



**HAL**  
open science

# HETEROGENEITE DES SOURCES DE DONNEES ET ENTREPOTS NUMERIQUES COLLABORATIFS UNIVERSITAIRES

Frédérique Peguiron, Odile Thiery

► **To cite this version:**

Frédérique Peguiron, Odile Thiery. HETEROGENEITE DES SOURCES DE DONNEES ET ENTREPOTS NUMERIQUES COLLABORATIFS UNIVERSITAIRES. VSST 2013, Oct 2013, Nancy, France. hal-01418315

**HAL Id: hal-01418315**

**<https://hal.science/hal-01418315>**

Submitted on 16 Dec 2016

**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.

# HETEROGENEITE DES SOURCES DE DONNEES ET ENTREPOTS NUMERIQUES COLLABORATIFS UNIVERSITAIRES

Frédérique PEGUIRON (\*), Odile THIERY (\*\*)

[Frederique.peguiron@univ-lorraine.fr](mailto:Frederique.peguiron@univ-lorraine.fr), [odile.thiery@loria.fr](mailto:odile.thiery@loria.fr)

(\*) UNIVERSITE DE LORRAINE - Bibliothèque universitaire de lettres et science humaines, 46 Avenue de la Libération - BP 33408 54015 NANCY CEDEX

(\*\*) UNIVERSITE DE LORRAINE - LORIA, Campus Scientifique BP 239 54506 VANDOEUVRE CEDEX

**Mots clefs :** collections numériques, entrepôt documentaire, modélisation utilisateur, norme documentaire

**Keywords:** digital collections, documentary data warehouse, user modelling, documentation standard

**Palabras clave :** colecciones digitales, almacen de documentos, modelo de usuario, documentacion estandar

## Résumé :

Les universités abritent un nombre considérable de documents révélateurs de son patrimoine par le biais de ses établissements documentaires universitaires. L'utilisateur universitaire pour satisfaire un travail pédagogique ou de recherche passe par des outils et des services, qui se veulent de plus en plus numériques et qui touchent également à l'arrivée rapide de nouvelles technologies et services. Les ressources financières sont limitées et les exigences de qualité d'information sont précises et importantes.

C'est pourquoi nous faisons dans cet article un certain nombre de propositions pour participer à l'existant numérique et en favoriser une appropriation collective. Nous verrons dans un premier temps le contexte du document numérique et notre périmètre universitaire, puis nous précisons ce que nous entendons par « objet numérique » et introduisons notre domaine des entrepôts collaboratifs. Ensuite nous aborderons les collections numériques universitaires ; enfin nos solutions et nos perspectives de recherche.

# 1. Introduction

Les universités abritent un nombre considérable de documents révélateurs de son patrimoine par le biais de ses établissements documentaires universitaires. Les bibliothèques numériques fonctionnent sur un mode parallèle. L'utilisateur universitaire pour satisfaire un travail pédagogique ou de recherche passe par des outils et des services qui ne répondent pas aux mêmes impératifs économiques. Les ressources financières sont limitées et les exigences de qualité d'information sont précises et importantes.

Il en ressort une proposition de bases numériques qui ne se parlent pas entre elles ou qui si elles se parlent entre elles le font au sein de collectivités répondant à des contraintes financières. Cependant toute une partie de la connaissance, librement accessible dans des bibliothèques et affranchie des droits d'auteur, peut venir enrichir la palette des ouvrages numérisés. Toutefois l'entreprise d'un tel chantier pose de nombreuses questions. Comment par exemple garantir l'unicité de l'objet ouvrage numérique d'origine et en préciser toutes ses variantes et aborde aussi le problème de l'hétérogénéité des données à numériser et à archiver.

Ces questions touchent à l'existant. Elles touchent également à l'arrivée rapide de nouvelles technologies et services qui, non pris en compte, viendront eux aussi se juxtaposer à l'existant à défaut de ne pas en permettre leur intégration ultérieure. C'est pourquoi nous réfléchissons ici comment être force de proposition pour participer à l'existant numérique et en favoriser une appropriation collective. Dans cet article nous verrons dans un premier temps le contexte du document numérique et notre périmètre universitaire, puis nous précisons ce que nous entendons par « objet numérique » et introduisons notre domaine des entrepôts collaboratifs. Le paragraphe 3 parlera des collections numériques universitaires. Le 4 présentera nos solutions, le paragraphe 5 présentera la plateforme PULSAR (Patrimoine Universitaire Lorrain des Savoirs Archivés). Enfin nous concluons sur nos perspectives de recherche.

## 2. Définir le problème et le contexte

### 2.1 Conditions préalables du contexte

Compte tenu de l'ampleur et de l'importance d'un chantier de numérisation, il est nécessaire d'avoir recours à une méthode pragmatique. Pour mener nos réflexions nous nous appuyons sur un certain nombre d'études [17] faisant l'état de l'art de la numérisation pour l'enseignement et la recherche [12]. Au

début des années 2000 ces études portent sur la qualité de la numérisation, le besoin de recourir aux normes et standards et l'aspect sociétal des contenus et des usages.

Deux constats apparaissent :

- la numérisation est très contrainte quant aux droits d'auteur,
- l'offre des bibliothèques numériques [3] va en s'accroissant.

Plutôt que de considérer ces deux éléments comme des freins à l'avancée d'un projet d'offre numérique [13] il est intéressant d'en faire des atouts pour nos réflexions.

### **2.1.1 Périmètre universitaire**

Le rapport d'ISAAC [9], sur « L'université numérique » évoque les enjeux de la numérisation dans l'enseignement supérieur. La qualité de service doit favoriser la réussite des étudiants et leur insertion dans la société de la numérisation. Le périmètre universitaire posé nous engage à rappeler brièvement le paysage documentaire au niveau de l'enseignement supérieur et de la recherche. Les universités abritent en leur sein des services communs de la documentation, des bibliothèques associées, des bibliothèques d'UFR, des bibliothèques d'instituts et de recherche. Ces établissements documentaires répondent à des logiques de lieux et de réseaux. Ils ont tous en commun le signalement de leurs collections et le souci d'offrir des services pour améliorer la visibilité et la valorisation du patrimoine universitaire à un échelon régional, national et international.

### **2.1.2 Les enjeux**

Un des enjeux est la fiabilité quant à la qualité des données en faveur d'un travail académique pour un investissement collaboratif. Les bibliothèques ne rassemblent pas tous les fonds documentaires. Plutôt que de regrouper physiquement les collections imprimées en un seul lieu, aujourd'hui la numérisation constitue un élément de réflexion raisonné pour rapprocher des objets numériques et bâtir des collections virtuelles. La dérivation de notices d'un réservoir institutionnel comme le SUDOC propose des fonctionnalités orientées « professionnel ». L'appropriation de textes numérisés dans un contexte universitaire implique des fonctionnalités orientées « usager ». Il est nécessaire que les collaborateurs d'un entrepôt numérique puissent prendre en compte des

informations supplémentaires à décrire autour de l'objet numérique et faire ces descriptions dans un environnement adéquat. Deux notions apparaissent : la spécificité de l'objet numérique sous plusieurs facettes (contenant et contenu) et le contexte de la collection (historique et domaine).

### **2.1.3 Fiabilité scientifique**

La technique de présentation de ces données, qui dépassent celle d'une notice standard, doit permettre une collaboration ultérieure entre les partenaires académiques pour, s'ils le désirent, articuler le résultat de leur recherche directement au corps du texte. Nous nous appuyons sur notre article [15] pour insister sur la notion d'autorité qui entre en jeu pour la production de connaissances à intégrer dans les réseaux sémantiques [11] à accès contrôlé.

## **2.2 Définition de l'objet et du document numérique**

Nous nous focalisons, dans ce papier, sur le type de document « ouvrage » appartenant à un fonds ancien ou à un fonds spécifique comme nous l'explicitons dans le paragraphe consacré aux fonds spécifiques. L'ouvrage numérisé prend la dimension d'objet numérique. Il répond alors aux critères de document numérique sous l'angle du traitement, de la diffusion, de l'exploitation, de l'enrichissement de données, d'annotations et de citabilité<sup>1</sup>.

A titre d'exemple, dans le cadre de l'histoire du livre ancien, nous sommes confrontés à la prise en compte des variations des impressions pour un même ouvrage. En effet la norme Z 44-074<sup>2</sup> de catalogage des livres anciens implique la création d'une nouvelle notice pour la moindre différence d'impression pour une même édition d'un ouvrage imprimé. Dans ce cas, où et comment, lors de la numérisation, restituer ce type d'informations au travers de messages clairs et réutilisables ?

---

<sup>1</sup> Citabilité : favoriser la citation des références bibliographiques

<sup>2</sup> Association française de normalisation. (1986). *Catalogage des monographies anciennes: Rédaction de la description bibliographique : Z 44-074*. Paris: AFNOR.

L'analyse de la situation actuelle des bibliothèques numériques au travers de projets parallèles rend compte des évolutions souhaitables pour passer d'objets éparpillés et incohérents à des objets numériques consultables à distance et présentés dans le contexte de la diversité possible d'un seul et même ouvrage en intégrant les éditions successives, les indications de corrections des éditions, des impressions, les variations des post faces. Tous les éléments en marge de l'édition comme les annotations, les dédicaces, les ex-libris participent de l'histoire de livre en complément de sa genèse.

### **2.2.1 L'identifiant unique de l'objet numérique**

Dans notre cas l'unicité de la notice catalographique implique l'unicité de l'ouvrage numérisé. Nos propositions sont fondées sur un standard pris en compte dans les projets de numérisation des ouvrages. Il s'agit du DOI et de l'ISBN-A [8].

L'ISBN (International Standard Book Number) est un numéro d'identification à 13 chiffres largement utilisé dans le commerce international du livre depuis plus de 35 ans. Il est assigné à travers un réseau d'agences internationales d'enregistrement ISBN. L'ISBN est utilisé pour identifier chaque publication unique soit sous la forme d'un livre physique ou matériaux connexes, tels que des livres électroniques, logiciels, supports mixtes etc.

Le DOI (Digital Object Identifier) offre un identificateur permettant une action pour une utilisation sur les réseaux numériques : c'est à la fois le mécanisme de nommage des ressources et un protocole de résolution des identifiants en adresses plus concrètes.

L'ISBN-A «ISBN action» est un service édité par le DOI, dans lequel un numéro ISBN existant est exprimé dans le système DOI. Le modèle de métadonnées DOI permet la mise à disposition d'informations compatibles et complémentaires à celles prévues par le système ISBN telles que des informations bibliographiques dans la norme ONIX [6].

En incluant l'ISBN dans la syntaxe DOI de manière standard, un DOI est obtenu par dérivation d'un numéro ISBN. L'agence internationale de l'ISBN et l'Agence Internationale DOI Foundation ont convenu d'une procédure. L'ISBN-A est le premier exemple d'un tel accord dans l'action.

En nous appuyant sur l'exemple précédent nous réfléchissons de façon collective vers quelle(s) solution(s) s'orienter : demander un ISBN-A pour l'objet numérique ou un DOI pour l'ouvrage numérisé ?

## 2.2.2 Codage, encodage et décodage des données dans les bibliothèques universitaires

Parmi tous les formats de codage, encodage et décodage des données utilisés dans les bibliothèques universitaires, nous nous attacherons à certains « ingrédients » maîtrisés par les professionnels au service du processus de numérisation d'ouvrages en tenant compte des besoins des utilisateurs. Dans notre scénario les ouvrages imprimés à numériser sont au préalable signalés dans SUDOC<sup>3</sup> à l'aide de notices décrites ou codées au format UNIMARC<sup>4</sup>.

Les accès à un objet numérique physiquement archivé sur un serveur sont multiples. Nous parlerons d'ubiquité des métadonnées d'accessibilité (cf. *figure 1*). Ces métadonnées d'accessibilité sont encodées en fonction des systèmes d'information : catalogue, moissonneur, outils bibliographiques, réseau sémantique.

L'atout majeur d'une notice normalisée en UNIMARC permet l'exportation des mêmes données en d'autres formats : Dublin Core<sup>5</sup>, XMP<sup>6</sup>, XML<sup>7</sup>. C'est au travers des protocoles d'échange de données que les métadonnées sont décodées pour satisfaire l'utilisateur en situation d'interrogation. Ainsi le protocole Z39.50<sup>8</sup> permet la récupération de notices et leur citabilité, OAI-PMH<sup>9</sup> permet le moissonnage des notices. Enfin notons que les formalismes RDF<sup>10</sup> et OWL<sup>11</sup> permettent la réutilisation des modèles et des contenus dans le cadre des réseaux sémantiques.

---

3 Système universitaire de documentation

4 UNIVersal MACHine Readable Cataloguing

5 Le Dublin Core est un schéma de métadonnées générique qui permet de décrire des ressources numériques ou physiques et d'établir des relations avec d'autres ressources : norme ISO 15836

6 Extensible Metadata Platform

7 Extensible Markup Language

8 Ce protocole a donné lieu à la norme américaine ANSI/NISO Z39.50 et aux normes ISO 23950. Son objectif est de permettre l'interrogation de bases de données, notamment bibliographiques

9 Open Archives Initiative Protocol for Metadata Harvesting

10 RDF : Resource Description Framework

11 OWL : Web Ontology Language

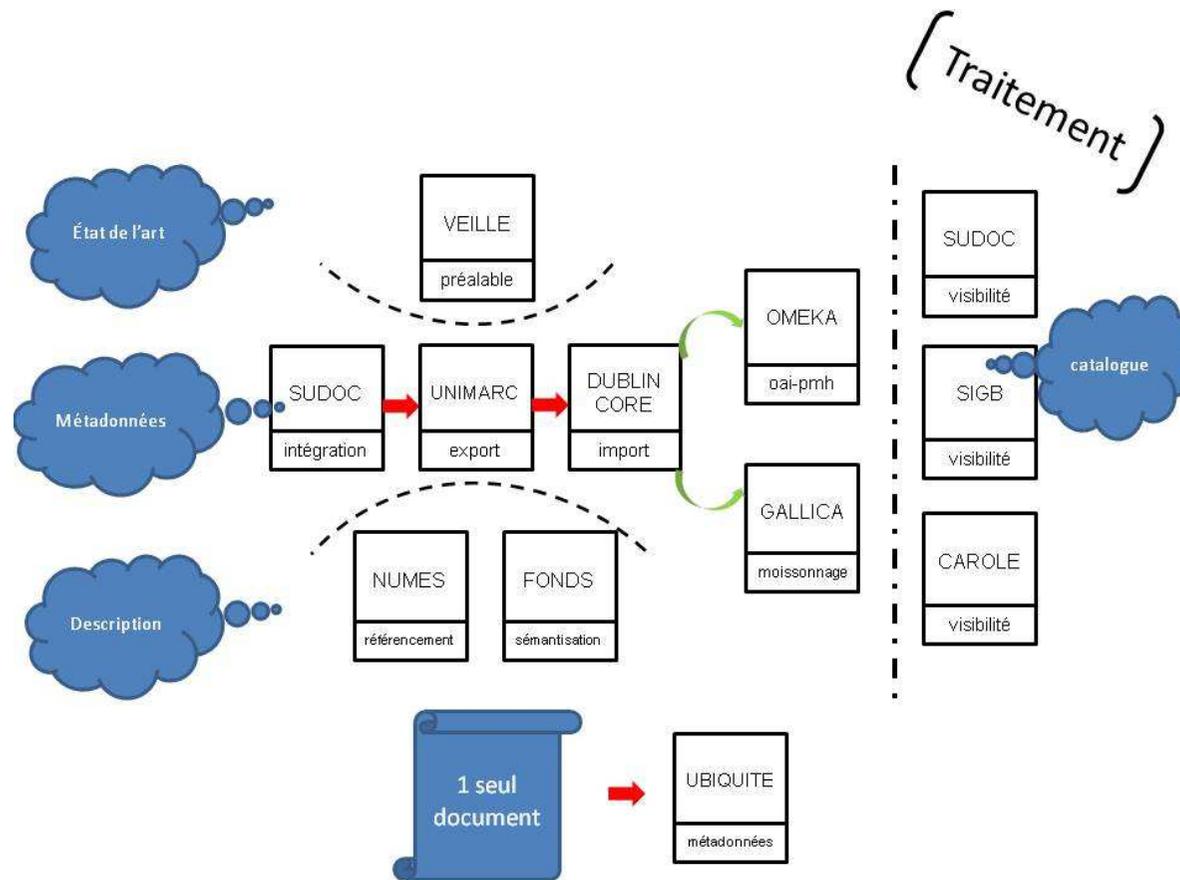


Figure 1 : Chaîne de traitement des métadonnées et de leur ubiquité

Reprenons notre représentation des utilisateurs [16] : RU=T,B,F,A

où T signifie type d'utilisateur, B pour les besoins exprimés, F pour les fonctions et A pour les activités selon :

T {professionnel des bibliothèques, enseignants-chercheurs}

B {accessibilité, interopérabilité, réutilisabilité}

F {lecture, écriture, exportation, récupération}

A {collaboration, créativité, annotation}

La représentation de l'utilisateur développée dans la thèse de Frédérique PEGUIRON [14], associée aux différents types de codage des données, schématise en *figure 2* les combinaisons possibles des données pour répondre aux besoins des utilisateurs, à partir d'un raisonnement en terme d'efficacité.

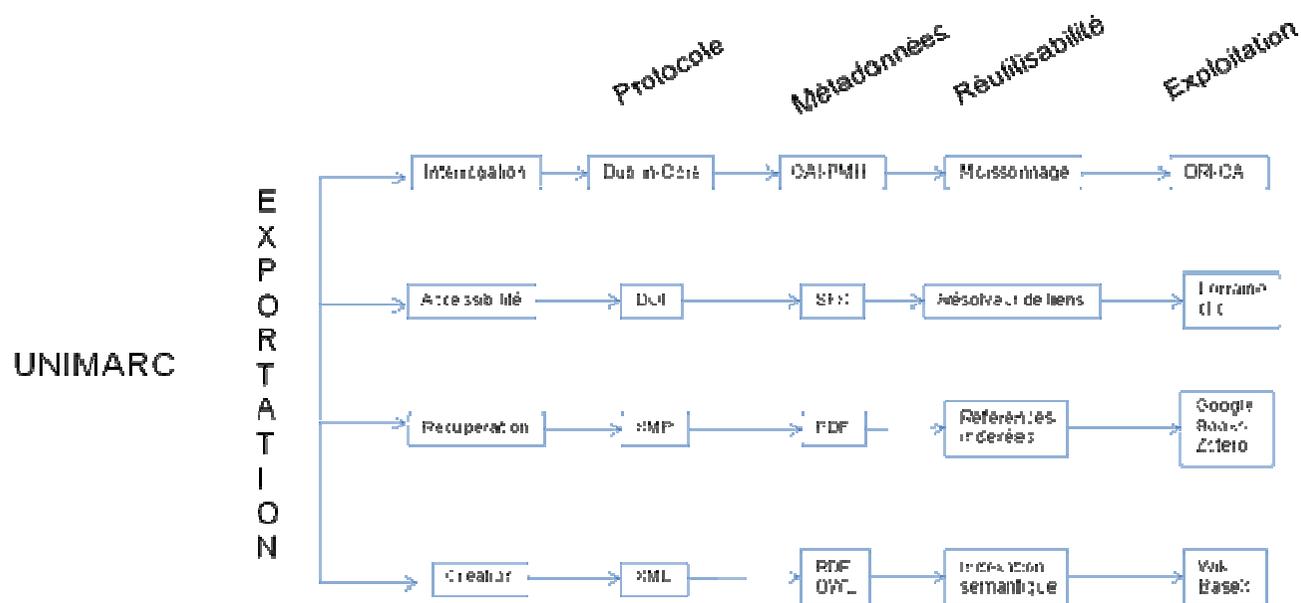


Figure 2 : Efficacité des données dans le cadre de leur ubiquité

A la suite de l'énumération des principaux enjeux d'un entrepôt numérique :

- la visibilité des fonds sur un plan international,
- la coopération documentaire entre différents organismes locaux, régionaux et nationaux,
- la valorisation numérique des collections patrimoniales,
- l'interopérabilité des bibliothèques numériques,
- la structuration et éditorialisation des corpus numérisés,
- les corpus numérisés considérés comme objets de recherche,

notre représentation des utilisateurs permet de replacer ses besoins relatifs à l'exploitation d'un document numérisé comme le figure l'illustration ci-dessous (figure 3).

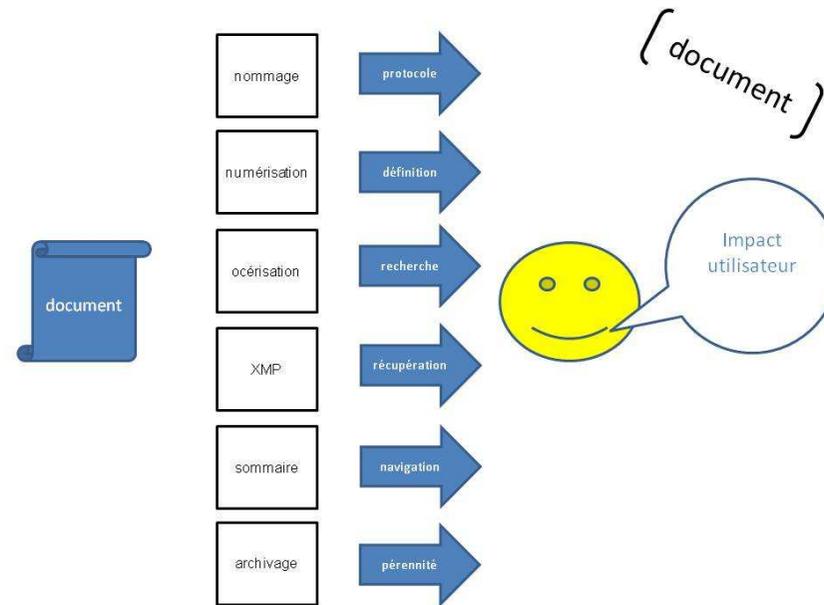


Figure 3 : Traitement du document numérisé et impact utilisateur

A côté des notions de qualité de numérisation, d'océrisation, de tables des matières qui constituent des pré-requis à tout projet d'entrepôt numérique nous ajoutons une valeur ajoutée quant à l'interopérabilité via les différents formats de métadonnées ainsi que la notion de citabilité permettant à l'utilisateur de récupérer les métadonnées environnantes et intégrées au document.

## **3. Collections numériques**

### **3.1 Partir de l'existant des bibliothèques numériques**

Entamer un projet implique de connaître parfaitement le fonds de l'établissement universitaire. La numérisation est étroitement liée au signalement et/ou à la rétro-conversion des fonds dans le catalogue national universitaire. Ce travail préalable permet de comparer l'existant de la bibliothèque déjà numérisé sur la toile et de raisonner en économie d'échelle. Après avoir vu les aspects normatifs inhérents aux collections numériques dans le précédent paragraphe, abordons maintenant les aspects techniques, institutionnels et organisationnels.

### **3.2 Aspects techniques et institutionnels**

Sur un plan national il existe des appels à projets de rétro-conversions émanant de l'ABES<sup>12</sup> pour alimenter le SUDOC et des appels à initiatives de numérisation émanant de la BNF<sup>13</sup>. Répondre à ces appels est essentiel car cela permet de reconsidérer les aspects techniques d'une numérisation et de placer l'établissement dans un cadre institutionnel. Répondre à un appel avec la BNF ouvre à la collaboration. La considération de l'interopérabilité des corpus numérisés prend une dimension réalisable. Les appels à candidature correspondent à la complétion de collections de la BNF ou l'INHA<sup>14</sup> dans les arts par exemple.

En parallèle aux appels à projets nationaux les bibliothèques universitaires abritent des fonds très hétérogènes, « anciens, remarquables et spécifiques ». Ces fonds sont parfois éclatés entre différents établissements et méritent d'être rapprochés, virtuellement, pour leur visibilité.

---

12 Agence bibliographique de l'enseignement supérieur

13 Bibliothèque nationale de France

14 Institut National d'Histoire de l'Art

A titre d'exemple la bibliothèque universitaire de lettres de Nancy dispose des fonds anciens regroupés sous l'appellation « Fonds de l'époque moderne » couvrant une période du 15<sup>ème</sup> au 18<sup>ème</sup> siècle. Elle dispose d'un fonds janséniste remarquable nommé « Fonds Tavenaux » et d'un autre fonds tout aussi remarquable « Fonds Paul Meyer » qui couvre le domaine de la philologie. Les fonds « Fonds Nazi », « Fonds Jeanne d'Arc », « Fonds de la Sarre », « Fonds Decaux », « Fonds Lajarrige » sont spécifiques à la région du Grand Est. Ces fonds sont objet de recherche au niveau de l'Université de Lorraine, au niveau national et international. A côté de considérations de recherche s'ajoutent des préoccupations de sauvegarde du patrimoine livresque du 19<sup>ème</sup> siècle. Celui-ci présente des problèmes quant à la conservation physique du document. Les produits de blanchiment du papier utilisés à l'époque a des conséquences sur la survie des pages. L'Allemagne, de son côté, prend en compte cet aspect de sauvegarde du patrimoine et met en œuvre des dotations pour procéder à la numérisation de ces collections. L'éventualité d'obtenir des subventions pour une numérisation systématique de tous les ouvrages du 19<sup>ème</sup> d'une bibliothèque universitaire n'est pas envisageable. Pour cette période et si les droits d'auteur et d'éditeur sont caduques, une pratique consiste à numériser les ouvrages demandés par le service du prêt entre bibliothèques.

### **3.3 Aspects organisationnels**

L'utilisateur peut prendre connaissance sur un plan national du recensement des corpus numérisés dans l'enseignement supérieur via l'inventaire NUMES<sup>15</sup>. Visuellement la cartographie des bibliothèques numériques dans les universités et les municipalités françaises via google [3] témoigne de l'avancée des projets des bibliothèques numériques. Le 12 janvier 2010 Marc TESSIER remet son rapport [18] sur « la numérisation du patrimoine écrit » au Ministre de la culture et de la communication. Ce rapport évoque la maîtrise du processus de numérisation face à Google qui constitue un élément d'importance dans le contexte économique de l'existant numérique pour les ouvrages. Il met l'accent sur l'hétérogénéité des projets en cours, les contraintes relatives à la numérisation de masse et l'intérêt de la valorisation du patrimoine numérisé. Cela renforce l'idée que les signalements de NUMES ou de la cartographie Google n'évoquent pas suffisamment les collaborations existantes ou à venir. Ils ne sont pas le lieu d'un travail intellectuel sur les collections numériques comme objet de recherche. Ils offrent un espace collaboratif pour le recensement ou la visualisation d'éléments d'information, mais ils ne sont pas conçus pour mettre en regard objet numérique et collection numérique. La photographie de l'existant en matière d'entrepôts numériques nous amène à faire des propositions pour aborder la collaboration des personnes autour de la réutilisabilité des données.

---

15 NUMES : inventaire en ligne des corpus numérisés et des projets de numérisation des établissements et organismes de l'Enseignement supérieur et de la Recherche

## 4. Propositions autour du principe de collaboration et de réutilisabilité

Christelle DI PIETRO [7] révèle « en juillet 2011, un sondage GFK montre que 52% des ebooks téléchargés proviennent de bibliothèques publiques (principalement la BNF, Gallica et Google Books), 77% des téléchargements sont des ebooks gratuits et 27% proviennent de sites pirates». On trouve sur la toile un nombre considérable de versions numérisées dont les établissements documentaires possèdent une version imprimée. Se pose la question de la réutilisabilité de l'objet numérique. Faut-il télécharger l'objet ? Faut-il procéder à un lien vers cet objet ? Nos propositions viendront après la description de pratiques dans un premier scénario au paragraphe 4.1. Le scénario revisité en 4.2 démontre les objectifs et les bénéfices tirés pour les concepteurs et les utilisateurs à la mise en œuvre d'entrepôt numériques collaboratifs. Le 4.3 décrit les modèles sous-jacents à la numérisation.

### 4.1 Un premier scénario

Un point de vue théorique, couplé à une phase de mise en pratique, contribue à rendre perfectible un flux de travail. Pendant de nombreuses années certaines bibliothèques ont porté l'accent sur leurs SIGB<sup>16</sup> locaux occultant l'importance du réservoir national SUDOC, souvent dans un souci de précaution. Cela veut dire qu'il n'y a pas eu une intégration systématique de tous les types de documentation dans le réservoir national universitaire. Pourquoi une telle façon de procéder ? L'intégration de la documentation dans le SIGB « devait » contribuer à mieux répondre aux besoins des utilisateurs notamment par l'ajout de notes propres à un établissement documentaire. Par ailleurs certains SIGB sont plus faciles d'utilisation que le SUDOC. Cette pratique aboutit à des désagréments au niveau de la visibilité d'un fonds, de la réutilisabilité des données et de l'interopérabilité puisqu'elle ne prend pas en compte le contexte national.

Les appels à projet de l'ABES pour des rétro-conversions de masse et les appels à contribution de la BNF pour des vagues de numérisation ont, de ce fait, mis au jour tout l'intérêt de privilégier une démarche nationale dans le signalement des collections. Ces deux opérations sont intrinsèquement couplées. Proposer une offre numérique implique de connaître le fonds dans sa globalité pour être en mesure de faire émerger des corpus et de les rapprocher d'autres collections réparties en France. Répondre à des appels implique une étude préalable concertée avec d'autres établissements documentaires universitaires ou municipaux. La numérisation physique est chronophage et suppose des compétences techniques élaborées. A titre d'exemple La Direction de la

---

<sup>16</sup> Système intégré de gestion de bibliothèque

Documentation et de l'Édition de l'Université de Lorraine, la Bibliothèque nationale de France et la recherche de l'Université de Lorraine coopèrent à partir d'un cadre régional [Université/Municipalité/Recherche] pour le moissonnage d'un corpus juridique régional via Gallica.

## 4.3 Les modèles sous-jacents à la numérisation

### 4.3.1 Modèle XMP

Un document au format .pdf permet d'encapsuler des métadonnées en XMP. Professionnels et utilisateurs sont concernés. Les professionnels interviennent dans le choix du modèle des métadonnées pour que l'utilisateur puisse récupérer les métadonnées automatiquement dans un logiciel bibliographique par exemple. Reprenons notre *figure 1* « Efficience des données dans le cadre de leur ubiquité ». La notice au format UNIMARC est réutilisable pour un autre système, cette fois, pour la récupération de données. Le format des données est transformable en XMP. Le professionnel encapsule les métadonnées XMP dans le document pdf. A l'interrogation, les métadonnées peuvent être récupérées automatiquement dans l'environnement bibliographique de l'utilisateur comme le permet ZOTERO [5] qui est une extension du navigateur Mozilla pour la gestion de bibliographies.

### 4.3.2 Les modèles sémantiques

Par une couche sémantique, le wiki sémantique permet d'ajouter aux pages la représentation des métadonnées et leurs relations sous forme d'annotations sémantiques. Le wiki sémantique [10] donne une signification aux liens qui relient les pages du wiki. Les descriptions sont directement intégrées au code source de chaque page. Les données sont stockées dans un entrepôt de données spécifiques. Le formalisme des wikis sémantiques utilise les balises RDF et le vocabulaire OWL. Les professionnels de l'information structurent les descriptions en s'appuyant sur des modèles pour en permettre l'appropriation par les lecteurs-contributeurs. Le déploiement d'un wiki sémantique implique le chargement de modèles [4] pour décrire les informations. Ces informations décrites sont alors réutilisables en tout point du réseau sémantique. L'utilisateur, en situation de récupération de métadonnées et en situation de contribution collaborative, dispose, dans un environnement bibliographique, de formats d'export de notices. Les plus connus sont les formats RDF, BibTex<sup>17</sup>, MODS<sup>18</sup>,

---

<sup>17</sup> BibTeX est un logiciel de gestion de références bibliographiques et un format de fichier conçu par Oren Patashnik et Leslie Lamport en 1985 pour LaTeX. Il sert à gérer et traiter des bases bibliographiques.

<sup>18</sup> MODS Metadata Object Description Schema

RIS<sup>19</sup>, TEI<sup>20</sup> auxquels nous pouvons ajouter un format qui nous occupe dans notre réflexion le Wikipedia Citation Templates [22]. La *figure 4* ci-dessous explicite la base de métas modèles qui regroupent les différents modèles chargés dans l'environnement.

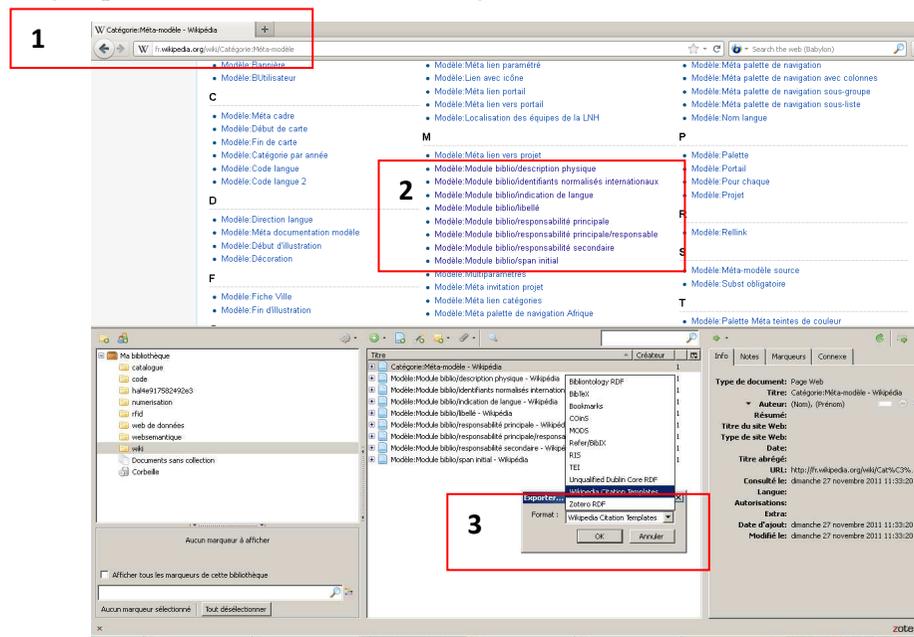


Figure 4 : Les métas modèles

19 RIS Research Information Systems  
20 TEI Text Encoding Initiative

Ce navigateur, enrichi du plugin ZOTERO, pointe sur la page méta modèle de wikipedia (1) où figurent les modèles bibliographiques par type de document (2). Le format Wikipedia Citation Templates (3) de ZOTERO ne sera exploitable dans l'environnement d'écriture collaborative que si le modèle bibliographique adéquat a été chargé dans le réseau sémantique.

L'utilisateur-contributeur appréciera de valoriser les notices dans l'environnement sémantique si le modèle permet de mettre en forme automatiquement sa bibliographie. L'utilisateur enseignant-chercheur dans un réseau sémantique participe à la description intellectuelle de l'objet numérique dans une collection numérique où seront indiqués :

- La genèse du projet et son nom,
- L'historique,
- Les domaines abordés,
- Les collaborateurs.

## **5. PULSAR : un exemple d'entrepôt numérique interopérable**

Pour poursuivre avec l'exemple autour du corpus juridique et illustrer notre démarche dans ce papier, la mise en œuvre d'un entrepôt tenant compte des consignes techniques de la BNF a été opérée à partir d'une plateforme ouverte OMEKA. Les ouvrages numérisés sont accessibles sur une plateforme nommée PULSAR [Patrimoine Universitaire Lorrain des Savoirs ARchivés] et dans le même temps moissonnés via Gallica.

Ci-dessous la *figure 5* met l'accent sur les besoins de l'utilisateur pris en compte en amont des réflexions et exprimés par la navigation dans le document, les recherches en plein texte, le guidage par des sommaires enrichis, des notices élaborées et la récupérabilité des métadonnées dans des outils appropriés et permettre leur citabilité dans des articles ou des espaces d'écriture sémantique.

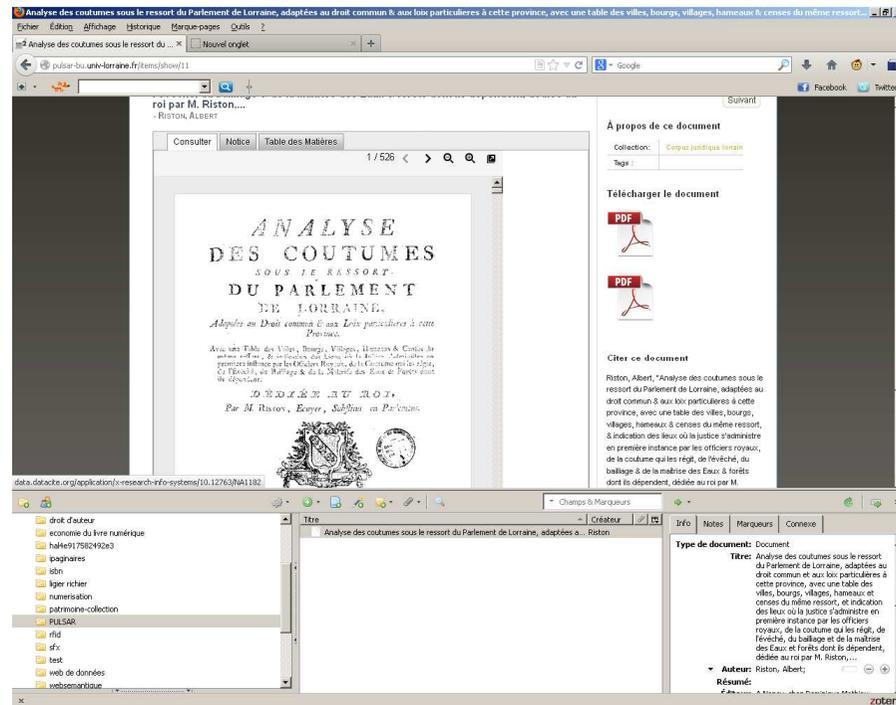


Figure 5 : Récupération des métadonnées par l'utilisateur final

## 6. Conclusion

Œuvrant dans un domaine où les technologies, les formats et les standards, les besoins et les pratiques des utilisateurs évoluent constamment, la numérisation en bibliothèque universitaire implique des procédures ajustables au fil du temps. La collaboration entre documentation et contenu académique

est en plein essor. C'est un enjeu central. Pourquoi ? Parce que créer de la connaissance sous forme de réseau sémantique est souvent aujourd'hui plus facile dans le milieu de l'information que dans les différentes disciplines universitaires. Seule la présentation d'un certain nombre d'expériences pourra entraîner l'adhésion de la communauté scientifique. Il est possible qu'une des clés soit l'appropriation par les enseignants-chercheurs à la fois pour la pédagogie, mais aussi en terme heuristique des ressources qu'offre la technologie proposée, si elle est pensée en adéquation avec le processus de numérisation.

La construction d'entrepôts collaboratifs numériques nous semble une bonne solution pour améliorer la qualité du service à l'utilisateur et renforcer la collaboration de recherche. Dans cet article nous avons montré les difficultés de tous ordres que soulève cette proposition, aussi bien sémantiques, techniques qu'organisationnels. Nonobstant ces problèmes il nous semble que c'est la solution d'avenir, à condition de prendre en compte l'utilisateur sous forme d'un modèle.

## 6. Bibliographie

- [1] BELAID A., EMPTOZ H. et VIGNAUX G., *Numérisation et valorisation des collections*, AS-CNRS 96 Rapport d'activités, 2004.
- [2] BnF - *Web sémantique, Web de données : présentation*, [En ligne : [http://www.bnf.fr/fr/professionnels/web\\_semantique\\_donnees/s.web\\_semantique\\_intro.html](http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html)]. Consulté le 14 juin 2013.
- [3] *Carte des bibliothèques numériques en France* - Google Maps, [En ligne : <http://maps.google.fr/maps/ms?ie=UTF8&oe=UTF8&msa=0&msid=205724028419509989164.00046ecee33cfb4e1330d>]. Consulté le 14 juin 2013.
- [4] *Catégorie : Méta-modèle* - Wikipédia. Available [En ligne : <http://fr.wikipedia.org/wiki/Cat%C3%A9gorie:M%C3%A9ta-mod%C3%A8le>]. Consulté le 14 juin 2013.
- [5] Center for History and New Media, *Guide rapide pour débiter*. [En ligne : [http://zotero.org/support/quick\\_start\\_guide](http://zotero.org/support/quick_start_guide)]. Consulté le 14 juin 2013.
- [6] DE SEPAUSY V., *Norme ONIX pour les métadonnées et un ISBN propre à chaque ebook*, 2010 [En ligne : <http://www.actualitte.com/actualite/lecture-numerique/acteurs-numeriques/norme-onix-pour-les-metadonnees-et-un-isbn-propre-a-chaque-ebook-16964.htm>]. Consulté le 14 juin 2013.
- [7] DI PIETRO C., *Marché du livre numérique*, ENSSIB.
- [8] DOI *System Factsheet* [En ligne : <http://www.doi.org/factsheets/ISBN-A.html>]. Consulté le 14 juin 2013.
- [9] ISAAC H., *L'université numérique*, Rapport au ministre de l'Enseignement supérieur et de la Recherche, MESR, 2008.

- [10] *Métadonnées pour WICRI, un réseau de wikis sémantiques pour les communautés de la recherche et de l'innovation* - Wicri Lorraine. [En ligne : [http://ticri.inpl-nancy.fr/wicri-lor.fr/index.php/M%C3%A9tadonn%C3%A9es\\_pour\\_WICRI,\\_un\\_r%C3%A9seau\\_de\\_wikis\\_s%C3%A9mantiques\\_pour\\_les\\_communaut%C3%A9s\\_de\\_la\\_recherche\\_et\\_de\\_l%27innovation](http://ticri.inpl-nancy.fr/wicri-lor.fr/index.php/M%C3%A9tadonn%C3%A9es_pour_WICRI,_un_r%C3%A9seau_de_wikis_s%C3%A9mantiques_pour_les_communaut%C3%A9s_de_la_recherche_et_de_l%27innovation)]. Consulté le 14 juin 2013.
- [11] MOIREZ, P., *le web semantique entre les mains des professionnels de l'infodoc.*, Archimag (Vincennes). 2011.
- [12] NOYER J.-M., *Synthèse du Programme de Numérisation pour l'Enseignement et la Recherche*, (PNER, 1999-2001) », 2001.
- [13] Patrimoine numérique, *Catalogue des collections numérisées* [En ligne : <http://www.numerique.culture.fr/pub-fr/index.html>]. Consulté le 14 juin 2013.
- [14] PEGUIRON F., *Application de l'Intelligence Economique dans un Système d'Information Stratégique universitaire : les apports de la modélisation des acteurs*, Thèse Université Nancy 2, 2006.
- [15] PEGUIRON F., *Observatoire intelligent : de l'information aux connaissances au travers de collections numériques*, in *Évolutions sociotechniques des bibliothèques numériques*, Hermès, 2011.
- [16] PEGUIRON F., THIERY O. *Modélisation des acteurs, des ressources documentaires : application à un entrepôt universitaire*, VSST'2006.
- [17] SOCIETE ORTECH, *L'état de l'art de la numérisation pour l'enseignement et la recherche : les acteurs de la numérisation et leurs programmes*, 2000.
- [18] TESSIER M., *Rapport sur la numérisation du patrimoine écrit*, 2010.
- [19] *The Digital Object Identifier System*, [En ligne : <http://www.doi.org/>]. Consulté le 14 juin 2013.
- [20] Vie de l'abd : [Communiqué], *Exceptions au droit d'auteur en faveur des bibliothèques et des archives*, IABD.
- [21] Vie de l'abd : [Communiqué] *Livres indisponibles et orphelins : quel enjeu pour les bibliothèques ?*, IABD... », [En ligne : <http://www.iabd.fr/2011/11/10/communiqu%C3%A9-livres-indisponibles-et-orphelins-quel-enjeu-pour-les-bibliotheques/>]. Consulté le 14 juin 2013.
- [22] Wikipedia:Citation templates - Wikipedia, the free encyclopedia. [En ligne : [http://en.wikipedia.org/wiki/Wikipedia:Citation\\_templates](http://en.wikipedia.org/wiki/Wikipedia:Citation_templates)] Consulté le 14 juin 2013.