- Relecture des fichiers METS de Dataland nécessaire

# Déclaration des liens : exemple pour l'IG LL 1189 page 436 (image 477), regeste n° 408

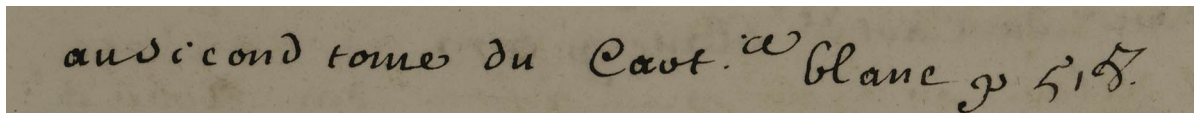
- Déclaration du fichier image :

```
<METS:file ID="inventairejpg_0477" MIMETYPE="image/jpeg">  
  <METS:FLocat LOCTYPE="URL" xlink:href="images/inventaire/tome1/inventairejpg_0477.jpg"/>  
</METS:file>
```

- Carte de structure physique :

```
<METS:div TYPE="page" LABEL="436" ID="tome1-page436">  
  <METS:mptr LOCTYPE="URL" CONTENTIDS="#notice407"/>  
  <METS:mptr LOCTYPE="URL" CONTENTIDS="#notice408"/>  
  <METS:mptr LOCTYPE="URL" CONTENTIDS="#notice409"/>  
  <METS:fptr FILEID="inventairejpg_0477"/>  
</METS:div>
```

- Carte de structure logique, encodée en se servant notamment des informations fournies directement par la notice de l'*Inventaire* :



```
<METS:div TYPE="notice" LABEL="n° 408" ID="notice408">  
  <METS:mptr LOCTYPE="URL" CONTENTIDS="#tome1-page436" xlink:title="inventaire"/>  
  <METS:mptr LOCTYPE="URL" xlink:title="mcartulaireblanc" xlink:href="mcartulaireblancmets.xml#tome2-  
page518"/>  
</METS:div>
```

## Déclaration des liens : exemple pour le *Cartulaire blanc* LL 1157 page 539 (image 575) : chapitre de Beaurain acte 4

- Dans la carte de structure logique :

```
<METS:div TYPE="chapitre" LABEL="beaurain">
<!-- etc. -->
  <METS:div TYPE="acte" LABEL="Beaurain 4">
    <METS:mptr LOCTYPE="URL" xlink:role="inventaire"
xlink:href="inventairemets1.xml#notice482"/>
    <METS:mptr LOCTYPE="URL" xlink:title="De Amaurico de Malo Repastu"
xlink:role="mcartblanc" CONTENTIDS="#tome1-page539"/>
    <METS:fptr>
      <METS:area BEGIN="#beaurain-acte4" FILEID="beaurain"
BETYPE="IDREF"/>
    </METS:fptr>
  </METS:div>
<!-- Etc. -->
</METS:div>
```

## Développement de l'application

- Choix du logiciel libre PLEADE (version 3.3.0.1), qui offrait deux des principaux groupes de fonctionnalités nécessaires (recherche et consultation de documents EAD, gestion et consultation de collections d'images)
- Installation adossée à une base de données MySQL
- Forte adaptation du logiciel (fichiers de configuration, sitemaps, XSLT, Javascript), en particulier :
  - pour les fonctionnalités liées aux fichiers TEI
  - pour la gestion des liens
  - pour la consultation des images
  - pour le système d'URLs
- Ouverture officielle de la version 1 de l'application dans quelques jours, à l'adresse <http://saint-denis.enc.sorbonne.fr/>
- [Démo]



## Perspectives

- Sur le plan scientifique :
  - de nouveaux moyens d'exploration du corpus à exploiter ;
  - éditions de nouveaux chapitres en cours et prévues, y compris par des chercheurs extérieurs à l'ENC ; espoir de les faciliter
  - intégration au corpus numérique des chartes originales ? Collaborations avec les AN ?
- Sur le plan informatique :
  - version 2 de l'application fin juillet, avec en particulier:
    - entrepôt OAI-PMH exposant en EAD et en TEI;
    - téléchargement des sources (au moins TEI et schéma)
  - divers tests en cours pour la génération de représentations graphiques des résultats de recherche (lignes de temps)
- Hors du strict périmètre de ce projet:
  - mutations (enrichissements divers) et réutilisations du corpus à prévoir (pour les fichiers TEI)
  - projet de portail sur les sources médiévales de l'Île-de-France
  - faire contribuer directement les étudiants?