



Réunion annuelle 2010 CESR, le 15 décembre 2010

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

- **Présents**

Ahmed Bensallah (LITIS, Université de Rouen / projet Impact), Mickaël Coustaty (L3i, Université de La Rochelle), Miguel Crespo Balzac (Digiscrib, La Riche), Geneviève Cron (projet Impact, BnF), Geneviève Guillemot (BnF, Réserve des livres rares), Dorothée Guillot (Polib, Université de Lille 3), Vanessa Le Rolle (Arkhênum), Frédéric-David Martin (Gallica, BnF), Valérie Neveu (Université d'Angers), Florent Palluault (Patrimoine écrit, Ministère de la culture), Thierry Paquet (LITIS, Université de Rouen / projet Impact), Delphine Usal (CN2SV), Richard Walter (TGE Adonis)

- **Région Centre**

Catherine Angevelle-Mocelin (BM, Orléans), Anne-Françoise Blot (BM, Orléans), Jean-Pierre Bouguier (DRAC Centre), Sylvie Fournioux (Livre au Centre, Vendôme), Michèle Prévost (BM, Tours)

- **IRHT**

Marie-Elisabeth Boutrou (section Humanisme)

- **Université François-Rabelais**

Pierre Aquilon (CESR), David Fiala (CESR, Ricercar), Stephan Geonget (CESR), Véronique Lacan (SCD Tours), Paul-Alexis Mellet (Univ. Cergy-Pontoise / CESR), Concetta Pennuto (CESR), Benoist Pierre (CESR), Jean-Yves Ramel (LI, Polytech Tours), Partha Roy (LI, Polytech Tours), Corinne Touchelay (SCD Tours)

- **CESR-BVH**

Samia Bendifallah, Laetitia Bontemps, Sandrine Breuil, Sébastien Busson, Enrico Cima, Lydie Danjean, Marie-Luce Demonet, Jorge Fins, Rémi Jimenes, Alice Nué, Toshinori Uetani.

- **CESR (étudiants)**

Tiphaine Boulay, Anaïs Colin, Lucile Griemard

- **Excusés**

Marie-Jeanne Boistard (BM Blois), Catherine Bony (BM Blois), Thierry Claerr (Ministère de la culture, service du livre et de la lecture), Thierry Delcourt (BnF), Anne-Sophie Delhaye (BnF), Nicole Dufournaud (EHESS, Paris), Pierre Gandil (Médiathèque de Grand Troyes), Guillaume Godet (BnF, Gallica), Bruno Guignard (BM Blois), Jérémy Herrant (PFE BVH, Polytech Tours), Isabelle Lamy (Musée Balzac), Marie-Hélène Lay (Ecrit.com, Poitiers), Thierry Leclair (BM Blois), Denis Maurel (Laboratoire d'Informatique, Polytech Tours), Raphaëlle Mouren (ENSSIB), Julie Pellegrin (Conseil général d'Indre-et-Loire, service des monuments et musées), Régis Rech (BM Tours), Sylvie Tisserand (Bibliothèque Henri IV, La Flèche)

• Acquisitions numériques des BVH

L'année 2010 aura permis une « mise à niveau » pour les traitements (contrôle qualité, structuration) des ouvrages numérisés au cours des années précédentes et la mise en place de nouveaux modèles pour la gestion et la diffusion des données, ce qui explique une progression lente en matière d'acquisition. Les 272 ouvrages d'Orléans ainsi traités au cours de l'année devraient être mis en ligne en 2011.

- **Fac-similés** : 160 ouvrages sont mis en ligne en 2010. Sur l'ensemble des 847 ouvrages numérisés, 534 sont accessibles en ligne, contre 374 en déc. 2009 (371 Tours, 272 Orléans, 86 Blois, 34 Châteauroux, 29 Vendôme, 29 Poitiers, 6 Châteaudun, 2 Romorantin, 1 Saumur, 1 Paris, 1 Budapest, 1 Vatican) ; plus de 30 000 éléments d'ornements extraits et indexés (bases de lettrines : OLDB, de portraits et de matériel typographique : BaTyR)

- **Epistemon, base textuelle** : 21 textes XML/TEI en ligne contre seulement 10 en 2009 (disponibles en téléchargement pdf), 16 html en ligne, 25 textes supplémentaires prévus pour mise en ligne au 1^{er} semestre 2011 ; 40 transcriptions en cours

- **Manuscrits** : *De minute en minute* comprend plus de 10 000 transcriptions ou analyses de minutes. 4500 minutes du XVe siècle ont été numérisées en collaboration avec les AD 37 et seront publiées, alignées à leur transcription. Les autres documents d'archives (livres de comptes) et manuscrits littéraires numérisés sont en cours de traitement.

- **Numérisations à venir** : Nogent-le-Rotrou, Loches, Bourges ?

• Accessibilité et propriété

- **Accès libre et gratuit à toutes les données** (y compris la structure TEI – en cours). La propriété intellectuelle pour toutes les données, métadonnées, bases de données et outils développés en partenariat (Agora, Retro du LI) qui sont libres de droits, est protégée avec la licence CC (Creative Commons : libre sans utilisation commerciale), mis à part le progiciel propriétaire Edit-tei (développé par la Société Digiscrib).

- **Identification des données** : Etude pour la mise en place d'identifiants « universels » et pérennes DOI, ark, handle... pour les contenus et notamment pour les transcriptions qui devraient également bénéficier d'un ISBN.

- **Archivage des données** : 1 copie en local (CESR) et 1 copie délocalisée (bibliothèques partenaires pour les fac-similés, chez le prestataire pour les bases textuelles), 1 copie délocalisée en projet au CINES (conservation et mise à jour, archivage pérenne) et IN2P3 (diffusion).

- **Balisage des métadonnées de catalogage** : Métadonnées de catalogues pour le moteur XTF (site BVH), compatibilité avec Gallica (OAI-PMH, Dublin Core), [Europeana](#) (OAI-PMH, norme ESE) et [Isidore](#) (moteur de recherche du TGE Adonis, CNRS).

• Communautés et publics visés

- **Publics** : Communautés de chercheurs et usagers en sciences humaines dans les domaines de la recherche, de la formation et de la culture

- **Intérêts** pour : le contenu (histoire, littérature, philosophie), le style (histoire de l'art), la langue, la forme (histoire du livre).

- **Détails des sources** : classiques de la Renaissance et ouvrages majeurs (Rabelais, Montaigne, Ronsard, etc.), sources religieuses, sources de la science, jurisprudence, textes politiques, ouvrages d'intérêt régional et programmes de recherche en cours.

• Diffusion : Evolution des besoins pour le site des BVH

- **Fonctionnalités du site des BVH à venir** : lecture en ligne et téléchargements multiformats (pdf, txt, mp3, epub, etc.), 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).
- **Passage à XML-TEI**, depuis 2007 : encodage des textes en TEI P5 et premier stage TEI ; en 2009, début de la bascule vers XTF (XML/ TEI) pour les fac-similés et les pdf ; et en 2010 : préparation de la bascule des autres bases de données (notices, feuilletage, sommaires, iconographie, minutes, manuscrits).
- **Intégration des fonctionnalités de visualisation et de navigation des BVH au moteur XTF** (J. Herrant, Projet de Fin d'Etudes - Polytech' Tours) : alignement du fac-similé avec sa transcription Epistemon, implémentation de Ajax Zoom, développement des interfaces, etc.
- **Affinement des métadonnées de typologie** (en projet) : la facette « genre » *Ex.* Jean Bodin, *La République*, 1576 - Domaine: histoire - Genre : Non-fiction, Théorie politique, Traité, Prose, Texte original (et non traduction).

• Production : Evolution des traitements pour la diffusion

- **Pré-balisage TEI** avec le progiciel EDIT-TEI (partenariat Digiscrib / BVH)
- **Vers une édition générique régularisée dite « patrimoniale »** : transcription diplomatique avec les abréviations, ligatures non développées et i/j, u/v non dissimilés. En janvier 2009 : automatisation de l'encodage des expansions de brévigaphes et de la dissimilation avec l'outil Dissimilog (T. Vincent) afin de permettre le double affichage en ligne (graphies originales/régularisées) ; en 2010 : une autre méthode est testée avec le progiciel Edit-TEI (Digiscrib) qui a ajouté deux modules de détildage et de dissimilation.
- **Evolution du corpus Epistemon** : l'objectif est d'atteindre 100 textes originaux en 2014, afin de permettre le traitement par des outils linguistiques de partenaires : Philologic (M. Olsen, Chicago), TXM (ICAR, Lyon), Analog (M.-H. Lay, Poitiers).

• Projets en partenariats

- **BAse de TYpographie de la Renaissance, BATYR (R. Jimenes)** : en cours de structuration. Cette base s'appuiera notamment sur des outils en cours de développement par le consortium NaviDoMass.
- **ANR NaviDoMass** : Indexeur de lettrines et [outil de recherche par l'exemple](#) (M. Coustaty - L3i, La Rochelle), pré-indexation d'images d'ornements typographiques (AGORA) et constitution automatique de spécimens de caractères (RETRO) et reconstitution des fontes (J.-Y. Ramel – LI RFAI Polytech' Tours), recherche sur les OCR (Ahmed Bensallah, Thierry Paquet - LITIS, Rouen)
- **« Google digital humanities research awards » 2010 (Europe)**
- **Using Pattern Redundancy for Text Transcription** (Projet 1) : porté par le Laboratoire d'Informatique de Tours.
- **Full-text retrieval and indexation for Early Modern French documents - Acquisition et indexation plein-texte pour les documents en français de la Renaissance** (Projet 2) : porté par l'équipe BVH du CESR et constitué de deux volets : **a)** Automatisation de la transformation en version patrimoniale, puis modernisée (par méthode n-grammes et dictionnaires), avec la société Digiscrib (Tours-La Riche) ; **b)** Outil de requête à partir de la variation graphique (application de règles linguistiques + dictionnaires), avec le laboratoire FORELL (Poitiers, Marie-Hélène Lay).

- **Projet de rétroconversion des *Catalogues Régionaux d'Incunables (CRI)***: en MARC/TEI (F. Palluault)

- **Entités nommées, base BUDE (IRHT, M.-E. Boutroue) et RDF**: après génération et extraction d'index via le moteur XTF dans le corpus XML/TEI Epistemon, et extraction des autorités à partir du catalogue (vedettes Rameau enrichies), le projet vise à lier les entités nommées des BVH aux fiches prosopographiques (biographies, bibliographies) de la [Base Budé](#) et faire apparaître les réseaux sociaux en RDF.

- **Les BVH et le Centre de ressources numériques TIPo : Textes Imprimés Patrimoniaux (TGE Adonis)**



Soutenue par le [TGE Adonis](#), l'équipe des BVH qui se situe à l'intersection de plusieurs communautés, met dorénavant à disposition des chercheurs ses compétences en tant que CRN (Centre de ressources numériques) « [TIPo](#) ». TIPo propose par exemple d'aider à : l'organisation de sessions de formation, la mise en route ou au suivi de projets, la production de tutoriels et de documentation pour des outils spécifiques.

- **Publication des actes notariés du XVe siècle des Archives départementales d'Indre-et-Loire (2011)** : collaboration des centres de ressources numériques du TGE Adonis BVH-TIPo et [TELMA](#) (IRHT, Orléans) :

- **BVH-TIPo** : Elaboration du modèle en XML/TEI, encodage fin des transcriptions et analyses issues de la base *De minute en minute* et traitements pour mise en ligne sous XTF ;

- **TELMA** : Expertise sur les actes médiévaux, création du formulaire de saisie pour le basculement BDD vers XML/TEI.

- **Projets BVH 2011-2012**

- 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
- Version 3 du [Manuel d'encodage TEI "Renaissance"](#) (en ligne depuis 2008)
- Développements de modèles en XML/TEI pour : la description d'imprimés, d'incunables, de manuscrits (éditions génétiques en TEI), d'archives et d'images (thésaurus [Iconclass](#), métadonnées des bases [ODLB](#) et BaTyR)

- **Formations dispensées et à venir**

- **Stage TEI formation continue** (SUFECO – CESR-BVH) – Novembre 2010, Tours :
- Stage TEI de niveau 1 : dispensé depuis 2008, reconduit en Novembre 2011 (durée allongée à 3 jours depuis 2010)

- Stage de niveau 2 : première édition (date à déterminer)

- **Modules de numérisation et stage TEI du master pro « Patrimoine écrit et édition numérique »** (traitement de l'image, initiation aux OCR., initiation à l'encodage XML/TEI de textes patrimoniaux de la Renaissance) et **Journée des professionnels** (ouverte au public, reconduite le 3 Mars 2011).

- **ANGD gestion numérique des sources de la recherche en sciences humaines et sociales** organisée par le CNRS et le TGE Adonis (gestion de projet) – Octobre 2010, Aussois (Savoie) : sera reconduite en Octobre 2011.