



BVH

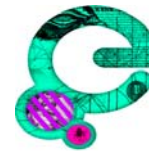
Le patrimoine numérique des
Bibliothèques Virtuelles Humanistes :
culture, formation et recherche



Marie-Luce Demonet, AG BVH

15 décembre 2010

<http://www.bvh.univ-tours.fr>



Acquisitions numériques

Depuis 2003

- 1) Actuellement environ 800 (**563** en ligne en décembre 2010 — 400 en déc. 2009) fac-similés d'ouvrages de la Renaissance (371 Tours, 6 Châteaudun, 29 Vendôme, **34 Châteauroux**, **86 Blois**, 29 Poitiers, 1 Budapest, Romorantin, Paris, Saumur ; 3/272 Orléans, en traitement). **Bourges en 2011?**
 - Avec leurs sommaires
 - Avec les illustrations extraites et indexées -Bases de lettrines, portraits, matériel typographique: plus de 30 000 éléments extraits
- 2) Epistemon, base textuelle : 19 pdf, 16 html, **21 xml/tei (10 en 2009)**
- 3) Manuscrits
 - De minute en minute (10 000), **Archives XVe des AD 37 (4500)**
 - Archives, livres de comptes, manuscrits littéraires

Les BVH-TIPo: Textes Imprimés Patrimoniaux et humanités numériques

- Les mots-clés des « digital humanities »:
 - Accessibilité
 - Protection (propriété intellectuelle)
 - Données réutilisables, partageables, transmissibles
 - Archivées de façon durable
- A l'intersection de plusieurs communautés

Accessibilité et propriété

- Accès libre et gratuit à toutes les données (y compris la structure TEI – en cours)
- Veille juridique et propriété intellectuelle pour:
 - Les fac-similés numériques, les images extraites
 - Les textes transcrits et publiés en html et xml
 - Les textes encodés en TEI
 - Les métadonnées, les bases de données
 - Le progiciel co-développé avec une société (Edit-tei)
 - Les applications en licence libre (AGORA, avec le LI de Tours)

= *libres de droits, avec licence CC (sauf Edit-tei)*



Les identifiants

1. Le nommage du fac-similé et de ses pages avec l'identifiant de l'institution:
 1. *B452346101_1533* pour les *Tragiques* (BM Orléans)
 2. *XUVA_Gordon1563_R65b_2* pour les *Discours* de Ronsard (Université de Virginie, Gordon Collection)
 3. *M0275_01* pour les *Essais* (1580) au Musée de Sologne de Romorantin
2. L'identifiant « universel »: DOI, ark, handle...
3. L'identifiant de la transcription: DOI, handle, ISBN (en projet)

Archivage local/ délocalisé

Données

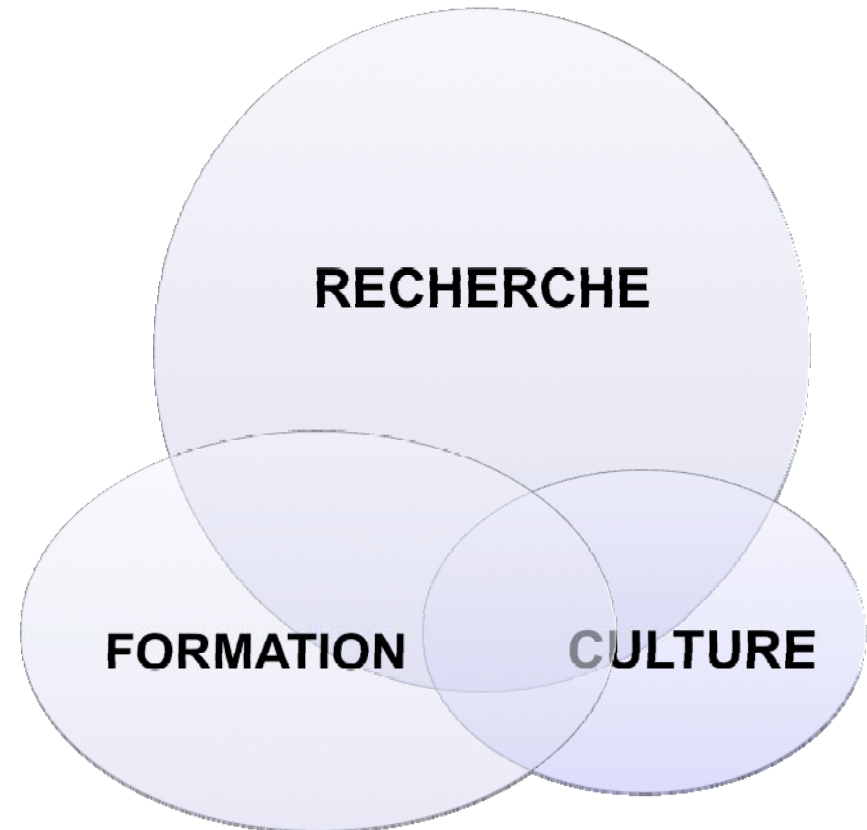
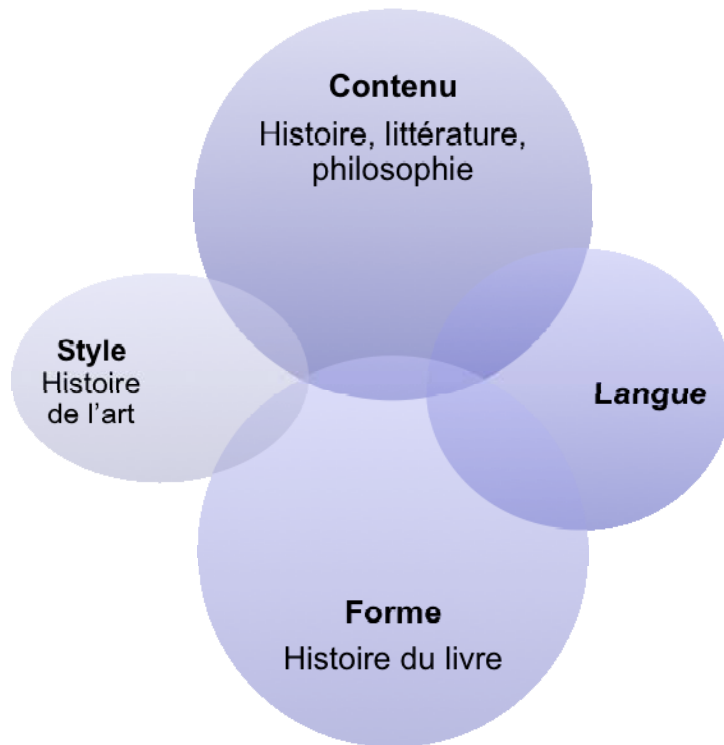
SVN
Externalisé
Pdf, textes

Copie 1
JPG
bibliothèque

Copie 2
JPG
CESR

Copie 3 JPG
Cines
(conservation
et mise à
jour)
IN2P3 (pour
la
(re)diffusion)

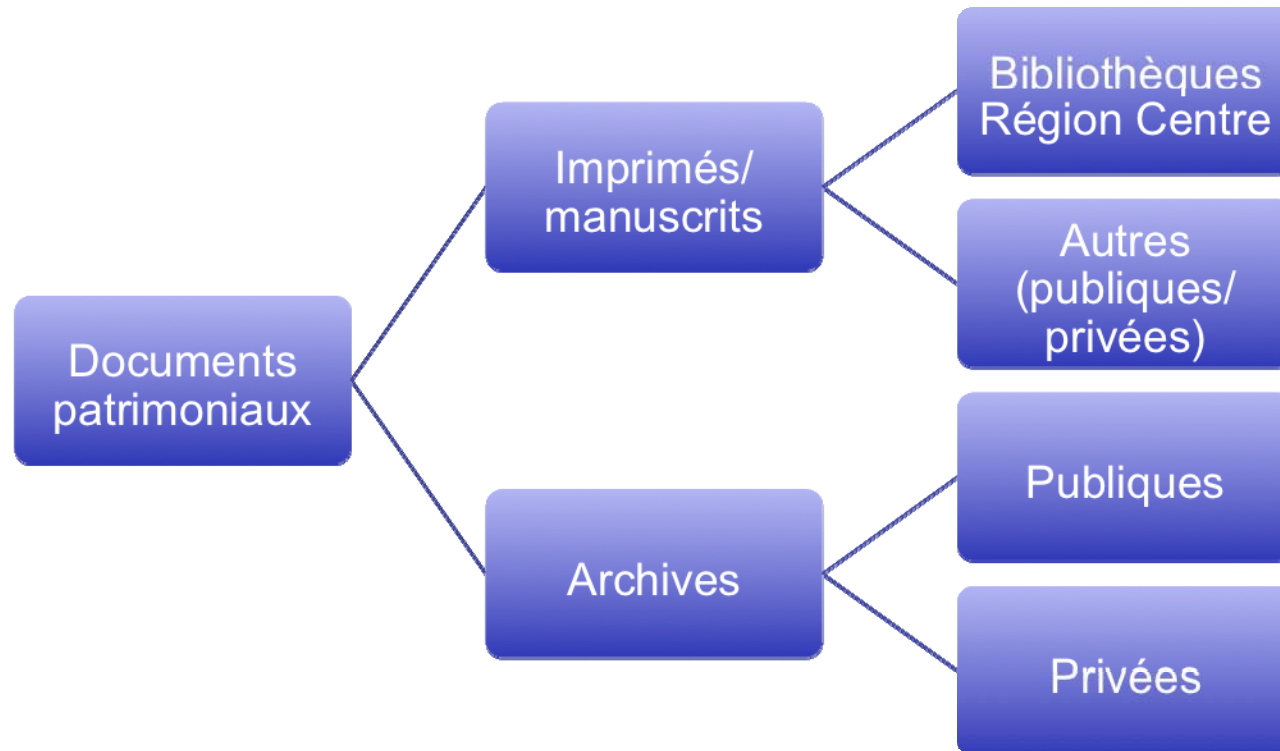
Communautés de chercheurs et d'usagers (sciences humaines)

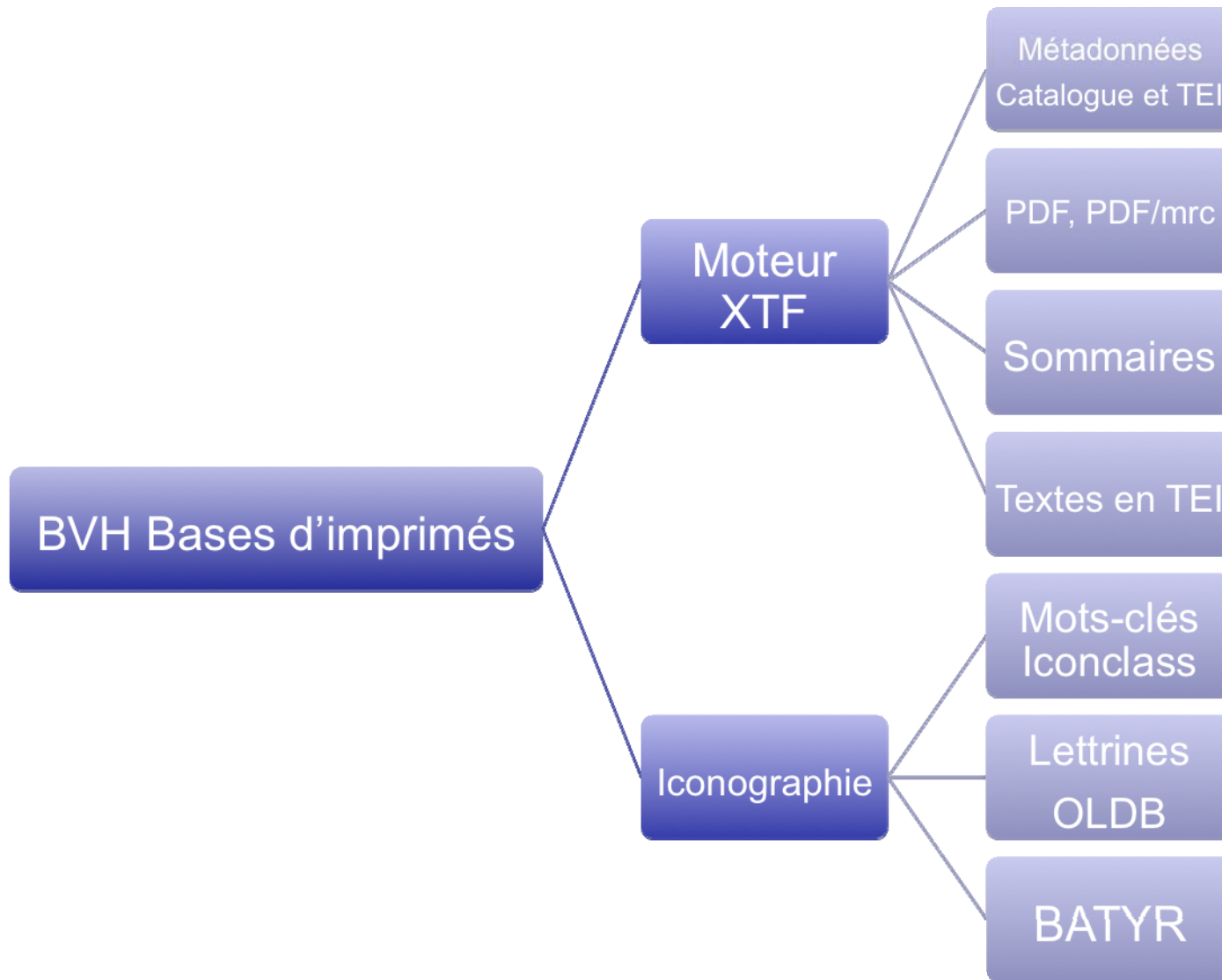


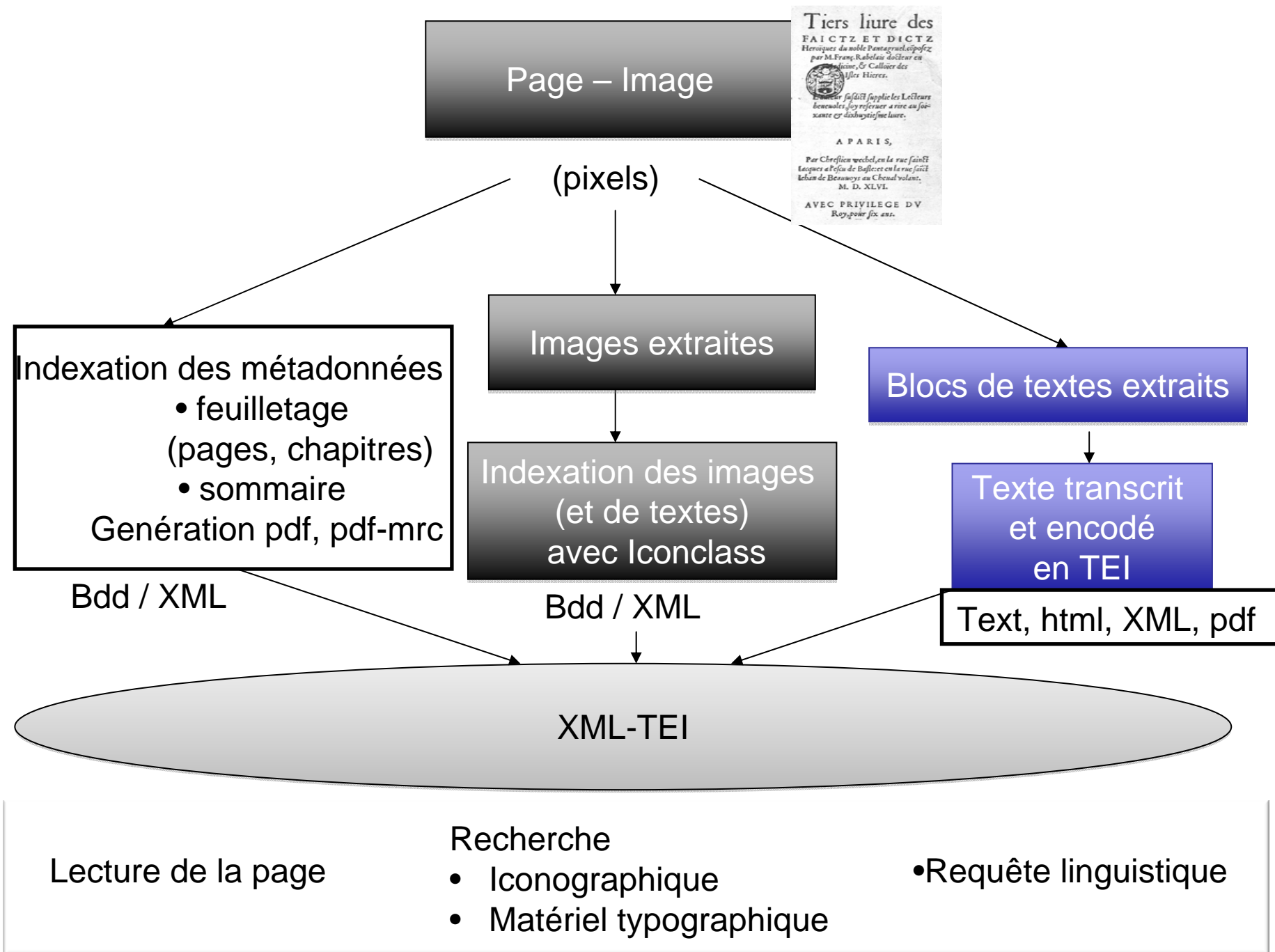
Contenus

- Classiques de la Renaissance et ouvrages majeurs (Rabelais, Montaigne, Ronsard...)
- Sources religieuses
- Sources de la science
- **Jurisprudence**, textes politiques
- Ouvrages d'intérêt régional et programmes de recherche en cours

Documents patrimoniaux

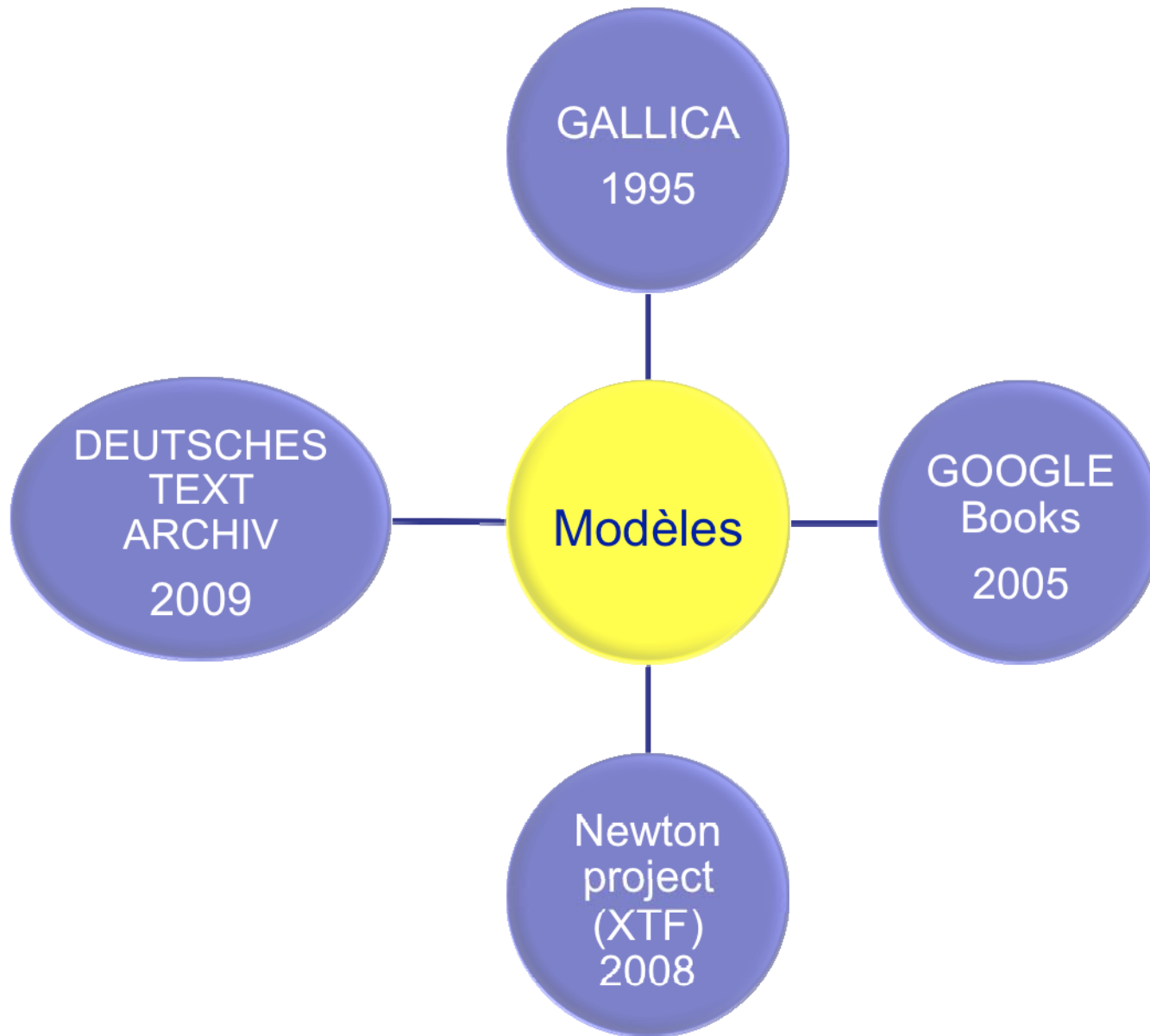






Évolution des besoins

- Lecture en ligne **et** téléchargements multiformats (pdf, txt, mp3, **epub**,...)
- Accès par mots-clés dans l'image, dans le texte et **web sémantique (langage RDF)**
- Annotation collaborative, folksonomies
- Accès **plein-texte** (par OCR brut/ corrigé)
- Choix des niveaux de transcription (diplomatique/ patrimonial/ **modernisé**)



Deustches Text Archiv : un modèle pour l'affichage texte / image ?

The screenshot shows the DTA website interface. On the left, a sidebar contains metadata for the book 'Die Kultur der Renaissance in Italien' by Jacob Burckhardt (PND), including the year 1860 and the publisher Schweighauser. A callout box points to the 'Download: HTML | XML' links with the text: 'Liens vers les métadonnées et le téléchargement des fichiers html et xml'. The central part of the page displays the book cover, which features the title 'Die Cultur der Renaissance in Italien.' and the author 'Jacob Burckhardt.' Below the cover, there is a small image of the book's title page. A callout box points to the XML code on the right, which is a snippet of the document's structure, including elements like <docTitle>, <docAuthor>, and <docImprint>. The text in the callout box reads: 'Affichage de l'image et/ou du texte en html ou en xml'.

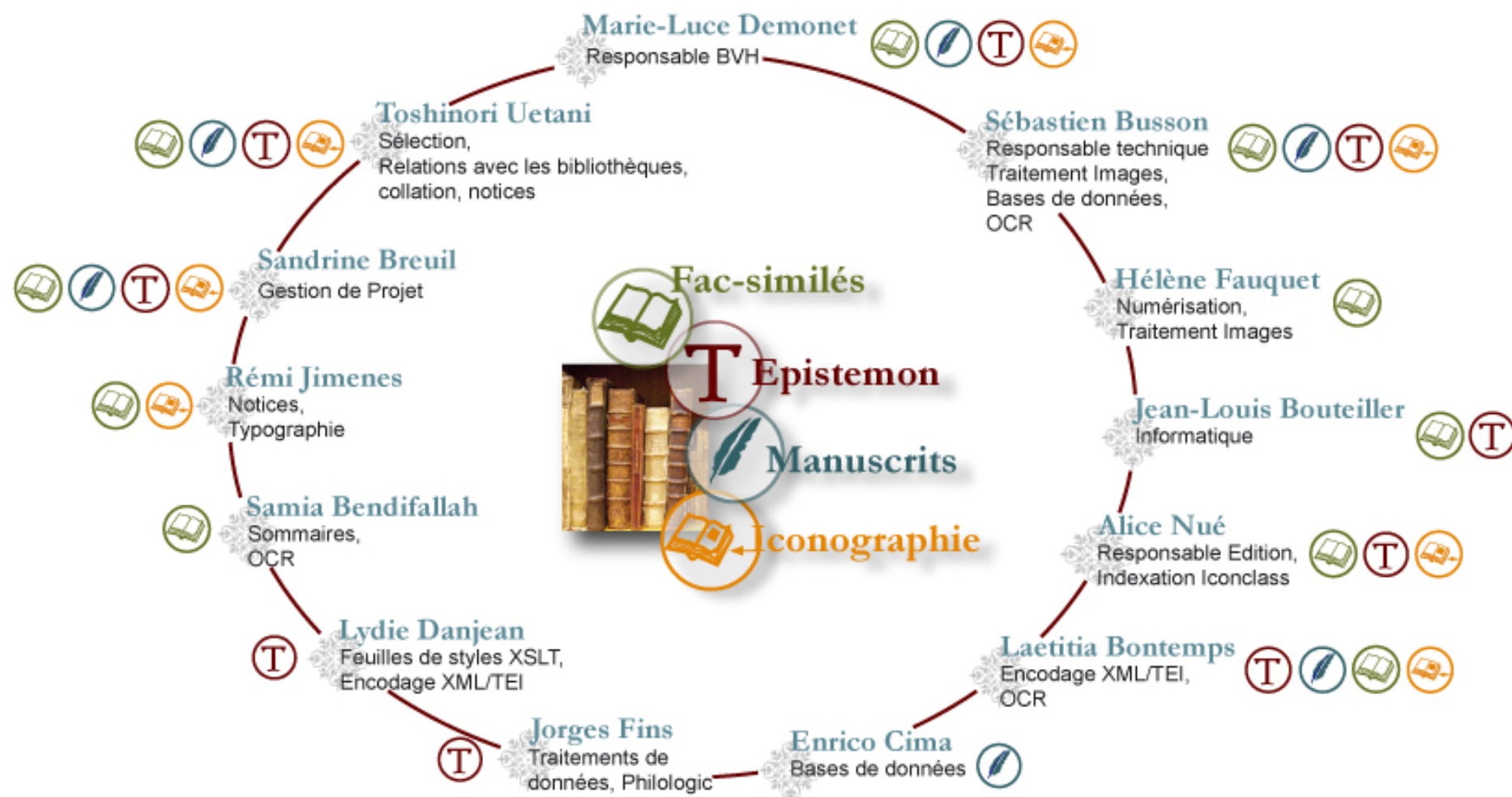
<http://www.deutschestextarchiv.de>

SCHÉMA DE PRODUCTION

BVH Répartition des ressources


Equipe des Bibliothèques Virtuelles Humanistes

<http://www.bvh.univ-tours.fr>



Métadonnées bibliographiques:

Les Tragiques (1616) d'Agrippa d'Aubigné sur Gallica

**Les tragiques donnez au public par le larcin de Promethee -Au Dezert, par
L.B.D.D. M.DC.XVI-1616** 

Notice complète

Titre : Les tragiques donnez au public par le larcin de Promethee

Auteur : Aubigné, Théodore Agrippa d' (1552-1630) [Ne voir que les résultats de cet auteur](#)

Éditeur : Au Dezert, par L.B.D.D. M.DC.XVI

Date d'édition : 1616

Type : monographie imprimée

Langue : Français

Format : [30]-391-[4] p. ; in-4

Format : application/pdf

Droits : domaine public

Identifiant : ark:/12148/bpt6k70069c

Source : Bibliothèque nationale de France, Rés. Ye-625

Relation : <http://catalogue.bnf.fr/ark:/12148/cb30035563g>

Provenance : Bibliothèque nationale de France

Thématique : Littérature francophone

Date de mise en ligne : 15/10/2007

[Fermer la notice complète](#)

Les métadonnées enrichies par les BVH (1)



Auteur : [Aubigné, Théodore Agrippa d'](#)

Titre : Les Tragiques

Imprimeur : Moussat, Jean

Libraire : L.B. D. D

Date : 1616

Format : 4°

Collation : [7, 1 bl., 7] f., 391 p. (quelques erreurs de numérotation), [2, 1 bl.] f.

Titre long : Les Tragiques. Donnez au public par le larcin de Promethee.

Adresse typographique : Au Dezert, par L. B. D. D. [Le Bouc Du Desert]. (Maillé : chez Jean Moussat) 1616

Langue : Français

Notes : La première édition anonyme des *Tragiques*. Elle est imprimée à Maillé, propriété acquise par Aubigné, sur les presses de Jean Moussat, qui ont aussi imprimé l'Histoire universelle.

Etat physique :

Université François Rabelais, Tours - UMR 6576 du CNRS – MSH de Tours
<http://www.cesr.univ-tours.fr/>

Les métadonnées enrichies par les BVH (2)

Reliure : Reliure parchemin.

Provenance :

Annotations :

Matière : [Littérature](#)

Bibliographie : Barbier (J. P.), *Ma bibliothèque poétique*, IV-1, Genève, 1998, n° 1.

Brunet (J.-Ch.), *Manuel du libraire et de l'amateur de livres*, 5e éd., I, col. 544 ; *Supplément*, I, col. 75.

Schrenck (G.), *Bibliographie des écrivains français : Agrippa d'Augigné*, Paris, 2001, n° 31.

Edition moderne :

Edition critique par J.-R. Fanlo, Paris, 2003,

Autres numérisations :

[Gallica](#) (l'éd. de 1616 ; l'éd. Ludovic Lalanne, Paris, Plon, 1857, rpt. Kraus, 1979).

Localisation : Bibliothèque Municipale Orléans

Cote : D1533 res

Numérisation : IRHT - Light Phase sur Contax 645 - 2003

Mise en ligne : 01/07/2003

Url : <http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=58>

© Bibliothèque Municipale Orléans

Passage à XML/ TEI

- 2007 : “TEI membership” et début de l’encodage des textes en TEI P5, premier stage TEI
- 2009 : début de la bascule vers XTF (XML/ TEI) pour les fac-similés et les pdf
- 2010: préparation de la bascule des autres bases de données (notices, feuilletage, sommaires, iconographie, minutes, manuscrits)

Balisage des métadonnées: notices des fac-similés et en-têtes Epistemon en TEI

Métadonnées de catalogage

1. Métadonnées de catalogues pour le moteur XTF (gestion de la majeure partie du site BVH)
2. Compatibilité CCFR (comme le catalogue des incunables) et OAI-PMH pour moissonnage par Gallica (fac-similés seulement)
3. Compatibilité [Europeana](#) (TEI/MARC/Dublincore), tous documents (y compris les transcriptions)
4. Compatibilité [Isidore](#) (moteur de recherche du TGE Adonis, CNRS), fac-similés seulement

Balisage des métadonnées en TEI: typologie

1. Typologie des recueils:
 1. Editorial: ex. de *l'Olive* de Du Bellay (1549)
 2. Commercial: Les *Discours des misères* et la *Continuation des Discours* de Ronsard
 3. Recueil factice: entités séparées
2. Corpus: ex. de l'ensemble des éditions de référence de Rabelais, du projet « Montaigne »
3. Corpus-collection: ex. des textes politiques des États Généraux de Blois

Métadonnées: la facette « genre » (en projet)

Jean Bodin, *La République*, 1576:

- Domaine: histoire
- Genre:
 - » Non-fiction
 - » Théorie politique
 - » Traité
 - » Prose
 - » Texte original (et non traduction)

Rétroconversion du *CRI*: MARC/ TEI (F. Palluault)

- ```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="info:lc/xmlns/marcxchange-v1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:tei="http://www.tei-c.org/ns/1.0"
xsi:schemaLocation="info:lc/xmlns/marcxchange-v1 marcxchange-tei-
beta.xsd info:http://www.tei-c.org/ns/1.0 tei_ms.xsd">
<record format="UNIMARC" type="bibliographic">

<!-- Label : zone obligatoire
Positions 0-4 : longueur de la notice. Normalement, valeur calculée par
l'ordinateur au moment de l'échange de données
Position 5 : statut de la notice. n = nouvelle notice
Positions 6-9 : codes d'application
Pos. 6 : type de document : a = texte imprimé
Pos. 7 : niveau bibliographique : m = monographie ; c = recueil factice
Pos. 8 : niveau hiérarchique : # = relation hiérarchique non définie
Pos. 9 : non définie : #
```



## Acquisition numérique du texte: les OCR

---

- Résultats obtenus avec FineReader (ABBYY), Omnipage avec/ sans dictionnaires spécifiques, performances du format d'Adobe-mrc; tests comparatifs
- Tests du logiciel BIT-Alpha (société Tomasi) à Wolfenbüttel et à Tours. Acquisition du logiciel par le CESR (04-12-2008)
- Développement de RETRO (U. de Tours, JY Ramel) après classement des clusters par AGORA

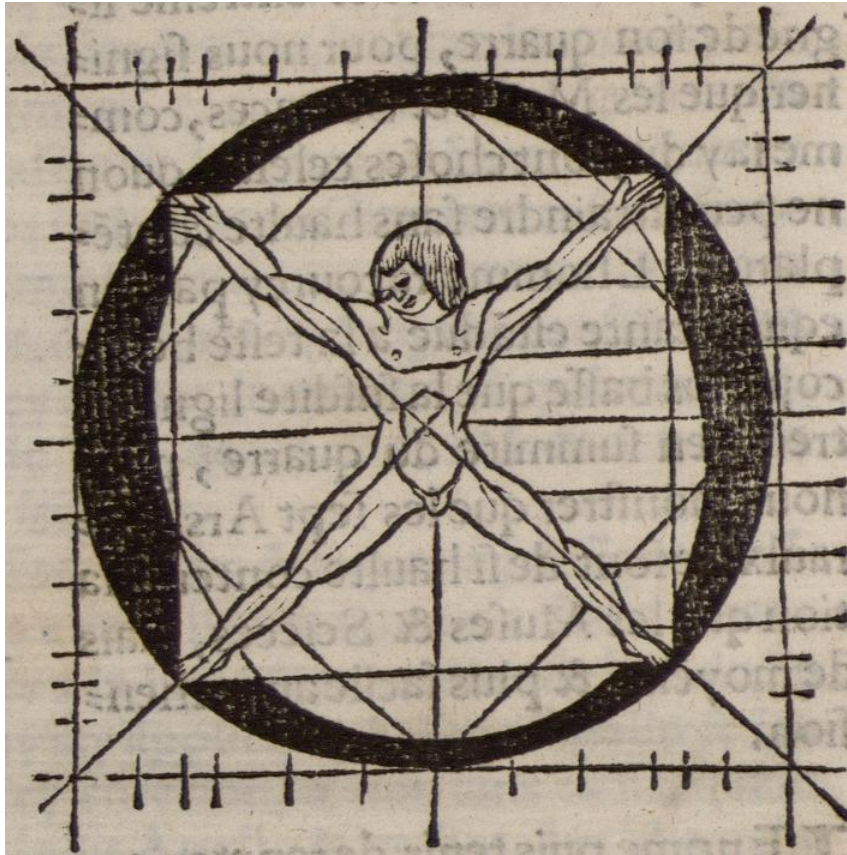
|                                 | Munich/<br>Google copy | Tours copy,<br>JPG/PDF MRC | Transcription |                                             |
|---------------------------------|------------------------|----------------------------|---------------|---------------------------------------------|
| Helicon                         | 1                      | 1                          | 1             | probably in<br>modern<br>dictionary         |
| enthusiasme                     | 0                      | 1                          | 1             | not in the<br>modern French<br>dictionaries |
| arbitr<br>[character<br>string] | 0                      | 5                          | 7             | recognition<br>problem                      |
| pantagruelion                   | 4                      | 7                          | 14            | pb de<br>reconnaissance                     |
| bouteille                       | 8                      | 11                         | 11            | présent ds dict.<br>du fr. mod.             |
| melancholie                     | 0                      | 1                          | 3             | absent des dict.<br>du fr. mod.             |
| beste                           | 0                      | 0                          | 3             | s/f et ligatures                            |
| bestes                          | 0                      | 0                          | 8             |                                             |
| befte                           | 2                      | 1                          | 0             |                                             |
| <b>beftes</b>                   | <b>7</b>               | <b>0</b>                   | <b>0</b>      |                                             |
| bestez                          | 0                      | 0                          | 1             |                                             |

Tests sur  
le *Tiers*  
*Livre 1546*

De l'extraction de l'élément figuré à  
l'image indexée  
(avec le consortium ANR Navidomass)

---

# Indexation des illustrations du *Champfleury* (G. Tory, 1529, exemplaire BM de Blois) avec Iconclass



## ICONCLASS

- 31A12 figure, inscrite dans un cercle ou un carré, avec les jambes écartées et les bras étendus vers le haut ou le côté (+ variante: vu de face)
- 49D342(SQUARE) quadrilatère régulier : carré
- 49D36 Cercle (planimétrie géométrie)
- 49D52 mesurer
- 49L12(O) écriture romaine; écriture basée sur l'alphabet romain (avec LETTRES) 49L1 Lettres, alphabet, texte
- 49D52 Tracer des figures, mesurer, calculer
- 49D362 Quadrature du cercle

Autres descripteurs : Leonardo da Vinci/ The Vitruvian Man/ Jean Perreal/ Jean de Paris

# AGORA

---

- Extraction automatique des éléments figurés du livre
- Préindexation automatique des types d'éléments et pré-balisage
- SOUS-BASES D'IMAGES
  - Galerie des marques typographiques
  - Galerie de portraits
  - Base emblèmes (indexée par Iconclass)
  - Bases lettrines et ornements (indexation propre)

- Bases lettrines et ornements (indexation propre), par Mickael Coustaty (Université de La Rochelle)

## BATYR: base de typographie de la Renaissance (en projet)

- Ornements gravés
  - Ornements de fonte
  - Polices de caractères
  - (Commémoration Garamont 2011)
  - Indexeur de lettrines
- 
- Pré-indexation d'images d'ornements typographiques (AGORA)
  - Constitution automatique de spécimens de caractères (RETRO) et reconstitution des fontes
  - [http://www.bvh.univ-tours.fr/materiel\\_typo.asp](http://www.bvh.univ-tours.fr/materiel_typo.asp)

# Améliorer l'OCR sur les imprimés anciens :

quelques pistes de travail (R. Jimenes)

## Étape 1 : Constituer des répertoires de polices de caractères transcrits manuellement, intégrés au logiciel RETRO

en s'appuyant sur le recensement des caractères typographiques employés en France au XVI<sup>e</sup> siècle, publié par H.D.L. Vervliet en 2010.

Il serait probablement inutile de constituer un répertoire intégrant la totalité des polices recensées, car les variantes graphiques d'une fonte à l'autre sont parfois très faibles.

⇒ il convient donc d'opérer une sélection pertinente des polices à intégrer.



## Caractères romains

| Gras                                                                                                                                                                                                                                                     | Maigres                                                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>tatis accesserit Senecæ, si tuo, h<br/>rit comprobatus, qui ut his ter<br/>Sigismundo, totiq; regno exin<br/>mum præstas, ita disciplinis qu<br/>tissimum exhibes Mœcenatem<br/>Opt. Max. bene fortunare dis<br/>ista fatalis lues, quæ per orbem</p> | <p>mier homme Adam, &amp; autres les<br/>cendants, enseignez de Dieu, ont<br/>choses par leur nom (comme est c<br/>se) c'est à dire proprement &amp; re<br/>encore que par la mesme permil<br/>l'orgueil eust engendré cõfusion<br/>&amp; diuersité de langues, &amp; que la</p> |

## Caractères italiques

| Gras                                                                                                                                                                                                                                                                    | Maigres                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>esse consuevit. Propter quòd, testi<br/>næ Romanæ uino abstinebant: q<br/>si esset, ut par est, modus in reb<br/>non est quicquam aliud uino mod<br/>etiam Persæ existimabant fomite<br/>tamentum ingenij uirtutisq; , si<br/>uino flagraret, qui, teste Strabo.</p> | <p>iore numero? Nostrates quiden<br/>litteris cum Latina matre, ut r<br/>conueniunt. Ergo rei litteraria<br/>uniuersas apud linguas &amp; anti<br/>prona uocis facilitate primum<br/>est a. Deus rerũ uerborũque pi<br/>merorũ quoque unitas, his a</p> |

## Étape 2 : automatiser la sélection des répertoires de fontes interrogés pour la transcription

en s'aidant des caractéristiques propres aux polices de caractères anciennes :

- **Le corps de caractère** : *pour une résolution donnée (300 dpi), compter la hauteur en pixels pour des blocs de 20 lignes, afin de déterminer le corps de la police.*
- **La graisse du caractère** : *après binarisation et suppression du bruit, calculer la densité de pixels noirs sur le nombre total de pixels.*

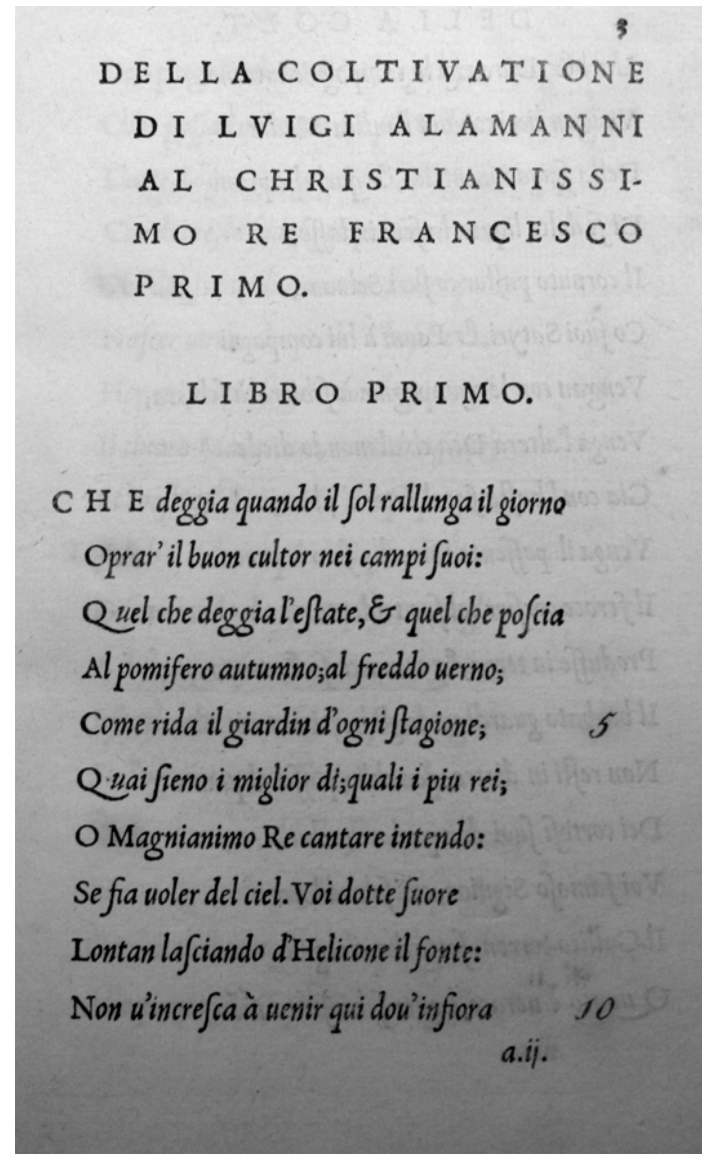
Pour améliorer la transcription du texte, Retro devra être capable de distinguer les différentes fontes employées (selon leur corps et leur graisse), et d'identifier le répertoire le plus approprié à la transcription de chacune de ces polices.

Dans les prochains mois :

- Sélection, dans le *Conspectus* de H.D.L. Vervliet, d'un corpus de fontes représentatif des grandes tendances graphiques au XVI<sup>e</sup> siècle, en fonction de leur corps, de leur graisse et de leur dessin
- Numérisation et transcription manuelle des spécimens sélectionnés, et complétés par des exemples supplémentaires tirés des BVH

## Écueils et problèmes posés

Pour déterminer automatiquement le corps d'une fonte, le logiciel devrait détecter les cas de polices interlignées



-La graisse apparente d'une fonte varie selon l'usure des types : le logiciel devra être capable d'en tenir compte ;

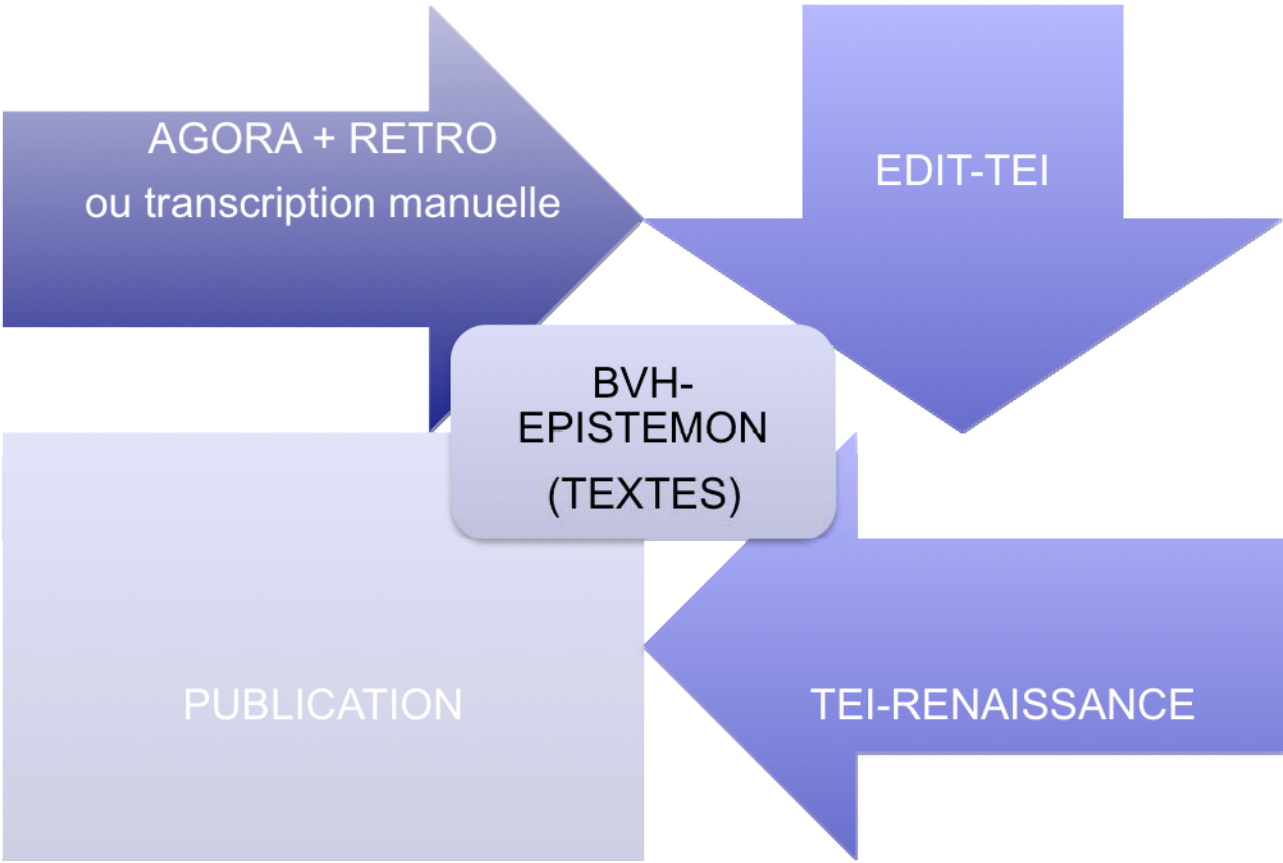
-Le seuil de densité (pixels noirs / pixels blancs) déterminant la graisse d'un caractère doit être modulé en fonction du corps de ce caractère : les lettres de petits corps sont souvent plus grasses que les lettres de gros corps.

-L'intégration à Unicode des caractères spéciaux du XVIe siècle n'est pas encore complète ;

-La sélection des fontes pour constituer les répertoires doit être effectuée selon des critères raisonnés (qu'il convient de déterminer dans une discussion entre historiens et informaticiens);

# De la transcription à l'encodage

---



## La TEI « Renaissance »

- Renaissance.dtd
- Renaissance.xsl
- Fichier type
  - header, qui contient
    - Les champs des notices
    - La localisation des facsimilés
    - Les auteurs
    - Les crédits
    - Les versions
- => affichage total ou partiel en tête du document

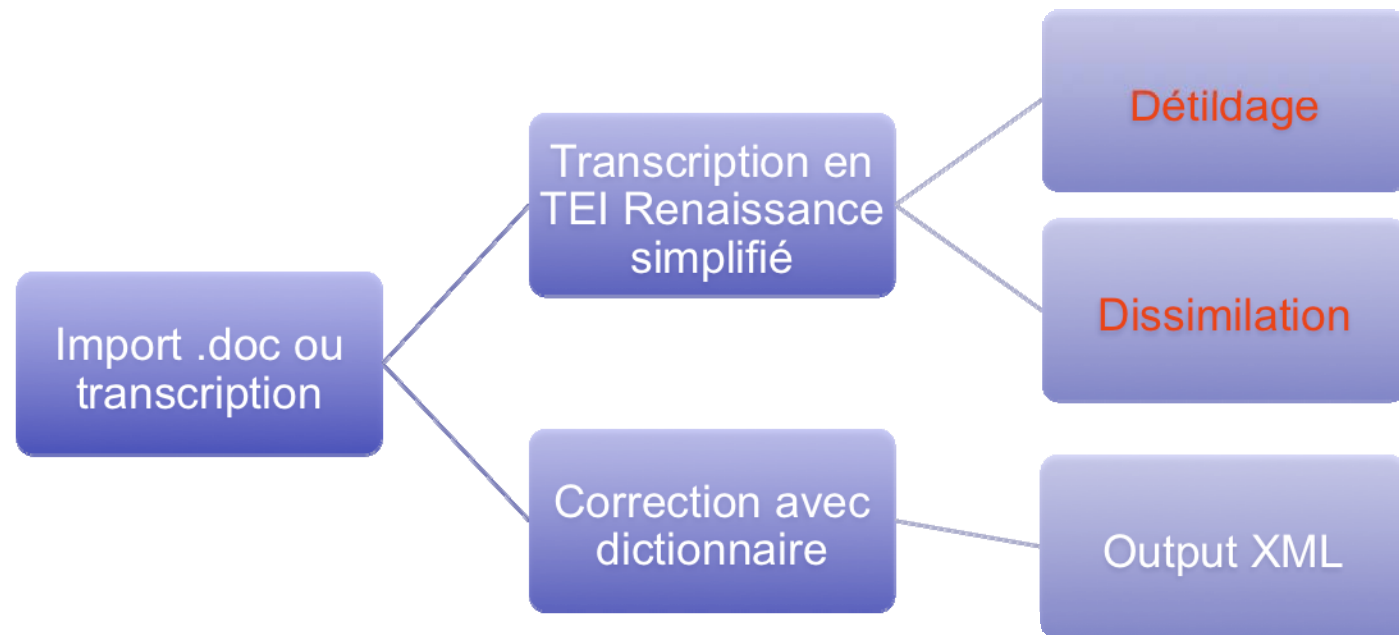


## Pré-balisage TEI avec l'outil EDIT-TEI

Pour les textes transcrits par opérateur, sans OCR  
(ou importés):

- Ajout de deux modules: détildage et dissimilation

# LE PROGICIEL EDIT-TEI (Digiscrib / BVH)



## Textes encodés

Bilan (décembre 2010) <http://www.bvh.univ-tours.fr/Epistemon/index.htm>

- 20 textes en TEI en ligne
- 25 textes encodés en TEI en cours de contrôle
- 15 en cours d'encodage ou de vérification
- 40 en cours de transcription

En cours: 10 000 minutes notariales déjà transcrites, à encoder, et un petit corpus de manuscrits

## Le corps du texte: encodage TEI « patrimonial »

- Constituer un Corpus en français de la Renaissance (principalement), avec traitement par Philologic (Chicago), TXM (Lyon), Analog (Poitiers)
- Objectif : 100 textes originaux en 2014 (sans restriction d'accès, graphies originales/régularisées)
- Encodage **patrimonial**: sélection de balises pour traitements automatiques

## Vers une édition générique régularisée

- Transcription diplomatique possible, avec les abréviations, ligatures et ij/ uv originaux
  - Janvier 2009: Expansion automatique des brévigraphes et dissimilation automatique avec l'outil Dissimilog (T. Vincent) par encodage avec <choice> et <reg> et <choice> pour le double affichage;
  - **2010: autre méthode avec Edit-Tei**

## Version originale/ régularisée

- Affichage html:
  - Avec la feuille de style 1: version originale
  - Avec la feuille de style 2:
    - Affichage de la version régularisée et pop-ups pour les corrections
- Double affichage géré par le moteur XTF: application en cours
- Vers une version modernisée (graphies)

## Développements en cours

- Affichage
  - Alignement fac-similé/ texte avec XTF, par J.Herrant (PFE du laboratoire d'informatique, Tours)
  - Installation du zoom sur image Ajax, génération d'images à la volée
  - Harmonisation des headers
- Encodage:
  - Reprendre tous les premiers textes encodés
  - Nettoyer les balises pour Philologic et les logiciels de données textuelles

## 2011-2012

- Veille sur les moteurs de recherche (TXM, DDC pour le projet DTA, Pinakes, Pleade (ENC)...)
- Projet lettrines (suite de l'ANR Navidomass)
- Recherches OCR (thèse BnF, Google award)
- Encodage
  - Alignement de versions différentes et de traduction
  - Traitement des signes de ponctuation et des caractères spéciaux (intégration des nouvelles balises TEI d'A. Lavrentiev)
- Versions oralisées
- Entités nommées et autorités Rameau



# Expansion

## *Contrat quadriennal 2012-2015 (étendu à 2016)*

- Intégration du master pro dans un ensemble professionnalisant “Patrimoine matériel et immatériel”
- Perméabilité master pro/ recherche
- Dépôt d'un projet de laboratoire d'excellence “Patrimoine Moyen Age et Renaissance”, avec accent sur le numérique (image/ texte/ son), Tours-Orléans-Poitiers-La Rochelle
- Intensification et internationalisation des stages de formation initiale et continue
- Souhaitée: une formation doctorale à double compétence (humanités – nouvelles technologies)

# Documentation

## Accès libre ou à la demande:

- Un *Manuel d'encodage TEI "Renaissance"*, en ligne depuis 2008, version 3 en préparation
- Manuel SVN (en préparation)
- Descripteurs d'images (Iconclass, ODLB - lettrines)
- Cahiers des charges (numérisation, interface web, moteur de recherche, normes de saisie)
- Modèles de métadonnées (description d'imprimés, d'incunables, de manuscrits (**éditions génétiques en TEI**), d'archives)

# Publications, participations aux colloques et ateliers en 2010

- Janvier 2010, présentation des BVH pour le *Bulletin de l'Association des Bibliothécaires de France*.
- Mars 2010, Tokyo, « Digital scholarship »
- Juin 2010, King's College, Digital Humanities conference (comme observateur)
- Octobre 2010, La Haye, workshop des data providers d'Europeana
- Octobre 2010, La Haye et Amsterdam, Europeana, Annual Meeting
- Novembre 2010, Paris, Numérisation et histoire, GIS érudition
- Novembre 2010, Strasbourg, ESF workshop sur les Digital Humanities
- Novembre 2010, Zadar (Croatie), congrès de la TEI
- Novembre 2010, Chicago, Digital Humanities and Computing Science
- Novembre 2010, Tours, Journée Incunables
- Décembre 2010, Lyon-Valpré, Université d'hiver Adonis

# Formations

---

- Juin 2010, TEI France (Lyon, ISH, MUTECH)
- Octobre 2010, ANGD (Aussois), gestion de projet
- Novembre 2010, Stage TEI formation continue (SUFCO), Tours, CESR
- Modules de numérisation du master pro « Patrimoine écrit et édition numérique »: traitement de l'image, initiation aux OCR
- Stage TEI du master pro, Tours, CESR (janvier 2011, augmentation de sa durée)
- 21 janvier 2011: présentation du module TEI pour les éditions génétiques
- Octobre 2011, ANGD (Aussois), gestion de projet
- SUFCO TEI niveau 1 et niveau 2 (mai et automne 2011?)
- Formation TEI à Aix-en-Provence, avec MEET (février 2011)?

# Les archives

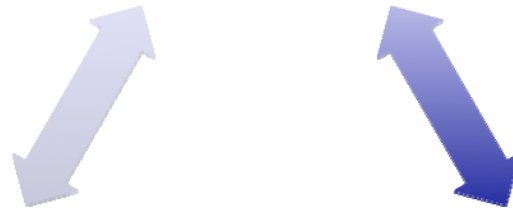
- Intégration des archives dans les BVH pour des projets spécifiques:
  - Disponibilité des 10 000 minutes notariales transcrites par P. Aquilon, B. Chevalier
  - Numérisation de toutes les minutes du XVe siècle (4500), Convention avec les AD 37, publication en cours
  - **Fonds d'archives des artistes tourangeaux (exposition 2012), en projet**
  - Archives d'auteurs du XVIe siècle (Rabelais, Ronsard, Montaigne, Béroalde de Verville...), en projet

# BVH-TIPO et TELMA

- Publication des actes notariés AD 37
  - BVH-TIPO
    - Transcriptions ou analyses
    - Modèles d'encodage TEI, Traitement par XTF
  - TELMA
    - Expertise sur les actes médiévaux
    - Formulaire de saisie pour le basculement BDD vers XML/ TEI
- Recrutement CDD en collaboration



Minutes

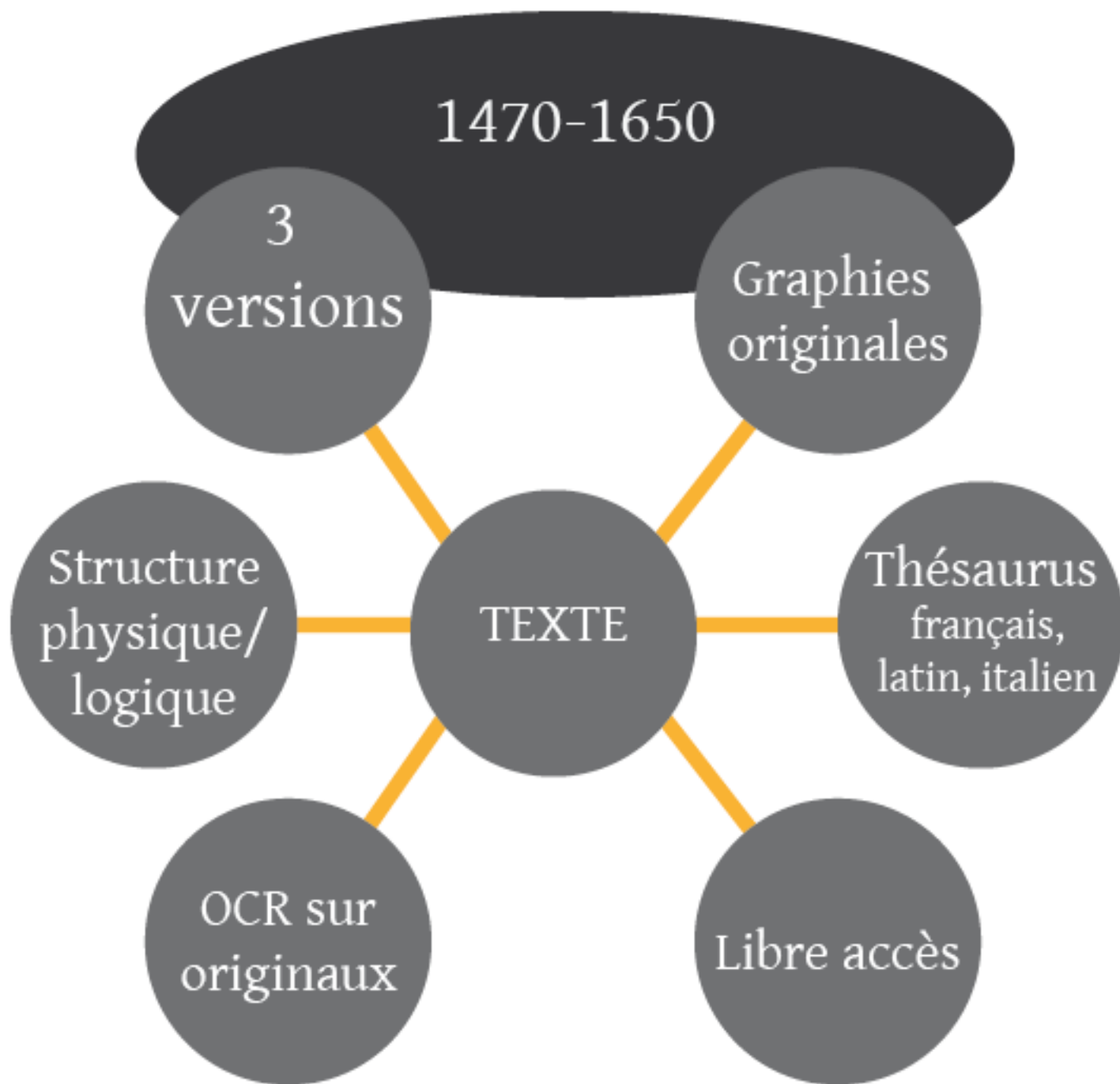


BVH-TIPo  
TEI



TELMA  
Formulaire

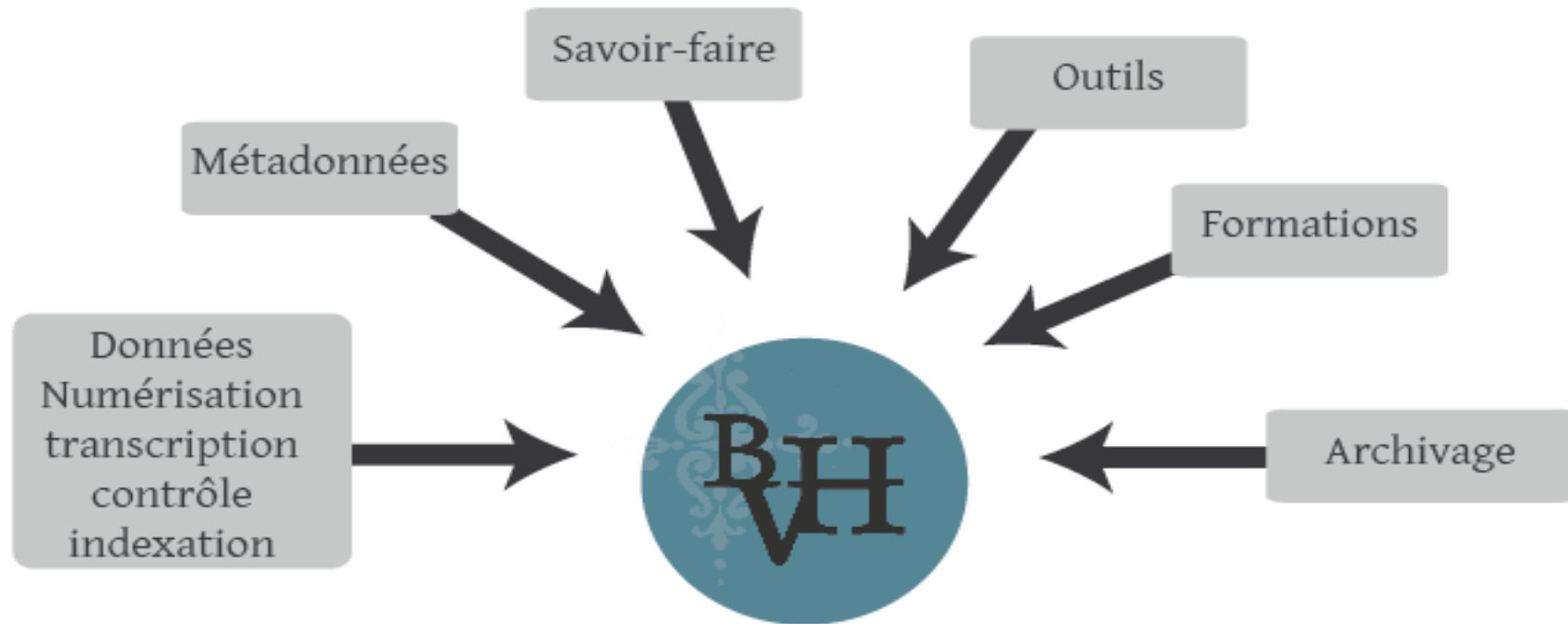




## Entités nommées, base BUDE (IRHT, M.-E. Boutroue) et RDF

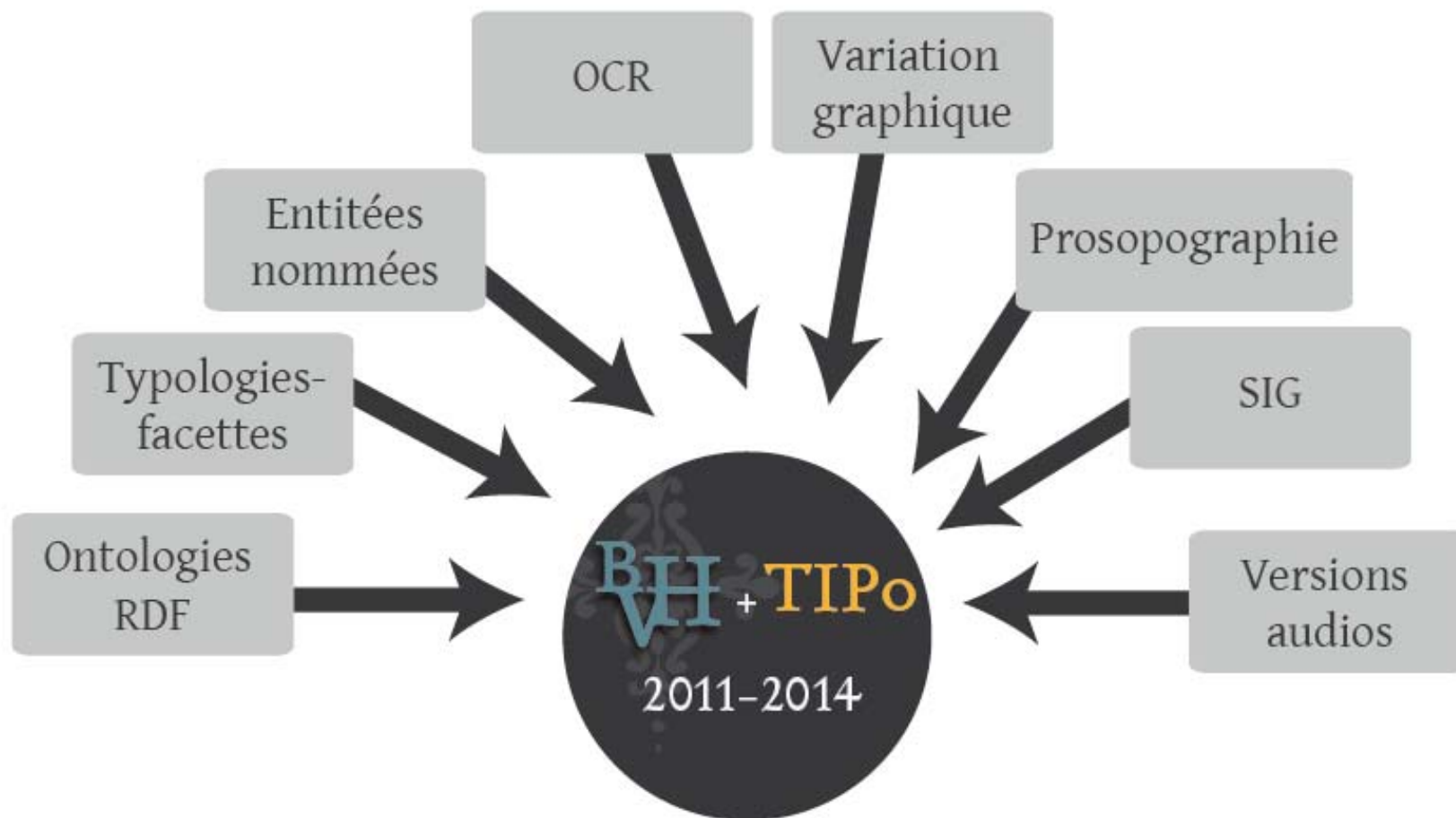
- Encodage TEI des noms propres (personnes, lieux, alias, personnes morales...) dans tous les textes
  - Génération et extraction d'index à partir d'XTF
  - Extraction des autorités à partir du catalogue (vedettes Rameau enrichies)
- Lien avec la Base Budé (Marie-Elisabeth Boutroue, IRHT)
  - Fiches prosopographiques de tous les humanistes (biographies, bibliographies)
  - Encodage des réseaux sociaux avec RDF (en projet)

# Les BVH et le Centre de ressources numériques TIPO (Adonis)

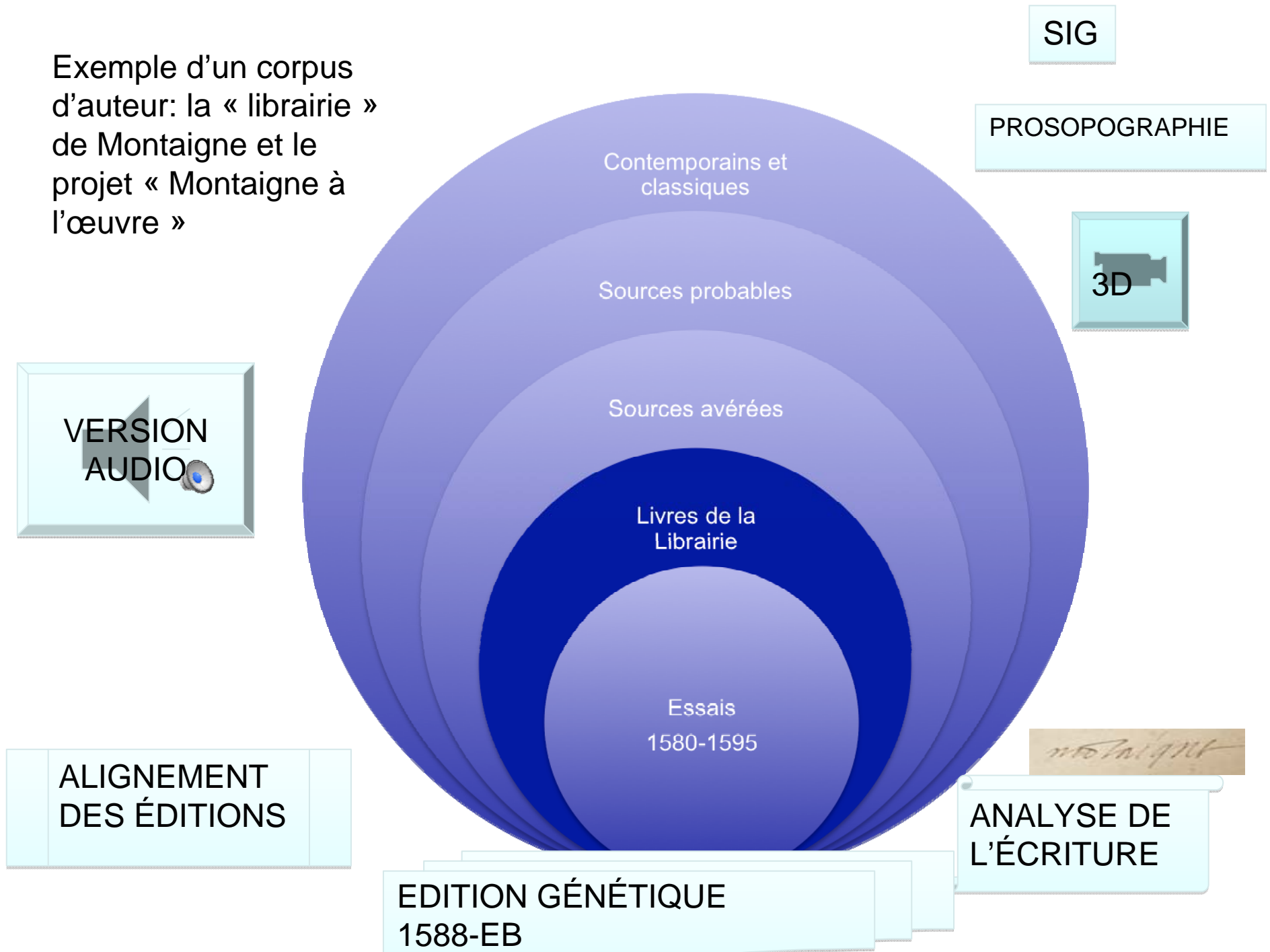


# Missions de BVH-TIPO

- Ressources humaines et outils
  - Organiser des sessions de formation (partenariat avec le MCC?)
  - Aider à la mise en route ou au suivi de projets
  - Produire des tutoriels et de la documentation pour les outils spécifiques
  - Produire les Wiki, liste de diffusion, newsletter, gestion de l'annotation collaborative
- Fonctionnement :
  - Missions et participations aux formations, aux congrès
  - Missions de conseil sur place
  - Actions de valorisation, traductions



Exemple d'un corpus d'auteur: la « librairie » de Montaigne et le projet « Montaigne à l'œuvre »



SIG

PROSOPOGRAPHIE

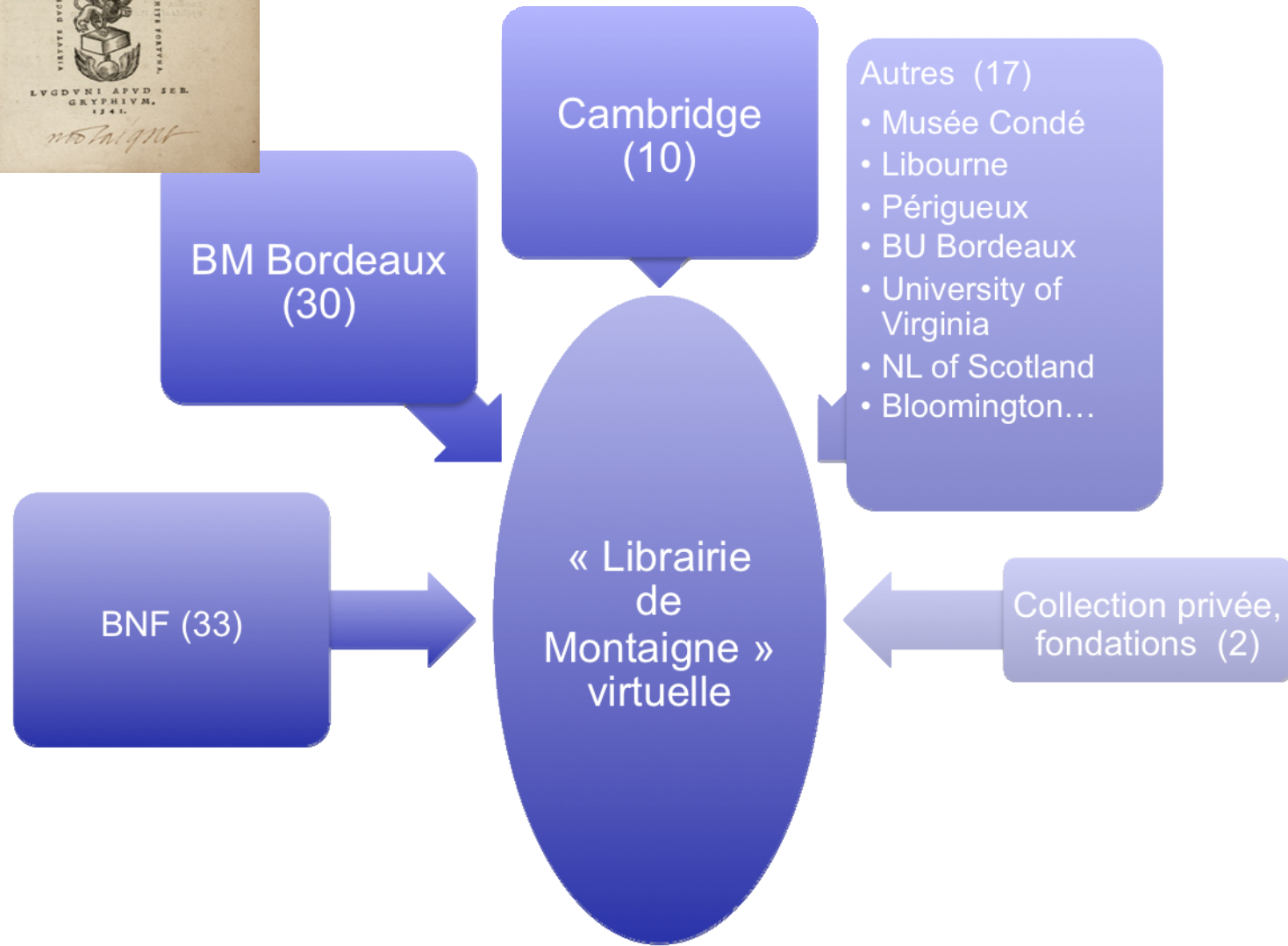
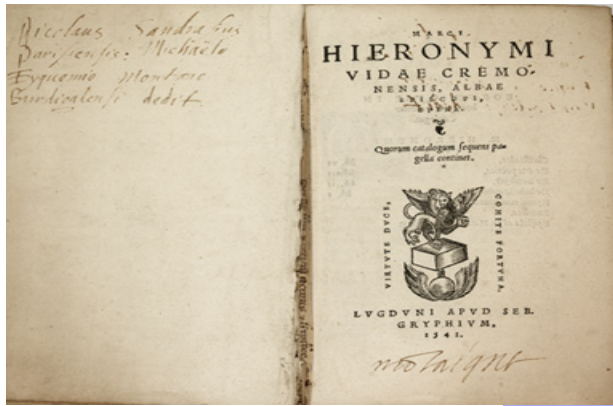
3D

VERSION  
AUDIO

ALIGNEMENT  
DES ÉDITIONS

EDITION GÉNÉTIQUE  
1588-EB

ANALYSE DE  
L'ÉCRITURE



Præterea, nisi materies æterna fuisset,  
Antehac ad nihilum penitus res quæque redissent,  
De nihiloque renata forent, quæcunque videmus.  
At quoniam supra docui nil posse creari  
De nihilo, neque quod genitum est, ad nil reuocari:  
Esse immortalia primordia corpore debent,  
Dissolui quod quæque supremo tempore possint,  
Materies ut suppeditet rebus reparandis.  
Sunt igitur solida primordia simplicitate,  
Nec ratione queunt alia seruata per æuum  
Ex infinito iam tempore res reparare.

DE NIHILOQUE RENATA] sic restitui à codicibus manuscr. ad hanc scripturam quasi manu ductus: in quibus scriptum est partim De nihiloque nata, partim, que ranta, partim que nata. ex qua postrema scriptura alij fecerunt, de nihilo que sita, ut habent omnes ferè vulgati codices: alij de nihilo quoque nata: quæ tamen probabilis lectio est. Sunt & qui legunt, De nihiloque parata forent. sed ita legendum esse, ut excudendum curavi, euincit locus similis infra eod. lib. de nihiloque renata vivecat copia rerum. DISSOLVI QVO] id est in quæ primordia res quæque genitæ sui interitus tempore dissolui queant. SUPREMO] ultimo. Horat. epist. ad Torq. Supremo te sole domi Torquate manebo. Idem ad Alb. Tib. Omnem crede diem tibi diluxisse supremum. SUPPEDITET] suppetat, satis sit, in prælo: vel parata sit, nam suppeditet hoc loco absolutè accipiendum est, ut sæpe alibi. SUNT IGITUR SOLIDA] concludit hunc locum hoc modo. Constant igitur corpora prima solida simplicitate, id est sunt solidissima, & simplicissima: atque ita sunt æterna. quod si non essent solidissima, interirent tandem, neque æterna essent. Si corpora prima interirent: vel non haberent res, vnde reficerentur, vel ex nihilo reficerentur: quod fieri non posse probatum est.

Denique si nullam finem natura parasset  
Frangendis rebus: iam corpora materiæ  
Usque redacta forent, euo frangente priore,  
Ut nihil ex illis à certo tempore posset  
Conceptum, summum ætatis peruadere florem.  
Nam quiduis citius dissolui posse videmus,  
Quàm rursus refici: quapropter, longa dici  
Infinitæ ætas anteaacti temporis omnis.

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*

*cederunt / 20 /*

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*

T. LVCRETII CARI  
DE RERVM NATVRA,  
LIBER SECVNDVS.



Vane, mari magno turbatibus æquora vêtis,  
E terra magnum alterius spectare laborem.  
Non quia vexari quæqu岸 incunda voluptas,  
Sed, quibus ipse malis careas, quia cernere suaue st.

Suaue etiam belli certamina magna tueri,  
Per campos instructa tua sine parte pericli:  
Sed nil dulcius est, bene quàm munita tenere  
Edita doctrina sapientum templa serena: *huc pertinet*  
Despicere vnde queas alios, passimque videre *212/269/375/*  
Errare, atque viam palanteis querere vitæ, *468/*  
Certare ingenio, contendere nobilitate,  
Nocteis atque dies niti præstante labore  
Ad summas emergere opes, rerumque potiri. *445/*  
O miseræ hominum mentes, ô pectora cæca:  
Qualibus in tenebris vitæ, quantisque periculis  
Degitur hoc æui, quodcunque est! nonne videtis  
Nil aliud sibi naturam latrare, nisi ut cum  
Corpore seiunctus dolor absit, mente fruatur  
Iucundo sensu, cura semota, metuque?  
Ergo corpoream ad naturam pauca videmus  
Esse opus omnino, quæ demant quenque dolorem.  
Delicias quoque uti nullas substernere possint:  
Gratius interdum neque natura ipsa requirit.  
Si non aurea sunt iuuenum simulacra per ædeis  
Lampadas igniferas manibus retinentia dextris,  
Lumina nocturnis epulis ut suppeditentur.

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*

*Si quædam in quibusdam  
sunt res quædam in quibusdam  
sunt res quædam in quibusdam*



# Atelier d'édition génétique

TEI Genetic Workshop: Encoding the documents, encoding the Genesis, par Elena Pierazzo & Malte Rehbein (Zadar, 10 novembre 2010)

Un nouveau module TEI pour les éditions génétiques (L. Bontemps, Lou Burnard)

Remplacer <text> par <document>

<TEI>

<teiHeader>

<facsimile>

<ge:document>

<text>

Elements:

@rotate, @ stage

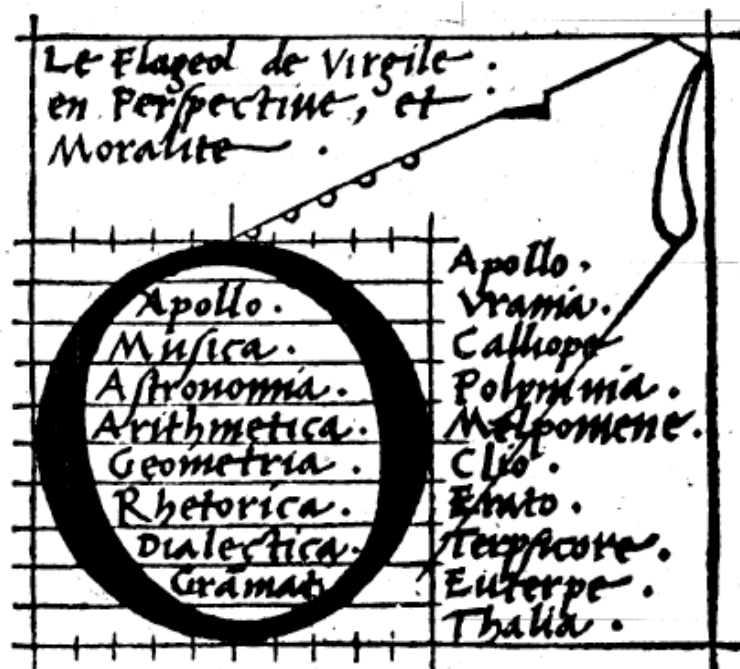
Montaigne, *Essais*,  
1588, « Exemplaire  
de Bordeaux »

*Vota puer soluit, que femina vouerat Iphis.*

Passant à Victry le François, ie peuz voir vn homme que l'Euesque de Soissons auoit nommé Germain, en confirmation, lequel tous les habitans de là ont cogneu, & veu fille, iusques à l'aage de vingt deux ans, nommée Marie. Il est à cett' heure fort barbu, & vieil, & ne fect point marié. Faisant, dict-il, quelque effort en sautant, ses membres virils se produisirent; & est encore en vsage entre les filles de là, vne chanson, par laquelle elles s'entraduertissent de ne faire point de grandes eniambees, de peur de deuenir garçons, cōme Marie Germain. Ce n'est pas tant de merueille, si ceste sorte d'accident se rencontre frequent: Car si l'imagination peut en telles choses, elle est si continuellement & si vigoureusement exercée en ce subiect, que pour n'auoir si souuent à rechoir en mesme pensée & aspreté de desir, elle a meilleur compte d'attacher & incorporer, vne fois pour toutes, cette virile partie aux filles. Les vns attribuent à la force de l'imagination, les cicatrices du Roy Dagobert & de Sainct François. On dict que les corps s'en-enleuent telle fois de leur place. Et Celsus recite d'un Prestre, qui rauissoit son ame en telle extase, que le corps en demouroit longue espace sans respiratio & sans sentiment. Il est vray semblable, que le principal credit des miracles, des visions, des enchantemens, & de tels effects extraordinaires, vienne de la puissance de l'imagination, agissant principalement contre les ames du vulgaire, ~~ou~~ <sup>plus</sup> ~~ou~~ <sup>moins</sup> de resistance. On leur a si fort faisi la creance, qu'ils pensent voir ce qu'ils ne voyent pas. Ie suis encore de cette opinion, que ces plaisantes liaisons <sup>des mariages</sup>, de quoy <sup>le mo</sup> de se voit si <sup>leur</sup> ~~plein~~, qu'il ne se parle d'autre chose, ce sont des impressions de l'apprehension & de la crainte. Car ie scay par experience, que tel de qui ie puis respondre, comme de moy mesme, en qui il ne pouuoit ~~choir~~ <sup>choir</sup> soupçon de foiblesse,

## Similarité d'écritures

L'écriture gravée de G. Tory  
(*Champfleury*, 1529)



Échantillons d'écriture du  
manuscrit anonyme,  
*Itinerarium Antonini*, 1511,  
BM d'Orléans

de

et

## Les “Google digital humanities research awards” 2010 (Europe)

- Proposition et visite à Tours (octobre 2010), dépôt des projets (1<sup>er</sup> novembre), notification (4 décembre), 50 000\$
- Projet 1 (porté par le Laboratoire d'Informatique de Tours):
  - Using Pattern Redundancy for Text Transcription
- Projet 2 (porté par le CESR-BVH):
  - Full-text retrieval and indexation for Early Modern French documents

## Projet 2 (CESR-BVH): Acquisition et indexation plein-texte pour les documents en français de la Renaissance

- Volet 1: Automatisation de la transformation en version patrimoniale, puis modernisée (par méthode n-grammes et dictionnaires), avec la société Digiscrib (Tours-La Riche)

*vifue > vifve > vive*

- Volet 2: outil de requête à partir de la variation graphique (application de règles linguistiques + dictionnaires), avec le laboratoire FORELL (Poitiers, Marie-Hélène Lay), à partir du mot « vif », recherche des adjectifs

*vif, vifs, vifz, vive, viue, vifues, vifves, vifues, ... =>  
recherche dans les bases hétérogènes*

# Collaborations nationales

- Toutes les bibliothèques de la Région Centre:
  - Argenton-sur-Creuse, Blois, Bourges, Chartres, Châteaudun, Châteauroux, Dreux, Loches, Montargis, Nogent-le-Rotrou, Orléans, Romorantin, Tours, Vendôme
- Archives municipales et départementales
- Bibliothèque nationale de France (pôle associé « Renaissance »)
- Poitiers BU et Médiathèque, BM Saumur
- Bibliothèque du Prytanée militaire de La Flèche
- **Fonds privés**
- Bibliothèque Mazarine, Sainte-Geneviève (en cours)
- Bibliothèque du Musée Condé à Chantilly (en cours)
- En discussion: Le Mans, Angers, **BU Toulouse ...**

## Collaborations nationales (laboratoires, entreprises)

- IRHT, section de l'humanisme
  - Consortium Navidomass (LI Tours – La Rochelle – Rouen – Nancy – Rennes), CRIP 5 (Paris V), LIRIS (INSA-Lyon), Informatique
  - Laboratoire Ligérien de Linguistique (Orléans-Tours)
  - FORELL B (Poitiers), Linguistique
  - GIS « Sources de l'érudition en Europe et en Méditerranée » (IRHT, Ecole des Chartes, LAMOP, LEM, LEM-EPHE, CESCUM, BnF, CIHAM, UPR 76, ARTEM-Nancy, UMR 8167, UMR 6125-Aix, FR 33, GAHOM-EHESS )
  - PPF, Fédération de recherche et Labex avec le CESCUM
  - Laboratoire ICAR, ENS-Lyon, Linguistique
  - MMSH Caen, pôle numérique
- Associations: Amis d'auteurs, maisons d'auteur...

### *Entreprises*

- Digiscrib (Tours-La Riche)
- Google
- Prestataires en numérisation, transcription, travaux informatiques

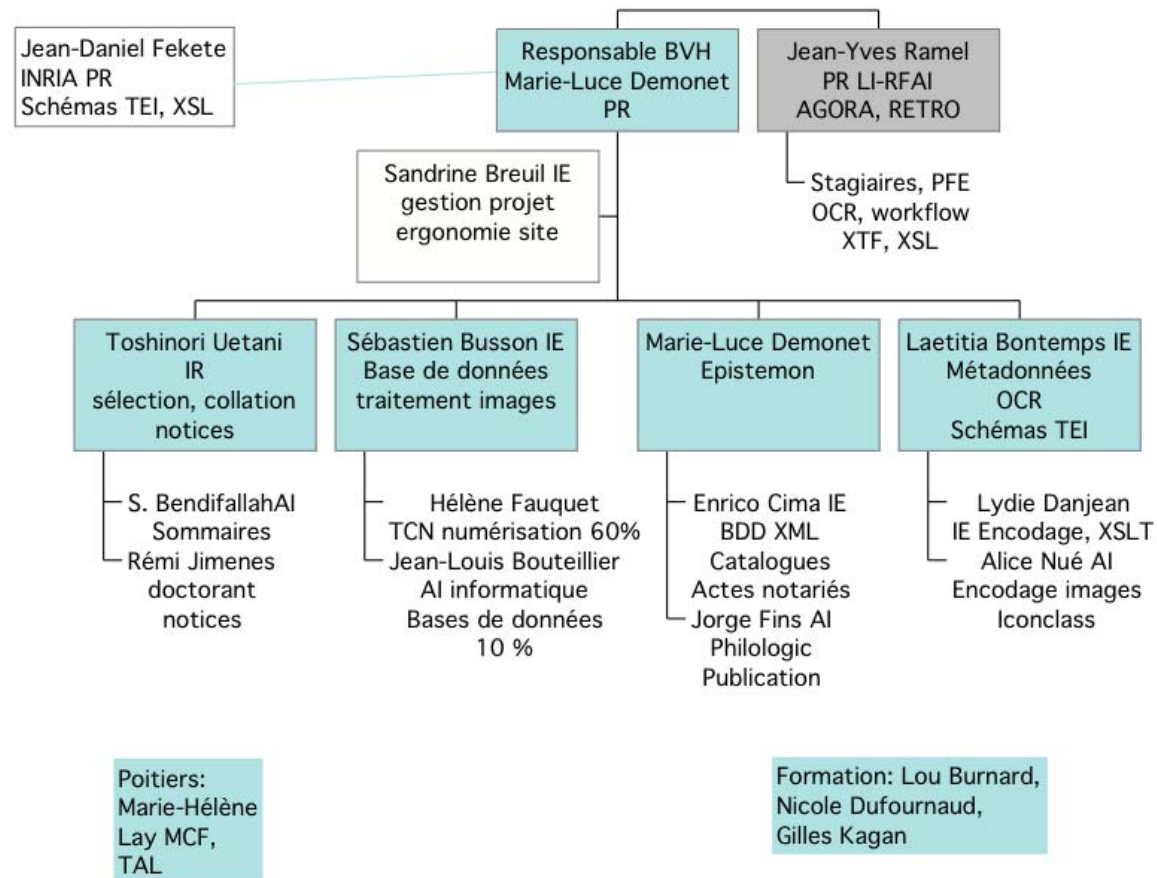
# Collaborations internationales

- Cambridge University Library (10 livres ayant appartenu à Montaigne, dont le Lucrèce annoté) – numérisation acquise (2010)
- Université de Chicago, PhiloLogic
- INTÉGRATION dans le CONSORTIUM EUROPEANA (métadonnées, web sémantique – ontologies/ mots-clés), en juin 2009
- Herzog August Bibliothek, Wolfenbüttel
- Glasgow University (livres d'emblèmes)
- Bibliothèque nationale de Hongrie
- University of Virginia Library
- Mnemosyne (Iconclass), The Hague
- **Projet « MONK »- Bamboo (consortium américain)**



Bibliothèques  
Etablissements  
partenaires

# Organigramme BVH



Prestataires  
serveur externe,  
SVN  
Editeur TEI  
Dissimilation

# CO-FINANCEMENTS BVH 2011

