



L'École nationale des chartes et les « humanités numériques » : premier bilan et perspectives

Florence Clavaud

directrice des nouvelles technologies et de l'informatique

École nationale des chartes (Paris, France)

florence.clavaud@enc.sorbonne.fr

Conférence à l'EBSI, Montréal (Québec, Canada), 24 février 2009

L'École des chartes (ENC) : un établissement d'enseignement et de recherche

- Fondée au XIX^e siècle, l'ENC donne aujourd'hui la formation scientifique nécessaire à de futurs chercheurs en histoire, et aux futurs conservateurs français (archives, bibliothèques, inventaire, Monuments historiques, musées...) ; un cursus exigeant sanctionné par le diplôme d'archiviste-paléographe et la soutenance d'une thèse ;
- elle abrite aussi une équipe de recherche pluridisciplinaire (enseignants-chercheurs) ;
- cœur des enseignements et de l'expertise : l'histoire et ses sciences auxiliaires (paléographie, diplomatique, langues anciennes et philologie, archivistique, histoire du livre et des media, histoire du droit, histoire de l'art...) ;
- une toute petite institution parisienne, engagée dans des partenariats avec les universités et des unités de recherche du CNRS, et connue un peu partout

L'informatique à l'ENC

- il y a quelques années, l'ENC a choisi de considérer l'informatique et le développement Web comme **une activité stratégique**, pour :
 - diffuser des ressources et outils de référence ;
 - publier les travaux réalisés par ses chercheurs et des chercheurs associés à partir de documents primaires ;
 - plus généralement, renouveler les moyens et les méthodes de la recherche en histoire et sciences auxiliaires de l'histoire
- > les **programmes quadriennaux 2006-2009 et 2010-2013** formulent clairement cette orientation

2004-2008 : une phase exploratoire et fondatrice...

- un **contexte favorable**, au sens où le champ des possibles est enfin très ouvert, et où quelques chercheurs de l'ENC sont fortement motivés ;
- lancement et réalisation des premiers projets mêlant étroitement dimensions scientifique et informatique ; le Moyen Âge y occupe une place prépondérante ;
- à cette occasion, après exploration, **définition des grands principes technologiques** et mise en place des moyens matériels et humains ;
- toujours à cette occasion, vérification du concept que l'édition électronique à vocation scientifique ou pédagogique, rigoureuse et sans compromis, est possible

Les lignes directrices alors établies

- rendre disponibles des documents inédits ou dont l'édition est ancienne, en mode texte ;
- compte tenu de l'effort à faire, le faire de manière professionnelle, et se donner les moyens de le faire une fois pour toutes, en tout cas de passer les années ; autrement dit de **pérenniser les corpus numériques** ; cela implique de respecter des normes, et l'utilisation de XML, en particulier de TEI (<<http://www.tei-c.org/>>, aujourd'hui dans sa version P5), est naturelle ;
- veiller à l'**accessibilité** des ressources (en particulier par des URL significatives et pérennes) ;
- offrir des services différents et/ou à valeur ajoutée par rapport aux publications imprimées, par ailleurs toujours l'objet de plusieurs collections, et parfois produites en parallèle ;
- publier en respectant les droits des détenteurs des documents primaires et des auteurs de l'édition ;
- le faire pour le cercle étroit des chercheurs et étudiants de l'ENC, et pour d'autres

Principales réalisations de cette période

- un **site à vocation pédagogique (THELEME)**, aussi le fruit des travaux d'étudiants, dont le périmètre va progressivement s'étendre (extension de l'époque médiévale à l'époque moderne) et dont le concept (dossiers documentaires) va servir de base pour développer un album de diplomatique européen ;
 - une **collection d'éditions électroniques (ELEC)** au sein de laquelle 8 numéros sont des bases de données de référence (applications PHP/MySQL), 10 consistent en éditions intégrales de corpus textuels anciens, majoritairement des textes médiévaux ;
 - un **centre de ressources numériques (TELMA)**, dont l'ENC est co-pilote avec l'IRHT, créé en 2006, positionné comme **opérateur technique et pôle d'expertise** ;
 - une **plate-forme technologique** sous-jacente aux applications d'éditions de textes, TELMA, couplant des moyens de navigation et de consultation, et un moteur de recherche, le tout pour documents TEI ; plate-forme développée par l'ENC à l'aide de logiciels libres
- (le tout accessible aussi depuis le site Web institutionnel de l'ENC,
<<http://www.enc.sorbonne.fr/>>)

Des acquis méthodologiques

- une équipe informatique composée de personnes au profil mixte (scientifique et technique), à l'écoute des chercheurs ;
- une expertise dans les technologies du développement Web normé et dans la mise en œuvre des grands modèles XML pour le document structuré ;
- une expérience de la conduite de projet informatique fondée sur la souplesse et la réactivité ;
- des liens avec d'autres opérateurs et pôles d'activité

Aujourd'hui : un contexte en pleine évolution

- Grâce en particulier à ces travaux, et à l'activité de multiples opérateurs techniques, les chercheurs français en SHS commencent à exprimer en nombre des besoins ;
- Après plusieurs années d'incertitude, un vrai réseau d'opérateurs techniques est en cours de constitution grâce à l'action et aux moyens du TGE ADONIS du CNRS (un des objectifs étant un métaportail de ressources) ;
- un indice fort : en France, le concept anglo-saxon de *Digital Humanities* ("humanités numériques") sort des petits cercles convaincus pour faire son entrée dans le vocabulaire des politiques

Du côté de l'équipe technique ENC

- d'une part des **problèmes classiques**, nés de la tension entre deux impératifs : mener à bien de nouveaux projets, tout en veillant à la maintenance et à l'évolution de l'existant ;
- d'autre part une forte **montée en charge** :
 - > plus de projets, en interne comme de l'extérieur, pour les quatre prochaines années ;
 - > donc beaucoup plus d'interlocuteurs ;
 - > des contenus plus divers (des types de documents primaires nouveaux et variés, des corpus parfois composites) ;
 - > des attentes fonctionnelles plus complexes
- **par essence des cas difficiles** : ceux où les corpus sont à inventer, où les besoins fonctionnels ne sont pas totalement définis, où les chercheurs ne sont pas toujours disponibles

Retour sur le scénario habituel

- discussion avec le chercheur : définition du périmètre du projet (les documents à traiter, les principales attentes fonctionnelles) en fonction des objectifs ;
- travail d'exploration technique, puis rédaction de cahier des charges (pour la sous-traitance) et/ou de spécifications (si l'ENC fait le travail elle-même) ;
- la production d'images numériques est sous-traitée ;
- en ce qui concerne l'informatisation du texte (éventuellement océrisation, dans tous les cas structuration), l'acteur peut être un sous-traitant, ou l'équipe informatique ENC, ou les deux ; c'est rarissime que ce soit le chercheur ;
- l'équipe ENC s'occupe de développer l'application Web, en l'adossant à la plateforme TELMA et en lui ajoutant d'autres composants éventuellement ;
- l'application est hébergée sur le serveur Web de l'ENC, qu'elle soit faite pour l'ENC ou pour une équipe extérieure ;
- les outils développés ou adaptés sont *open source* et même libres, mais restent à la seule disposition de l'ENC et de l'IRHT (pour ce qui est de TELMA)

Réécrire le scénario ?

Conditions de production des corpus numériques

- **Faire entrer le chercheur dans le cercle**, en tout cas lui en donner les moyens, en lui offrant des outils conviviaux lui permettant de convertir ses fichiers en documents structurés normés ; ainsi :
 - il aura la maîtrise totale de sa production scientifique ;
 - en la produisant en XML il portera peut-être un autre regard sur son corpus numérique, de nouvelles idées pourront naître, ou de nouvelles manières d'utiliser les normes ;
 - l'équipe technique pourra se concentrer sur d'autres activités, développement, mais aussi conseil, assistance, formations...
- > c'est l'objet d'une réponse au récent [appel à projets ADONIS](#), conjointement avec les équipes du programme MUTECH (<<http://www.mutec-shs.fr/>>), et avec l'intention de travailler avec d'autres équipes étrangères

Réécrire le scénario ?

Le cœur applicatif : outils de navigation et de recherche

- en ce qui concerne la **plate-forme TELMA, pour faciliter son utilisation et accélérer les publications** :
 - passer à une architecture modulaire où moteur de recherche et moyens de navigation sont nettement séparés ;
 - mutualiser, pour la navigation, tout ce qui peut l'être (logique de transformation de documents TEI P5 suffisamment générique) ;
 - rendre la recherche plus performante, en enrichissant les moyens de recherche documentaire et en améliorant les moyens de la fouille de texte ;
 - documenter et si possible et pertinent, distribuer le tout sous licence libre.
- > développements en cours ; mise en production prévue pour la mi-2009, pour de nouveaux corpus et au moins une partie des anciens

Réécrire le scénario ? Éditions en ligne

- Aller plus loin :

- mettre en place de vraies solutions pour que le corpus numérique évolue après sa publication (interfaces de dépôt de fichiers, par corpus entier ou par item) ;
 - accorder plus de place, voire la place principale, aux séries d'images du document primaire ;
 - compléter les moyens de navigation, selon les projets, par des accès cartographiques (géoréférencement), par l'utilisation de lignes de temps, de fichiers d'autorité, par la génération de graphes, d'ontologies... (cela demandera d'enrichir le balisage XML) ;
 - revoir l'ergonomie des éditions électroniques réalisées en 2006-2007 ;
 - explorer des idées visant à **rompre avec les interfaces proches de l'édition imprimée**
- ### - et l'édition imprimée ?
- faire du document XML le format d'entrée de la chaîne d'impression ? Dans ce cas il faudra en particulier chercher des moyens de convertir en XML des thèses composées en LaTeX.

Réécrire le scénario ? Services

- faire en sorte que les corpus publiés soient accessibles directement depuis des applications distantes (portails documentaires ; moteurs de recherche du Web), extractibles sur demande dans divers formats pour servir à d'autres usages : **interopérabilité** (amélioration de l'entrepôt de métadonnées OAI-PMH existant, ajout de marqueurs RDFa, identifiants uniques pour les ressources comme ceux conformes aux spécifications [ARK...](#))
- proposer aux chercheurs eux-mêmes, pour leur propre travail, des services et/ou des **briques logicielles récupérables**, de fouille de texte, d'utilisation de ressources lexicographiques, de traitement statistique de l'écriture...
 - > objectifs secondaires de deux projets de recherche, un sur le latin médiéval, un sur l'écriture ancienne

Pour y arriver : s'organiser mieux, partager

- gagner en efficacité, capitaliser les connaissances : planifier et coordonner les projets, documenter les développements, utiliser des méthodes et outils de développement collaboratif (systèmes de gestion de versions de fichiers)
- formaliser un minimum les relations avec les équipes de chercheurs externes (conventions définissant droits, rôles et responsabilités de chacun)

Par ailleurs :

- d'autres équipes œuvrent dans les mêmes directions, par ex. le Cluster 13 en région Rhône-Alpes, le CCH à Londres ou l'université d'Oxford, l'université de Chicago, etc.
 - des outils libres faits par d'autres peuvent s'avérer très utiles (comme [PhiloLogic](#) utilisé à l'ENC pour la fouille de texte) ;
 - des groupes de travail internationaux existent pour faire avancer les normes et outils ([TEI Manuscripts Special Interest Group](#), projet européen [ENRICH](#), [Charters Encoding Initiative](#)...)
- > se rencontrer, échanger, ou plus : collaborer dans le cadre de projets informatiques, en partageant les tâches

Recherche, développements, valorisation,...enseignement

Toujours pour construire un cercle :

- travailler avec et pour les «chercheurs de demain», en fait les étudiants :
 - faire entrer la culture informatique dans la formation scientifique initiale des étudiants archivistes-paléographes de l'ENC ;
 - former des jeunes de façon plus intensive aux nouvelles technologies (Master 2 « nouvelles technologies et histoire », voir : <http://www.enc.sorbonne.fr/parcours-master.html>) ;
- développer une offre de formation permanente ;
- rencontrer régulièrement les chercheurs (écoles thématiques comme celle organisée en octobre 2008 : <http://www.cn2sv.fr/ecole-sources-num/index.php/Accueil>, séminaires)

