



# L'Impact des Humanités Numériques dans la Cohérence d'un entrepôt numérique - une triangulaire entre chercheurs, spécialistes de l'IST et informaticiens

Frédérique Peguiron

## ► To cite this version:

Frédérique Peguiron. L'Impact des Humanités Numériques dans la Cohérence d'un entrepôt numérique - une triangulaire entre chercheurs, spécialistes de l'IST et informaticiens. DTUC '20: Digital Tools & Uses Congress, Oct 2020, Virtual Event Tunisia, France. pp.1-8, 10.1145/3423603.3424008 . hal-03081951

**HAL Id: hal-03081951**

**<https://hal.science/hal-03081951>**

Submitted on 18 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# L'Impact des Humanités Numériques dans la Cohérence d'un entrepôt numérique

une triangulaire entre chercheurs, spécialistes de l'IST et informaticiens

Frédérique Peguiron-Demoulin

Université de Lorraine

Nancy, France

frederique.peguiron@univ-lorraine.fr

## ABSTRACT

Les humanités numériques ou *digital humanities* font partie d'une culture qui rassemble bibliothécaires, ingénieurs, enseignants et chercheurs. Leurs pratiques s'imbriquent pour élaborer et construire ensemble des projets numériques dans les disciplines des Sciences humaines et sociales ou des Arts et des Lettres. Cet article montre comment l'écriture numérique collaborative, pierre angulaire de la construction du patrimoine numérique, participe également de l'édition numérique. Un raisonnement par anticipation, démontre comment les bibliothécaires en lien avec les informaticiens, peuvent venir en appui aux chercheurs, pour construire des projets de recherche numériques. Un cas pratique illustre l'aspect théorique et montre par un retour d'expérience, comment passer d'un environnement classique Codex vers un environnement hypertexte aux relations sémantiques. Les perspectives de l'article donnent des pistes sur les compétences à acquérir dans le cadre de nouvelles filières d'enseignement à inventer autour du «génie numérique».

## CCS CONCEPTS

• **Applied computing** → **Arts and humanities**; *Sound and music computing*; **Digital libraries and archives**; *Document management and text processing*; • **Information systems** → **Digital libraries and archives**; *Extensible Markup Language (XML)*; **Collaborative and social computing systems and tools**; **Wikis**; **Specialized information retrieval**; • **Human-centered computing** → **Interaction paradigms**;

## KEYWORDS

Digital humanities, Digital libraries, Digital datawarehouse, user modeling, semantization

### ACM Reference Format:

Frédérique Peguiron-Demoulin. 2020. L'Impact des Humanités Numériques dans la Cohérence d'un entrepôt numérique: une triangulaire entre chercheurs, spécialistes de l'IST et informaticiens. In *Digital Tools & Uses Congress (DTUC '20)*, October 15–17, 2020, Hammamet, Tunisia. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3240117.3240128>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

DTUC '20, October 15–17, 2020, Hammamet, Tunisia

© 2020 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-6451-5/18/10...\$15.00

<https://doi.org/10.1145/3240117.3240128>

## 1 INTRODUCTION

Les humanités numériques ou *digital humanities* [11] font partie d'une culture qui rassemble bibliothécaires, ingénieurs, enseignants et chercheurs. Leurs pratiques s'imbriquent pour élaborer et construire ensemble des projets numériques dans les disciplines des Sciences humaines et sociales, des Arts et des Lettres[8]. Ce présent article met en éveil les champs d'actions possibles et les compétences nécessaires des acteurs des systèmes d'informations numériques pour prendre part aux constructions transdisciplinaires numériques dans le cadre de bibliothèques universitaires. Nous montrons comment l'écriture numérique, pierre angulaire de la construction du patrimoine numérique, participe également de l'édition numérique. Un raisonnement par anticipation, démontre comment les bibliothécaires en lien avec les informaticiens, peuvent venir en appui aux chercheurs, pour construire des projets de recherche numériques. Cet article offre une partie théorique à partir de concepts et de leurs relations pour poser des hypothèses qui font entrevoir les convergences techniques entre types d'acteurs pour la construction de systèmes d'informations numériques inter-opérables d'une part et collaboratif d'autre part. Un cas pratique illustre ensuite le raisonnement théorique et montre par un retour d'expérience le passage d'un environnement classique Codex vers un environnement hypertexte avec des relations sémantiques. Les perspectives de l'article donnent des pistes sur les compétences à acquérir dans le cadre de nouvelles filières d'enseignement à inventer autour du «génie numérique»[13].

## 2 ENGAGEMENT POUR LES HUMANITÉS NUMÉRIQUES EN BIBLIOTHÈQUE UNIVERSITAIRE DE LETTRES ET SHS

### 2.1 Exposition de premiers travaux de recherche sur les Humanités Numériques

**Un environnement sémantique comme plateforme expérimentale à l'écriture numérique**

Une première approche des Humanités Numériques par les définitions a permis d'en dresser un état de l'art pour apprendre par les aspects historiques[1] l'évolution de la *Literary and linguistic computing* vers les *Humanities computing* puis l'apparition du terme *Digital humanities*. Ces travaux entrepris en 2015 par une étudiante en L3 d'Information et Communication, aboutissent à un rapport de stage[23] sur le rôle des bibliothèques universitaires dans la

co-construction du savoir dans les Humanités Numériques<sup>1</sup>. Y sont mentionnés les principaux centres d'Humanités Numériques et institutions en Europe, aux États-Unis, dans le monde avec une focale sur quelques projets français emblématiques. On y prend connaissance de l'impact des *Digital Studies* et de l'OpenEdition en faveur des Humanités numériques en bibliothèques universitaires.

### Exploitation, appropriation et exploration d'un environnement sémantique par les chercheurs

D'abord utilisé comme un système d'information d'écriture dans un réseau d'hypertextes en utilisant des modèles et des relations sémantiques [21], ce même environnement permet une seconde approche en 2015 pour l'exploration de corpus via un système d'information orienté exploration et curation de données. La compatibilité entre ontologies Wiki et règles de curation serveurs[12] contribue à la mise en œuvre d'explorateur de corpus. A titre d'exemple, le serveur d'exploration sur les humanités numériques[25] ramène après interrogation dans un contexte hétérogène, les pays qui travaillent sur les Humanités Numériques dans le monde par un mécanisme d'identification géographique[4] à partir des mentions d'Universités.

### Les humanités numériques comme champ d'investigation pour l'innovation numérique

Ces 2 contributions, impliquant des membres de la cellule Innovation Partenariale de l'Université de Lorraine, aboutissent en 2015 à l'organisation d'un *brunch* qui a vocation à déployer des programmes pour renforcer et accroître l'impact économique et social de l'Université de Lorraine. Il est articulé autour de trois axes forts : les mutations industrielles, l'entrepreneuriat et les transformations sociétales. Le Brunch Humanités Numériques, conduit par la cellule Innovation Partenariale, rassemble alors des partenaires du Centre Dramatique National Nancy-Lorraine, du laboratoire ATILF – Lorraine, de Nancy Numérique, de la pédagogie numérique de l'université et des bibliothèques universitaires<sup>2</sup>. Les personnes en présence constituent un réseau de partenaires pour réfléchir en 2016 à l'ouverture d'un parcours transversal en L2. Le groupe propose une 1ère maquette d'enseignements en Humanités Numériques en 2017 qui évolue à partir de 2018 vers des Unités d'Enseignement Humanités Numériques du L2 au Master. Dans cette deuxième version, la bibliothèque universitaire de lettres, sciences humaines et sociales, prend une part active pour le montage des programmes et des enseignements autour du patrimoine numérique.

## 2.2 Problématique posée pour la construction d'un entrepôt numérique orienté acteurs

Il s'agit d'explicitier comment dépasser le cadre de définitions des *Digital Humanities* au moyen d'un raisonnement à base d'une résolution de problèmes. Il s'agit d'en faire émerger des concepts et poser des hypothèses, afin d'évoluer vers une herméneutique[17] des Humanités Numériques. Nous observons que de nombreuses bibliothèques numériques ne dialoguent pas entre elles : elles répondent à des logiques de lieux, de politiques ou bien encore de réseaux. Quels

sont les points de convergence entre les acteurs pour la construction d'un entrepôt numérique commun ? Cette problématique est dans la lignée d'un article présenté lors du colloque VSST'2013[27] : Hétérogénéité des sources de données et entrepôts numériques collaboratifs universitaires[22].

## 2.3 Modélisation des acteurs

Lors de nos travaux précédents nous avons présenté et validé un modèle de représentation des acteurs en situation de recherche d'informations qui a fait l'objet d'un article[20]. Il nous permet d'aborder les Humanités Numériques à partir de la représentation de l'utilisateur dans un système d'information numérique :

RU	Représentation de l'utilisateur
(T)	Type d'acteurs
(B)	Besoins des acteurs
(F)	Fonctions des acteurs
(A)	Activités des acteurs

Table 1: Représentation de l'utilisateur :  $RU = (T, B, F, A)$

Dans un contexte universitaire[8], observons les types d'acteurs impliqués dans la construction d'un entrepôt numérique à savoir les enseignants-chercheurs, les professionnels de l'Information Scientifique et Technique et les informaticiens.

## 3 PHASE PRÉPARATOIRE AUX HUMANITÉS NUMÉRIQUES

### 3.1 Enseignant-chercheur

Dans notre raisonnement, il s'agit de trouver comment atteindre l'utilisateur final «enseignant-chercheur» pour lui permettre de travailler dans une situation optimale en s'appuyant sur les Humanités Numériques. Un travail remarquable a été entrepris par un étudiant[24] en deuxième année de Master de Lettres Modernes, parcours Métiers du livre pour une application d'Istex à la philologie dont un extrait explicite la démarche :

*«LorExplor se propose de constituer une bibliothèque Open Source de composants XML pour construire des systèmes de recherche, des serveurs d'exploration de corpus et des chaînes de curation de données. L'objectif est de donner les moyens aux documentalistes de construire des plates-formes intermédiaires analysant de 100 000 à 1 000 000 de documents pour répondre à des besoins spécifiques»*

Ces besoins spécifiques étaient notamment la prise de connaissance d'un domaine par les enseignants-chercheurs, pour entamer des travaux de recherche.

Le développement d'un serveur d'exploration sur la philologie[28] a permis d'effectuer une analyse des acteurs de ce domaine par exploration de corpus, et de l'affinage des résultats par filtrage informatique, en développant les compétences associées.

#### Concept 1 - construction des savoirs

Naît de cette expérimentation un 1er concept, qui est pour le système d'information la construction des savoirs. Les spécialistes de l'Information Scientifique et Technique ou bibliothécaires travaillent avec une communauté de chercheurs hors les murs amenés à explorer et exploiter un système d'information numérique.

<sup>1</sup>Une synthèse est proposée dans une plateforme expérimentale au service de la recherche sous la forme d'un wiki sémantique : Les humanités numériques par définitions[31].

<sup>2</sup>On peut le retrouver en ligne sous l'intitulé : Participation au Brunch - Humanités numériques[9].

### 3.2 Professionnels de l'Information Scientifique et Technique

Les professionnels de l'Information Scientifique et Technique sont à la tête d'un patrimoine universitaire à gérer. Leurs fonctions touchent à la description de ce patrimoine et à sa valorisation. Les décisionnaires doivent prendre en compte principaux besoins des utilisateurs finaux :

- Le repérage de sources et de données,
- Le traitement des données,
- La constitution de bases de données[16],
- L'écriture recourant à des cartes heuristiques,
- L'écriture collaborative,
- La communication scientifique<sup>3</sup>.

Les humanités en lettres, sciences humaines et sociales constituent un milieu numérique de bout en bout qualifié d'*e-science* à toutes les étapes du travail de chercheur. Les humanités numériques décloisonnent les disciplines en suscitant des échanges sur les technologies utilisées en commun sur le principe des *cultural studies*[3].

#### Concept 2 - exploitation des savoirs

Par une approche transdisciplinaire, apparaît un 2ème concept à prendre en compte pour l'utilisateur final en amont de la conception du système d'information à savoir, une mise en œuvre pour la faisabilité de l'exploitation des savoirs. Ce processus est rendu possible par l'investissement des informaticiens.

### 3.3 Informaticiens

Les informaticiens travaillent avec une communauté de partenaires spécialisés dans l'ingénierie computationnelle. Ils sont amenés à être les architectes du Système d'Information Numérique. Comment les informaticiens recourent-ils à l'ingénierie computationnelle pour la mise en œuvre d'un système d'information numérique ? Rappelons qu'ils doivent avoir une parfaite maîtrise des besoins, des fonctions et des activités des 3 types d'acteurs identifiés dans notre étude. En effet, les enseignants-chercheurs travaillent avec la communauté des étudiants. Ils sont en situation d'exploration d'informations dans le cadre de leur recherche. Cette information compilée est mise à disposition par les moyens des supports d'Information Scientifique et Technique (IST). Le chaînon informatique favorise une capitalisation de ces connaissances et leur transfert dans un cadre pédagogique.

#### Concept 3 – mise à disposition des savoirs

Les 3 types d'acteurs sont reliés par les savoirs, d'où émerge un 3ème concept qui touche à la mise à disposition des savoirs sous forme numérique.

### 3.4 Phase 1 du raisonnement conceptuel

Ces 3 concepts permettent de dessiner un schéma conceptuel d'où émerge une hypothèse.

- C1 Construction des savoirs
- C2 Exploitation des savoirs
- C3 Mise à disposition des savoirs

Les relations mettent en relief la transformation des savoirs en connaissances au travers d'une dynamique entre acteurs dont les actions sont corrélées par la satisfaction des contributeurs.

- Rel. 1 {C1, C2} : Capitalisation des connaissances
- Rel. 2 {C2, C3} : Description valorisation d'un patrimoine
- Rel. 3 {C3, C1} : Architecture d'un système d'information numérique

#### Hypothèse I - contributeurs des savoirs

Nous posons notre hypothèse I : la construction d'un entrepôt numérique collaboratif doit prendre en compte les fonctions des chercheurs, des professionnels de l'information Scientifique et Technique et des informaticiens car ils sont tous 3 les contributeurs des savoirs comme le représente le schéma suivant «Phase préparatoire aux Humanités Numériques».

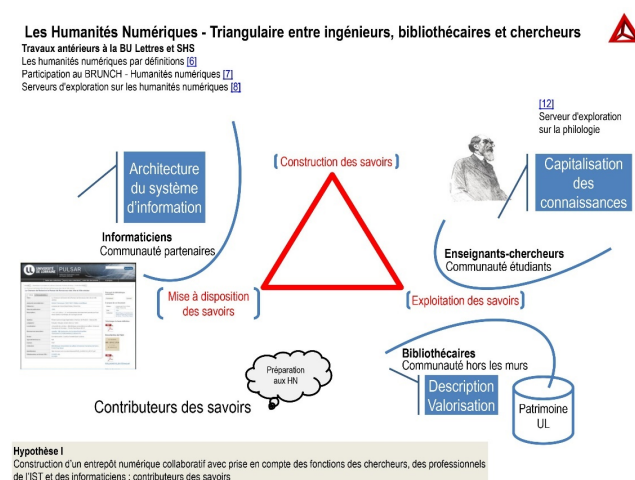


Figure 1: Phase préparatoire aux Humanités Numériques

## 4 ARTICULATION AUX HUMANITÉS NUMÉRIQUES

Cette seconde partie s'appuie sur l'article[22] présenté lors du colloque VSST'2013 et qui présentait un cas pratique de numérisation dans un environnement classique Codex. Nous y avons étudié le traitement des métadonnées et l'efficacité des données dans le cadre de leur ubiquité. Partant de l'objet numérique, nous démontrions l'intérêt de raisonner en termes de collections numériques universitaires en les contextualisant. L'exemple choisi est tiré d'une mise en œuvre d'un entrepôt tenant compte des consignes techniques de la Bibliothèque Nationale de France pour proposer une plateforme ouverte<sup>4</sup>. Les objets numérisés y sont accessibles sur une plateforme nommée PULSAR (Patrimoine Universitaire Lorrain des Savoirs ARchivés[22]) et dans un même temps moissonnés via Gallica<sup>5</sup>.

<sup>3</sup>Sous les formes de blogs, carnets de recherche, dépôt dans une archive ouverte, présentation à l'aide de divers logiciels ...

<sup>4</sup>Voir <https://omeka.org/>

<sup>5</sup>Voir <https://gallica.bnf.fr/edit/und/bibliotheques-de-luniversite-de-lorraine>

#### 4.1 Orientation services usagers

Les enseignants ont des besoins à prendre en compte en amont des réflexions par les concepteurs autour des services. Ces services concernent :

- La recherche d'information via les métadonnées au travers des catalogues (SUDOC, Ulysse, PULSAR, Gallica...),
- La recherche en texte intégral,
- La recherche par identifiant DOI,
- La récupération et la citabilité par les outils à implémenter et paramétrer dans les systèmes d'information pour permettre l'ajout de plugins par l'utilisateur.

Les services autour des informations font apparaître la notion d'ubiquité des données facilitatrices du travail d'exploration des connaissances pour l'acteur enseignant-chercheur.

##### Concept 1 – édition ouverte du savoir

Les besoins exprimés par les utilisateurs sont:

- la navigation dans le document,
- les recherches en plein texte,
- le guidage par des sommaires enrichis,
- des notices élaborées,
- la récupérabilité des métadonnées dans des outils appropriés pour en permettre leur citation.

Tous ces aspects sont autant de paramètres à prendre en compte pour la construction de plateformes d'édition ouverte du savoir.

#### 4.2 Rapprochement des services

Rendre exploitable et appropriable un entrepôt numérique par les chercheurs constitue un paramètre à prendre en compte par les professionnels de l'IST pour le rapprochement de services à partir des objets numériques qui composent un fonds patrimonial numérisé. La visibilité et la valorisation œuvrent au rapprochement des services. Ce rapprochement est rendu possible par :

- L'enrichissement des données,
- Le traitement des données selon les normes et standards comme par exemple l'encodage des données selon différents formats de description (Dublin Core, EAD, TEI),
- La réutilisabilité des données.

**Concept 2 – Système ouvert du patrimoine** Le rapprochement des services rend l'interopérabilité des données et des systèmes ouverts patrimoniaux impérative. C'est une condition nécessaire au dialogue entre systèmes d'information numérique. La construction de réseaux de portails (ORI-OAI) participe, selon les acteurs à :

- l'apprentissage cognitif,
- l'agrégation des informations,
- la dématérialisation des données et des programmes,
- la valorisation du patrimoine numérique.

#### 4.3 Rapprochement des objets numériques

L'appropriation des usages et pratiques des enseignants-chercheurs et des professionnels de l'Information Scientifique et Technique par les informaticiens constitue un gage de réussite pour le développement d'un système d'information numérique. Un audit pose les différentes phases d'un processus de digitalisation pour dessiner

un *workflow* qui offre une focale pour repérer les possibles automatisations de certaines étapes. L'ingénierie computationnelle prend place à ces différents niveaux d'automatisation dont on peut citer :

- L'automatisation de la digitalisation,
- L'automatisation de l'OCRisation,
- L'automatisation de l'intégration des métadonnées dans l'objet numérique,
- L'automatisation des exports/imports des notices décrites selon différents formats : txt, csv, xml.

Le rapprochement des objets numériques fait apparaître que la maîtrise des sources ouvertes et de leur protocole participe à la construction d'un entrepôt d'objets numériques. Un entrepôt OAI-PMH[15] permet l'exposition et le moissonnage des données.

##### Concept 3 – Écosystème d'information ouvert

La garantie de l'édition ouverte du savoir par les enseignants-chercheurs associés au travail des professionnels de l'IST favorise la construction d'un système patrimonial ouvert. Les informaticiens s'emploient à mettre en œuvre un écosystème d'information ouvert, 3ème concept de réflexion autour des systèmes d'information et de leur articulation aux Humanités Numériques.

#### 4.4 Phase 2 du raisonnement conceptuel

Les concepts reliés entre eux au travers de la Figure 2, permettent d'aboutir à une seconde question hypothétique.

- C1 Édition ouverte du savoir
- C2 Système ouvert du patrimoine
- C3 Écosystème d'information ouvert

Les relations entre les différents concepts font apparaître la notion de standards ouverts[6] pour la construction des informations numériques.

- Rel. 1 {C1, C2} : Ubiquité des informations numériques
- Rel. 2 {C2, C3} : Interopérabilité des systèmes d'information numériques
- Rel. 3 {C3, C1} : Sources ouvertes du patrimoine numérique

##### Hypothèse II

Par déduction, la seconde hypothèse touche au rôle des standards ouverts[5] dont les Humanités Numériques constituent la cheville ouvrière d'un entrepôt numérique orienté services, comme le représente la Figure 2 :

#### 5 AU CŒUR DES HUMANITÉS NUMÉRIQUES

Une illustration par un retour d'expérience permet de mettre en relief les différentes phases théoriques explicitées précédemment. Elles reposent sur un raisonnement à partir des usages, des pratiques et des objectifs des 3 acteurs identifiés en amont de la construction d'un système d'information numérique. Un travail de recherche sur plusieurs années a été entrepris autour d'un fonds spécifique d'ouvrages. Le Fonds Paul Meyer[10], du nom de son donateur est la propriété de la bibliothèque universitaire de lettres, sciences humaines et sociales de l'Université de Lorraine. Paul Meyer (1840-1917) était un philologue, ancien directeur de l'École des Chartes, qui travaillait plus particulièrement sur l'histoire romane. Les travaux de recherche sont menés par des protagonistes de l'Université de Lorraine représentés par des chercheurs de l'ATILF<sup>6</sup>,

<sup>6</sup>Voir <https://www.atilf.fr/>

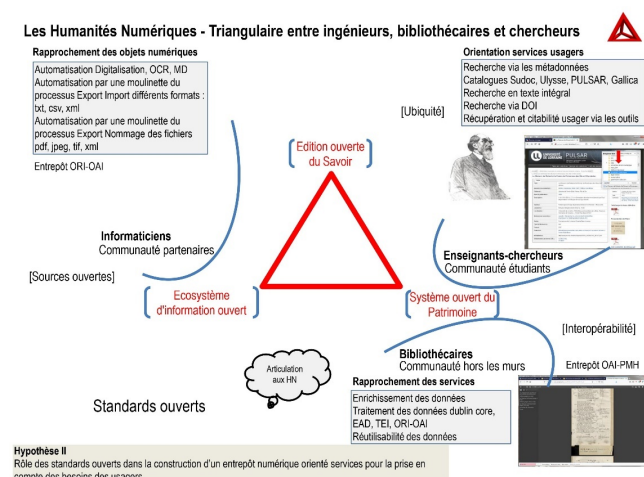


Figure 2: Articulation aux Humanités Numériques

des étudiants du Master en Lettres Modernes - parcours Métiers du livre, des informaticiens du LORIA<sup>7</sup> et des professionnels de l'IST de la Direction de la Documentation et de l'Édition<sup>8</sup>. Les 1ères recherches tournaient autour de la valeur ajoutée, tirée de la description d'un patrimoine universitaire remarquable[11] dans un environnement sémantique.

## 5.1 De l'écriture sémantique au serveur d'exploration

Ces mêmes protagonistes se posent la question de savoir comment un système informatique est à même de fournir des résultats intéressants pour la recherche. Il désirent penser leur insertion dans une infrastructure de connaissance en ligne. D'où l'idée d'utiliser le même espace MediaWiki pour y développer des serveurs d'exploration qui sont des systèmes ouverts. La flexibilité de ce type de serveur leur permet de s'adapter aux spécificités de chaque corpus hébergé. Cela ouvre de nombreuses possibilités quant au traitement des informations que l'on peut en extraire. Par exemple le serveur d'exploration sur la Chanson de Roland[29] intéresse en tout premier lieu les enseignants-chercheurs pour découvrir sur la toile les communautés de chercheurs qui s'y intéressent, et découvrir leurs angles d'approche.

### Concept 1 – Données de la recherche

Dans notre contexte, les activités des chercheurs sont rendues possibles via un concept relatif aux données de la recherche qui permet :

- L'appropriation de la plateforme par les chercheurs,
- L'exploration de corpus,
- La réflexion sur la transposition de la comparaison du texte initial avec des éditions multiples,
- L'exploitation des annotations pour la construction d'un texte intelligible.

<sup>7</sup>Voir <https://www.loria.fr/fr/>

<sup>8</sup>Voir <https://bu.univ-lorraine.fr/>

## 5.2 Description sémantique d'un fonds patrimonial

Depuis 2010, la bibliothèque universitaire de lettres, sciences humaines et sociales en lien avec le LORIA, développe un MediaWiki approprié par les professionnels de l'information scientifique et technique et les enseignants-chercheurs pour y décrire des collections patrimoniales en s'appuyant sur des modèles de descriptions au service d'une écriture sémantique. C'est par exemple le cas dans la description du Fonds Paul Meyer[10]. L'utilisation de l'outil Wiki sémantique rend visible les travaux de Paul Meyer et a une grande portée dans le monde de la recherche.

### Concept 2 – Ubiquité des données

L'outil wiki fonctionne avec une syntaxe particulière, fondée sur l'utilisation de codage et de balises des données. Cela permet des mises en pages travaillées intégrant tableaux et images. Le wiki sémantique permet de créer des liens de sens entre différents mots qui constitue le processus de sémantisation. Cela permet par exemple de procéder à des requêtes sémantiques pour compter le nombre d'occurrences d'un mot ou bien encore de regrouper des variantes orthographiques de mots. Notre environnement sous forme de média wiki fait apparaître un concept relatif à l'ubiquité des données pour :

- La conception et l'exploitation des données,
- La réutilisabilité des données dans les modèles de description,
- La construction d'un Web sémantique grâce au langage RDF[14].

## 5.3 Du codex à la sémantisation : exploitation d'annotations

### La sémantisation de la Chanson de Roland[18]

Paul Meyer, lors de son exercice en qualité de directeur de l'École des Chartes, entreprit des travaux non achevés sur la Chanson de Roland. Ses recherches consistaient à en comparer différentes éditions au manuscrit d'Oxford[30]. Il souhaitait *in fine* en proposer une version au plus proche du texte initial en annotant un exemplaire d'une édition de 1869 par Francisque Michel. Il décéda sans pouvoir aboutir ce travail colossal. La bibliothèque universitaire de lettres, sciences humaines et sociales, en lien avec des chercheurs de 2 laboratoires<sup>9</sup> propose de relever le défi de finaliser le travail. Il s'agit d'offrir une version de la Chanson de Roland avec l'exploitation posthume des annotations de Paul Meyer. Ce travail s'appuie sur une palette d'outils offerts par les Humanités Numériques pour montrer le résultat attendu par Paul Meyer.

Rappelons, que la plateforme PULSAR [Patrimoine Universitaire Lorrain des Savoirs ARchivés] met à disposition des objets numériques, dont La Chanson de Roland. De plus, l'exposition de ses métadonnées en permet le moissonnage via Gallica dans un environnement codex. L'idée originale est de tirer parti de cette version digitalisée de La Chanson de Roland pour la porter dans un MediaWiki dédié à l'ingénierie des connaissances pour les communautés de la recherche et de l'innovation.

### Concept 3 – Interopérabilité des données

<sup>9</sup>L'ATILF, pour le travail culturel sur le texte et le LORIA, pour le travail sur le traitement informatique du texte.



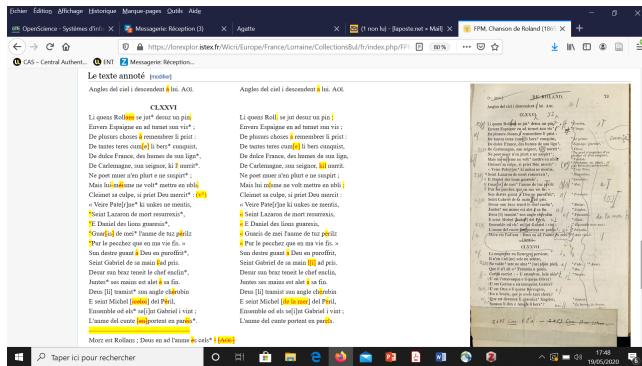


Figure 3: Sémantisation de La Chanson de Roland

Ce travail consiste à reproduire sur le wiki sémantique, le texte manuscrit de Paul Meyer pour la transcription de ses annotations. Lors de la phase de digitalisation de la version imprimée, les ajouts manuscrits marginaux ou sur papier collé à la page ou bien encore les paperolles, se sont révélés indéchiffrables pour l'informatique. L'anticipation a constitué le cœur de l'innovation à ce niveau par l'interopérabilité des données pour un travail *a posteriori* rendu possible grâce à l'export des pages digitalisées en différents formats comme par exemple :

- Le passage du format pdf en xml,
- L'exploitation du format xml par un script pour la localisation des mots dans la page,
- La mise en œuvre automatique de la mise en page sémantique identique au texte initial.

#### 5.4 Phase 3 du raisonnement conceptuel

Ces étapes témoignent d'une progression des possibilités offertes par un espace sémantique. Cela permet de passer d'un processus de description à l'exploration de corpus, pour passer du codex à la sémantisation au moyen d'une palette d'outils offerts par les Humanités Numériques. Par un récapitulatif des concepts, nous posons notre Hypothèse III :

- C1 Données de la recherche
- C2 Ubiquité des données
- C3 Interopérabilité des données

Les relations mettent en relief la transformation des savoirs en connaissances mise au travers d'une dynamique entre acteurs dont les actions sont corrélées par la satisfaction des contributeurs.

- Rel. 1 {C1, C2} Exploration des données de la recherche
- Rel. 2 {C2, C3} Sémantisation d'un fonds de recherche
- Rel. 3 {C3, C1} Objet numérique augmenté

#### Hypothèse III – Écriture numérique

La construction d'un entrepôt numérique s'appuie sur une écriture numérique partagée au service de l'intelligence collective que nous représentons en figure 4.

### 6 RÉOLUTION DE LA PROBLÉMATIQUE

Notre présentation a porté sur le rôle des acteurs dans le projet d'entrepôts numériques universitaires en le fondant sur le cas pratique d'un objet numérique. Par les concepts et les hypothèses qui

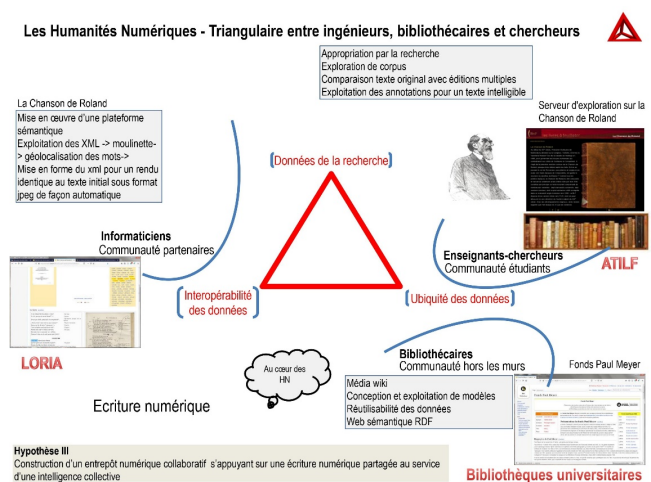


Figure 4: Au cœur des Humanités Numériques

en découlent, nous avons abordé les compétences nécessaires des acteurs enseignants-chercheurs, des professionnels de l'IST et des informaticiens pour la numérisation d'un fonds ancien. Les travaux de recherche sur l'objet numérique «La Chanson de Roland» ont montré une numérisation augmentée du texte, par un processus de sémantisation, via un Wikimédia. La plateforme collaborative a été appropriée par les chercheurs, rendue exploitable par les spécialistes de l'IST et construite par les informaticiens. Partant d'une problématique, qui pose la question de l'hétérogénéité des bibliothèques numériques sur la toile, notre raisonnement aboutit à des points de convergences. Ces derniers sont le fait de l'expertise métier propre aux acteurs et les **Compétences Communes** par le biais des données, des savoirs et des systèmes ouverts pour la **Cohérence** d'un espace numérique. Nous le représentons par l'acronyme  $C^3$  pour les 3 types de compétences identifiées :

$$C^3 \Leftrightarrow \text{Compétences}^3$$

- (1) Compétences pour la transformation des données en savoirs par l'acteur chercheur
- (2) Compétences pour une vision globale du patrimoine et du numérique par l'acteur spécialiste de l'IST
- (3) Compétences pour une maîtrise des systèmes complexes par l'acteur informaticien

Le schéma en Figure 5 fait une synthèse des concepts émergeant à chaque étape du processus de construction d'un entrepôt numérique par type d'acteurs et de de compétences nécessaires.

### 7 CONCLUSION ET PERSPECTIVES

Cet article a présenté les nouvelles compétences à acquérir au travers d'un cas pratique autour du patrimoine universitaire digitalisé, de l'écriture numérique et de l'édition numérique. Notre objet de recherche «La Chanson de Roland» témoigne, pour la mise en œuvre d'un entrepôt numérique dans un environnement sémantique, d'une porosité entre lesdites compétences des différents acteurs, qu'ils soient chercheurs, informaticiens ou designers de l'IST. Tous mettent en commun leurs compétences pour bâtir des projets dans le domaine des Humanités Numériques ; une coopération à l'œuvre

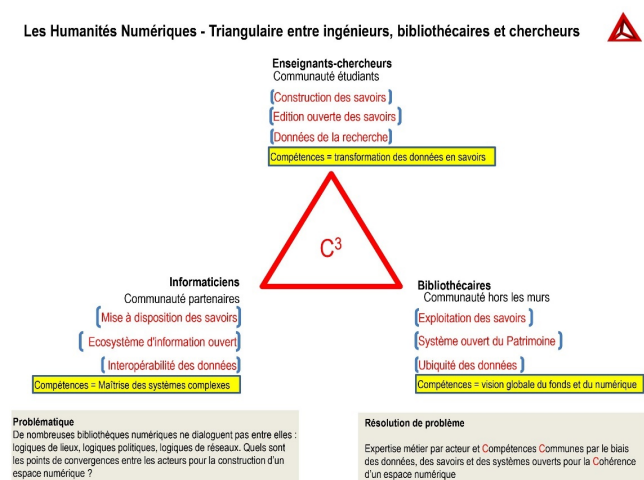


Figure 5: Résolution de la problématique

tant dans les projets de recherche, dans les projets de numérisation des patrimoines universitaires et dans l'enseignement. Par ce retour d'expérience, nous avons voulu montrer la portée de l'évolution des outils computationnels qui contribuent au développement des Humanités Numériques. Une véritable révolution voit le jour dans les nouveaux modes d'enseignement et dans la conduite de la recherche mise en regard avec les compétences à acquérir. Dans la Figure 6 nous présentons la poursuite de nos recherches par la génération d'une partition de musique[26] de La Chanson de Roland intégrée au Wikimédia[2]. Notre espace expérimental touche la communauté des chercheurs en musicologie et place l'étudiant dans une situation d'apprentissage. Ce dernier intervient directement dans l'environnement sémantique[7] à partir de LilyPond<sup>10</sup>, outil puissant et flexible, adapté aux systèmes de notation. Il apprend à utiliser les modèles MediaWiki, c'est à dire à formuler ainsi le résultat attendu pour placer les mots et les notes aux bons endroits. Nos

Figure 6: Sémantisation d'une partition de musique de La Chanson de Roland

perspectives de recherche touchent à des questions autour de l'enseignement des Humanités Numériques. Cela permet d'offrir aux étudiants la possibilité de s'investir dans la recherche autour des enjeux herméneutiques des Humanités Numériques grâce à des plateformes expérimentales.

<sup>10</sup>Voir <https://lilypond.org/>

## ACKNOWLEDGEMENTS

Merci à

- Isabelle Turcan, Professor of French Literature at the University of Lorraine for sharing her research work on the Paul Meyer Fund
- Jacques Ducloy, engineer emeritus at the University of Lorraine for the design and availability of a network of semantic wikis
- Anne-Pascale Parret, Director of Direction and Documentation for supporting the organization of a day around Digital Humanities as part of research week at the University of Lorraine in January 2020: Working with Digital Humanities[19].

## REFERENCES

- Aurélien Berra. 2015. Pour une histoire des humanités numériques. *Critique* 819-820, 8 (2015), 613. <http://www.cairn.info/revue-critique-2015-8-page-613.htm>
- Maurice Bouchor. 2020. La chanson de Roland, chanson (Bouchor) — Wicri Musique. (2020). [https://loexplor.istex.fr/Wicri/Musique/fr/index.php?title=La\\_chanson\\_de\\_Roland\\_chanson\\_\(Bouchor\)](https://loexplor.istex.fr/Wicri/Musique/fr/index.php?title=La_chanson_de_Roland_chanson_(Bouchor))
- Maxime Cervulle, Jade Lindgaard, Éric Macé, Éric Maigret, Angela McRobbie, David Morley, and Éric Neveu. 2014. *Cultural Studies : Genèse, objets, traductions*. Éditions de la Bibliothèque publique d'information, Paris. <http://books.openedition.org/bibpompidou/1612>
- Ghislaine Chartron. 2018. Géopolitique de l'open access. In *International Colloquium on Open Access*. Moroccan Institute of Scientific and Technical Information, Rabat, Morocco, 7–21.
- Ghislaine Chartron and Joachim Schöpfel. 2017. Dossier Open access et Open science en débat. *Revue française des sciences de l'information et de la communication*, 11, Société Française des Sciences de l'Information et de la Communication. (2017). <https://journals.openedition.org/rfsic/2868>.
- Ghislaine Chartron and Joachim Schöpfel. 2017. Libre accès aux publications et sciences ouvertes en débat. *Revue française des sciences de l'information et de la communication* 2017, 11, Article 1 (2017), 10 pages. <https://journals.openedition.org/rfsic/3331>
- Stéphane Chaudiron, Bernard Jacquemin, and Eric Kergosien. 2018. L'apport du web sémantique à la sauvegarde du patrimoine immatériel. Les cas du textile, de la musique et de la mine. In *HumaNum. Information, communication et humanités numériques. Enjeux et défis pour un enrichissement épistémologique. Actes du 23e colloque bilatéral franco-roumain en Sciences de l'information et de la communication*. Accent, Cluj-Napoca, Romania, 311–328. <https://hal.archives-ouvertes.fr/hal-02356933>
- Marin Dacos and Pierre Mounier. 2015. *Humanités numériques*. Research Report. Institut français. 9782354761097 pages. <https://hal.archives-ouvertes.fr/hal-01228945>
- Université de Lorraine. 2015. Les Brunchs : édition 2014/15. (2015). <http://www.univ-lorraine.fr/content/les-brunchs-edition-2014>
- Université de Lorraine. 2019. Fonds Paul Meyer — Collections Bul. (2019). [https://loexplor.istex.fr/Wicri/Europe/France/Lorraine/CollectionsBul/fr/index.php?title=Fonds\\_Paul\\_Meyer](https://loexplor.istex.fr/Wicri/Europe/France/Lorraine/CollectionsBul/fr/index.php?title=Fonds_Paul_Meyer)
- Dictionnaire du Moyen Français. 2018. DMF - Site du laboratoire ATILF. (2018). <http://www.atilf.fr/spip.php?rubrique174&idfirst=739>
- Jacques Ducloy. 2020. Systèmes d'information encyclopédiques édités par les scientifiques. (March 2020). <https://www.openscience.fr/Systèmes-d-information-encyclopediques-edites-par-les-scientifiques>
- Dinah Galligo. 2016. Humanités numériques : nouvelle discipline ou méthodologie des Sciences humaines et sociales ? (June 2016). [http://blogs.sciences-po.fr/prospectibles/2016/06/08/Prospectibles\\_blog\\_de\\_la\\_Cellule\\_Pro prospective\\_de\\_la\\_Bibliothèque\\_de\\_Sciences-Po](http://blogs.sciences-po.fr/prospectibles/2016/06/08/Prospectibles_blog_de_la_Cellule_Pro prospective_de_la_Bibliothèque_de_Sciences-Po)
- RDF Working Group. 2014. Resource Description Framework 1.1 Semantics. (2014). <http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/> Page Version ID: 169943370.
- Open Archives Initiative. 2017. Open Archives Initiative Protocol for Metadata Harvesting. (2017). <https://www.openarchives.org/pmh/>
- Gérald Kembellec, Orélie Desfriches-Doria, and Marie Gisbert. 2020. Bibliographies de Critiques d'art francophones — Un projet interdisciplinaire d'Humanités Numériques. *Revue ouverte d'ingénierie des systèmes d'information* 1, 3, Article 4 (June 2020), 22 pages. <https://www.openscience.fr/Bibliographies-de-Critiques-d-art-francophones-Un-projet-interdisciplinaire-d>
- Jean-Guy Meunier. 2017. Humanités numériques et modélisation scientifique. *Questions de communication* 2017, 31 (Sept. 2017), 19–48. <http://journals.openedition.org/questionsdecommunication/11040>



- [18] Francisque Michel. 2017. Chanson de Roland. In *fonds Paul Meyer*. BUL Collection, INIST. [https://loexplor.istex.fr/udl-bul-col.fr/index.php/Chanson\\_de\\_Roland](https://loexplor.istex.fr/udl-bul-col.fr/index.php/Chanson_de_Roland)
- [19] Frédérique Peguiron. 2020. *Retour sur la Journée d'Etude : Travailler avec les Humanités Numériques*. Technical Report. University of Lorraine. 13 pages.
- [20] Frédérique Peguiron and Odile Thiery. 2004. Modéliser l'acteur dans le système d'information stratégique d'une université. (2004), 13 pages.
- [21] Frédérique Peguiron and Odile Thiery. 2007. Analyses multidimensionnelles de contenus documentaires dans un ENT universitaire au service de l'acteur enseignant-chercheur. In *Actes du 10<sup>ème</sup> Colloque International sur le Document Electronique*. Europia, Rabat, Maroc, n.p. <http://odel.irevues.inist.fr/cide/?id=156>
- [22] Frédérique Peguiron and Odile Thiery. 2013. Hétérogénéité des sources de données et entrepôts collaboratifs universitaires. In *Veille Stratégique Scientifique et Technologique*. IRIT, Nancy, France, 19.
- [23] Claire Sinigaglia. 2015. *Quelle est l'innovation liée au rôle des bibliothèques universitaires dans la co-construction du savoir dans les humanités numériques ?* Stage réalisé à la Bibliothèque universitaire du Campus lettres et sciences humaines de Nancy (54), s.d. Frédérique Peguiron, responsable de la Bibliothèque universitaire. Bibliothèque universitaire de lettres sciences humaines et sociales, Nancy.
- [24] Miklós Takach. 2015. *Application d'Istex à la philologie*. Stage réalisé à la Bibliothèque universitaire du Campus lettres et sciences humaines de Nancy (54), Sous la tutelle de Frédérique Peguiron, responsable de la Bibliothèque universitaire. Bibliothèque universitaire de lettres sciences humaines et sociales, Nancy.
- [25] Wicri Ticri. 2015. Serveurs d'exploration sur les humanités numériques. (2015). [https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Humanit%C3%A9s\\_num%C3%A9riques](https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Humanit%C3%A9s_num%C3%A9riques)
- [26] Julien Tiersot. 1909. *Chants populaires pour les écoles. Poésies de Maurice Bouchor. Mélodies recueillies et notées par Julien Tiersot*. Hachette, Internet Archive, San Francisco, CA. <http://archive.org/details/smchantspopulaires00tier>
- [27] VSST'13 2013. VSSTColloque International V.S.S.T.'2013, Veille Stratégique Scientifique et Technologique, Nancy, 23 - 25 octobre 2013. (2013). <https://atlas.irit.fr/PIE/VSSST/VSSST-2013.html>
- [28] Wicri. 2015. Serveur d'exploration sur la philologie — Wicri Linguistique. (2015). [https://loexplor.istex.fr/Wicri/Linguistique/fr/index.php/Serveur\\_d\\_exploration\\_sur\\_la\\_philologie](https://loexplor.istex.fr/Wicri/Linguistique/fr/index.php/Serveur_d_exploration_sur_la_philologie)
- [29] Wicri. 2016. Serveur d'exploration sur la Chanson de Roland — Wicri France. (2016). [https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Serveur\\_dexploration\\_sur\\_la\\_Chanson\\_de\\_Roland](https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Serveur_dexploration_sur_la_Chanson_de_Roland)
- [30] Wikisource. 2016. *Manuscrit d'Oxford*. Wikimedia, San Francisco. [https://fr.wikisource.org/wiki/La\\_Chanson\\_de\\_Roland/Manuscrit\\_d%E2%80%99Oxford](https://fr.wikisource.org/wiki/La_Chanson_de_Roland/Manuscrit_d%E2%80%99Oxford)
- [31] Wikri. 2015. Humanités numériques — Wicri France. (2015). [https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Humanit%C3%A9s\\_num%C3%A9riques](https://loexplor.istex.fr/Wicri/Europe/France/fr/index.php/Humanit%C3%A9s_num%C3%A9riques)