

## Avant-propos destiné aux évaluateurs

Cet article propose une réflexion qui s'appuie en partie sur des témoignages issus d'un parcours professionnel relativement visible. L'auteur est un retraité qui n'a pas besoin d'être évalué pour des raisons de carrière.

De plus, le cadre d'application traité, un wiki du réseau Wicri sur la Chanson de Roland, est immédiatement identifiable.

Surtout, l'article propose un mode rédactionnel qui privilégie la lecture en ligne. Le texte sera recopié et annoté sur le site Wicri. En particulier, les notions qui ne sont pas universellement connues sont explicités par un lien vers la partie encyclopédique (ou sur d'autres wikis du réseau).

Un regard sur la version en ligne est indispensable pour une évaluation correcte. Avant le résultat de l'évaluation ce lien est :

[https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS\\_\(2023\)\\_Ducloy\\_\(proposition\)](https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS_(2023)_Ducloy_(proposition))

Pour cet ensemble de raisons, les contraintes du double aveugle sont impossibles à respecter et conduiraient à un résultat illisible. Nous avons donc décidé de ne pas appliquer les mécanismes d'anonymisation.

Si l'article est accepté le lien de la page sera le suivant :

[https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS\\_\(2023\)\\_Ducloy](https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS_(2023)_Ducloy)

Le texte de l'article sera le même que celui de la version acceptée (définitive). Il sera enrichi par des liens ou des images actives. Des notes de bas de pages pourront être supprimées pour être remplacées par des liens.

# Humanités assistées par ordinateur, un exemple avec la *Chanson de Roland*.

*Computer Assisted Humanities with the Chanson de Roland.*

Jacques Ducloy,

Laboratoire Paragraphe, Université Paris 8.

**Résumé.** Fournir un résumé en français. La taille du résumé peut éventuellement être telle que la première page puisse contenir le résumé et les mots-clés, aussi bien en français qu'en anglais.

**Mots-clés.** Chanson de Roland, Humanités numériques, Semantic MediaWiki.

**Abstract.** Fournir un résumé en anglais.

**Keywords.** Chanson de Roland, Digital Humanities, Semantic MediaWiki.

**Version en ligne.** Une version annotée et avec des liens actifs est visible ici : [https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS\\_\(2023\)\\_Ducloy](https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS_(2023)_Ducloy)

## 1 Introduction

A partir des premiers résultats d'un projet de bibliothèque numérique sur la *Chanson de Roland*, nous proposons des réflexions sur l'appropriation des technologies numériques pour la valorisation du patrimoine culturel.

Cette analyse est rédigée par un retraité qui s'appuie sur 50 ans d'expérience sur de grandes réalisations : le dictionnaire du Trésor de la langue française et les bases de données bibliographiques Pascal et Francis. En effet, après une phase vécue comme prestigieuse, elles ont été abandonnées en raison de difficultés techniques. Pour proposer des voies de redressement nous avons travaillé sur deux technologies complémentaires : l'ingénierie XML d'une part, les wikis sémantiques de l'autre.

Appliquées initialement aux analyses de publications scientifiques, elles se sont avérées très performantes sur la gestion et la publication de données diversifiées dans les humanités numériques. Par exemple, un wiki dédié à la musique rassemble des références bibliographiques, des articles réédités en mode hypertexte sémantique, des œuvres musicales (avec des partitions codées en LilyPond), des manuscrits et leurs transcriptions. Si les articles sont relativement indépendants les uns des autres, un noyau encyclopédique permet de naviguer dans cet ensemble.

Par des concours de circonstances, nous avons travaillé sur la *Chanson de Roland*. Cette thématique, apparemment très spécialisée, est en réalité un point d'entrée pour l'exploration d'un ensemble de poésies épiques avec des développements sur plus de 10 siècles d'histoire, de littérature, de musique, de linguistique, et qui plus est, à un niveau international (et multilingue). Cette thématique introduit également une différence fondamentale avec notre antériorité : une grande majorité de documents sont en dépendance étroite les uns avec autres.

Nous avons donc décidé d'expérimenter une vaste bibliothèque numérique où l'on puisse réaliser l'ensemble des actions liées à la recherche, depuis la transcription des données jusqu'à la diffusion de connaissances vers le grand public. Cette infrastructure est également utilisable pour des formations professionnelles destinées aux professionnels du soutien de la recherche et aussi pour les conservateurs ou les chercheurs impliqués dans les humanités numériques.

Dans cet article, nous présenterons nos motivations pour ces travaux et les solutions envisagées autour de l'Information Scientifique et Technique (IST). Nous montrerons ensuite comment elles s'appliquent aux humanités numériques et plus particulièrement dans la valorisation du patrimoine écrit.

## 2 Grandeur et décadence de l'IST en France

Notre projet est motivé par 50 ans d'activité dans l'IST. Ce témoignage comporte des assertions qui ne font pas l'unanimité mais qui sont importantes à connaître pour comprendre nos motivations et les options techniques retenues.

### 2.1 Se dégager du complexe inhibitif de rigueur pour aborder le numérique

Dans le contexte du Plan Calcul (1966), des initiations à l'informatique ont été créées dans les écoles d'ingénieur. Avec des collègues, nous sommes lancés dans le calcul numérique assisté par ordinateur avec les langages Algol ou Fortran.

Cette démarche n'était pas anodine. En effet, en 1956 à Nancy, Jean Legras, le fondateur de l'IUCA<sup>1</sup> écrivait (Legras 1956) : « *L'ingénieur, le physicien se trouvent souvent*

---

<sup>1</sup> Institut Universitaire de Calcul Automatique, un service commun pour les facultés et laboratoires universitaires sur Nancy en 1965.

*devant les problèmes que les mathématiciens classiques n'ont pas pu résoudre. Il leur faut alors, ou renoncer à l'emploi de l'outil mathématique, ou **utiliser des méthodes moins strictes, que réprouvent les mathématiciens, mais qui sont seules capables de les dépanner.*** ».

Pour illustrer un véritable changement de paradigme, il ajoutait : « *Il est alors indispensable que l'ingénieur, le physicien et tous ceux qui s'occupent de mathématiques appliquées, soient capables **de se dégager du complexe inhibitif de rigueur que leur a imposé leur éducation, et qu'ils osent se lancer à l'aventure : la vérification expérimentale sera là pour leur crier casse-cou le cas échéant.*** ».

Cette remarque sur le **complexe inhibitif de rigueur** nous paraît également de plus en plus fondamentale en 2023 pour dominer les humanités numériques.

### **Un premier exemple dans la documentation en 1973**

En 1970, après un DEA en analyse numérique, j'ai démarré ma carrière comme assistant à Nancy (pendant un an) où j'ai notamment enseigné le langage Fortran. Puis j'ai intégré l'IUCA comme ingénieur système (en passant une thèse en compilation en parallèle). En 1973, j'ai été invité à assurer une initiation au langage COBOL pour les étudiants de l'IUT Carrières de l'Information à Nancy. Cette option avait été retenue comme une école de rigueur par mes prédécesseurs issus de l'informatique de gestion. L'écriture d'un programme COBOL était particulièrement rébarbative<sup>2</sup>. Par exemple, COBOL ne manipulait que des données de taille fixe. Une notice bibliographique devait donc être distribuée sur quelques dizaines de cartes perforées. Compte tenu de mon expérience en analyse numérique, il me paraissait impossible de motiver les étudiants dans ces conditions. Comme le compilateur Fortran de l'ordinateur ICL 1901 de cet IUT disposait d'une extension pour manipuler des chaînes de caractères, j'ai décidé de me dégager du **complexe inhibitif de rigueur** porté par COBOL pour montrer aux étudiants, en utilisant Fortran, comment faire pour réaliser des filtres dans des corpus bibliographiques.

## **2.2 Une première mondiale dans les humanités numériques avec le TLF**

La langue française, dans sa dimension historique est un héritage fondamental. En 1960, Paul Imbs, également à Nancy, avait lancé un projet sur 20 ans pour la réalisation informatique d'un dictionnaire de langue, le Trésor de la langue française. Le CNRS avait acquis l'ordinateur français le plus puissant de cette époque un Gamma 60<sup>3</sup> de la compagnie Bull. Mais la programmation devait se faire dans un langage machine assez acrobatique, elle était donc inaccessible aux chercheurs en sciences humaines (ou même en calcul numérique). Le CNRS a donc appelé des informaticiens de haut niveau pour réaliser les développements. La compagnie Bull avait également affecté des ingénieurs pour cette vitrine technologique.

Malheureusement, cette équipe a eu une durée de vie limitée au stade initial. Dans les années 73, leurs programmes étaient devenus totalement obsolètes avec un ordinateur plus moderne, l'Iris 80, construit par la Cii<sup>4</sup>. Mais les experts étant partis et la transition a été très difficile.

Dans les années 80, Jacques Dendien, a rejoint le TLF pour y développer des services de haut niveau comme Frantext ou le TLFi (le TLF accessible par Internet). En dépit de ces succès en 95, l'expérience difficile du management de la production

<sup>2</sup> Les professionnels avaient introduit la catégorie analyste pour sous-traiter la programmation à des programmeur-codeurs.

<sup>3</sup> Cette puissance était en fait très modeste. En effet la mémoire centrale était de 130 K (octets) complétée par un tambour de 100 K. Le stockage de données utilisait exclusivement des bandes magnétiques (pas de disques).

<sup>4</sup> Compagnie Internationale pour l'Informatique

avait été tellement mal vécue par le CNRS qu'il a renoncé à engager une mise à jour du TLF. Le TLFi qui était le dictionnaire français de référence avec un immense succès sur le Web dans les années 2000 est maintenant supplanté par Wiktionnaire, technique et juridiquement piloté à San Francisco.

### **2.3 Une référence mondiale en 1975 : Pascal sur Cyclades avec MISTRAL**

En 1970, la Cii a développé MISTRAL, un système de recherche d'information pour placer la France en position mondiale dans l'IST. Compte tenu de la présence du TLF, la Cii nous a naturellement invité à acquérir ce progiciel.

Les étudiants de L'IUT ont été les pionniers à Nancy. En 1973, la première version ne fonctionnait qu'avec des bandes magnétiques (6 dérouleurs) et elle ne pouvait pas être utilisée en travaux pratiques. En revanche, en 1974, une nouvelle version, disque cette fois permettait déjà des extractions avec des équations booléennes. En parallèle, PIUCA, grâce au TLF, étant devenu site pilote pour tester les nouvelles versions du système SIRIS 8 (et de Mistral), les étudiants ont bénéficié de conditions exceptionnelles pour l'époque. Par petits groupes ils pouvaient créer leur propre base (avec un thésaurus) et lancer des recherches en temps partagé.

Forts de cette première expérience, nous avons ensuite informatisé le BALF<sup>5</sup>, associé au TLF. Ce bulletin existait sous la forme de notices bibliographiques. Avec un informaticien de l'INIST nous avons réalisé un transcodage (de mémoire assez simple) et généré une base Mistral. Nous avons organisé des séances de formation où je me souviens d'avoir insisté sur l'intérêt de la navigation dans le thésaurus.

En même temps, grâce aux bonnes relations que Claude Pair avait avec l'IRIA, nous avons pu disposer d'un strapontin dans les réunions techniques du réseau Cyclades, la préfiguration française de l'Internet, pour y être finalement connecté en 78, Nancy.

Mais la plus grande performance se situait du côté du CDST qui avait réussi, grâce notamment à l'impulsion de Nathalie Dusoulier à créer la base Pascal à partir des bulletins signalétiques du CNRS. Elle avait choisi d'utiliser le format ISO 2709 qui venait d'être créé (en 1973) dont la manipulation était assez complexe mais qui garantissait une compatibilité internationale. Avec une production qui était déjà de 400.000 références par an, la base Pascal a pu être accessible sur le réseau Cyclades sous le logiciel MISTRAL.

Malheureusement cette position d'excellence a été de courte durée. Dans les années 80, le réseau Cyclades a été arrêté. Le logiciel Mistral n'a pas été repris par le groupe Bull. L'équipe MISTRAL a rejoint la société TéléSystèmes pour y créer les services Questel.

De plus, forte de ce succès, Nathalie Dusoulier a mené une carrière dans les bibliothèques de l'ONU, Genève et New York. Elle a notamment assuré la fédération numérique de l'ensemble des bibliothèques de l'ONU. Le CDST est devenu très dépendant, en amont du savoir-faire de la société Jouve pour la constitution des bases de données et de la société Questel pour les services en ligne. Cette situation a causé de nombreux problèmes de gestion qui ont conduit les pouvoirs publics à la création de l'INIST.

### **2.4 Stations de travail Unix pour l'ingénierie XML**

Dans les années 80, les ordinateurs Multics ont remplacés les Iris 80. Avec la Cii basé à Louveciennes, nous avons des relations privilégiées avec les experts (Siris 8 ou MISTRAL) ou avec les équipes Iria (Cyclades). Multics étant géré à Phoenix, le

---

<sup>5</sup> Bulletin Analytique de la Langue Française.

travail à l'IUCA n'avait plus le même intérêt. J'ai alors rejoint une équipe sur un projet nommé ANL pour Association Nationale du Logiciel.

Ce projet était piloté par l'Agence de l'Informatique (ADI) et le CNRS avec comme partenaires le CNET, l'INRIA et le Ministère de la Recherche. Suite à la réalisation d'un inventaire de logiciels issus de laboratoires public, l'ANL est devenu un Groupement Scientifique pour la valorisation informationnelle des logiciels issus de la recherche. J'en ai pris la direction en 1981. Nous avons constitué un inventaire (informatisé) d'un millier de dossiers logiciels et nous organisons des expositions de logiciel en France et à l'international.

Nous étions en première ligne pour repérer des logiciels innovants pour le traitement de l'information technique. Ainsi, en 82-83 nous pouvions générer des catalogues et alimenter un serveur (sous le logiciel Texto). Un virage très important a été pris avec le pilotage des actions SM 90 par l'ADI. La SM 90 était issu des études du CNET pour concevoir une station de travail sous le système Unix qui commençait son expansion. L'ANL a alors été sollicitée pour faire un inventaire des logiciels français disponibles sous unix, avec le montage de démonstration. Notre inventaire numérique est devenu une matière première pour de nombreux tests de logiciels. En particulier, les équipes travaillant sur les compilateurs de compilateurs commençaient à appliquer leurs outils initialement conçus pour des programmes structurés aux documents. Du côté de l'équipe technique ANL nous avons donc fait une utilisation intensive d'analyseurs lexicaux (lex) pour adapter nos données à des logiciels d'intelligence artificielle (Lisp ou Prolog).

Coup de tonnerre, en 1987, Alain Madelin décide la dissolution de l'ADI qui assurait 50% du soutien de l'ANL. Je me suis alors rapproché de l'INIST. Débauché par Goéry Delacôte et sous la direction de Nathalie Dusoulier<sup>6</sup>, j'ai assuré au départ la direction Informatique. L'INIST avait hérité d'un schéma directeur basé sur un système intégré avec un SGBD relationnel sur un mainframe IBM. Cela ne me paraissait pas adapté à l'indexation des bases bibliographiques mais raisonnable pour les services de fournitures de documents. Or, Nathalie Dusoulier tenait à un système dédié pour la bibliothèque. Elle m'a invité à plonger dans les normes de catalogue, et plus précisément dans l'étude du format Unimarc sous la norme ISO 2709<sup>7</sup>. J'ai ainsi découvert que, malgré mon expérience documentaire antérieure j'avais tout à découvrir en bibliothéconomie ! L'INIST a donc fait l'acquisition, pour la bibliothèque, un système Geac qui a donné entièrement satisfaction.

Grâce aux relations issues de l'ANL, j'ai découvert (début 89) la norme SGML qui me paraissait bien adaptée à la norme ISO 2709. Bien avant MarcXml de la Library Of Congress, Nous avons alors développé une boîte à outil (iLib) pour le développement rapide d'applications. Avec un mécanisme préfigurant XPath nous avons démarré par des filtrages de corpus ISO 2709. Puis en s'inspirant des chaînes du TLF et de l'architecture MISTRAL nous avons spécifié des modèles SGML pour les données internes (fichiers inverses par exemple). Nous avons développé des modules (en langage C) qui permettaient de générer des systèmes de recherche avec des mécanismes de classification, dénommés serveur d'exploration.

## 2.5 Le démantèlement des missions stratégiques du CNRS en IST

Goéry Delacôte m'avait donné comme mission à l'informatique de redonner à moyen terme l'indépendance technologique (numérique donc) de l'INIST. L'action SGML entrainait dans cette stratégie, mais dans un climat souvent très conflictuel. En

<sup>6</sup> Qui avait été rappelée par le CNRS pour la création de l'INIST à Nancy.

<sup>7</sup> Plus connue sous l'appellation MARC. D'un point de vue informatique, une notice MARC est un ensemble de petits arbres où toutes les données structurelles sont variables.

effet, de nombreux cadres à qui le CNRS avait demandé de rester pour assurer une continuité de services étaient très impliqués dans les relations avec les sous-traitants et voyaient comme une menace la stratégie d'indépendance.

De plus, en 1991, nouveau coup de tonnerre, Goéry Delacôte, en conflit avec la Direction Générale, quitte le CNRS pour aller diriger l'Exploratorium de San Francisco. Parmi les raisons du conflit, l'INIST avait créé une filiale INIST Diffusion pour commercialiser les bases Pascal et Francis et les prestations de fourniture de documents. En dépit de la bonne qualité du nouveau service de FDP, le marché n'a pas suivi. Le CNRS a créé un Groupe INIST, piloté de fait par la filiale (et plus précisément par son chiffre d'affaires). Pour cela le CNRS a recruté des cadres plutôt issus du secteur de la vente en ligne et favorable à un retour à un modèle informatique centralisé avec maintien des aspects techniques à la sous-traitance. Le CNRS a fait machine arrière en 2000 (j'ai alors été rappelé comme directeur des produits et services). Mais un nouveau changement de direction de l'INIST est intervenu en 2004 dans un contexte difficile à la direction générale. Une nouvelle stratégie trop dépendante de la filiale s'est révélée catastrophique sur le long terme pour les bases Pascal et Francis qui ont finalement été démontées. Pendant ce temps-là, les américains, et notamment la NLM avec qui nous faisons jeu égal en dans les années 70 jusqu'au premières années de l'INIST a plus que doublé sa production et possède de fait un monopôle stratégique.

### **3 Une stratégie de mobilisation générale avec le projet Wicri**

Un an après le démarrage de Wikipédia, en 2002, le moteur MediaWiki apporte un ensemble d'innovations fondamentales qui vont devenir réellement disponibles à partir 2006. En désaccord avec la politique de l'INIST, j'ai alors rejoint la DRRT Lorraine où nous avons lancé une expérimentation sur l'usage des wikis sémantiques pour la promotion des résultats de la recherche. Nous avons monté un démonstrateur avec un wiki pour la région Lorraine et d'autres sur les priorités du CPER (eau, bois et forêts, sols urbains...). Puis, souvent pour des raisons démonstratives le réseau s'est enrichi pour atteindre maintenant 150 wikis.



**Figure 1.** *Le réseau Wicri en 2023*

La nature des wikis s'est diversifiée autour des publications scientifiques avec la revue les mots de l'agronomie (avec l'INRAE) ou les wikis des colloques CIDE ou H2PTM. Pour être transparent, j'ai aussi travaillé la faisabilité d'une réponse pilotée par des coopération entre les laboratoires de recherche face aux monopoles américains dans les bases bibliographiques. En même temps, à l'occasion d'une action sur la Renaissance en Lorraine, nous avons commencé à rééditer des ouvrages anciens, ouvrant ainsi une dimension héritage culturel. Nous avons fait un premier essai en 2010 avec un ouvrage de Lepage sur le Palais Ducal de Nancy. Il a été réédité avec une organisation hypertexte, en reprenant notamment une gravure

avec des renvois dans le texte de l'ouvrage où des annotations orientent le lecteur vers des explications complémentaires dans la souche encyclopédique. Pour des expérimentations multimédia, nous avons également investi sur la musique avec le langage LilyPond et nous avons réédités d'articles encyclopédiques (Jean-Jacques Rousseau par exemple) et d'extraits du TLF.

La figure 2 montre en vert sur la partie gauche l'ensemble des documents numériques qui cohabitent sur un wiki. Les catégories et liens sémantiques permettent de faire interagir divers modèles ontologiques (avec par exemple celui de Mistral). Enfin les modèles et modules permettent de programmer tout ce qui est spécifique à un domaine donné<sup>8</sup>. Une équipe de recherche qui investit en formation devient totalement autonome.

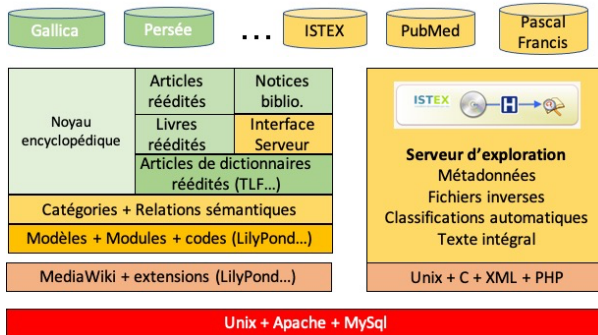


Figure 2. Éléments d'un ensemble wiki + serveur d'exploration

Aidés par un financement ISTEEX, nous avons réalisé le couplage d'un wiki avec notre approche dite des serveurs d'exploration. Comme évoqué plus haut, elle a démarré à l'INIST en 1991. Une équipe animée par Xavier Polanco disposait d'un ensemble de programmes pour la détection des fronts de recherche. Ils avaient été élaborés dans le cadre de thèses, avec une très grande incompatibilité entre eux. Ils avaient souvent été programmés avec les données en mémoire – ce qui limitait la taille des corpus. En imitant le TLF, qui utilisait intensivement le programme de tri standard, nous avons pu traiter de gros corpus. La boîte à outils iLib était limitée par un format SGML approximatif dédié à la norme ISO 2709. Elle a cependant permis la réalisation du système Stanalyst et, comme en 1976 avec Mistral, le retour à l'INIST d'un système de recherche documentaire pour la totalité de Pascal.

En même temps, au Loria à partir de 1993, j'ai développé une nouvelle version (Dilib). Elle préfigurait le modèle DOM du format XML. Elle permettait de mener des études de classification sur des corpus toujours volumineux mais avec des sources de plus en plus diversifiées (Medline, Dublin Core). Tout en faisant évoluer les versions dans une stratégie de compatibilité avec le W3C une action patrimoniale a été menée avec la base Biban (Base iconographique et bibliographique sur l'Art Nouveau). Lors de mon retour temporaire à l'INIST en 2000, j'ai emmené Dilib qui a été notamment utilisée pour un programme de formation (mutation technologique) et pour la création d'un service de veille et d'édition numérique. Une dizaine d'année plus tard, avec le programme LorExplor financé par ISTEEX, de nouvelles versions ont vu le jour avec une innovation assez fondamentale. A l'INIST le paramétrage des applications et le nettoyage des données (curation) relevait du bricolage informatique. Avec la version LorExplor les wikis ont été

<sup>8</sup> Par exemple, sur Wikipédia, les outils géographiques sont réalisés par les contributeurs.



utilisés pour la navigation (cartes dynamiques), le paramétrage et la curation des données. Près de 150 serveurs manipulant de 1000 à plus de 20.000 documents, dans tous les domaines couverts par Wicri ont été développés.

## 4 Roland au combat pour le patrimoine numérique

L'innovation n'étant pas un long fleuve tranquille, et sans insister à nouveau sur les détails, l'approche Wiki ne fait pas l'unanimité au sein des services de support à la recherche. A la fin du programme ISTEEX, l'INIST, qui avait été mandatée et financée par ISTEEX pour héberger le réseau Wicri et les serveurs d'exploration, a refusé d'engager une discussion sur une collaboration éventuelle. Bénéficiant de très bonnes relations avec la BUL de Nancy je me suis rapproché du laboratoire Paragraphe, pour tester l'ensemble du potentiel Wicri dans des actions de recherche au niveau d'un laboratoire. A 75 ans, je me suis mis dans la peau d'étudiants de master démarant des thèses en musicologie, philologie ou médiévistique. Je suis aussi bibliothécaire, éditeur, et informaticien, où plus précisément humaniste capable de programmer des modules récursifs, comme un chimiste résout des équations aux dérivées partielles.

### 4.1 La défaite de Roncevaux

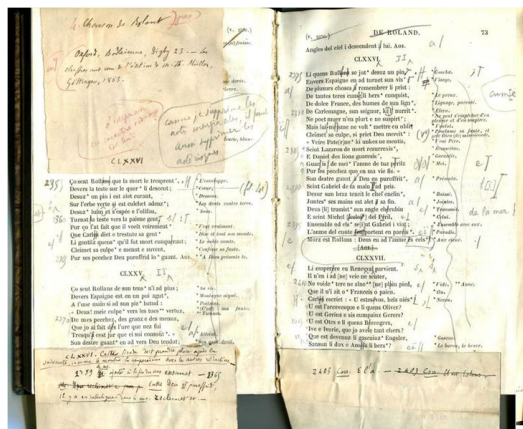
Le 15 août 778, de retour d'Espagne, Charlemagne perd son arrière-garde, tombée, à titre de représailles, sous le feu des troupes des seigneurs basques dont il a attaqué les possessions. Lors de la bataille de Roncevaux, l'arrière-garde est écrasée, provoquant la mort de nombreux braves de l'entourage de Charlemagne, dont celle de Roland, préfet de la Marche de Bretagne.

Tels sont les faits racontés par Éginhard au chapitre neuvième de sa *Vita Karoli Magni* (Vie de Charlemagne), et rappelés par Léon Gautier dans son édition populaire de 1895.

### 4.2 Premières étapes du projet

#### *Un stage d'une filière Métier du livre pour un ouvrage annoté*

En 2014, suite à nos travaux sur la réédition d'articles anciens, nous avons été sollicités par Isabelle Turcan pour accompagner un étudiant d'une filière "Métiers du livre" pour numériser un ouvrage annoté. Plus précisément, il s'agissait d'une édition critique d'un manuscrit (celui d'Oxford) publiée en 1869 par Francique Michel. Cet ouvrage était annoté par Paul Meyer.

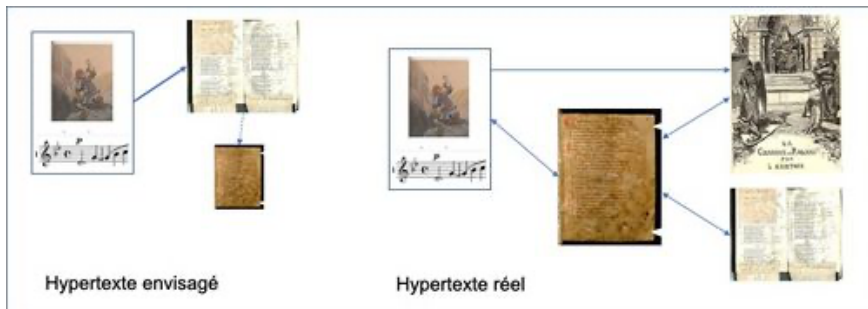


Le démarrage a été très rapide avec un expert en numérique (votre serviteur) à mi-temps pour former le stagiaire et développer quelques modèles MediaWiki (pour réaliser la mise en page). L'étudiant a commencé à produire des résultats pertinents en quelques jours. A la fin du stage, toutes les pages annotées avaient été traitées et une partie conséquente de l'ouvrage avait été transcrit en code wiki. Nous avons alors profité de cette première étape pour faire une démonstration (à destination des philologues) sur l'utilisation des wikis sémantiques en annotant les variantes d'orthographe dans le nom des personnages.

### ***Un stage apparemment anodin, mais décisif***

En mai 2021 un nouveau stage a conduit à mettre en place un projet de plus conséquent en visant un nouveau public : celui des choristes. En effet dans le cadre de travaux sur une bibliothèque numérique en musique, nous avons localisé une suite pour chœur et orchestre basée justement sur le manuscrit d'Oxford. Nous avons demandé aux stagiaires de mettre en relation les vers de l'oratorio avec le texte de Francisque Michel, en introduisant, à titre d'illustration, des facsimilés de feuillets du manuscrit.

Après un démarrage plutôt satisfaisant sur les premières strophes, des incohérences de numérotation de vers sont rapidement apparues. En effet, Gilles Mathieu avait travaillé à partir d'une autre édition critique (Léon Gautier). Le modèle hypertexte s'est donc fondamentalement modifié, avec 2 éditions critiques à prendre en compte. Le manuscrit devient alors le composant fondamental pour l'organisation numérique.



Il a donc fallu modifier en profondeur le modèle initial. Dans notre réflexion sur l'usage du numérique, ce premier problème, découvert au bout de quelques jours de développement, nous a semblé particulièrement démonstratif. En quelques mois, nous disposions d'un ensemble déjà démonstratif. Lors d'un séminaire de travail (avec des philologues) sur le fonds Paul Meyer, nous avons mis en évidence des faits peu connus. Par exemple, nous avons localisé l'ouvrage cible des annotations de Paul Meyer. Nous avons commencé à identifier la communauté des chercheurs qui, autour de l'École des Chartes et de la revue *Romania* examinaient et commentaient des éléments fortement localisés (un mot dans une laisse donnée...).

Nous avons donc décidé d'analyser le potentiel de cette thématique pour un projet conséquent de bibliothèque numérique.

### **4.3 Une bibliothèque numérique aux objectifs multiples**

Il nous a donc paru intéressant d'étudier une infrastructure numérique utilisable par des chercheurs pour leurs investigations et pas seulement pour la diffusion des résultats. La valorisation du fonds Paul Meyer offrait un premier axe de développement. Paul Meyer n'est pas seulement l'annotateur de Francisque Michel, il a été directeur de l'École des Chartes en 1882 et un des fondateurs des revues *Romania* et de la *Revue critique d'histoire et de littérature*.

Le réseau Wicri contient des sites basés sur des collections thématiques de publications de recherche. La thématique ici retenue introduit une contrainte nouvelle. En effet, les données sur lesquelles s'appuient ces publications sont des manuscrits qui sont numérisés et structurés en mode hypertexte sur le même site.

Du côté musical, la perspective d'un concert ouvre deux axes complémentaires. Cette œuvre ayant été peu interprétée, nous voulons faciliter les « négociations » entre le compositeur et un directeur artistique. Nous avons donc entrepris de transcrire en code LilyPond l'ensemble des partitions pour être capable de générer de nouvelles partitions en quelques jours. Nous avons aussi cherché à donner des consignes de prononciation correctes pour les vers du manuscrit. Ceci implique un alignement très précis sur les vers du manuscrit en remontant vers un glossaire qui peut donner des indications.

#### **4.4 Des documents hétérogènes, très diversifiés et très interconnectés**

##### ***Combien de mètres de rayonnage pour la Chanson de Roland ?***

Dans la bibliothèque des lettres de l'Université de Lorraine, la Chanson de Roland occupe trente centimètres de rayonnage dont une dizaine pour les trois tomes (2 940 pages) de Joseph Duggan (Duggan 2005). La requête « "Roland" "Chanson" "Charlemagne" », sans les citations, sur Google Scholar donne environ 14.000 références, soit une centaine de mètres... La mise en place d'une bibliothèque significative n'est donc pas une entreprise anodine. Le modèle général d'un wiki sur le réseau Wicri donne déjà une base d'organisation.

##### ***Le manuscrit d'Oxford***

Le manuscrit d'Oxford occupe une place particulière au cœur du dispositif, avec 3 types de « pages wikis » :

- Chaque facsimilé de page (144 au total) donne lieu à une page de description plutôt destinée à la gestion. Par exemple, les images extraites (comme une lettrine) pour alimenter un article mentionnent leur appartenance à ce facsimilé. Réciproquement MediaWiki gère les liens inverses. Il est ainsi possible de connaître tous les extraits et de savoir où ils ont été utilisés.
- Chaque feuillet (72) possède sa page de description (avec insertion des images de facsimilé recto et verso). Citons ici une spécificité de ce manuscrit. Le philologue allemand Edmund Stengel a édité une version critique en 1878 où chaque page imprimée recouvre exactement le contenu d'une page du manuscrit. Les pages feuillets vont permettre au chercheur de confronter une page du manuscrit avec une interprétation de référence.
- Chaque laisse (396 avec notre repérage) donne lieu à la création d'une page wiki où sont reproduites les facsimilés des pages correspondantes dans le manuscrit. Nous avons choisi une version de référence (Gautier 1872) afin d'associer à chaque couplet sa transcription et sa traduction. Nous verrons que chacune de ces pages wiki héberge d'autres informations de provenance diverse.

Quelques-uns seront traité dans leur intégralité. Par exemple, le premier stage ne portait que sur la partie de l'ouvrage de Francisque Michel qui était sa version critique du manuscrit d'Oxford. Il traite également deux autres manuscrits :

Les vers sont généralement identifiés par des numéros (de 1 à 4401). Nous avons été amenés à créer des pages vers qui sont des redirections au sens wiki.

Dans d'autres articles, nous avons signalé les problèmes rencontrés par la diversité de numérotations de vers et de laisse par différents auteurs. Ces faits sont mentionnés dans chaque page laisse.

### ***Les autres manuscrits***

Chaque manuscrit donne lieu à une étude particulière.

Quelques-uns sont en cours de traitement dans leur intégralité. Par exemple, le premier stage ne portait que sur la partie de l'ouvrage de Francisque Michel qui était sa version critique du manuscrit d'Oxford. Il traite également deux autres manuscrits :

- Le manuscrit dit de Paris dans son intégralité (6828 vers décasyllabiques répartis en 375 lisses monorimes). Mais ce manuscrit est incomplet, tout le début a été égaré.
- Le début du manuscrit dit de Châteauroux (85 lisses, 1332 vers) correspondant à la partie manquante dans le manuscrit de Paris. Le manuscrit complet fait 8201 vers sur 452 lisses.

Pour ces deux manuscrits, le travail d'alignement avec celui d'Oxford est fondamental pour comprendre l'histoire de ce poème. Ce traitement pose en fait une multitude de petits problèmes.

D'autres manuscrits comme ceux dits de Cambridge et de Venise 4 donneront lieu au même type de traitement. Citons également le manuscrit de Conrad qui est en allemand avec de très intéressantes illustrations qui sont absentes sur les autres.

De nombreux textes établissent des comparaisons avec d'autres manuscrits du moyen âge ou même de la Renaissance qui sont traités de façon partielle.

Citons également des manuscrits pour des partitions (Charpentier).

## **4.5 Des pratiques également diversifiées**

### ***Apports et limites du libre accès à l'information culturelle***

La gestion des droits d'accès à l'information avec un haut niveau de confidentialité introduit un très haut niveau de contraintes qui entraînent des développements coûteux et complexes. Nous ne mettons donc en ligne que des documents libres de droits. Tout le monde peut lire les textes et seuls des contributeurs dûment identifiés peuvent intervenir sur les contenus.

## **5 Bilan et perspectives**

En 1991, en me confiant la direction du Département Recherche et Produits Nouveaux de l'INIST, Goéry Delacôte m'avait invité à travailler sur la Station de travail du chercheur.

### **5.1 Gérer les réticences individuelles**

Pour la réflexion engagée dans cet article sur les humanités numériques, la maîtrise des applications en science humaine demande une forte expertise en informatique théorique (structures de données, etc). Dans les sciences de

l'ingénieur, un expert doit savoir résoudre des systèmes d'équations aux dérivées partielles. Dans les sciences humaines, un expert doit savoir écrire un programme récursif de navigation dans les arborescences.

## **5.2 Aspects institutionnels**

## **6 Conclusion**

Tout article doit contenir une conclusion qui sera d'alinéa 1 (Titre 1) le texte de type paragraphe.

## **7 Bibliographie**

Legras, J. (1956). *Résolution des équations aux dérivées partielles*, Dunod 1956

## **8 A insérer**

A l'occasion de manifestations grand public comme les « fêtes de la science » nous avons été confronté à un phénomène de pertes de racines culturelles en quelques générations :

En 1881, était officiellement désignée comme l'un des textes classiques à l'usage des élèves de seconde,

En 1906, dans une revue de grande diffusion pour la jeunesse, on peut trouver une bande dessinée sur Roland.

Dans les années 50 à 60, la Chanson faisait encore l'objet d'études dans les lycées. Elle était également présentée dans les cours d'histoire pour les cours élémentaires. Le manuel d'Histoire de France diffusé par Nathan en 1955 consacre 2 pages (sur 80) à Roland (autant que pour Charlemagne, Louis XIV fait mieux avec 4 pages).