



Assemblée générale 2011

Les BVH : programme de recherche en « humanités numériques »

Partie 1

Marie-Luce Demonet et l'équipe des BVH
Lundi 5 décembre 2011
CESR, Tours



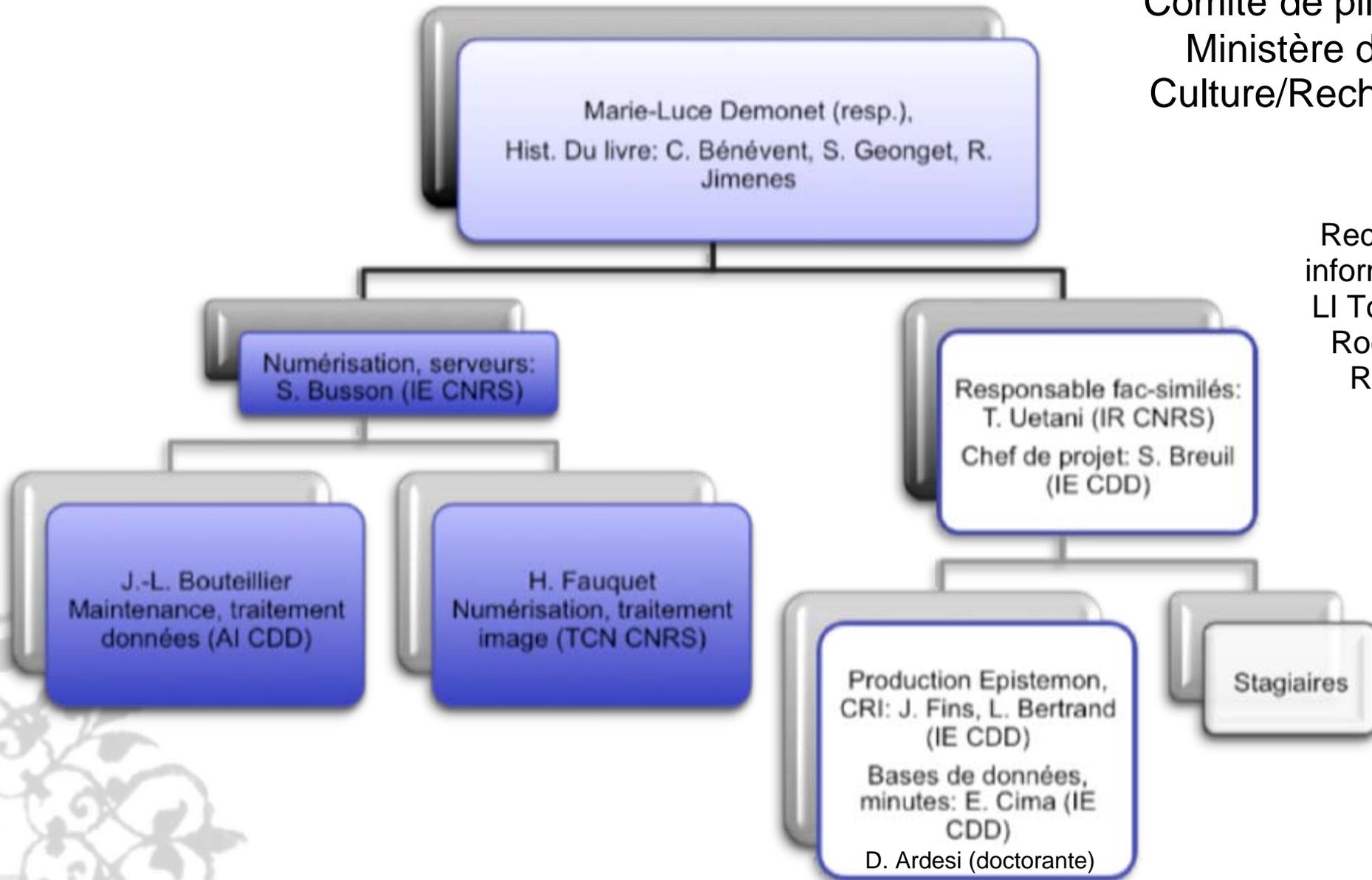
Les BVH dans les programmes de recherche au CESR

- Les programmes de numérisation au CESR (en cours) – Axe « Renaissance » du CPER, projet PADOVA (jusqu'en 2013)
- Labellisation par le CNRS de la Fédération de recherche CESR-CESCM
- Les programmes numériques du CESR (Labex Patrimar/ Equipex Biblissima)
- Labex PATRIMAR:
 - Ricercar, IRHT, BVH, CESCM: Personae (prosopographie) APR régional
- Equipex Biblissima: IRHT, EHESS, EnC, BnF, CIHAM (Lyon), BVH
- Laboratoires d'informatique impliqués: L3I La Rochelle, LISI et SIC Poitiers, LI Tours (Patrimar), laboratoires du Campus Condorcet (Biblissima)

Organigramme

Comité de pilotage:
Ministère de la
Culture/Recherche

Recherche
informatique:
LI Tours, La
Rochelle,
Rouen



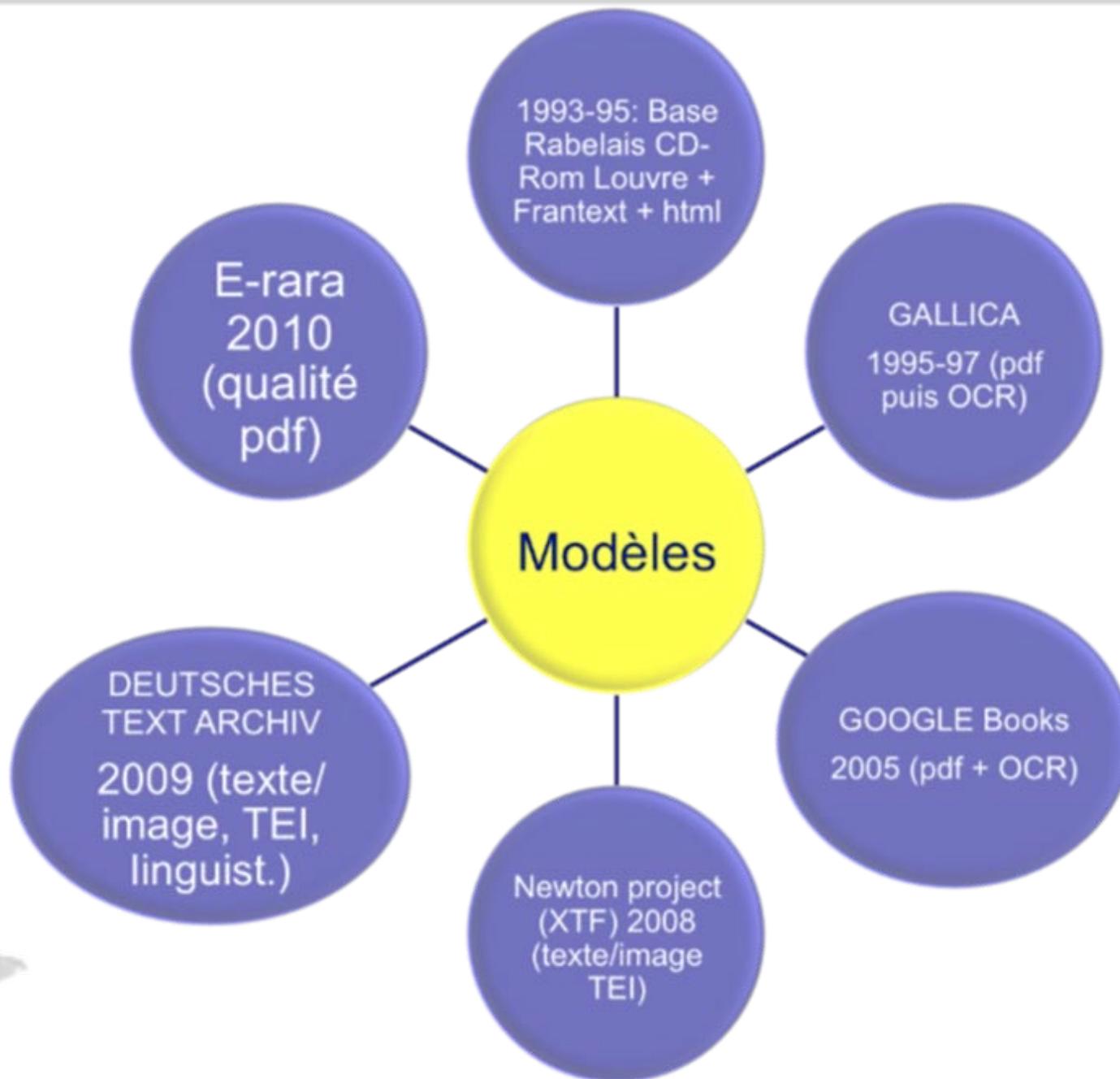
Des bases multi-formats

- Actuellement environ 920 fac-similés numérisés
 - 607 titres en ligne au 2 décembre 2011 — 400 en déc. 2009 (383 Tours, 7 Châteaudun, 19 Vendôme, 38 Châteauroux, 108 Blois, 29 Poitiers, 6 CAMBRIDGE, 1 à Budapest, Romorantin, Paris, Saumur, Nogent-le-Rotrou, Bologne... ; 21/ 272 Orléans, en traitement, 3 en 2010). **Projet de Bourges (100 à 150 ouvrages), déposé en novembre 2011, appel à projet du MCC**
 - Avec leurs sommaires
 - Avec téléchargement pdf/ et pdf-mrc
 - Avec les illustrations extraites et indexées (23 0000)
 - Bases de lettrines, portraits, matériel typographique: plus de 30 000 éléments extraits
- Archives, livres de comptes, manuscrits littéraires
- 302 192 téléchargements de fichiers pdf/4 615 000 téléchargements tous fichiers confondus (page web, images, pdf, ...) pour les 11 mois de 2011.

La base textuelle Epistemon

- Textes en français de la Renaissance, transcriptions inédites
 - Résorption du corpus antérieur en html (10 en 2011 sur 19 en 2010)
 - Encodage en XML-TEI depuis 2007
 - 46 textes, 36 en XML-TEI (10 en 2009, 21 en 2010), avec téléchargement des formats pdf, tei, html
- Manuscrits transcrits
 - De minute en minute (12 000 actes notariés), dont 4 500 documents d'archives du XVe siècle des AD 37 – Transcriptions ou analyses
 - Manuscrits du fonds Frotté (comptes de Marguerite de Navarre - début de transcription)

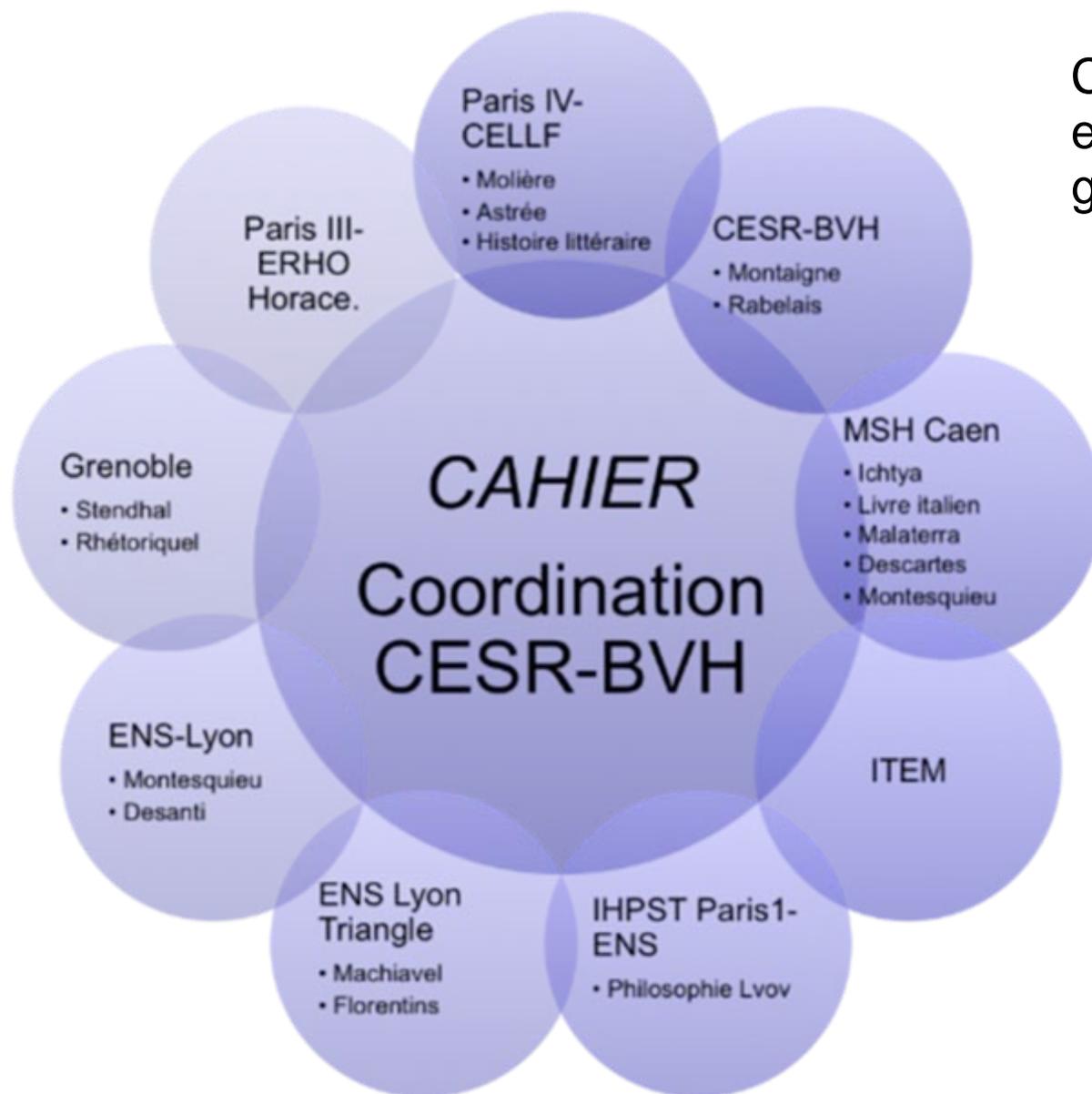
Les modèles



Les BVH dans les TGIR (Très Grandes Infrastructures de Recherche)

- BVH-TIPo dans le TGE Adonis: Textes Imprimés Patrimoniaux et humanités numériques – 2010-2011
- Appliquer les principes des « digital humanities »:
 - Accessibilité (Open data)
 - Protection (propriété intellectuelle)
 - Données réutilisables, partageables, transmissibles
 - Archivage durable
 - A l'intention des communautés de chercheurs
- Production ou Production *et* Services?
- En cours depuis juin 2011: redéfinition à partir des consortiums Corpus
- Participation des BVH au VCC3 (Scholarship content) de DARIAH pour les données textuelles

Les BVH dans DARIAH: le consortium CAHIER



Création d'un
entrepôt commun
géré par Adonis?

Les BVH et GALLICA

- Prolongation de la convention de Pôle associé « Renaissance » 2006
- Moissonnage des métadonnées des fac-similés (pas encore des transcriptions)
- Révision de la convention, incluant la recherche
- Association aux travaux du projet européen IMPACT (Improving Access to Text)
 - Recherches sur les OCR
 - Thèse en commun avec Rouen et BnF
 - Recherches sur la variation graphique du français de la Renaissance

Les BVH dans EUROPEANA

- Convention depuis juin 2009 (data content provider) Université de Tours/ Europeana
- Moissonnage depuis décembre 2009
- Participation aux sessions plénières 2009, 2010, et ateliers du CCPA
- Membre du Council of Content Providers (prochaine réunion: 6 décembre 2011)
- Nouvelle convention sur la disponibilité des métadonnées (signature novembre 2011 par le président de l'Université)
- Recommandations de la Communauté européenne sur la numérisation (28 octobre 2011)
- Feuille de route 2012 dans le cadre de la “New Renaissance”:
 - ressources terminologiques
 - guides de bonnes pratiques
 - déclaration sur les métadonnées utilisées
 - registre d'utilisateurs centralisé, données statistiques partagées

Financements

- Numérisation et traitement des données (hors recherche): budget annuel moyen

Numérisation Blois 2009 : 108 ouvrages (extern.)	Transcriptions	Sommaires	Serveur, SVN, développement Edit-Tei
20000	36000	17000	15000

Budget annuel moyen : 88 000 €

CO-FINANCEMENTS BVH 2012 (acquis ou attendus)

	CESR récurrents Univ/CNRS	CPER Etat	CPER Région	CPER CNRS	PPF Poitiers-Tours	APR Région ReNom	Adonis	Google	MCC hist du livre	Cahier	IUF	MCC Incunables	Feder
2011	1000	10000	40000	10000	10000	-	20000	35000	19000	3000	15000	30000	0
2012	1000	0	40000	?	0	35000	0	?	19000 ?	?	15000	0	30000

- 2012: entre 140 000 et 150 000 (la différence supprime 1 poste)
- 2011: 184 000
- *Projets en attente de résultats*: Labex Patrimar, Equipex Biblissima, ANR-Corpus Montaigne, Numérisation Bourges.
- *A noter*: sur 8 dépôts ANR, 7 refus (dont 3 comme porteur). Projets européens: 3 dépôts (en informatique) comme partenaire, 3 refus
- *APR Régional*: 3 dépôts, 1 refus.



Dissémination :
documentation, publications,
formations



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, OLDB - 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)

Publication et dissémination

- Novembre 2010, Tours, Journée Incunables
- Décembre 2010, Lyon-Valpré, Université d'hiver Adonis
- Janvier 2011, Journée édition génétique avec l'ITEM, Tours
- Juin 2011, Journée Patrimoine écrit, La Rochelle
- Juin 2011, congrès des DIGITAL HUMANITIES, Stanford (présentation du projet Montaigne et des premiers résultats obtenus avec la bourse Google)
- Juillet 2011, visite du manager Europe-universités de Google
- Juillet 2011, journée de travail avec l'EnC et l'AtiLF, avec ICAR-Lyon
- Octobre 2011, Séminaire de Caen
- Décembre 2011, Council of Content Providers, Europeana, Rotterdam
- Décembre 2011, Présentation des travaux réalisés avec les bourses Google (Centre culturel, Paris)
- En projet: participation au congrès des DIGITAL HUMANITIES, Hambourg, juillet 2012 – à la session plénière d'Europeana, Londres, juin 2012.

Formations

- Octobre 2011, ANGD (Aussois), gestion de projet
- Journée formation (suivie) sur XSLT, Lyon
- Novembre 2011, Stage TEI formation continue (SUFCO), Tours, CESR
- Organisation de l'atelier du consortium CAHIER, décembre 2011, Paris-Sorbonne
- 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 2012)
- Formation sur les entrepôts OAI-PMH (Adonis et Consortium CAHIER), printemps 2012



Les Bibliothèques Virtuelles Humanistes

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



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