

Structuration de données, recherche et édition

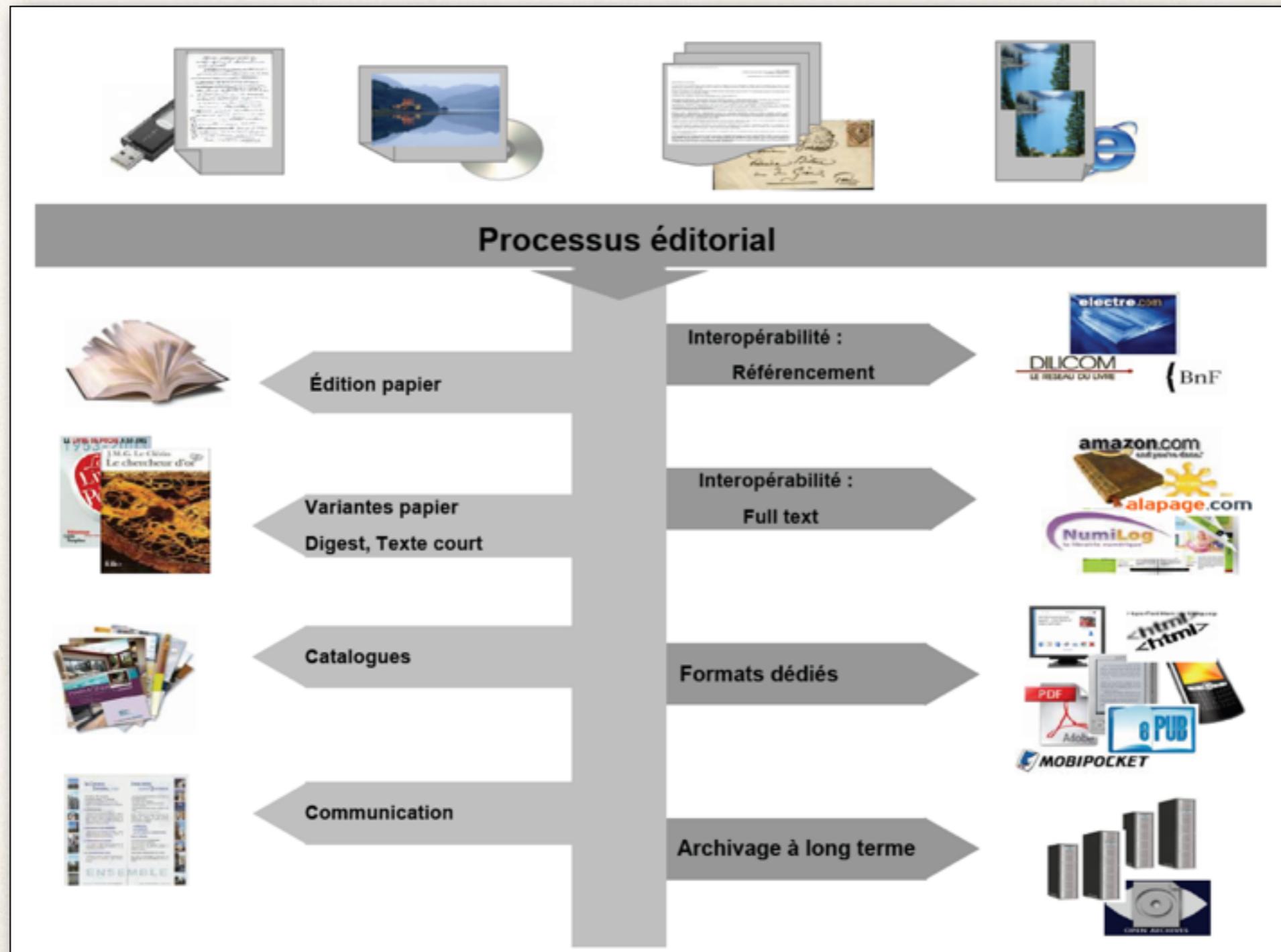
École d'été Biblissima

Pierre-Yves Buard – Pôle Document Numérique, MRSH Caen
Université de Caen Normandie

2 septembre 2016 – Avranches

Contexte éditorial

Convergence numérique



Objectifs et principes

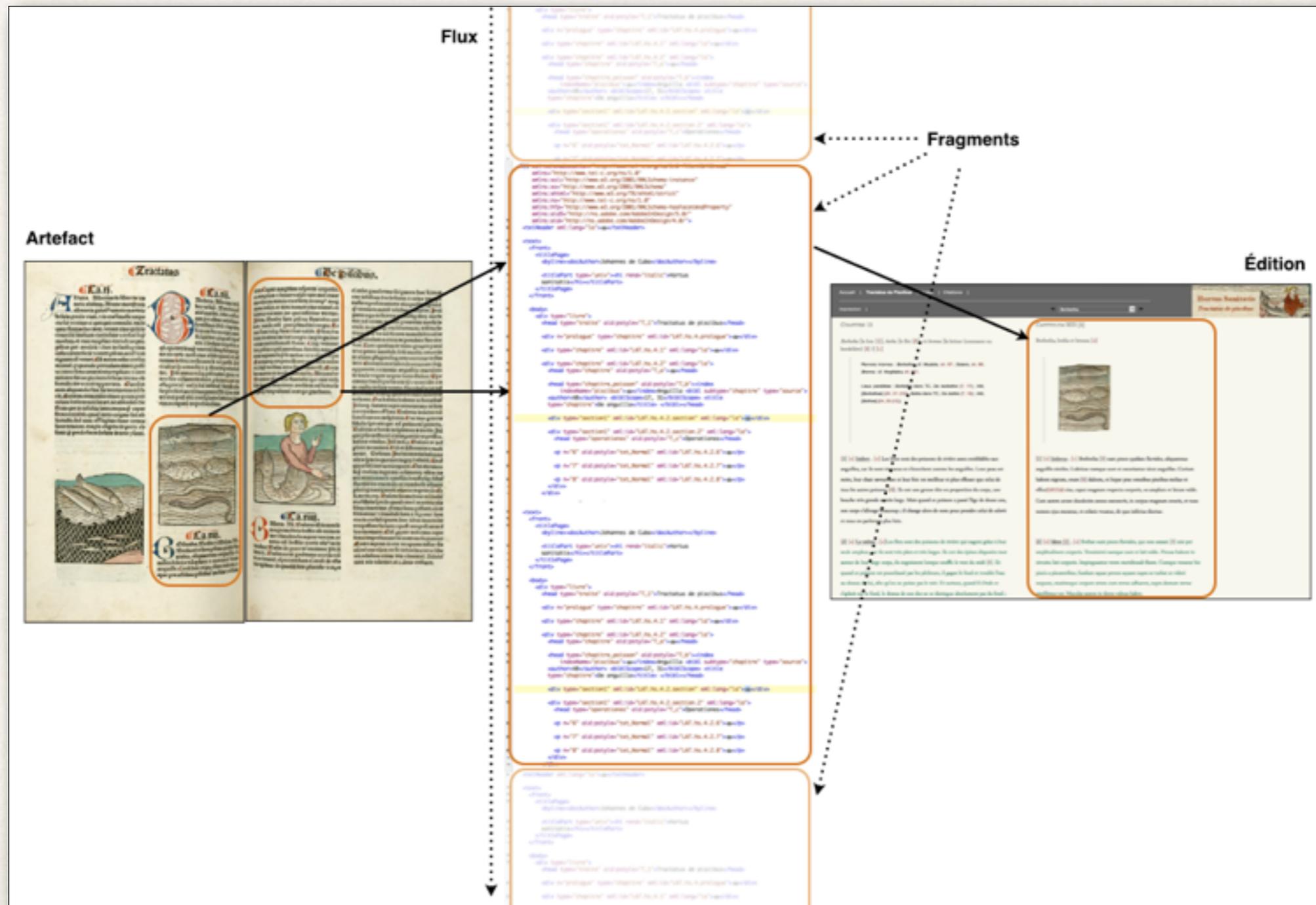
... sue contiguitate regioni. simul ad quosdam uenerunt parochos. ...
... cum ipse abba loci suo retulisset antistita: uberrimas la-
... potenti deo uicere retulit. qui pro lapsis nature fragilitate ter-
... sistenciam sibi ministrorum dignatur prebere suffragium. H-
... decebat ueneratione sumptis a loco pignoris quo beatus ar-
... sui memoriam fidelibus commendauerat. partem scilicet tuba-
... archangeli in monte gangano supra altare qu-
... posuit. & partem marmoris supra quod stetit. cu-
... usque in eodem loco super octam uestigia. iam dictis scrib-
... trum locum referenda patrocinia contradidit: conditione-
... ra uidelicet ut quos una anglice reuelationis societate
... quoque eternaliter neceret caritatis conuicio. Summi in-
... repedantes: post multa itineris spatia ad locum quo digressi
... so die quo fabrica completa est in monte iam dicto in occidu-
... quasi nouum ingressi sunt orbem: que primum uicium dent

Objectifs

- * Articuler la valorisation du patrimoine et la recherche scientifique
- * Valoriser les fonds traités
- * Préserver la qualité éditoriale (pour tous les supports)
- * Améliorer l'indexation des documents (permettre des recherches avancées)
- * Automatiser la production des formes diffusées et des bases de connaissances
- * Respecter les normes internationales (circulation des données et des outils)
- * Assurer l'archivage à long terme

Principes

Flux et fragments de textes



Principes

Flux de textes – balisage éditorial

- * niveau de balisage suffisant pour l'édition :
versification, titres, paragraphes, etc.
- * peu de profondeur (divisions, puis deux niveaux maximum : paragraphes et caractères)

```
<div type="chapitre">
  <div type="section" xml:id="AFR.1" xml:lang="afr">
    <lg xml:id="AFR.1.1">
      <l n="1" aid:pstyle="txt_Original_Vers" xml:id="vers1"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/"><pb ed="A"
          n="1r" />Molz pelerins qui vunt al munt<note type="marginal"
          xml:id="AFRftn1"> <emph aid:cstyle="typo_Italique">B</emph>: Les
          bones gens qui vunt au mont <emph aid:cstyle="typo_Italique">(voir
          infra), avec majuscule initiale de grande taille.</emph>
        </note></l>

      <l n="2" aid:pstyle="txt_Original_Vers" xml:id="vers2"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Enquierent
        molt, et grant dreit unt,</l>

      <l n="3" aid:pstyle="txt_Original_Vers" xml:id="vers3"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Comment
        l'igliese fut fundee</l>

      <l n="4" aid:pstyle="txt_Original_Vers" xml:id="vers4"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Premierement,
        et estoree.</l>

      <l n="5" aid:pstyle="txt_Original_Vers" xml:id="vers5"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Cil qui lor
        dient de l'estoire</l>

      <l n="6" aid:pstyle="txt_Original_Vers" xml:id="vers6"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Que cil
        demandent, en memoire</l>

      <l n="7" aid:pstyle="txt_Original_Vers" xml:id="vers7"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Ne l'unt pas
        bien, ainz vunt faillant</l>

      <l n="8" aid:pstyle="txt_Original_Vers" xml:id="vers8"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">En plusors
        leus, et mespernant.</l>

      <l n="9" aid:pstyle="txt_Original_Vers" xml:id="vers9"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Por faire la
        apertement</l>

      <l n="10" aid:pstyle="txt_Original_Vers" xml:id="vers10"
        xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">Entendre a cels
        qui escient</l>
```

Principes

Flux de textes – balisage scientifique

- ❖ exemple de structure complexe
- ❖ segments, indications bibliographiques, plusieurs types de notes, etc.
- ❖ structure aussi profonde que nécessaire

```
<p n="1" next="2" aid:pstyle="txt_Normal"
xml:id="LAT.hs.4.1.1"><bibl subtype="paragraphe" type="source">
<author>VB</author> <biblScope>17, 29, 1</biblScope> </bibl><index
indexName="marqueurCitation">
<term type="orig">Ex Libro de naturis rerum</term>

<term type="norm">TC</term>

<term type="identifie">TC</term>
</index>. <seg type="citation"><bibl subtype="fragment"
type="source"> <author>TC</author> <biblScope>7, 9</biblScope>
<note type="sources" xml:id="LATnote18">Vincent de Beauvais suit
fidèlement Thomas de Cantimpré.</note> </bibl><index
indexName="piscibus">
<term>abremon</term>
</index>Abremon<note type="philologie"
xml:id="LATnote19">Kitchell & Resnick 1999, 1660, n. 49, dans
leur commentaire au chapitre d'Albert le Grand (AM 24, 9 (10))
consacré au poisson <hi rend="italic">abremon</hi>, suggèrent
d'expliquer le nom <hi rend="italic">abremon</hi> par une
corruption de l'expression <hi rend="italic">ab arena</hi>, «_qui
vient du sable_». Un étymon latin, même séduisant, est délicat à
justifier pour un terme qui, de toute évidence, s'inscrit, avant
le latin, dans une tradition gréco-orientale.</note>, alias<note
type="apparat" xml:id="LATnote20">alas <hi
rend="italic">1536.</hi></note> <index indexName="piscibus">
<term>abarenon</term>
</index>abarenon, est piscis, ut dicit Aristoteles, multum
fecundus ovis, nec aliorum<note type="apparat"
xml:id="LATnote21"><hi rend="italic">post</hi> aliorum <hi
rend="italic">hab.</hi> ea <hi rend="italic">1491 VB</hi>.</note>
more contactu<note type="apparat" xml:id="LATnote22">contractu <hi
rend="italic">1491</hi>.</note> maris<note type="philologie"
xml:id="LATnote23">Le traducteur médiéval de l'<hi
rend="italic">Hortus sanitatis</hi> a compris <hi
rend="italic">maris</hi> comme le génitif de <hi
rend="italic">mas</hi>, «_le mâle_», en voyant peut-être dans ce
passage une allusion à la fécondation externe des œufs de la
femelle par la laitance du poisson mâle. Mais ce mode de
reproduction, très fréquent chez les poissons ovipares, était bien
connu des Anciens, et on comprendrait mal en quoi l'athérine
ferait figure d'exception. Il faut donc plutôt interpréter <hi
rend="italic">maris</hi>, comme le génitif de <hi
```

Éditions multimodales

Caractéristiques

- * plusieurs modes d'accès aux textes
 - * plusieurs supports
 - * plusieurs solutions de lecture sur un seul support
 - * articulation de plusieurs supports et de plusieurs solutions d'accès aux textes (sélection de textes en fonction des supports, multiplication des parcours de lecture, etc.)
- * stable sur tous les supports pour permettre la citation

Méthodes et outils

... sue contiguitate regioni. simul ad quosdam uenerunt parochos. ...
... cum ipse abba loci suo retulisset antistiti: uberrimas lau-
... potenti deo uicem retulit. qui pro lapsis nature fragilitate ter-
... sistenciam sibi ministrorum dignatus prebere suffragium. H-
... decebat ueneratione sumptis a loco pignoris quo beatus an-
... sui memoriam fidelibus commendauerat. partem scilicet tuba-
... ipse memorat archangeli in monte gangano supra altare qua-
... straxerat posuit. & partem marmoris supra quod stetit. cui-
... usque nunc in eodem loco super octaua uestigia. iam dictis scrib-
... trum locum referenda patrocinia contradidit: conditione
... ta uidelicet ut quos una anglice reuelationis societate
... quoque eternaliter neceret caritatis conexio. Summi in-
... repedantes: post multa itineris spatia ad locum quo digressi
... so die quo fabrica completa est in monte iam dicto in occidentem
... quasi nouum ingressi sunt orbem: que primum uicium dent

Méthodes

Principes (1)

- * intégration du XML pour l'ensemble des productions
 - * solution adaptée aux grandes fonctions de l'éditeur scientifique (diffusion, production de formes référençables, traduction d'une structure logique en une forme intelligible adaptée à un (ou des) support(s) donné(s))
 - * rationalisation des pratiques autour du document numérique
 - * séparation fond / forme
 - * notion centrale de document structuré pérenne et archivable (indépendant des logiciels)
 - * niveau de balisage *éditorial*

Méthodes

Principes (2)

- * préserver la qualité éditoriale quel que soit le support de diffusion : culture du secrétariat d'édition
- * un seul flux de production pour tous les supports (*Single Source Publishing*)
- * une seule méthode de travail (conservation des outils habituels du secrétaire d'édition)

Méthodes

Outils

- ❖ **Production**

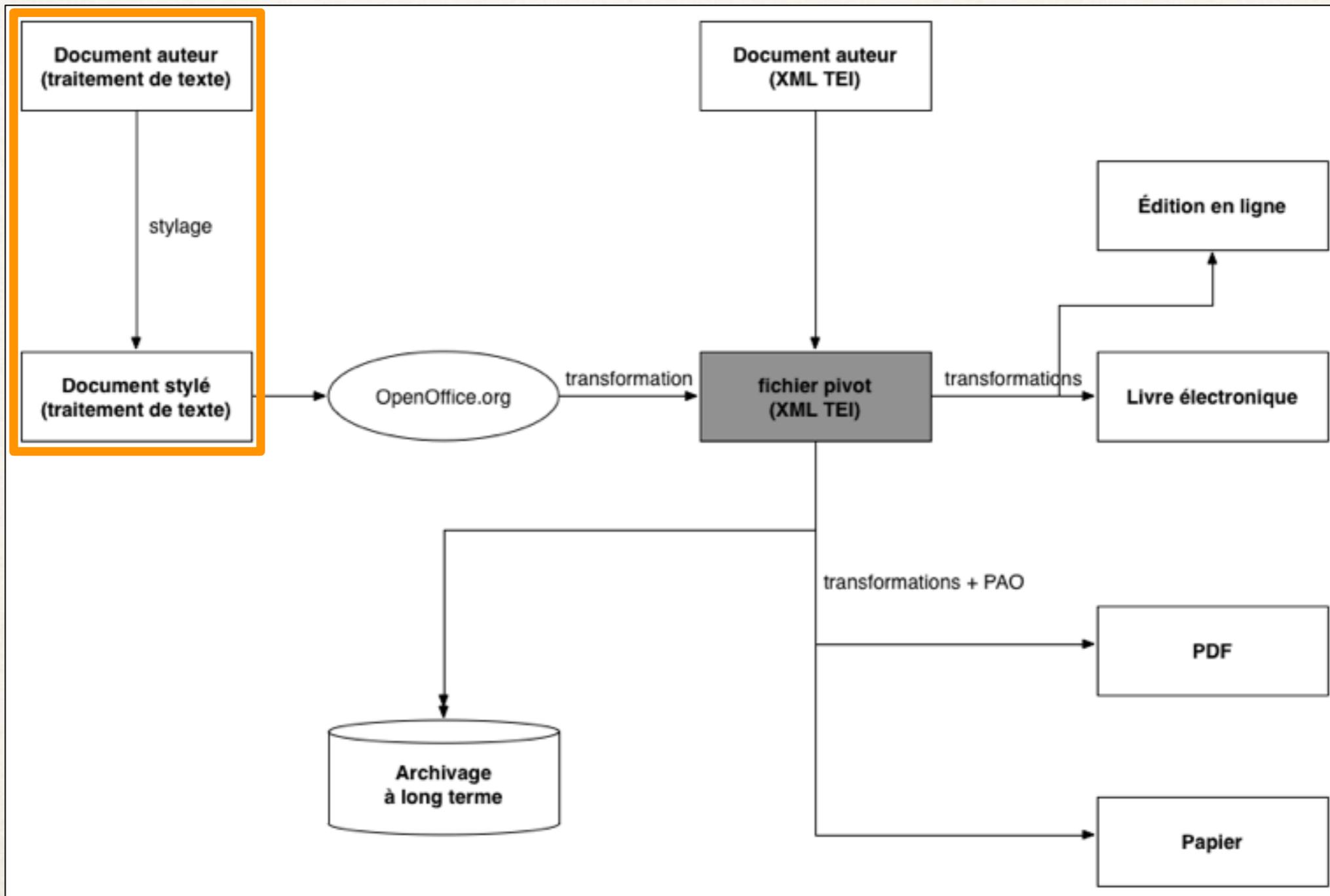
- ❖ Traitement de texte (styles et feuilles de styles)
- ❖ OpenOffice (traitement de texte et surtout outil de conversion)
- ❖ XSLT (*Extended Stylesheet Language Transformation*)
- ❖ Éditeur XML (XMLmind XML Editor, Oxygen, XMetal, Morphon, Millefeuille)
- ❖ Logiciel de PAO (Indesign, Xpress, FrameMaker)

- ❖ **Diffusion / exploitation / exploration**

- ❖ Langage de scripts (applescript, javascript)
- ❖ PHP, XSLT, CSS
- ❖ xquery, XML:DB (BaseX, eXist), Philologic, etc.

Méthodes

Traitement de texte



Méthodes

Traitement de texte (2)

Stylage systématique du texte

- ❖ objectif initial : systématisation des formes
- ❖ étiquetage de tous les éléments constitutifs du texte (paragraphes et séquences de caractères)
- ❖ c'est une étape de pré-structuration

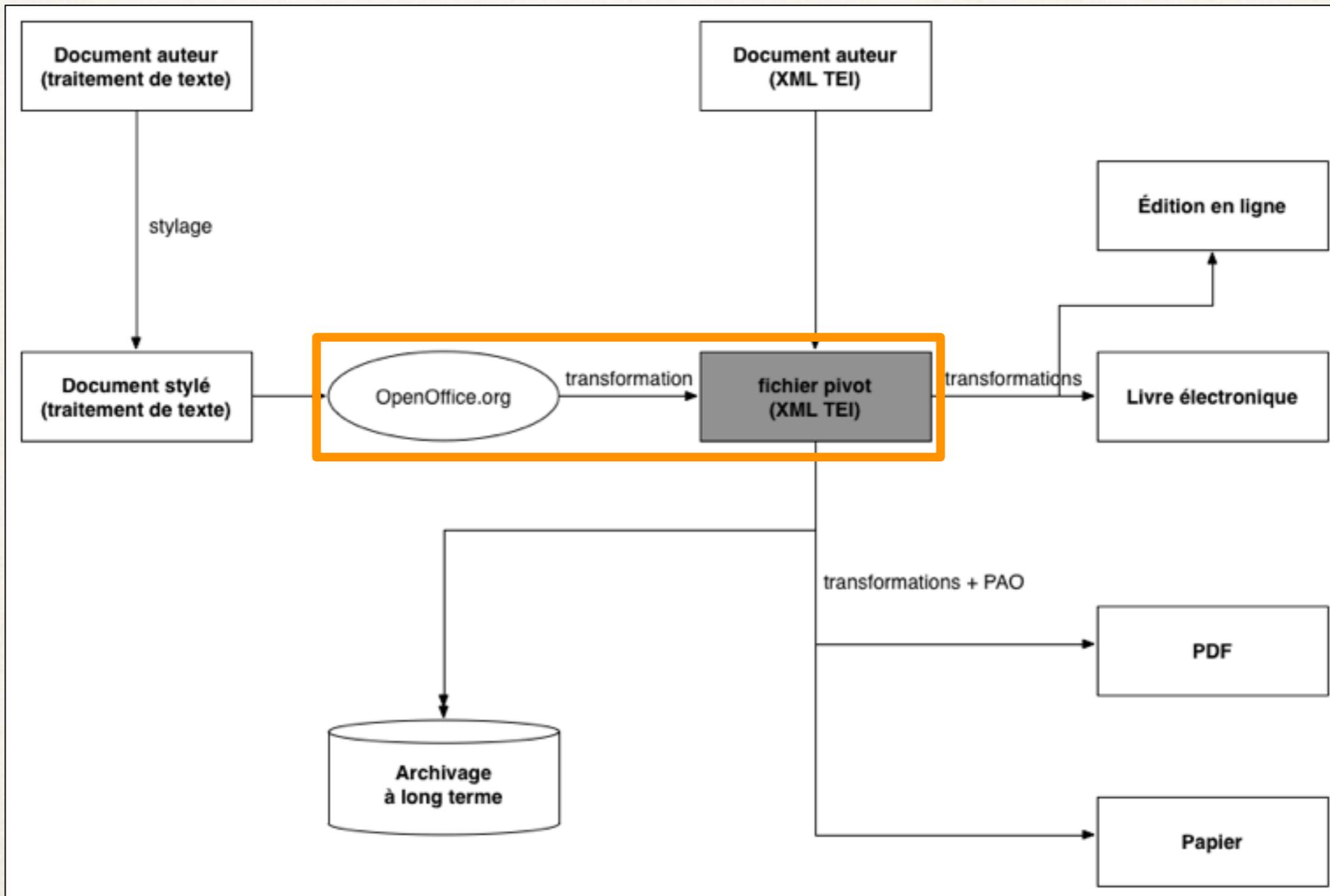
Méthodes

Traitement de texte (3)

Titre 1	Désignation et reconnaissance : le concours pour «chercher un terme appellatif qui remplace celui de CANUT» dans <i>L'Écho de la fabrique</i>
PE Auteur	Marie-France Piquet
PE Paragraphe sa	Contrairement à ce que l'on pourrait attendre, le mot <i>canut(s)</i> est peu employé dans <i>L'Écho de la fabrique</i> , et si la lecture de la version numérisée du journal en offre des occurrences nombreuses, c'est en raison surtout des notes de l'équipe éditoriale où les expressions «journal des canuts», «représentants des canuts», «revendications des canuts», «prud'homme canut», «lecteurs canuts», «intérieurs canuts»... sont régulièrement attestées, conformément à l'usage contemporain du terme.
PE Paragraphe al	Ceux que nous nommons maintenant <i>canuts</i> , ne se désignaient pas volontiers eux-mêmes sous ce terme, car il s'agissait alors d'un «sobriquet» qu'ils ressentaient comme une injure, d'une désignation humiliante qu'ils ont évité d'employer. L'exploration du texte numérisé permet même de soutenir que le mot <i>canut(s)</i> aurait été presque absent de <i>L'Écho de la fabrique</i> , si la question de cette dénomination n'avait été au centre d'une importante réflexion dont le journal a gardé la trace.
PE Paragraphe al	Des exemples de désignation stigmatisante d'un groupe social ne sont pas rares qui s'accompagnent souvent d'un retournement axiologique du désignant à la faveur d'un changement de rapport de forces. C'est ainsi que durant la Révolution l'expression «sans-culottes» a d'abord été une insulte, un «désignant socio-politique «ordurier» et «aristocratique»» avant de devenir une appellation assumée et revendiquée. Chaque cas participe cependant d'une situation singulière, et celui de <i>canut(s)</i> présente le grand intérêt d'avoir été explicitement formulé par les désignés eux-mêmes qui ont organisé au sein de <i>L'Écho de la fabrique</i> un étonnant «concours» pour trouver «un terme appellatif qui remplace celui de CANUT (en capitales dans le texte)».
PE Paragraphe al	C'est donc à ce mot qui fait aujourd'hui référence aux acteurs d'une révolte spécifique à l'histoire économique, sociale et culturelle lyonnaise, regardée souvent comme l'une des premières révoltes ouvrières du XIX ^e siècle, à son usage dans le journal de ceux qu'il désigne, à la mise en œuvre du concours et aux enjeux d'une telle désignation que cette étude est consacrée.

Méthodes

Conversion et édition XML



Méthodes

Conversion et édition XML (2)

Export XML OpenOffice.org vers XML TEI

- * OpenOffice est utilisé comme plate-forme d'exportation
- * exploitation des styles de paragraphes et de caractères pour produire les éléments XML
- * transformation XSLT pour réaliser la conversion en XML TEI
- * niveau de balisage éditorial

Méthodes

Conversion XML

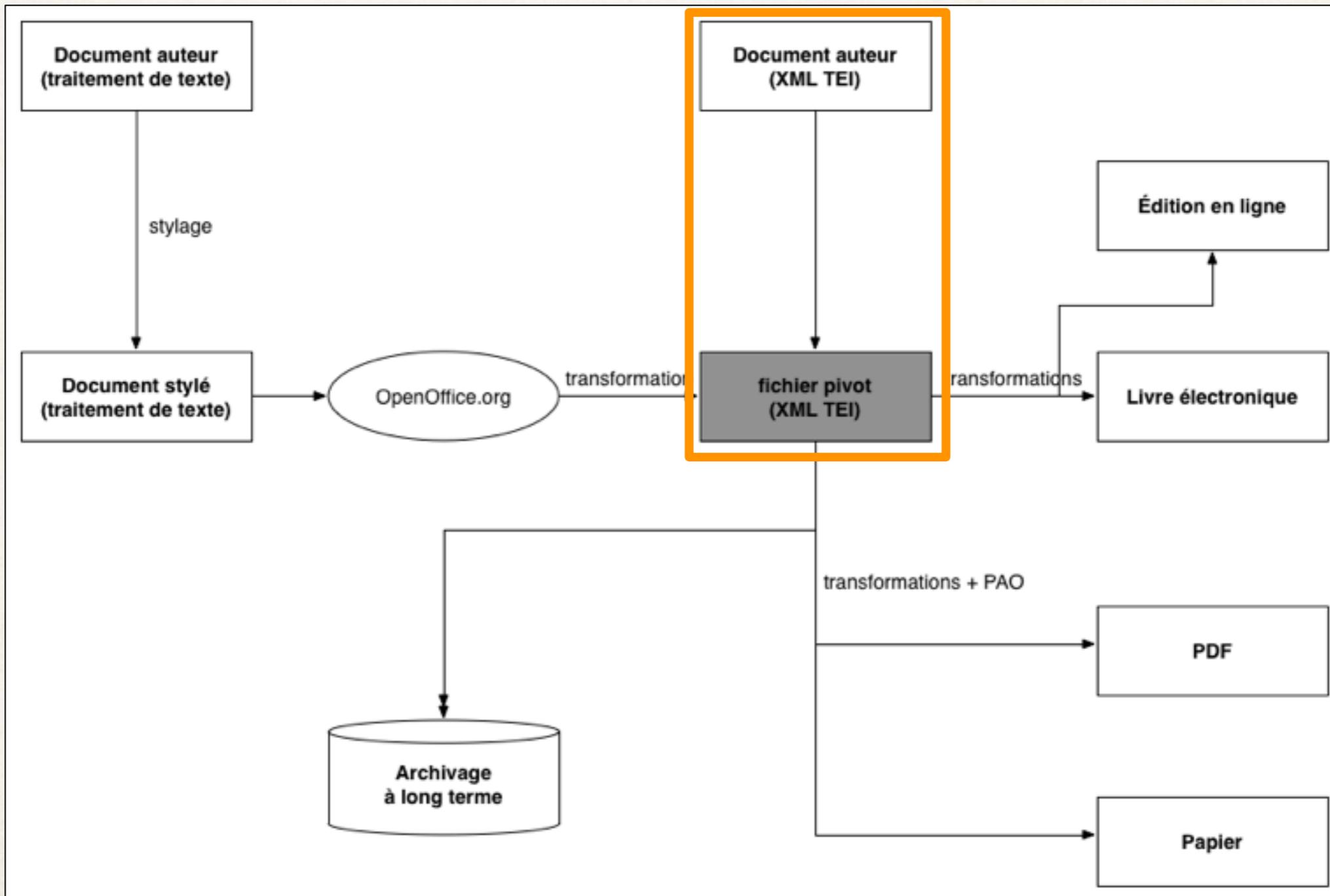
```
<text:h text:style-name="heading 2" text:level="2">
Installation des moines</text:h>
<text:h text:style-name="heading 3" text:level="3">
1. </text:h>
<text:p text:style-name="Standard">
1. La province de la Seconde Lugdunaise<text:span text:style-name="HH-Caractère de note de ba
<text:footnote text:id="ftn0">
<text:footnote-citation>
1</text:footnote-citation>
<text:footnote-body>
<text:p text:style-name="Footnote">
<text:s/>
Division administrative d'origine romaine, datant du début du IV<text:span text:style-name="T1
e</text:span>
siècle ap. J.-C.</text:p>
</text:footnote-body>
</text:footnote>
</text:span>
, qui s'appelle aujourd'hui la Normandie, passe à juste titre pour occuper le premier rang par
<text:footnote text:id="ftn1">
<text:footnote-citation>
2</text:footnote-citation>
<text:footnote-body>
<text:p text:style-name="Footnote">
<text:s/>
<text:span text:style-name="T2">
delectabilitas</text:span>
: ce substantif, inconnu à l'époque classique et à l'époque tardive, semble une création de l
<text:span text:style-name="T3">
delectabilis</text:span>
« agréable », « charmant », « séduisant ».</text:p>
</text:footnote-body>
</text:footnote>
</text:span>
des forêts, la douceur<text:span text:style-name="HH-Caractère de note de bas de page1111111
<text:footnote text:id="ftn2">
<text:footnote-citation>
3</text:footnote-citation>
<text:footnote-body>
<text:p text:style-name="Footnote">
```

```
<head aid:pstyle="T_1_Section" xml:id="FRintro.1.0">Installation des
moines</head>
<div type="section" xml:id="FRintro.1.1" xml:lang="fr">
<head aid:pstyle="T_3_Chapitre" xml:id="FRintro.1.1.0">1.</head>
<p aid:pstyle="txt_Original_Prose_Numero"
xml:id="FRintro.1.1.1">1. La province de la Seconde Lugdunaise<note
xml:id="FRftn1">Division administrative d'origine romaine, datant du
début du <hi aid:cstyle="typo_SC_regular" rend="small-caps">IV</hi><hi aid:cstyle="
J.-C.</note>, qui s'appelle aujourd'hui la <index indexName="Lieux">
<term>Normandie / Normannia</term>
</index>Normandie, passe à juste titre pour occuper le premier
rang parmi toutes les autres provinces que la Gaule renferme à
l'intérieur de ses frontières, du fait que, comme cela a été bien
reconnu, non seulement elle se suffit à elle-même par l'abondance
considérable de toutes ses ressources, mais elle fournit aussi aux
provinces voisines des contributions non négligeables. En effet la
salubrité de l'air, la fécondité d'un sol riche, la fertilité des
vignes, le charme<note xml:id="FRftn2">delectabilitas: ce
substantif, inconnu à l'époque classique et à l'époque tardive,
semble une création de l'auteur à partir de l'adjectif
<emph>delectabilis</emph> «&#160;agréable&#160;», «&#160;charmant&#160;»,
«&#160;séduisant&#160;».</note> des forêts, la douceur<note
xml:id="FRftn3">apricitas: nom formé à partir de l'adjectif
<emph>apricus</emph>, «&#160;ensoleillé&#160;», «&#160;exposé au soleil&#160;»; tern
utilisé par Columelle (7, 4, 5: <emph>apricitas diei</emph>), Pline
l'Ancien (<emph>Nat</emph>. 6, 46) et le géographe du <hi
aid:cstyle="typo_SC_regular" rend="small-caps">III</hi><hi
aid:cstyle="typo_Exposant" rend="sup">e</hi>&#160;siècle Solin (61,
1). Ce mot a été repris par la suite par des auteurs chrétiens comme
Salvien de Marseille (<emph>Ecc1</emph>. 4, 35) et Rurice de Limoges
(<emph>Epist</emph>. 1, 17) au <hi aid:cstyle="typo_SC_regular"
rend="small-caps">V</hi><hi aid:cstyle="typo_Exposant"
rend="sup">e</hi>&#160;siècle.</note> des bois et des vergers,
l'agrément des jardins et des plantations salutaires, l'abondance
des matériaux de toutes sortes<note xml:id="FRftn4">metallorum
congerie: <emph>metallum</emph> signifie «&#160;matériaux de toutes
sortes&#160;» (métaux, marbre, pierres, terres particulières) que renferme
le sous-sol de Normandie. C'est en ce sens que Walafrid Strabon l'a
```

XML OpenOffice.org → XML TEI

Méthodes

Conversion et édition XML (3)



Méthodes Édition XML

```
<TEI xsi:schemaLocation="http://www.tei-c.org/ns/1.0 file:HS/HS.xsd"
xmlns="http://www.tei-c.org/ns/1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:xhtml="http://www.w3.org/TR/xhtml1/strict"
xmlns:ns="http://www.tei-c.org/ns/1.0"
xmlns:hfp="http://www.w3.org/2001/XMLSchema-hasFacetAndProperty"
xmlns:aid5="http://ns.adobe.com/AdobeInDesign/5.0/"
xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/">
<teiHeader xml:lang="la">
</teiHeader>
<text>
<front>
<titlePage>
<byline><docAuthor>Johannes de Cuba</docAuthor></byline>
<titlePart type="univ"><hi rend="italic">Hortus
sanitatis</hi></titlePart>
</titlePage>
</front>
<body>
<div type="livre">
<head type="traite" aid:pstyle="T_1">Tractatus de piscibus</head>
<div n="prologue" type="chapitre" xml:id="LAT.hs.4.prologue">
</div>
<div type="chapitre" xml:id="LAT.hs.4.1" xml:lang="la">
</div>
<div type="chapitre" xml:id="LAT.hs.4.2" xml:lang="la">
<head type="chapitre" aid:pstyle="T_a">
<head type="chapitre_poisson" aid:pstyle="T_b"><index
indexName="piscibus">
</index>Anguilla <bibl subtype="chapitre" type="source">
<author>VB</author> <biblScope>17, 31</biblScope> <title
type="chapitre">De anguilla</title> </bibl></head>
<div type="section1" xml:id="LAT.hs.4.2.section" xml:lang="la">
<div type="section1" xml:id="LAT.hs.4.2.section.2" xml:lang="la">
<head type="operationes" aid:pstyle="T_c">Operationes</head>
<p n="6" aid:pstyle="txt_Normal" xml:id="LAT.hs.4.2.6">
<p n="7" aid:pstyle="txt_Normal" xml:id="LAT.hs.4.2.7">
<p n="8" aid:pstyle="txt_Normal" xml:id="LAT.hs.4.2.8">
</div>
</div>
```

The screenshot shows a web browser displaying the rendered XML content. The page title is "Tractatus de piscibus". The main text is in Latin and includes several sections with associated notes and apparatuses. The text is rendered with various styles, including bold and italic, and includes a table of attributes on the right side.

Tractatus de piscibus

|||||Prohemium [Note LATnote2 (apparat) : prohemium — procedemus non hab. 2536.]

[[AM 24 (1)] Jam [Note LATnote3 (apparat) : jam — natatibus hab. HS ex AM] in praecedenti tractatu [Note LATnote4 (apparat) : libro AM] locuti sumus de volatilibus, quae in genere animalium videntur esse [Note LATnote5 (apparat) : videntur esse per perfectiones hab. AM] perfectiora natatilibus.] [[Nota HS] Ideo nunc in isto quarto tractatu, Deo opitulante, dicemus de natatilibus sive animalibus aquaticis, hoc est de piscibus.] [[VB 17, 1] Nam [Note LATnote6 (apparat) : nam — volando hab. HS ex FR] sicut [Note LATnote7 (apparat) : nam sicut : sicut ergo FR] aer ornatus est suis animalibus — scilicet [Note LATnote8 (apparat) : idem FR] volatilibus, quae possunt ex eo respirare propter crassitudinem humoris et quorum corpora potest [Note LATnote9 (apparat) : possit FR] sustinere —, sic et aqua suis, id est natatilibus. Quae quidem [Note LATnote10 (apparat) : quidem non hab. FR] pinnulis suis innituntur [Note LATnote11 (apparat) : innituntur Proba situat FR] aquae natando, sicut aves pennis suis [Note LATnote12 (apparat) : per suis hab. aeri FR] volando.] [[Nota HS] De piscibus igitur dicturi secundum ordinem alphabeti et per capitula sicut in praecedentibus procedemus.]

Capitulum I [Note LATnote13 (apparat) : caput 1 2536.]

#Abremon, #achandes, #accipender [Note LATnote14 (apparat) : accipenser 2536 ut semper], #albirem et #alphoram [VB 17, 29 De abremon [Note LATnote15 (apparat) : abremon VB] et achande et accipendro, albitrez [Note LATnote16 (apparat) : albitrez vel albiterum VB ut semper albiter VB ut semper], alphoraz [Note LATnote17 (apparat) : alphas vel alporum VB ut semper.]]

Attri...	Valeur
aid:c...	-
aid:...	-
aid:t...	-
aid:t...	-
aid:t...	-
copyOf	-
corr...	-
excl...	-
n	-
next	-
prev	-
rend	italic
rend...	-
sam...	-
select	-
synch	-
xml...	-
xml:id	-
xml:l...	-
xml:...	-
xsl:n...	-
xsl:s...	-
xsl:t...	-

Méthodes

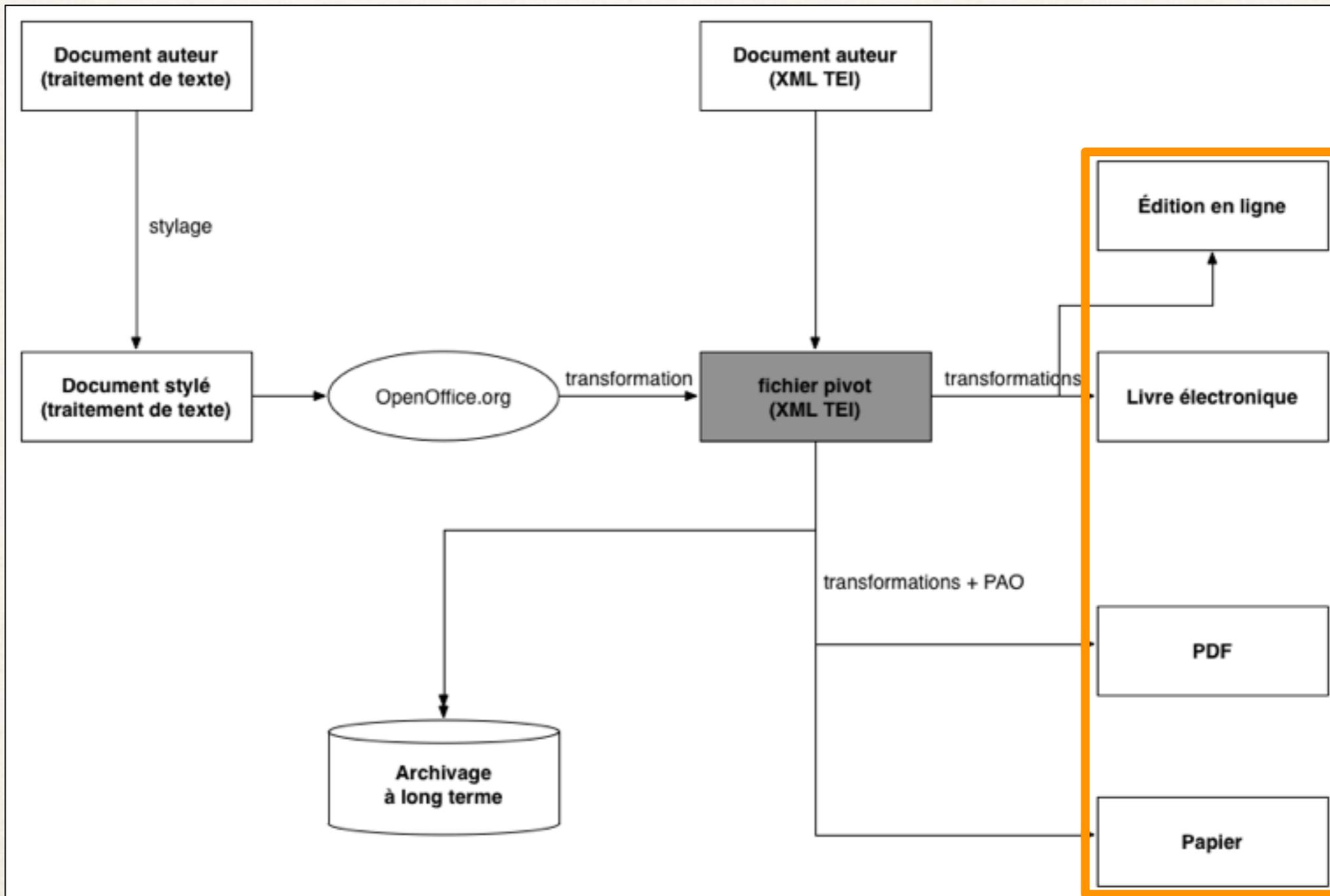
Édition XML (2)

Enrichissement de la structure et corrections éditoriales

- ❖ le fichier XML TEI produit présente toutes les qualités nécessaires pour un archivage pérenne (indépendance technique, encodage des caractères en unicode)
- ❖ c'est le fichier qui doit contenir toutes les corrections (établissement définitif du fond)
- ❖ les interfaces de travail sont entièrement paramétrables (commandes XML et CSS)

Méthodes

Formes de diffusion



Méthodes

Formes de diffusion

- * toutes les formes de diffusion sont produites à partir du fichier XML : soit automatiquement par des feuilles de transformation XSLT (XHTML) soit par une importation dans un logiciel de PAO (papier, PDF).
- * exploitation contextuelle des structures XML pour produire des formes intelligibles sur des supports donnés
- * exports pour revues.org (CLEO) ou CAIRN

Méthodes

Formes de diffusion – Éditions en ligne

Accueil Sommaire Index Bibliographie Recherche plein texte

Mode d'emploi Sigles et abréviations Fac-similés << Ch. 16 - Cancer >> Options d'affichage

Hortus Sanitatis
Tractatus de piscibus

CHAPITRE 16

Cancer [le crabe ; le homard ; l'écrevisse^[1]] [•] [•] [•] [•]

CAPITULUM XVI^[1]

Cancer [•] [•] [•] [•]

Renvois internes : Cancer : cf. Circhos, ch. 21 ; Karabo, ch. 48 ;
Locusta marina, ch. 52.

Lieux parallèles : TC, De cancris (7, 19) ; AM, [Cancer] (24, 31
(22)).



[1] [•] D'après le *Liber de naturis rerum*. [•] Le cancer est pourvu de pattes, de bras et de pinces, qui lui servent de mains. Il marche en arrière et il ne peut jamais avancer droit devant lui^[2]. [•] Les mâles possèdent deux aiguillons entre le ventre et la queue, qui manquent chez les femelles^[3]. Les femelles qui portent des œufs sous leur ventre guérissent des morsures de serpents^[4]. [•] Au moment de l'accouplement, le mâle grimpe sur le dos de la femelle qui, lorsqu'elle sent au mouvement du mâle sur son dos qu'il veut s'accoupler, s'incline sur le côté et se retourne face à lui, et c'est ainsi que l'accouplement se fait^[5]. [•] Les cancri [les langoustes] restent cachées pendant les cinq mois d'hiver. Au printemps, elles se débarrassent de leur vieille carapace, comme font les serpents^[6]. [•] Un cancer à qui on a donné à boire du lait vit plusieurs jours sans eau. [•] Lorsqu'il est vieux, on trouve dans sa tête deux pierres de couleur blanche pailletées de rouge, qu'on prétend dotées d'une si grande vertu que, données en boisson, elles guérissent les élancements du cœur.

[1] [•] Ex Libro de naturis rerum^[2]. [•] Cancer pedes habet et brachia forcipisque pro manibus. Retrogradus est nec unquam ante faciem suam ambulare novit. [•] Mares duas habent spinas inter ventrem et caudam, quibus feminae carent. Quae ova in ventre habent ictibus serpentum^[3] medentur. [•] Coitus tempore mas ascendit in dorso super feminam. At illa, motum ejus sentiens a tergo coire volentis, inclinata super latus ad ipsum se convertit, sicque [Prisus/ve] coitus completur. [•] Latent in hieme mensibus quinis. Veris tempore senectutem exuunt velut anguis. [•] Cancer lacte potatus sine aqua vivit multis diebus. [•] Cum autem senuerit, duo lapides albi coloris intermixto rubore inveniuntur in ejus capite ; quos tantae virtutis esse dicunt ut in potu^[4] dati cordis punctiones sanant.

Méthodes

Formes de diffusion – Parcours de lecture (1)

ms210

Revelatio

Introductio monachorum

ms211a

Liber de apparitione

Revelatio

ms211b

Liber de apparitione

Revelatio

Introductio monachorum

Miracula I-II

De translatione

De scuto et gladio

Miracula III-X

Miracula XI

ms212

Revelatio (variante ms212)

Introductio monachorum
(variante ms212)

De scuto et gladio

Miracula VII-X

Miracula XI

Miracula III-VI

ms213

Liber de apparitione

Revelatio

Revelatio adjonction 2

Introductio monachorum

Miracula I-VI

De translatione

Miracula VII-X

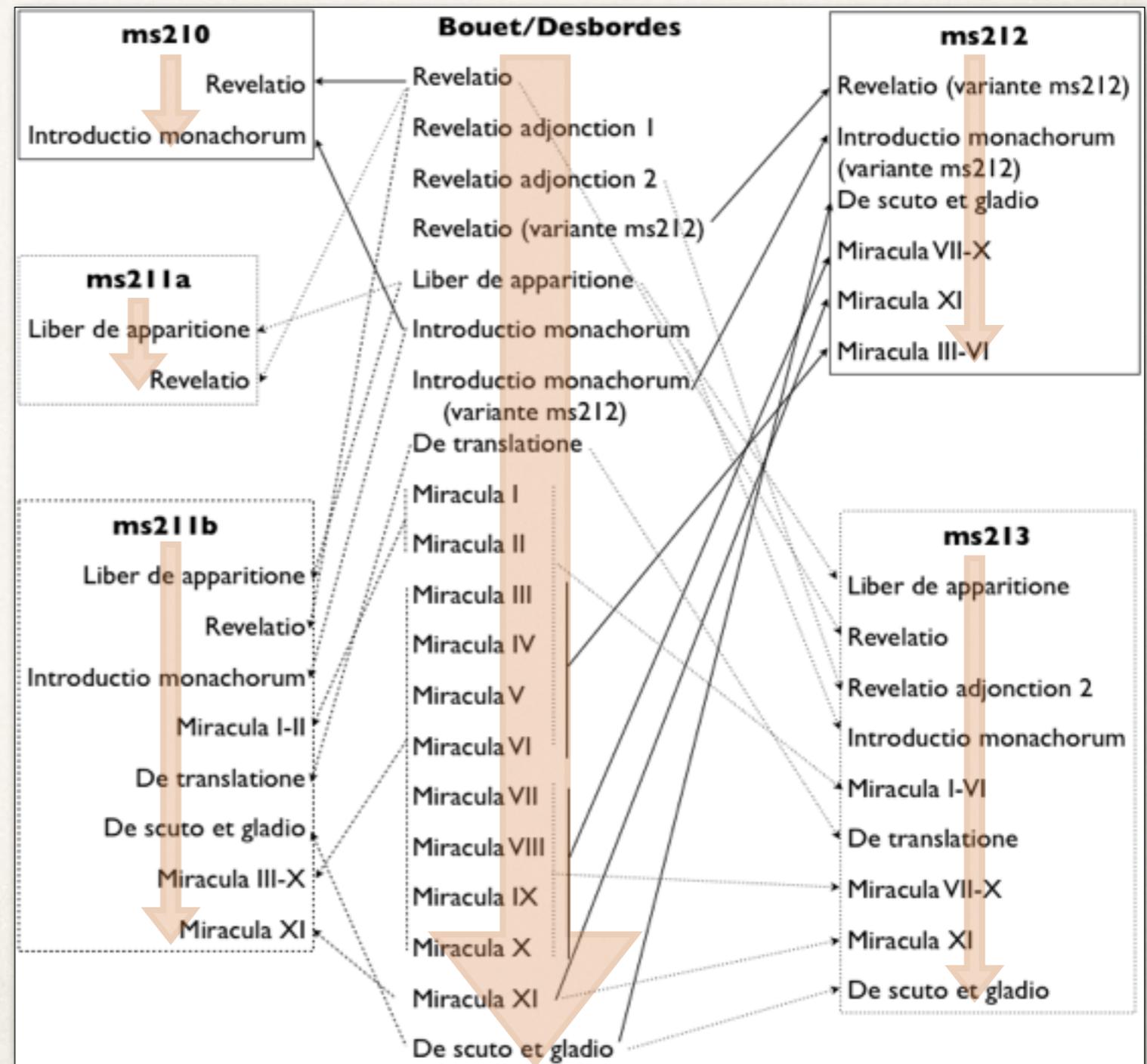
Miracula XI

De scuto et gladio

Méthodes

Formes de diffusion – Parcours de lecture (2)

- ❖ Constitution du flux des chroniques latines
- ❖ Parcours de lecture de la collection de textes selon P. Bouet et O. Desbordes
- ❖ Plusieurs parcours de lecture



Méthodes

Formes de diffusion – PAO

Importation du fichier XML dans un logiciel de PAO

- * on automatise un maximum de tâches de mise en page à partir de la structure XML
- * au moyen de scripts
- * établissement d'une forme
- * exploitation des espaces de noms d'Indesign pour chaque élément devant faire l'objet d'une mise en forme systématisable :
- * ajout de l'espace de noms : `xmlns:aid="http://ns.adobe.com/AdobeInDesign/4.0/"`
- * Pour les paragraphes : attribut *aid:pstyle*
- * Pour les caractères : attribut *aid:cstyle*

Méthodes

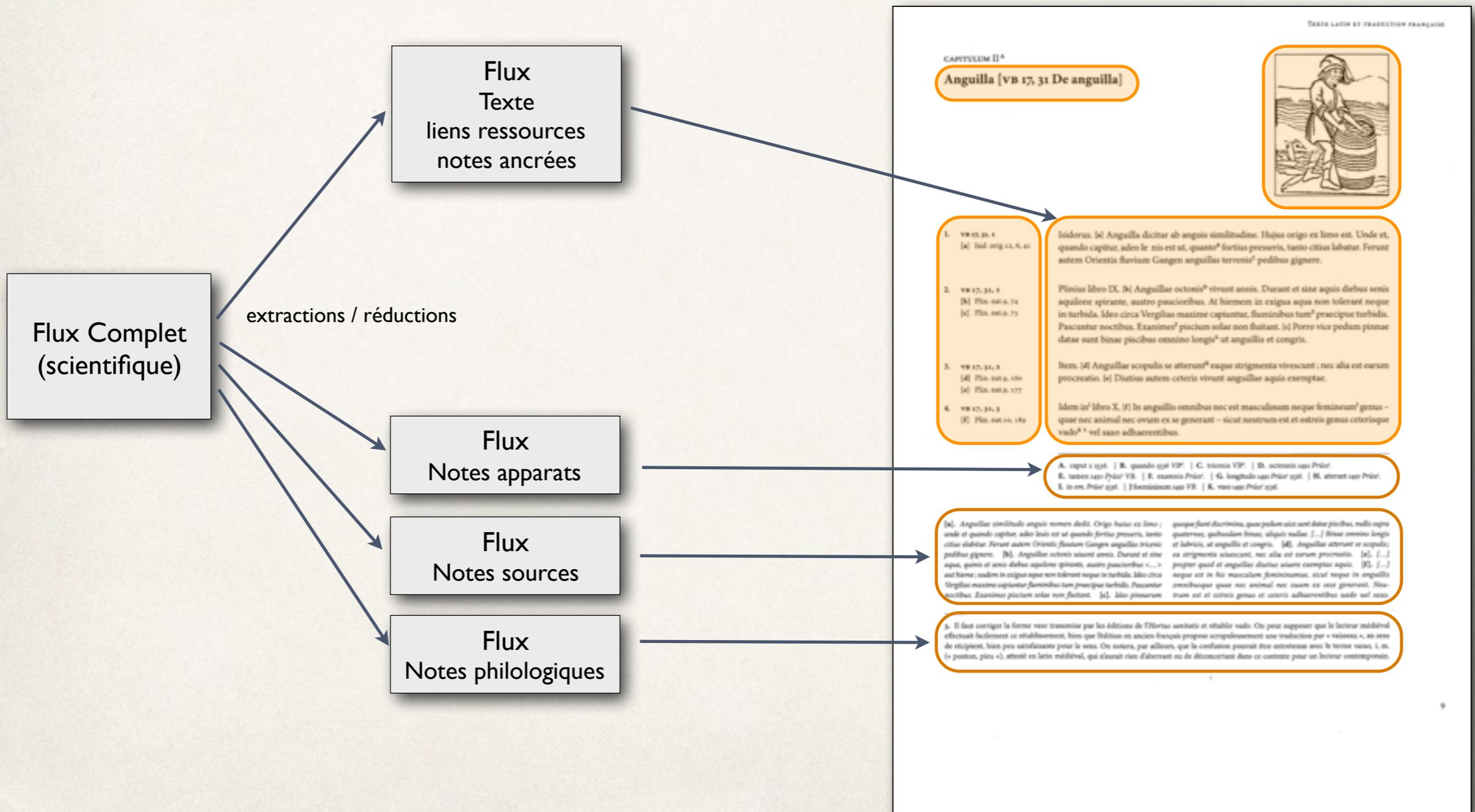
Formes de diffusion – PAO

The image shows a digital manuscript viewer interface. On the left, there is a sidebar with a tree view of the document structure, including elements like 'lang = fr-FR', 'xmlns:ad = http://ns.adobe.com/...', 'h1Header', 'text', 'font', 'body', 'emph', 'div1', 'div2', 'div3', and 'div4'. The main area displays a page from a manuscript, titled 'LE RINGE DU MONT SAINT MICHEL'. The text on the page is in French and discusses the customs of the monks of Mont Saint-Michel. On the right side of the page, there is a list of marginal notes, each with a small icon and a reference number. The notes are organized into a list that corresponds to the text on the page. The interface also shows a ruler at the top and a scroll bar on the left.

Ici toutes les notes marginales sont placées automatiquement à partir du flux XML

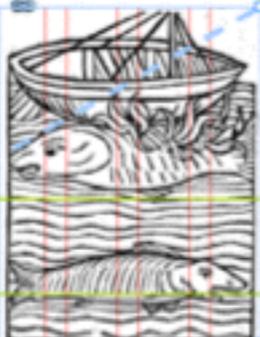
Méthodes

Séparation des flux + réduction du niveau d'encodage



Méthodes

Formes de diffusion – PAO

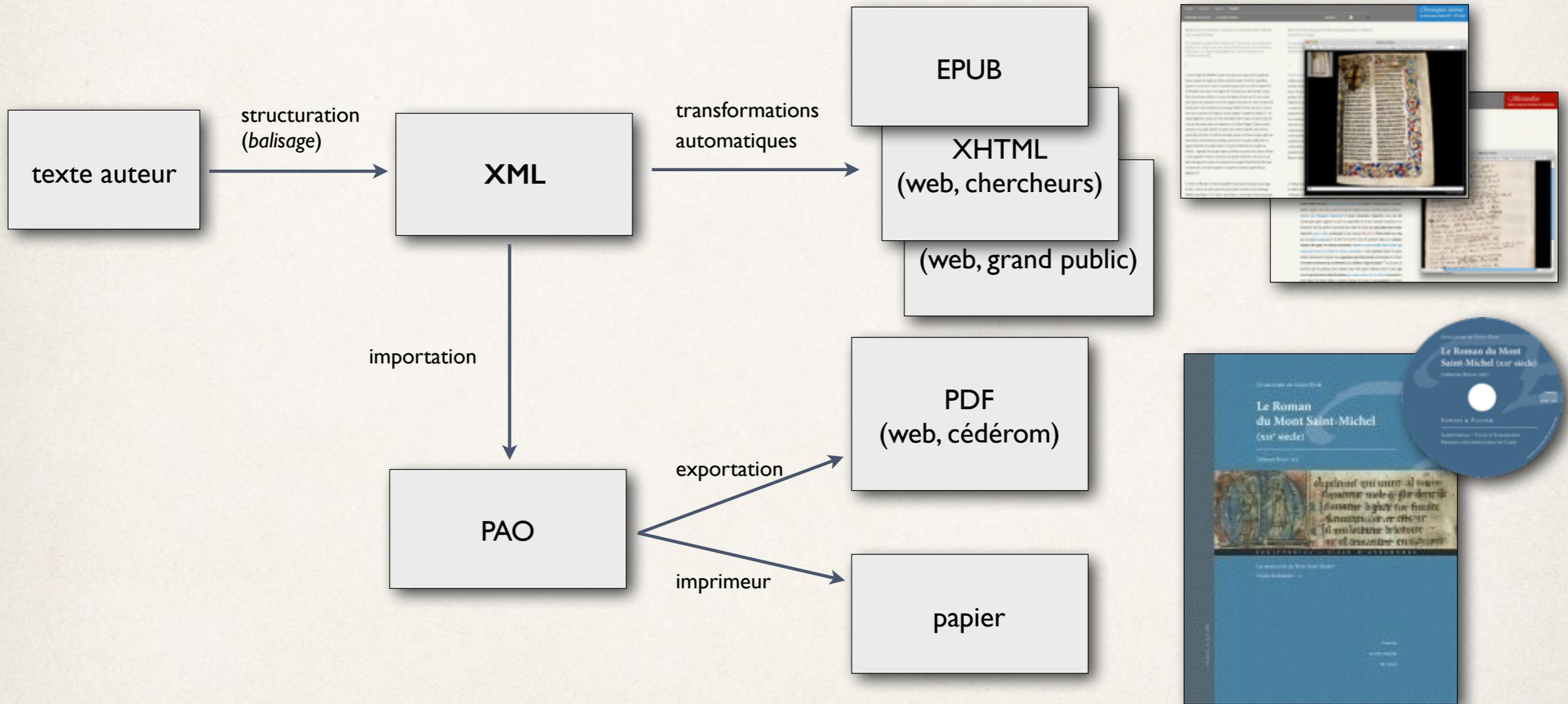
HORTUS SANCTI VINCENTII	CAPITULUM V ^m	TEXTE LATIN ET TRADUCTION FRANÇAISE
<p>CHAPITRE 5^m Aspidochelon [= l'aspic-tortue =] et aurata [l. dorade] VR 17, 33 De aspidochelone et aurata</p>	<p>Aspidochelon^m et aurata VR 17, 33 De aspidochelone et aurata</p>	
<p>1^o – aspic. <i>Aspidochelon</i> est le nom donné en grec à un monstre marin qu'on appelle en latin <i>aspidochelone</i> [aspic-tortue]. C'est un grand citracé dont le cuir fait penser au gravier des plages. Il arrive donc souvent que des marins confondent son dos, quand il émerge au-dessus des flots, avec une île: ils y plantent leurs grappins pour amarrer leur bateau, y allument un feu pour faire cuire leur repas. Mais lorsque l'animal sent l'odeur et la chaleur du feu, il s'éloigne étourdi alors les bateaux par le fond. Voici un autre trait de sa nature: quand il a faim, il ouvre la bouche et il s'en exhale comme une odeur suave. Les poissons plus petits, par l'odeur alléchés, convergent dans sa gueule, et quand il a la bouche bien pleine, il la referme sur ses poies et les avale¹.</p> <p>Le même. La dorade doit son nom au fait que sa tête a la couleur de fer.</p>	<p>1^o Physiologus. [a] <i>Aspidochelone</i>^m Graeco dicitur belua quaedam in mari quae Latine dicitur <i>aspidochelone</i>. Et autem citra magnas habens super costas suas tanquam tabulam quod est iuxta litus maris. Unde plerumque, elevata de suo super undas, a navigantibus nihil aliud creditur esse quam insula. Itaque applicantes dentes^m super eam^m, alligant naves^m, accendant focos ad coquendum cocum^m ille^m vero, ubi sentit odorem [b] et calorem^m [c] ignis, [d] movetur^m et recessit, et [e] mergit naves in profundum. Alia quoque huius beluae natura haec est: quando erit, aperit os suum et quasi quendam odorem bene olerentem exhalat de ore suo, quae in sentientes minores pisces confluat in os ejus, quod ut impletum fuerit, claudit et eos transglutit.</p> <p>[f] Item, [g] <i>Aurata</i> dicitur ex eo quod in capite suo auri colorem^m habet.</p>	
<p>1^o – aspic. <i>Aspidochelone</i> combine avec l'absence des traits qu'on associe traditionnellement à la balaine (l'animal marin géant et le faucibus d'homme) et des caractéristiques propres à la tortue d'une part, et à l'aspic d'autre part. Ainsi, par sa première nature, l'animal est d'abord une «tortue», dont le caractère le plus remarquable est son poids pour un coque, c'est le thème le plus traditionnel de l'île trompeuse avec ses prétextes (Homerus d'Alexandre 3, 17; Strabon XV, 1, 12) et ses nombreux avatars, avec en particulier, pour le Moyen Âge occidental, l'île enchantée de <i>la Navigatio de saint Brendan</i> (par exemple, <i>Le voyage de saint Brendan</i>, v. 440-450, in <i>Navigatio sancti Brendani abbatis</i>, p. 10, dans «<i>Navigatio sancti Brendani</i>» éd. C. Selmer, Notre-Dame, University of Notre-Dame Press (Publications in Medieval Studies, 16), 1929, p. 10-11). Par sa seconde nature, le monstre magique est aussi un aspic. Le chapitre du <i>Physiologus</i> que consacre à l'aspic (<i>Physiologus</i> 33, voir Zacher 2004, 276-283) est éclairant, il raconte comment la femelle ouvre la bouche, où plonge le mâle, et la referme sur le corps du plâtre rose, preuve que, ainsi que nous l'avons vu, c'est aussi le cas de <i>la Navigatio de saint Brendan</i> (par exemple, <i>Le voyage de saint Brendan</i>, v. 440-450, in <i>Navigatio sancti Brendani</i>, p. 10, dans «<i>Navigatio sancti Brendani</i>» éd. C. Selmer, Notre-Dame, University of Notre-Dame Press (Publications in Medieval Studies, 16), 1929, p. 10-11). Par sa seconde nature, le monstre magique est aussi un aspic. Le chapitre du <i>Physiologus</i> que consacre à l'aspic (<i>Physiologus</i> 33, voir Zacher 2004, 276-283) est éclairant, il raconte comment la femelle ouvre la bouche, où plonge le mâle, et la referme sur le corps du plâtre rose, preuve que, ainsi que nous l'avons vu, c'est aussi le cas de <i>la Navigatio de saint Brendan</i> (par exemple, <i>Le voyage de saint Brendan</i>, v. 440-450, in <i>Navigatio sancti Brendani</i>, p. 10, dans «<i>Navigatio sancti Brendani</i>» éd. C. Selmer, Notre-Dame, University of Notre-Dame Press (Publications in Medieval Studies, 16), 1929, p. 10-11).</p>	<p>1^o – aspic. <i>Aspidochelone</i> combine avec l'absence des traits qu'on associe traditionnellement à la balaine (l'animal marin géant et le faucibus d'homme) et des caractéristiques propres à la tortue d'une part, et à l'aspic d'autre part. Ainsi, par sa première nature, l'animal est d'abord une «tortue», dont le caractère le plus remarquable est son poids pour un coque, c'est le thème le plus traditionnel de l'île trompeuse avec ses prétextes (Homerus d'Alexandre 3, 17; Strabon XV, 1, 12) et ses nombreux avatars, avec en particulier, pour le Moyen Âge occidental, l'île enchantée de <i>la Navigatio de saint Brendan</i> (par exemple, <i>Le voyage de saint Brendan</i>, v. 440-450, in <i>Navigatio sancti Brendani</i>, p. 10, dans «<i>Navigatio sancti Brendani</i>» éd. C. Selmer, Notre-Dame, University of Notre-Dame Press (Publications in Medieval Studies, 16), 1929, p. 10-11). Par sa seconde nature, le monstre magique est aussi un aspic. Le chapitre du <i>Physiologus</i> que consacre à l'aspic (<i>Physiologus</i> 33, voir Zacher 2004, 276-283) est éclairant, il raconte comment la femelle ouvre la bouche, où plonge le mâle, et la referme sur le corps du plâtre rose, preuve que, ainsi que nous l'avons vu, c'est aussi le cas de <i>la Navigatio de saint Brendan</i> (par exemple, <i>Le voyage de saint Brendan</i>, v. 440-450, in <i>Navigatio sancti Brendani</i>, p. 10, dans «<i>Navigatio sancti Brendani</i>» éd. C. Selmer, Notre-Dame, University of Notre-Dame Press (Publications in Medieval Studies, 16), 1929, p. 10-11).</p>	

Conclusion et perspectives

INCIPIT REVELA
TIO ECCLESIE DEI
MICHAELIS ARCHAN
GELI. IN OMNIBUS QUI
DICITUR TUOBA
IN OCCIDENTIS PAR
TIB: SUB
CHILDERTO REGE
FRANCORUM ET
LUTBERTO REGE

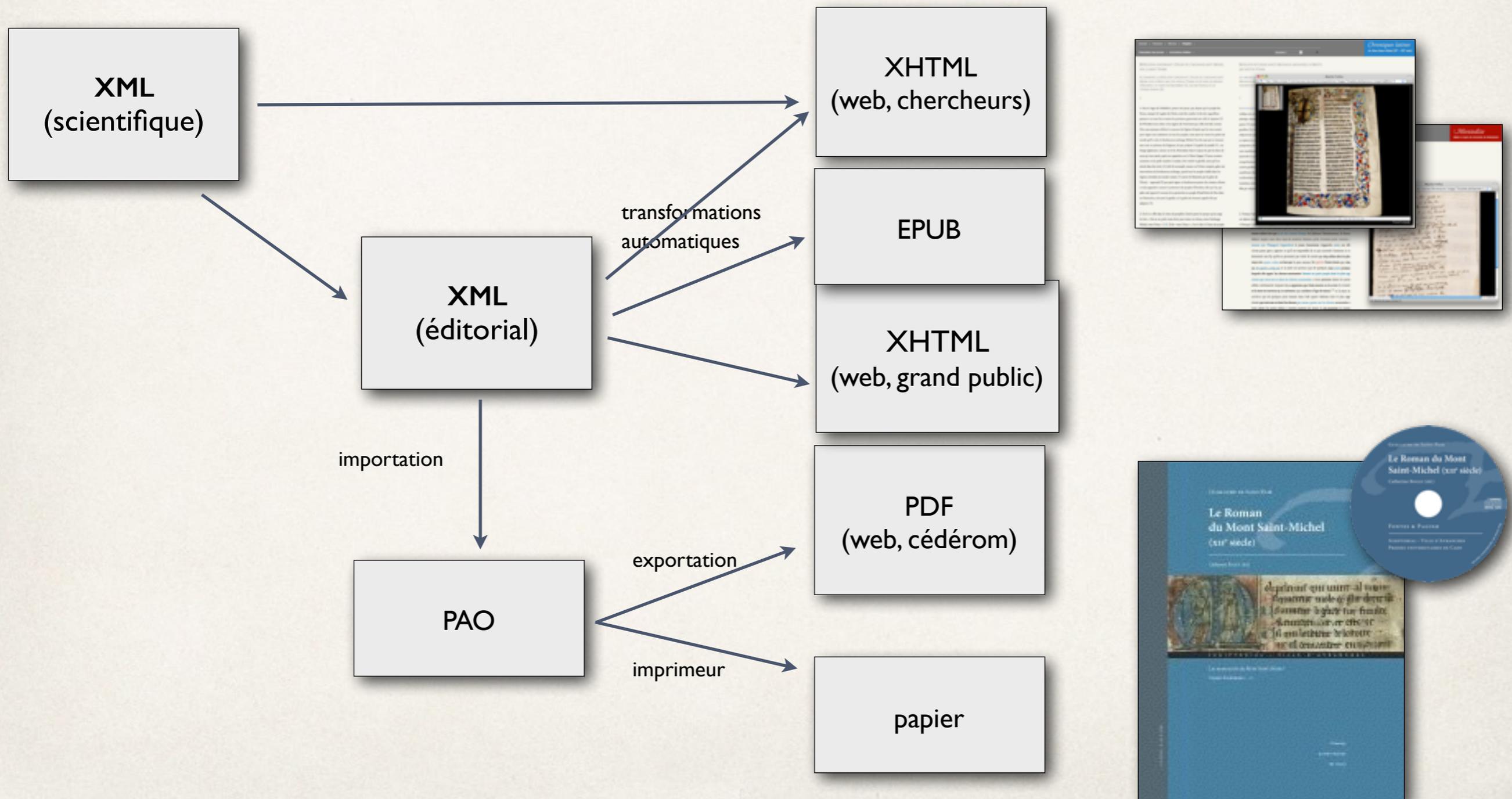
Single Source Publishing

Balisage éditorial



Single Source Publishing

Balisage scientifique + balisage éditorial

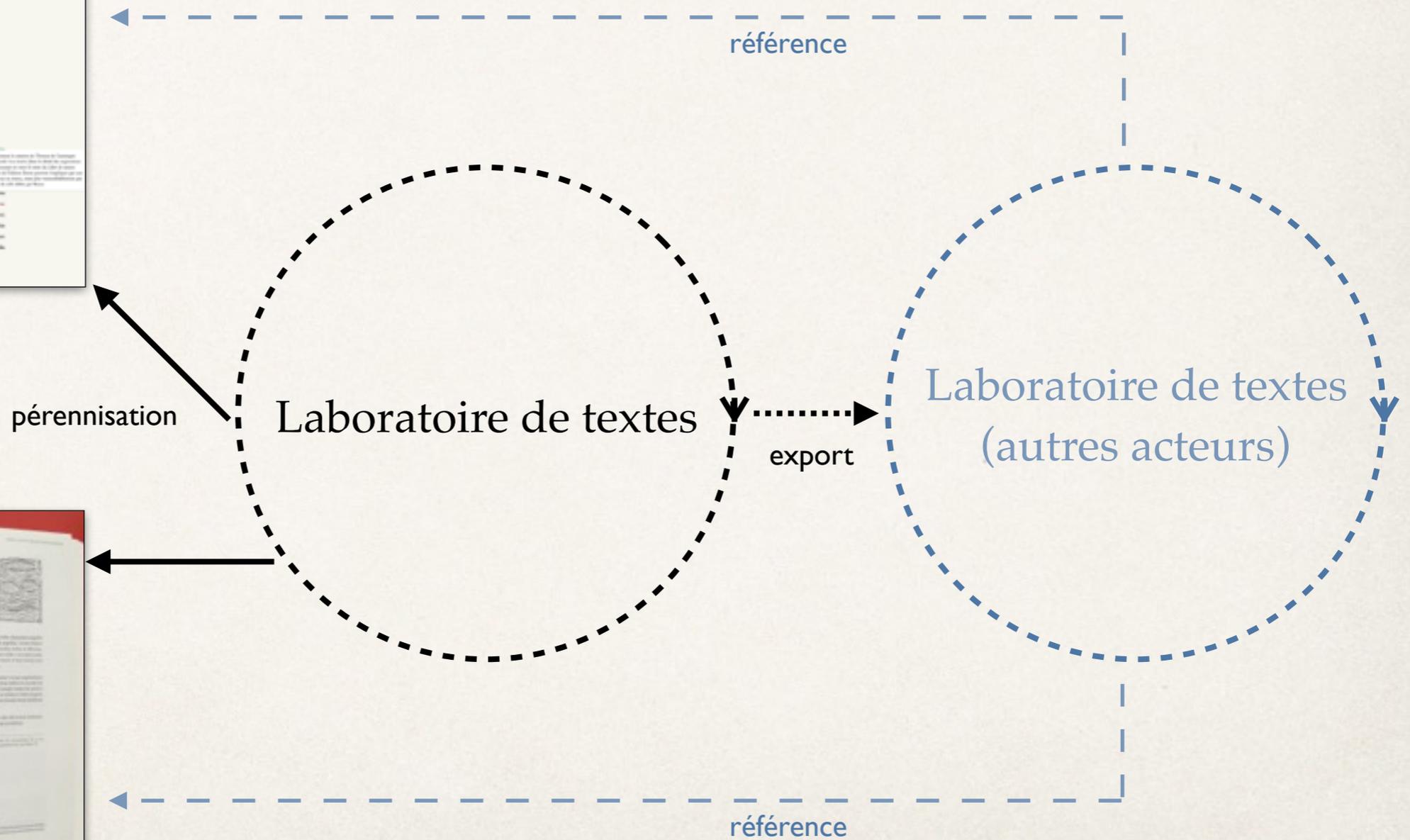
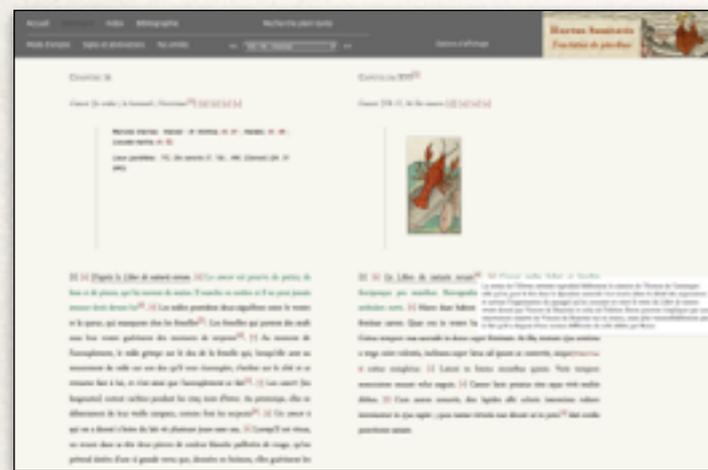


Organisation

Articulation – recherche / édition / ingénierie

- * collaboration constante entre les Presses universitaires de Caen et le Pôle Document Numérique de la MRSH
- * collaboration avec les chercheurs en amont de la publication (soutien à la recherche)
- * développement d'outils (CERTIC), définition de protocoles, de normes, de procédures pour le travail éditorial ou pour une recherche tournée vers l'édition
- * du cas le plus simple (revue) au plus complexe (édition de sources)
- * la qualité des outils de valorisation est fortement liée à la participation des chercheurs

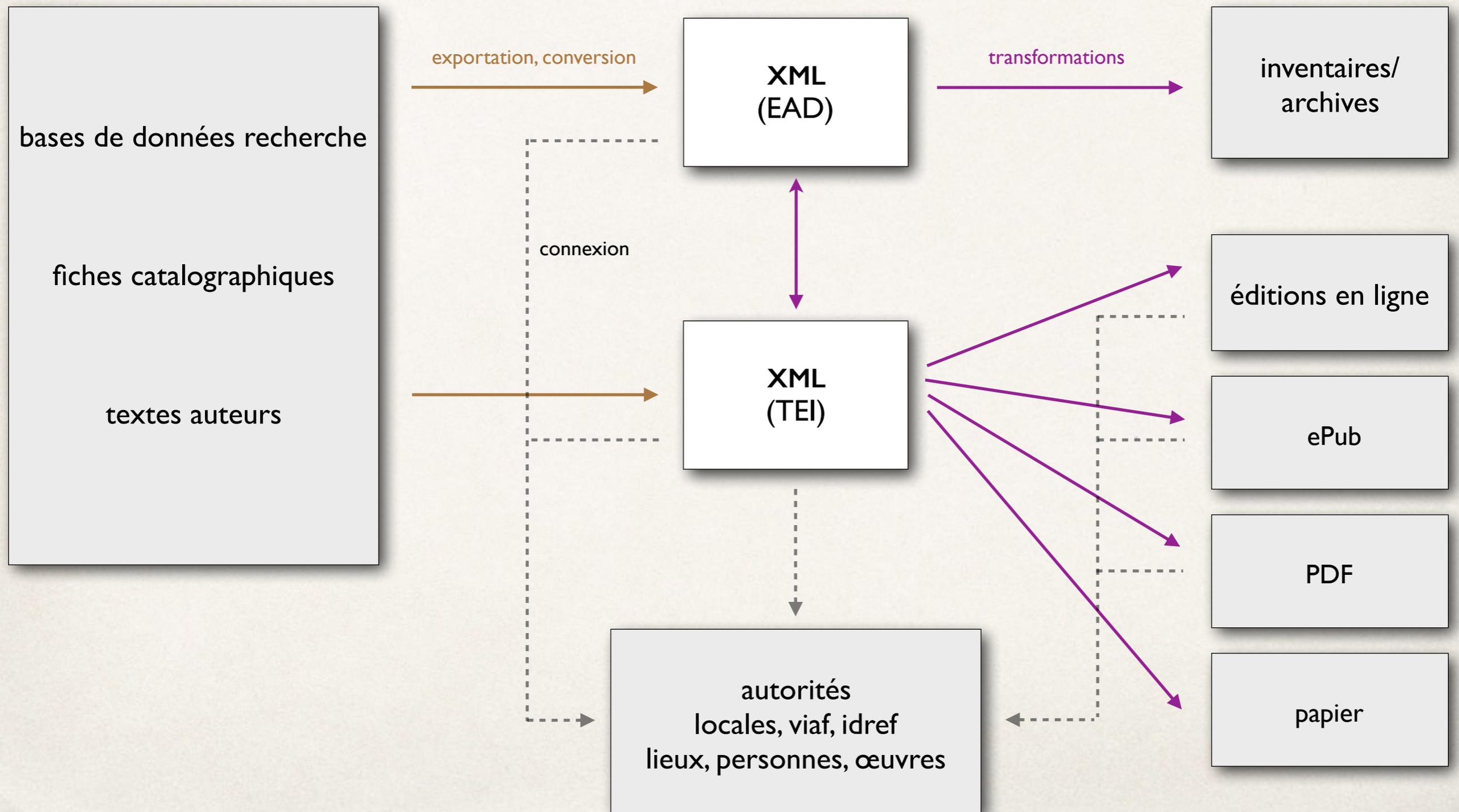
Organisation



Outils

- ❖ Word, feuilles de style
- ❖ Open Office
(norme Open Document)
- ❖ Éditeur XML (utilisation,
développement d'interfaces)
- ❖ InDesign (fonctions
d'intégration de flux XML,
langage de scripts)
- ❖ Postscript, PDF
- ❖ XSLT
- ❖ CSS
- ❖ Notions d'XHTML

Normes et standards



Humanités numériques

- ❖ Apporter la possibilité de l'expérimentation aux humanités (laboratoire de textes)
 - ❖ des solutions de balisage et d'annotation
 - ❖ des exploitations scientifiques et éditoriales
 - ❖ la possibilité de reproduire les expériences (en fournissant les flux et les outils)
- ❖ Permettre au chercheur d'étudier ses sources indépendamment des formes de diffusion envisagées pour l'édition
- ❖ Questionner le rapport du chercheur à son objet (les sources anciennes)

Perspectives

Des flux de textes aux bases de connaissances (1)

- * Exploitation de l'annotation des chercheurs constituée pendant les travaux d'édition pour
 - * proposer des solutions d'interrogation
 - * fournir des outils de visualisation des textes étudiés
- ➔ en tirant profit du gain d'expressivité des ontologies et des graphes pour l'interrogation, la visualisation ou l'accès au texte

Perspectives

Des flux de textes aux bases de connaissances (2)

<bibl><title /><author /><publisher /></bibl>

publisher $\xrightarrow{\text{est \acute{e}diteur de}}$ bibl

publisher $\xrightarrow{\text{est \acute{e}diteur de}}$ bibl $\xrightarrow{\text{a pour titre}}$ title

Perspectives

Logicisme

- ❖ Logicisme (J.-C. Gardin)
 - ❖ lecture rapide des raisonnements scientifiques
 - ❖ publication organisée des données mobilisées
 - ❖ partage des connaissances
 - ❖ élaboration progressive de bases de connaissances thématiques
- ❖ Deux projets :
 - ❖ arkeotek journal (<http://thearkeotekjournal.org>) (Valentine Roux, MAE)
 - ❖ Rapport de fouille de Rigny (Elisabeth Zadora-Rio, CITERES, Julia Roger, Jérôme Chauveau)

Perspectives – méthode

Construction des relations (1)

The screenshot displays a TEI editor interface. On the left, the XML code for a section titled 'section1_eglises.xml' is visible, showing a structure with nested elements like <div>, <head>, <div subtype="proposition">, <figure>, and <p>. The central area shows a text block with the heading 'PO_1. Sous l'église actuelle, deux bâtiments orientés selon le même axe qu'elle ont été découverts en fouille.' Below the text is an aerial photograph of a church and a floor plan diagram with a red line indicating an excavation trench. The right sidebar contains an 'Include' panel and an 'Attributs' panel with a table of attributes and values.

```
</titlePage>
</front>
<body>
  <div type="chapitre" xml:id="main_Div">
    <div type="section1" xml:id="sec1_1">
      <head style="T_2" subtype="level2">
        Identification des édifices
      </head>
      <div subtype="proposition"
        xml:id="section1PO_1">
        <head style="T_3" subtype="level3">
          PO_1. Sous l'église actuelle, deux
          bâtiments orientés selon le même
          axe qu'elle ont été découverts en
          fouille.
        </head>
        <figure>
          <graphic url="icono/1_Eglises/hr/"
          </figure>
        <p style="txt_Normal">
          Commentaire :
        </p>
        <p style="txt_Normal">
          <hi rend="italic"
            style="typo_Italique">
            La fouille à l'intérieur de l'église
            actuelle a été limitée à une
            tranchée de 3m x 8m. Elle a été
            poursuivie jusqu'à une
            profondeur moyenne de 1,70m
            sous le dallage actuel.
          </hi>
        </p>
        <p style="txt_Normal">
          <hi rend="italic"
            style="typo_Italique">
            Les trois bâtiments successifs, du
            plus ancien au plus récent, ont
            été désignés par les lettres X, Y
            et Z.
          </hi>
        </p>
      </div>
    </div>
  </body>
</div>
```

PO_1. Sous l'église actuelle, deux bâtiments orientés selon le même axe qu'elle ont été découverts en fouille.

Commentaire :

La fouille à l'intérieur de l'église actuelle a été limitée à une tranchée de 3m x 8m. Elle a été poursuivie jusqu'à une profondeur moyenne de 1,70m sous le dallage actuel.

Les trois bâtiments successifs, du plus ancien au plus récent, ont été désignés par les lettres X, Y et Z.

PO_2. Les plans des murs retrouvés en fouille peuvent être identifiés comme des chevets d'église successifs.

Attribut	Valeur
a:cstyle	-
a:pstyle	-
copyOf	-
corresp	-
decls	-
exclude	-
n	-
next	-
prev	-
rend	-
rendition	-
sameAs	-
select	-
style	txt_Normal
synch	-
xml:base	-
xml:id	-
xml:lang	-
xml:space	-
xsi:noNamespac...	-
xsi:schemaLocat...	-
xsi:type	-

Perspectives – méthode

Construction des relations (2)

```
☐ <div type="relation" xml:id="rel_1">
  ☐ <head style="T_4" subtype="level4">
    ANTECEDENTS:
  </head>
  ☐ <ptr subtype="is_based_on" target="#section1P0_1" type="logiciste"/>
  ☐ <ptr subtype="is_based_on" target="#section1P0_2" type="logiciste"/>
</div>
```

Perspectives – méthode

Création de l'ontologie

- ❖ Constitution d'une ontologie de base avec un minimum d'*a priori*

```
<owl:Class rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
<owl:ObjectProperty rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#is_based_on">
  <rdfs:domain rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <rdfs:range rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <owl:inverseOf rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#is_base_for"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#is_base_for">
  <rdfs:domain rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <rdfs:range rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <owl:inverseOf rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#is_based_on"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Titre">
  <rdfs:domain rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Niveau">
  <rdfs:domain rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Position">
  <rdfs:domain rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
</owl:DatatypeProperty>
```

Perspectives – méthode

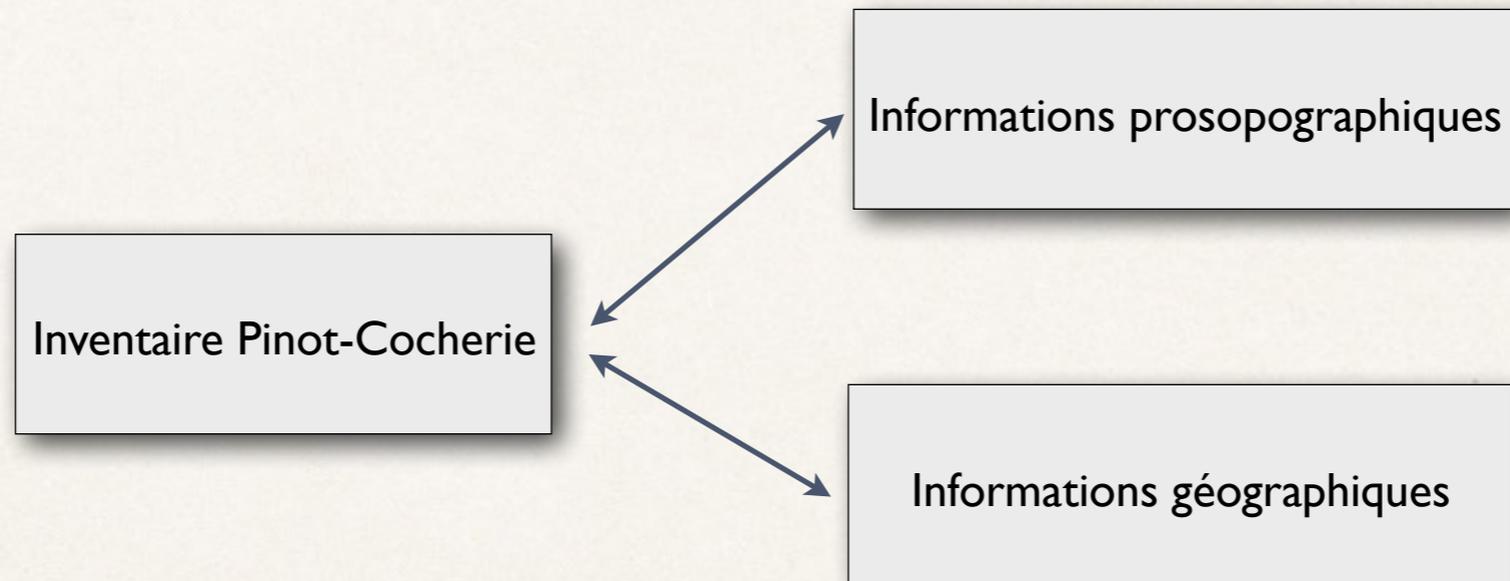
Peuplement de l'ontologie

- ❖ Peuplement de l'ontologie avec les individus identifiés par les chercheurs et des relations entretenues par ces individus

```
<owl:NamedIndividual rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_1">
  <rdf:type rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <pddn:Titre>P0_1. Sous l'église actuelle, deux bâtiments orientés selon le même axe qu'elle ont été découverts en fouille.</pddn:Titre>
  <pddn:Niveau>P0</pddn:Niveau>
  <pddn:position>1</pddn:position>
  <pddn:is_base_for rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P1_1"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_2">
  <rdf:type rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <pddn:Titre>P0_2. Les plans des murs retrouvés en fouille peuvent être identifiés comme des chevets d'église successifs.</pddn:Titre>
  <pddn:Niveau>P0</pddn:Niveau>
  <pddn:position>2</pddn:position>
  <pddn:is_base_for rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P1_1"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P1_1">
  <rdf:type rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <pddn:Titre>P1_1. Trois églises se succèdent sur le site de Rigny.</pddn:Titre>
  <pddn:Niveau>P1</pddn:Niveau>
  <pddn:position>3</pddn:position>
  <pddn:is_based_on rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_1"/>
  <pddn:is_based_on rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_2"/>
  <pddn:is_base_for rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P2_1"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_3">
  <rdf:type rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <pddn:Titre>P0_3. L'orientation des trois églises présente un décalage important par rapport à l'axe est/ouest habituel.</pddn:Titre>
  <pddn:Niveau>P0</pddn:Niveau>
  <pddn:position>4</pddn:position>
  <pddn:is_base_for rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P1_2"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P0_4">
  <rdf:type rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#Proposition"/>
  <pddn:Titre>P0_4. L'orientation des églises ne s'explique pas par une adaptation à la topographie du site.</pddn:Titre>
  <pddn:Niveau>P0</pddn:Niveau>
  <pddn:position>5</pddn:position>
  <pddn:is_base_for rdf:resource="file:/Users/pyb/Documents/Pole/XMLetRDF/logicisme.xml#section1P1_2"/>
</owl:NamedIndividual>
```

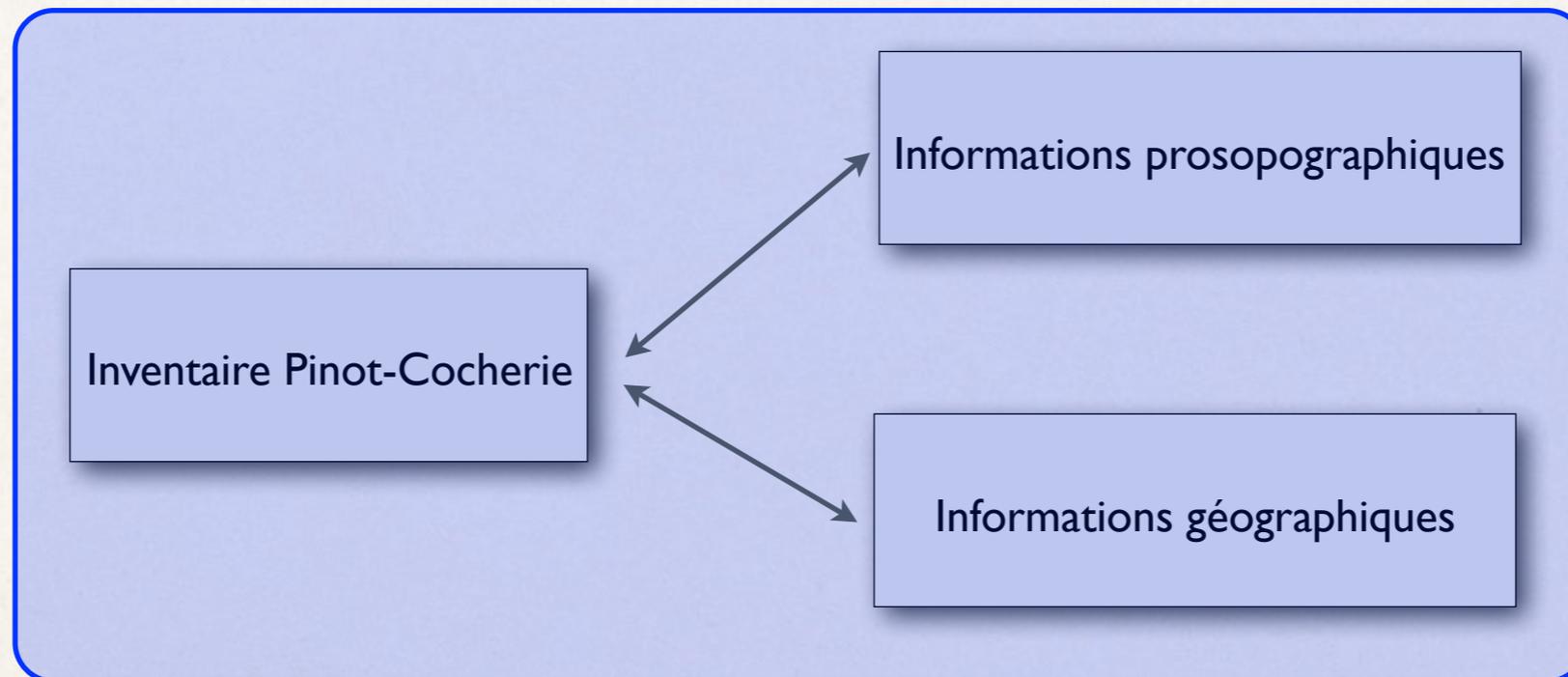
Perspectives – méthode

Gestion de flux



Perspectives – méthode

Gestion de flux > Bases de connaissances



Perspectives – outils

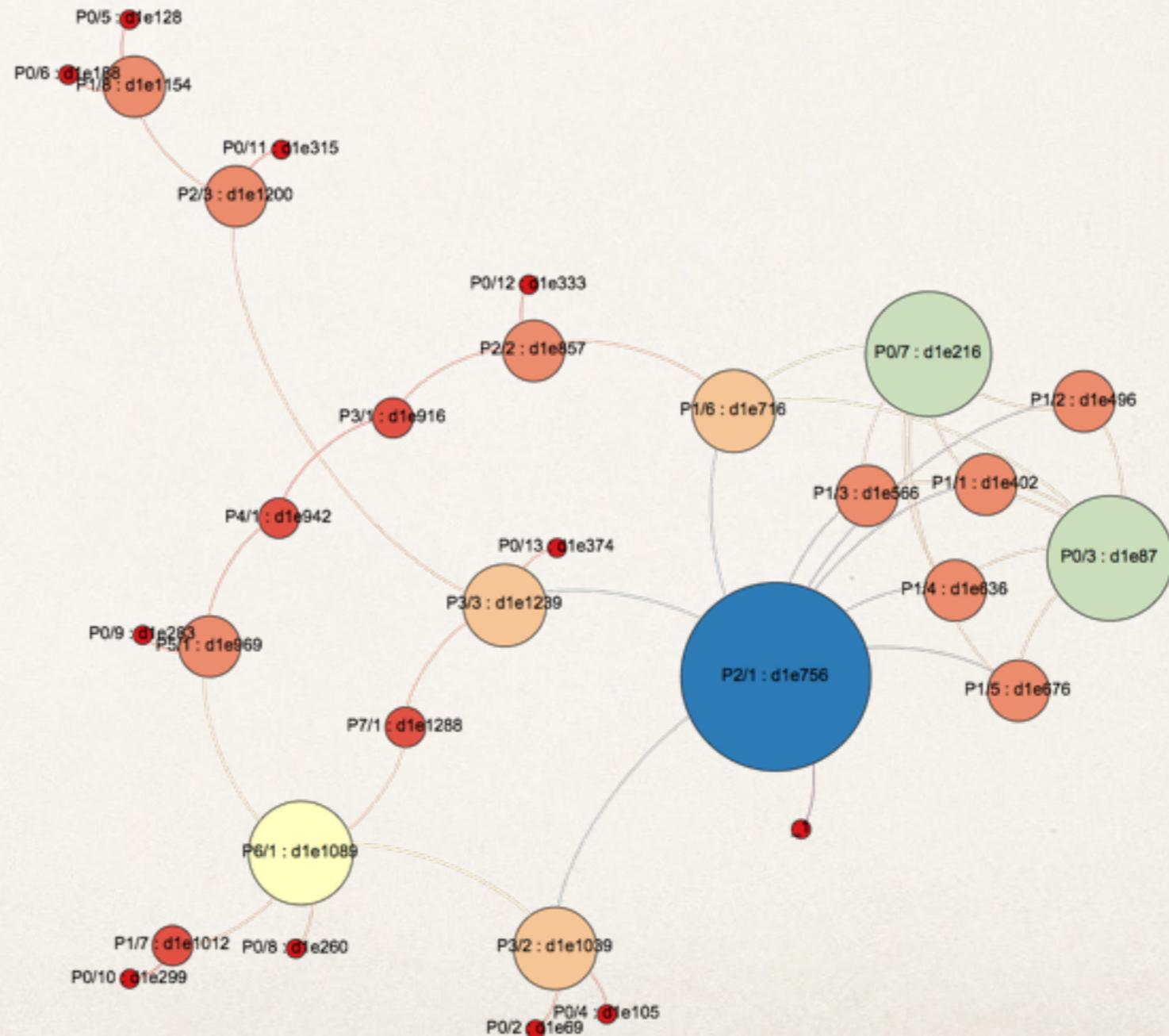
Production / conversion – XSLT

- * XML TEI pour les textes en entrée
 - * RDF/XML pour les graphes
- ➔ utilisation importante de XSLT pour les transformations

Outils

Visualisation et exploration interactive – Gephi (1)

Revue : <http://www.thearkeotekjournal.org>



Outils

Visualisation et exploration interactive – Gephi (2)

