

# Mise en valeur de collections numériques réparties : de l'information aux connaissances

Frédérique Peguiron (1), Odile Thiery (2)  
[Odile.Thiery@loria.fr](mailto:Odile.Thiery@loria.fr), [FrederiquePeguiron@loria.fr](mailto:FrederiquePeguiron@loria.fr)

SITE-LORIA Nancy Université, B.P. 239 54506 Vandoeuvre, France (1) et (2)

**RÉSUMÉ.** *L'idée novatrice est de travailler dans un esprit fédératif et de relier un objet documentaire ou culturel à tout un environnement. Les liaisons se font au travers d'un nouvel outil social et particulièrement adapté à l'esprit transdisciplinaire et transrégional. La construction de liens pour rapprocher virtuellement des collections réparties sur un plan géographique tient compte des projets existant pour s'articuler avec eux. Cette plateforme donne les moyens aux chercheurs de retrouver un maximum de liens entre les projets et les objets académiques de recherche. Plutôt que de se concentrer uniquement sur les objets numériques il est intéressant de les replacer dans leur environnement de façon à valoriser des fonds entiers de collections thématiques et cela en sortant d'un cadre institutionnel. Les métas connaissances pour la représentation des connaissances du domaine d'application relèvent de la méta base chargée de contenir les structures des bases métiers. Les connaissances ou les informations contenues dans la méta base portent sur les attributs nécessaires pour définir un schéma global des relations entre les objets numériques et les acteurs.*

**MOTS-CLÉS :** *Collections numériques, Entrepôt documentaire, Métadonnées, patrimoine universitaire et culturel, valorisation de la recherche.*

**ABSTRACT..** *The idea is to work in a spirit of federation and link a document or object to a cultural environment. The connections are made through a new social tool. This tool is adapted to the spirit transdisciplinary and trans. The bridge-building takes account of existing projects. This platform provides the means for researchers to find a maximum of links between projects and objects of academic research. It is interesting to place digital objects in their environment to enhance thematic collections outside an institutional framework. Knowledge or information contained in the basic concern metadata attributes to define an overall pattern of relationships between digital objects and actors.*

**KEYWORDS:** *Digital collections, documentary Warehouse, Metadata, academic and cultural heritage, promotion of research.*

## 1. Introduction

La mise en valeur des collections numériques est fondamentale et est un enjeu majeur pour l'enseignement supérieur et la recherche. De plus en plus de services communs de documentation universitaires se réunissent en réseaux afin de concevoir des bibliothèques virtuelles.

Ainsi définir une politique de partage de contenus documentaires au sein des universités est un défi dans un contexte pluriel [44] : contexte culturel, contexte économique, contexte de performance, contexte de réformes, contexte technologique, contexte de l'évolution des pratiques scientifiques, contexte international et contexte politique.

**Contexte culturel :** Les établissements universitaires abritent un patrimoine informationnel universitaire et culturel riche et varié tant du point de vue de la forme, du fonds que des niveaux des publics auquel il est destiné.

**Contexte économique :** La révolution technologique liée au boum de l'Internet, au nomadisme et à la miniaturisation pousse à des réflexions pour être au plus près des usagers et surtout à l'écoute de leurs besoins.

**Contexte de performance :** Les résultats de l'enquête de Shanghai dont on peut critiquer les critères de performance, comme par exemple la présence d'un prix Nobel faisant monter « artificiellement » le score d'une université ont fait l'effet d'un électro choc et ont eu des conséquences sur le regroupement des universités.

**Contexte de réformes :** De nouveaux objectifs et modes de gouvernance sont instaurés par les réformes, initiées en 2002 avec le LMD, poursuivies en 2007 avec l'ANR, l'AERES et finalement par la loi LRU en faveur de l'autonomie des universités.

**Contexte technologique :** Nous sommes à l'intersection de deux situations. D'une part une certaine maturité des espaces numériques de travail et d'autre part une mutation technologique qui favorise les espaces dits « web sociaux ».

**Contexte de l'évolution des pratiques scientifiques :** Pour caractériser cette évolution marquée par le terme e-Science, Tony Hey ancien directeur du programme de même nom au Royaume Uni évoque un véritable changement de paradigme. Les collections ne sont plus seulement un outil de consultation ou de dépôt mais le nouvel espace de travail du chercheur qui construit des connaissances au sein même de ces collections.

**Contexte international :** Cette évolution arrive dans un paysage de recherche mondialisé où les programmes à portée internationale se multiplient autour des infrastructures de la recherche qui deviennent des cyberinfrastructures (voir en Europe la feuille de route ESFRI qui fait écho aux rapports Atkins ou Welshons de la NSF). Une collection ne peut plus être évaluée dans un contexte local mais dans sa capacité à s'insérer dans un réseau international.

Et enfin un **contexte politique :**

Avant de définir le cadre géographique, thématique et institutionnel de nos propositions rappelons qu'à l'heure où nous écrivons cet article Nathalie Kosciusko-Morizet vient d'être nommée nouvelle secrétaire d'Etat au Développement de l'économie [33] pour transformer en actions concrètes les 154 propositions du plan France numérique 2012, présenté en octobre dernier par son prédécesseur, Eric Besson qui s'était engagé à accélérer la mutation de notre économie vers les nouveaux modèles de diffusion des contenus [5]. A côté de l'accès universel aux réseaux numériques E. Besson insiste sur le développement de la production et de l'offre de contenus numériques, la diversification des usages et services et la rénovation de la gouvernance et de l'écosystème de l'économie numérique. Outre le lien de plus en plus évident entre l'écologie et les technologies de l'information, l'une des tendances phares de l'année 2008 est la montée en puissance du «Green IT» qui a probablement valu la nomination de Nathalie Kosciusko-Morizet. Ces contextes sont autant de paramètres qui influent et mettent en relief les objectifs des utilisateurs conjointement aux objectifs des établissements pour une meilleure satisfaction de l'utilisateur et une meilleure gestion organisationnelle de contenus numériques. De ce fait nous raisonnons selon différents niveaux : conceptuel, pragmatique et expérimental. Les idées que nous livrons ici devraient aider à une meilleure articulation entre les informations et la politique scientifique dans les stratégies d'établissements de recherches avec les dispositifs universitaires et culturels. Dans cet article le paragraphe 2 présentera les projets en cours, leur contexte et leurs frontières, le 3 les différents types de sources documentaires. Puis dans le 4 nous parlerons du processus d'aide à la décision. Le paragraphe 5 introduira une description normalisée des ressources documentaires. Le 6 notre modèle de l'utilisateur  $RU = (T,B,F,A)$ . Enfin une conclusion viendra clore le propos sur nos perspectives de recherche.

## 2. Projets en cours

Il nous faut répondre à la question :

Comment pouvons nous réfléchir à la mise en valeur de collections numériques réparties ? Notre problématique s'articule autour des collections et des institutions du caractère innovant et opérationnel des propositions en termes de gestion des documents numériques.

Les collections **universitaires** et **culturelles** dans le paysage institutionnel sont référencées au travers de corpus ou de grands catalogues informatisés à l'aide de méta données relatives aux normes et standards en vigueur.

### 2.1. Acteurs académiques universitaires et culturels

Une bonne approche consiste à observer et recenser les acteurs autour d'une thématique sur un plan local, régional et transfrontalier.

Ainsi dans un cadre documentaire on assiste à un rapprochement institutionnel autour d'une thématique sur un plan national. On peut citer l'Université de Lille 1 qui développe depuis plusieurs années une bibliothèque numérique dénommée IRIS<sup>1</sup>, qui prend la suite de GRISEMINE. Cette bibliothèque comprend entre autres une collection en histoire des sciences. Cette bibliothèque n'est pas partie seule sur ce projet, qui émane directement des problématiques de recherche et des besoins du centre d'histoire des sciences et épistémologie de Lille 1, "savoirs, textes, langage" et "Laboratoire Painlevé", et cela en relation avec des chercheurs d'établissements répartis à l'échelon national : à Nantes, Toulouse Sabatier, Tours et Nancy 2. Les universités veulent développer une politique documentaire qui contribue à l'excellence de la recherche scientifique et à l'excellence pédagogique, donc à l'attractivité de ses établissements. Les objectifs se fondent sur l'analyse des besoins des utilisateurs : qualité et exhaustivité de l'offre documentaire pour la pédagogie et la recherche. Dans ce cadre les bibliothèques des SCD (Service Commun de Documentation) et les bibliothèques d'UFR (Unité de Formation et de Recherche) des universités abritent des fonds rares et précieux en relation étroite avec la pédagogie des universités et avec les axes de recherche des laboratoires de recherche. Compte tenu de la variété des fonds disciplinaires les projets de numérisation permettent de mettre en synergie des composantes documentaires d'instituts pour faire émerger des fonds disciplinaires à numériser et à valoriser et ainsi à coopérer autour de pratiques communes. Au niveau local on trouve une grande diversité des structures : au niveau de l'enseignement et de la recherche des services communs de documentation mais aussi des services de bibliothèques de campus voire des bibliothèques d'UFR ou de petites unités de recherche, des centres de documentation de laboratoire de recherche. On trouvera par ailleurs des établissements à vocation culturelle enfin des sites institutionnels. Toutes ces unités disposent par ailleurs de services d'archives. L'objectif des regroupements est de constituer des ensembles thématiques disjoints mais avec des points communs. Au niveau régional sont constitués des inventaires numériques régionaux qui permettent le rapprochement de laboratoires de plusieurs universités d'une région autour d'un espace disciplinaire.



### 2.2.A l'échelon transfrontalier

Par exemple, en septembre, l'Union Européenne a donné le feu vert à des projets transfrontaliers. Ces projets ont pour but de fonder les assises d'un espace intégré d'enseignement supérieur dans une grande région en tant que partenaires méthodologiques.

## 3. Typologie des sources

Pour adapter des descriptions de collections numérisées ou à numériser afin de les mettre en valeur dans les différents systèmes d'informations où elles sont susceptibles de figurer nous recensons des ressources documentaires variées :

- Acquisition : achat d'un établissement, achat collaboratif entre plusieurs instituts, des dons à concevoir dans leur ensemble. Cela peut être une bibliothèque personnelle d'un enseignant.
- Thème : des disciplines, domaines
- Forme : des imprimés, des manuscrits, des ouvrages
- Temps : des périodes couvertes
- Fonds : des fonds rares et précieux, des collections
- Géographie : contours géographiques

<sup>1</sup> <https://iris.univ-lille1.fr/dspace/>

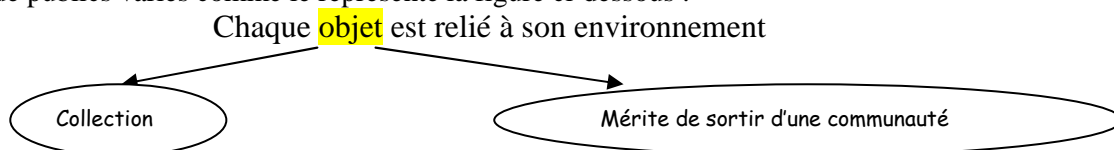
Dans le meilleur des cas ces sources documentaires sont signalées au travers de catalogues informatisés locaux (SIGB), nationaux (SUDOC), catalogues de manuscrits (CGMBPF) ou (CALAMES). Toutes ces sources sont à considérées selon trois angles : contenu, contenant et contexte et deviennent signifiantes en terme de collections.

### Collections

Outre le fait qu'un grand nombre de ces objets informationnels ne soient pas numérisés il est difficile d'évaluer les collections. Par évaluer nous entendons se mettre à un niveau macroscopique :

- Pour évaluer les collections dans leur ensemble
- Pour exprimer la spécificité de leur fonds
- Pour identifier les institutions qui les possèdent

Dans notre problématique, en nous plaçant à un niveau méta après avoir qualifié chaque objet, nous le mettons en relation avec son environnement pour le faire sortir d'une communauté et le valoriser au bénéfice de publics variés comme le représente la figure ci-dessous :



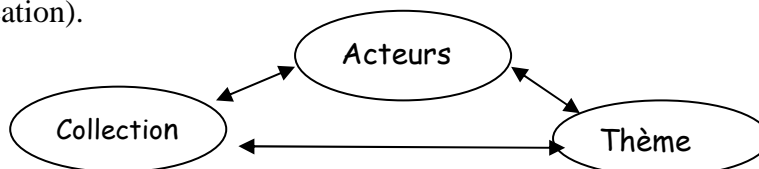
**Figure 1 : Objet-collection-communauté**

Les ressources constituent des informations banalisées. Nous désirons mettre en valeur les points forts d'une collection ou d'un ensemble de collections selon plusieurs points de vue **académique, thématique** en dépassant les cadres **institutionnels** pour satisfaire des publics variés autour d'enjeux économiques.

- ⇒ Comment pouvons nous rapprocher des collections sur plan géographique ?
- ⇒ Comment pouvons nous tisser un réseau numérique ?

Par les sources informationnelles ou grandes collections se dégagent des fonds spécifiques. Par une attaque thématique on identifie les acteurs qui traitent ces objets où qui mènent des recherches. Recherches étant de disciplines très différentes. Elles peuvent toucher :

- au contenu (thématiques),
- au contenant (informatique, sciences cognitives, sciences de l'information et de la communication).



**Figure 2 : Environnement des connaissances**

Toutes ces données numériques constituent un patrimoine culturel universitaire à valoriser en faveur de l'économie de la connaissance.

## 4. Processus d'aide à la décision ou d'aide à la connaissance

A l'inverse des systèmes d'information classiques où il faut préciser un modèle d'application avant de se lancer dans une opération, nous proposons une méthode adaptative qui repose sur la modélisation des objets numériques et des descriptions de collections qui viennent compléter la modélisation des acteurs qui ont fait le fruit de travaux de recherche publiés et qui seront évoqués au paragraphe 6.

### 4.1 Passage de l'information à la connaissance

Les universités disposent de toute une palette d'environnement et d'outils qui ont acquis une certaine maturité et stabilité. A côté de plateformes de numérisation on peut citer en tout premier lieu les grands catalogues qui gagneraient à devenir des catalogues enrichis par le biais d'une politique concertée et coopérative en proposant des collections numériques primaires. En second les universités disposent d'outils de dépôt et de diffusion de la production scientifique pour valoriser la recherche et

la publication en ligne. Ces plateformes tendent vers une interopérabilité avec les plates-formes nationales de dépôt (STAR, HAL, des cours en ligne, etc.). Enfin ces outils sont pour la plupart intégrés à l'espace numérique de travail de l'Université en lien avec les projets TICE. Comme nous l'avons exprimé précédemment une même collection peut en effet être décrite dans ces différents environnements, avec des caractéristiques différentes. Par exemple, une collection concernant un point particulier de la contribution d'un physicien français entre 1870 et 1914 n'aura pas le même environnement relationnel dans un observatoire relatif à la société française ou européenne à cette époque, dans une histoire de la physique ou dans l'inventaire des collections académiques de la Lorraine. Nous voulons également dépasser la notion d'inventaire systématique pour nous attacher à mettre en valeur les points originaux d'une collection parmi un ensemble. Ceci implique une cohabitation d'informations rédactionnelles et structurées.

S'agissant des sources documentaires les informations primaires sont les œuvres directes des producteurs (ou auteurs), les informations secondaires et tertiaires sont des transformations en « modèles réduits » des informations primaires, pour alimenter des bases documentaires. Pour caractériser correctement les informations issues des données de base et permettre de produire de la connaissance, une première étape, maintenant communément admise, est de caractériser les données par des métadonnées c'est-à-dire des données sur les données.

## 4.2 Les méta données dans un contexte documentaire

### 4.2.1 Définition

Selon Catherine Dhérent, [14] les métas données sont constituées de contenu structuré ou non qui peut être des mots, des formules, des analyses spectrales... Elles sont créées de façon automatique ou manuelle et peuvent avoir plusieurs niveaux de complexité. Les métas données sont faites pour être lues soit par l'humain, soit par la machine.

### 4.2.2 Quelques exemples de typologie

- *Externes aux ressources et sous forme papier* : une notice papier pour remplir un formulaire dont les données sont saisies par informatique ;
- *Externes aux ressources et stockées dans une base de données* : un catalogue bibliographique décrivant des ouvrages numérisés ;
- *Internes et encapsulées dans la ressource* : en-tête d'un document dans une DTD (Document Type Définition, document permettant de décrire un modèle de document SGML ou XML) ;
- *Métas données de gestion administrative de gestion de droits et métas données techniques* : elles permettent de donner du sens à des codes et de créer des ontologies, elles définissent les droits d'utilisation des ressources et permettent la préservation à long terme des ressources ;
- *Métas données de description* : elles permettent la recherche des ressources les plus pertinentes sur le web et expliquent comment les ressources sont exploitables.

En résumé ces métas données permettent la gestion et la manipulation de ressources documentaires. Les métas données sont toujours créées dans un but précis. Dans un même organisme, certains documents pourront ainsi être créés avec des jeux de métas données minimaux ou maximaux selon l'objectif (durée de conservation, usages, droits...), il faut noter qu'une partie doit être créée en même temps que le document lui-même et que d'autres métas données peuvent être créées tout au long de la vie du document.

## 4.3 Les métas données pour les bases de connaissance

Le système dispose ainsi des connaissances sur l'adéquation partielle des réponses du système au besoin de l'utilisateur [55] car le système ne connaît réellement ce besoin que par une estimation basée sur les requêtes. Au lieu de calculer ce besoin, nous proposons d'en intégrer la représentation dans le modèle de l'utilisateur, ce qui constitue l'originalité de notre proposition. Cela revient à stocker parmi les **métadonnées** du système, une représentation explicite de la structure des différentes bases

métiers. Notre préoccupation principale est donc la prise en compte de l'acteur dans la construction puis l'exploitation d'entrepôts de données. Les métas données sont dans ce contexte des informations sur les données indispensables à une exploitation efficace d'un entrepôt de données. Le terme de métas données est, nous l'avons dit, couramment défini comme "des données concernant les données". Les métas données décrivent le schéma d'un entrepôt de données et les données individuelles. Dans l'entreposage de données, les métadonnées sont classées d'après leur **objet** et le **public** auquel elles s'adressent. La donnée est forcément liée à d'autres objets du système d'information ; il est donc également nécessaire de représenter, de décrire et de stocker ces **interactions** avec d'autres données. Notre préoccupation principale est donc la prise en compte de l'acteur dans la construction puis l'exploitation d'entrepôts de données. C'est elle qui permet d'être garant de la cohérence fonctionnelle et technique et de la qualité des informations fournies aux utilisateurs métier par le biais de l'entrepôt de données. Les métas données permettent de décrire l'ensemble des règles, des définitions, des transformations et des processus associés à une donnée. On peut donc, par ce moyen, suivre l'évolution des données depuis le système source jusqu'à leur restitution. On est capable, dès lors, d'avoir une vision unifiée des données, transverse à tous les métiers de l'organisation. L'objectif de la modélisation, puis de la description est de favoriser la visibilité d'un patrimoine culturel tout en préservant une expression simple des informations pouvant répondre aux questions suivantes : quelles sont les caractéristiques de la ressource ? Comment classer cette ressource ? Comment mettre en relation le profil de la ressource et le profil utilisateur ?

## 5. Description normalisée des ressources documentaires

### 5.1 Normes et standards

Les normes et standards [43] apportent de nombreux éléments de réponses sur plusieurs points : le domaine, le matériel, l'interopérabilité des plates-formes, le type de médias, l'interface homme – machine, la description des contenus, l'architecture du système, les technologies collaboratives, le vocabulaire, les informations sur le participant, la description des compétences, la propriété intellectuelle, la qualité. Nous voyons que ces domaines constituent des classes d'objet. Cette mise en relation est réalisable, grâce aux métas données de l'entrepôt de données.

#### 5.1.1 Inventaires normés

Ces types de documents sont gérés par des systèmes qui leur sont propres, accessibles au travers de leur propre moteur de recherche ou par des moteurs de recherche fédérés. En ce qui concerne les catalogues de bibliothèque le format utilisé pour la description des notices tend vers le format MARC, avec des sous parents : LC-MARC, UNIMARC, US-MARC. En ce qui concerne la classification et l'indexation, on s'aperçoit qu'il est impossible d'avoir recours à un système unique. Quant à l'indexation, on voit les limites du recours à un thésaurus unique. Ces systèmes de classification et d'indexation ne suffisent pas à tous les impératifs de description pour les ressources documentaires.

#### 5.1.2 Plateforme de dépôt ORI OAI2

Dans le cadre d'objets numériques une plateforme de dépôt ORI OAI [6] permet de disposer de documents numérisés accessibles en ligne et des descriptions de ces documents. Le protocole OAI-PMH3 est un moyen d'échanger sur Internet des métadonnées entre plusieurs institutions, afin de multiplier les accès aux documents numériques. Il permet d'accroître la visibilité des collections numériques sur Internet, de reconstituer virtuellement des corpus à partir de ressources accessibles sur différents sites, d'alimenter des portails thématiques. L'OAI-PMH définit deux types d'acteurs : les fournisseurs de données, qui déposent leurs métadonnées sur un serveur web appelé "entrepôt", et les fournisseurs de service qui collectent ces données pour les intégrer à l'index de leurs propres bibliothèques numériques. Un même établissement peut jouer les deux rôles, diffuser ses métadonnées et moissonner celles des autres. Chaque description de document inclut une URL qui pointe vers sa

---

2 Outil de Référencement et d'Indexation - Open Archives Initiative

3 Open archives initiative - protocol for metadata harvesting = protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes

version numérique. Tout document est décrit selon le format de métadonnées généralistes Dublin Core4 non qualifié ou dans un autre format encodé en XML. Un entrepôt OAI peut être structuré en sous-ensembles (en anglais, "sets"). Cette structuration est facultative. Elle permet le moissonnage sélectif sur un ensemble donné de notices, sans avoir à récupérer toutes les notices contenues dans l'entrepôt. Il est possible de créer des ensembles par **thèmes**, en s'appuyant par exemple sur les grandes classes Dewey, ou bien par **types** de documents (livres, photographies, documents d'archives, périodiques etc.), ou encore par fonds ou collections. Cet élément constitue pour nous un véritable **verrou** au service de notre raisonnement autour de grandes **collections** relatives à un **thème** et/ou en fonction des **types** d'objets documentaires.

#### Exemple 1 :

Il est possible de récupérer uniquement les notices de Gallica

- ⇒ des ouvrages et périodiques portant sur le droit (set gallica:3:34)
- ⇒ des photographies et lots de photographies (set gallica:images:photographies)
- ⇒ des partitions du fonds Philidor (set gallica:philidor:partitions)

Le protocole OAI ayant été d'abord conçu pour le signalement des documents numériques librement accessibles en ligne, il est possible de structurer l'entrepôt en fonction des **usages** possibles sur le document. On distingue l'approche "numérique" (signalement des documents numériques) de l'approche "bibliographique" qui consiste à signaler également des ressources non disponibles en ligne.

#### Exemple 2 :

La BnF a fait le choix d'utiliser un entrepôt OAI pour décrire et signaler les documents numériques (OAINUM) et un autre entrepôt pour exposer l'ensemble de l'information bibliographique (OAIAT).

#### Exemple 3 :

Les sets peuvent être hiérarchisés. Dans ce cas, les deux-points (":") sont utilisés pour signaler la relation d'appartenance.

- ⇒ gallica:images:photographies indique que le set "photographies" appartient à l'ensemble "images", lui-même inclus dans l'ensemble "gallica".

Une même notice peut appartenir à plusieurs sets. Là encore nous verrons ultérieurement dans la suite de cet article l'importance de pouvoir qualifier les usages faits d'un objet documentaire numérique. Auparavant il est utile de préciser le contexte de la naissance de la mission TICRI.

#### 5.2.2 Le contexte : la mission TICRI [16]

Les TICE ont été fortement fédératrices de rapprochement de communautés d'enseignants, d'étudiants et de personnels administratifs autour d'objets informationnels pour satisfaire leurs besoins. Par ailleurs sur l'Internet du grand public est apparu un outil le wiki qui a été approprié par ses utilisateurs dépassant toutes les prévisions quant à son succès dans l'esprit de web social. Nous nous apercevons que les jeunes chercheurs sont de grands consommateurs des espaces collaboratifs. Citons en quelques uns Del.icio.us qui permet le partage de favoris décrits avec des tags. Yoono offre un moteur de recherche collaboratif. BitTorrent est un logiciel d'échange de fichier Point à Point. Au travers de MySpace les membres partagent les mêmes centres d'intérêt par leur page personnalisée. Netvibes fonctionnant sur le même principe propose des fonctionnalités supplémentaires pour l'agrégation d'e-mails, de fils RSS et d'apposition de tags. Ces deux constatations ont poussé la DRRT2 Lorraine avec le soutien du CNRS à mettre en œuvre un Prototypage de relations intercommunautaires qui se « rencontrent sur le terrain ».

Le sigle TICRI signifie : Technologies de l'Information Communication pour la Recherche et l'Innovation. Il a été construit par analogie avec l'acronyme TICE qui a généralisé l'usage des environnements numériques de travail (ENT) à destination de l'enseignement supérieur. TICRI a vocation à proposer une appropriation d'un environnement numérique par les chercheurs dans un esprit collaboratif.

Cette mission a été motivée par le constat d'une situation difficile de la recherche française par rapport à l'appropriation des infrastructures et informations numériques. Le début de l'introduction du rapport Salençon J. et Moatti A. 2008 [52] est parfaitement explicite sur ce point. Le recours aux technologies

---

4 Le format Dublin Core non qualifié définit 15 éléments de métadonnées, dont chacun est facultatif et répétable.

du web sémantique permet un transfert de l'information vers la connaissance ; cet environnement d'aide à la décision ou d'aide à la connaissance dépasse le cadre des inventaires pour mettre en regard Collections ↔ Acteurs ↔ Institutions.

### 5.3.2 Expérimenter les wikis

Les premiers contacts pris dans le cadre de la mission TICRI ont mis en évidence un besoin de sensibilisation au nouveau contexte de l'information numérique scientifique. La DRRT2 Lorraine a pensé à donner aux groupes de travail un outil collaboratif de type wiki. Ce choix initial permet déjà d'approfondir certains aspects liés à la visibilité des activités de recherche et des résultats de l'innovation. En effet, les universités sont sensibilisées aux indicateurs de performance et une analyse sommaire de leur visibilité sur des sites coopératifs montre que la France est particulièrement mal lotie de ce point de vue. La mission TICRI s'est donc interrogée sur un positionnement par rapport aux solutions de type Wikipédia.

Enfin, avec l'introduction de mécanismes formels, notamment l'approche sémantique, les wikis deviennent des systèmes d'information relativement universels. Des expérimentations ont démarré autour de la faisabilité de fonctions bibliographiques ou d'observatoire.

### 5.4.2 Premières observations

Préoccupés par les questions relatives aux collections numériques c'est dans cet esprit que nous nous sommes rapprochés de la mission TICRI pour monter un observatoire culturel académique distribué, où plus exactement un ensemble d'observatoires à caractère régional, institutionnel ou thématiques. Comme nous l'avons exprimé précédemment une même collection peut en effet être décrite dans ces différents environnements, avec des caractéristiques différentes. Par exemple, une collection concernant un point particulier de la contribution d'un physicien français entre 1870 et 1914 n'aura pas le même environnement relationnel dans un observatoire relatif à la société française ou européenne à cette époque, dans une histoire de la physique ou dans l'inventaire des collections académiques de la Lorraine. Nous voulons également dépasser la notion d'inventaire systématique pour nous attacher à mettre en valeur les points originaux d'une collection parmi un ensemble. Ceci implique une cohabitation d'informations rédactionnelles et structurées. Les wikis sémantiques apportent un élément de réponse technologique à cette question.

Nous avons d'ailleurs démarré en Lorraine une expérimentation sur un réseau de wikis sémantiques sur laquelle nous voudrions nous appuyer appelée WICRI.

L'outil WIKI a été choisi pour ses deux caractéristiques :

1. à savoir une appropriation aisée par ses contributeurs
2. une qualification documentaire des informations

Pour répondre à QUI fait quoi dans l'Université ? la plateforme WICRI5 traite de façon sémantique des données numériques en respectant les métadonnées RDF.

### 5.5.2 Wikis sémantiques

Dans ce réseau, un même objet par exemple, une description de projet d'un pôle de recherche, n'est pas décrite de façon identique sur un wiki à vocation universelle (type wikipédia), sur ceux des différents laboratoires participants, ou dans le wiki du pôle. Ces objets partagent bien des propriétés communes mais qui s'expriment différemment. Par exemple, la relation d'un acteur de type laboratoire et un projet sera implantée différemment dans le wiki du laboratoire, dans celui du pôle où il est clairement identifié et dans celui d'un autre laboratoire ou seules les autres universités sont connues (et non leurs composantes). Cet exemple fait apparaître le besoin d'outils d'assistance à la gestion de la distribution d'objets sur un réseau (et la nécessité de s'appuyer sur des métadonnées et sur des vocabulaires normalisés). Nous voudrions donc pouvoir mettre en scène des descriptions de collections dans différents contextes:

1. Un wiki "à vocation universelle" (WICRI)
2. Des wikis institutionnels (par exemple : laboratoire, MSH, Bibliothèques Universitaires et inventaire culturel académique régional)
3. Des wikis thématiques (les collections de l'histoire de la physique et les collections de la fin du XIX<sup>ème</sup>)

---

5 [http://maquettewicri.loria.fr/fr.wicri/index.php5?title=Wicri:DRRT\\_Lorraine](http://maquettewicri.loria.fr/fr.wicri/index.php5?title=Wicri:DRRT_Lorraine)

4. Enfin au niveau d'une collection où sont décrits les éléments de la collection. En effet ces descriptions d'objets numériques vont pouvoir servir d'illustration qui accompagnera les descriptions de collection.

La constitution de ces wikis passent par des phases de modélisations : modélisation des acteurs, modélisation des collections.

## **6. Mise en valeur d'un patrimoine académique par la visualisation des liens virtuels reliant objets-institutions-collections**

### **6.1 Modélisation des acteurs et/ou institutions**

La modélisation rendue nécessaire s'appuie sur la modélisation de la dimension humaine représentée par les différents acteurs pour concevoir un wiki sémantique. La modélisation reflète une image réelle du système. Pour être performant, un système d'information doit tenir compte des besoins spécifiques de chaque acteur. Dans [46] nous avons fait émerger des besoins, des fonctions et des activités propres à des types d'acteurs d'un système d'information pour proposer de représenter l'utilisateur ainsi  $RU = (T, B, F, A)$ , où : T est le type d'acteur, B sont les besoins, F sont les fonctions et A sont les activités des acteurs. Voici comment qualifier leurs types, leurs besoins, leurs fonctions et leurs activités.

- T = {personne, institution, domaine}. Par ce biais l'acteur est relié à son institution mettant en relief une ou des thématiques
- B = {s'inscrire, être identifié, être visible} en tant que contributeur.
- F = {écrire, créer, intégrer, vérifier, contribuer, corriger} fonctions qui évoluent en parallèle des nouvelles fonctionnalités offertes par le système.
- A = {déposer, explorer, interroger, analyser, synthétiser, annoter} témoignent de l'évolution des activités prises en compte dans le cadre des technologies web social.

Cette représentation de l'utilisateur aboutit au modèle formel RUBI<sup>3</sup> {Représentation des Utilisateurs et de leurs Besoins en Information lors de l'Interrogation après Identification}. Au travers des contributions se tissent des expressions entre les acteurs de même types ou de types différents en relation avec des domaines relatifs à des collections. Des niveaux institutionnels ou des niveaux thématiques sont ainsi mis en évidence par les recherches et les productions des acteurs : Qui écrit avec qui ? Sur quoi ? Ce principe de coopération dynamique favorise un passage de l'information à la connaissance.

### **6.2 Modélisation des collections par les objets numérisés et/ou à numériser**

A un moment où de nombreux projets prennent forme autour de la numérisation il est intéressant de réfléchir à la modélisation d'objets à numériser. Aborder le processus de numérisation par une approche qualitative permet de poser des questions sur :

- Les usages faits des produits numérisés : formation, recherche, valorisation, technologie, culture
- La valorisation des objets : les types d'indicateurs, les types de chercheurs
- La qualification des droits juridiques : mentions des juristes, identification des problèmes juridiques, prise en compte des droits, propriété intellectuelle, soutien au niveau juridique
- Les propriétés de l'objet : source de recherche, format, forme
- Le contexte de la ressource : publics cibles, domaines, densité sémantique, couverture

La modélisation des acteurs en relation avec leur institution et la modélisation des objets numériques en relation avec leur collection dégagent des classes d'objets décrits à l'aide de descripteurs. Ces données descriptives constituent des métas données. C'est bien par l'intermédiaire de ces métas données que nous pouvons favoriser un modèle d'intégration au sein d'un wiki sémantique à la destination d'acteurs identifiés. La finesse d'élaboration des métas données autour des collections et des acteurs avec la prise en compte de leur contexte permet de mettre en valeur un patrimoine académique par la visualisation des liens virtuels reliant objets-institutions-collections.

## 7. Conclusion : Une dynamique de recherche expérimentale

Pour réellement permettre l'identification de fonds particuliers et leur appropriation par divers publics le projet prend appui sur une thématique privilégiée. Ce choix n'exclue en rien d'autres thèmes mais crée une hiérarchie en fonction de l'état d'avancement du projet. Nous développons un outil qui est en même temps objet de recherches et plateforme d'actions en direction du grand public. Sur un plan intellectuel le caractère innovant repose sur la visualisation des liens sémantiques. Il relie à divers environnements des objets numériques pour apporter une valeur ajoutée à la recherche d'information. Parallèlement aux données numériques décrites à l'aide de leurs métadonnées et à leur évaluation nous nous intéressons à la métamodélisation d'un environnement WIKI sémantique. Ce projet rapproche des chercheurs du monde informatique, du monde de la cognition, des spécialistes des thématiques retenues et des spécialistes du patrimoine culturel. Le réseau WICRI initié à un plan régional pourrait nous aider à monter un consortium au niveau national en cas de succès. Ceci en réponse aux questionnements sur les répercussions institutionnelles induites par la recherche au travers de patrimoines culturels.

## 8. Bibliographie

1. Afolabi B., *La conception et l'adaptation de la structure d'un système d'intelligence économique par l'observation des comportements de l'utilisateur*, Thèse Université Nancy 2, Mars 2007.
2. Afolabi B., Thiery O. *Using Users' Expectations to Adapt Business Intelligence Systems*, in G. Budin, C. Swertz et K. Mitgutsch (Eds.), *Advances in Knowledge Organization (ISKO)*, vol. 10, Ergon Verlag, Würzburg, Austria. p. 247 - 254. June 2006.
3. Amerouali Y., *Metadata et Profil Utilisateur*, 2000 [www.cais-acsi.ca/proceedings/2000/amerouali\\_2000.pdf](http://www.cais-acsi.ca/proceedings/2000/amerouali_2000.pdf).
4. Atkins D, *Blue-Ribbon Advisory Panel on Cyberinfrastructure*, Rapport de la « National Science Foundation », 2003 < [http : //www.nsf.gov/od/oci/reports/toc.jsp](http://www.nsf.gov/od/oci/reports/toc.jsp)>
5. Besson E. 2008, [http://www.premier-ministre.gouv.fr/chantiers/croissance\\_847/developpement\\_economie\\_numerique\\_france\\_59678.html](http://www.premier-ministre.gouv.fr/chantiers/croissance_847/developpement_economie_numerique_france_59678.html)
6. Bibliothèque Nationale de France. *Guide d'interopérabilité OAI-PMH pour un référencement des documents numériques dans Gallica*. 2008
7. Boyer A., Nominé B., *Managing new educative technology in a medium size university*, In 20th ICDE World Conference on Open Learning and Distance Education. (Düsseldorf, Germany), 2001.
8. Chartron G., *Les chercheurs et la documentation numérique : nouveaux services et usages*, Paris, Cercle de la librairie, 2002.
9. David A., *Modélisation de l'utilisateur et recherche coopérative d'information*, HDR, Nancy 2, 1999.
10. David A., Sidhom S., *Intégration de la démarche d'Intelligence Économique dans l'architecture fonctionnelle d'un système d'information*, CERIST, 2004.
11. David A., Thiery O., *Prise en compte du profil de l'utilisateur dans un système d'information stratégique*, in veille stratégique scientifique et technique - VSST'2001, Barcelone Octobre 2001.
12. Day, Michael (UKOLN). *Mapping Dublin Core to UNIMARC*, 1999 [en ligne] [http://www.ukoln.ac.uk/metadata/interoperability/dc\\_unimarc.html](http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html) (consulté le 30 septembre 2008)
13. Delefosse M., Bouaziz F., *Gestion des métas données dans un entrepôt de données : état de l'art et implantation dans les logiciels actuels*, projet bibliographique M2 Miage ACSI, décembre 2008.
14. Dhérent C., Journée d'information AFNOR CG 46, 7, Bibliothèque nationale de France, juin 2005.
15. Desnos J.-F., *Projet Entrepôt de données*, 2002, [http://www.amue.fr/Telecharger/seminaire\\_pilotage\\_mars2002/J.F.Desnos.pdf](http://www.amue.fr/Telecharger/seminaire_pilotage_mars2002/J.F.Desnos.pdf).
16. Ducloy J., Hermann A., Mussot P., Thomesse J.-P. WICRI : *un réseau de wikis sémantiques pour les communautés de la recherche et de l'innovation*. AMETIST 2009.

17. Duffing G. , Thiery O., *Gestion et qualité de l'information stratégique : une approche par les risques des systèmes décisionnels*, Congrès AFME2008, Grenoble, mars 2008.
18. EPCC ARTECA 2008, *Les amphis de France 5*. NUMLOR#.1 [http://www.canalu.tv/producteurs/les\\_amphis\\_de\\_france\\_5/dossier\\_programmes/art/culture\\_et\\_europe/numlor\\_1](http://www.canalu.tv/producteurs/les_amphis_de_france_5/dossier_programmes/art/culture_et_europe/numlor_1)
19. *Esup portail : Environnement numérique de travail d'accès intégré aux services pour les étudiants et le personnel de l'enseignement supérieur*, <http://www.esup-portail.org/>.
20. Franco J.-M., *Le data warehouse : le data mining*, Paris, 1997.
21. Franco J.-M., *Le Datawarehouse : objectifs, définitions, architectures*. Eyrolles, 1997
22. Gabriel A., Ohayon, E., *Les outils décisionnels : description de l'offre commerciale et Open Source*. 2007 [http://www.loria.fr/~peguiron/html/profile\\_fset.html](http://www.loria.fr/~peguiron/html/profile_fset.html)
23. Gabriel A., Ohayon E., *Projet bibliographique : guides d'utilisations des outils décisionnels*. 2007 [http://www.loria.fr/~peguiron/html/profile\\_fset.html](http://www.loria.fr/~peguiron/html/profile_fset.html)
24. Goglin J.-F., *La construction du datawarehouse : du datamart au dataweb*, Paris, 2001.
25. Green IT Business et Solutions IT - *Toute l'actualité; de l'internet et du marché; IT* - ZDNet.fr <http://www.zdnet.fr/actualites/0,39051260,3800005277t-4000084997q-1+10,00.htm>
26. Inmon-William.-H., *Building the data warehouse*, New York, 2002.
27. Tony Hey *eScience and Digital Scholarship, séminaire INRIA*, [En ligne] <http://www.msr-inria.inria.fr/events-news/tony-hey-seminar> (Consulté février 2008 )
28. Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P., *Fundamentals of Datawarehouse*, Springer 2002.
29. Kimbal R., Caserta J., *The data warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, 2004
30. Kislin P., David A., Peguiron F., *Caractérisation des éléments de solutions en recherche d'information : conception d'un modèle dynamique dans un contexte décisionnel*, ISKO2003, Grenoble.
31. *L'Essentiel d'Unified Modeling Language (UML)*, [http://madchat.org/coding/other/CSI\\_UML\\_2003.pdf](http://madchat.org/coding/other/CSI_UML_2003.pdf).
32. Le Henaff D., Ducloy J., Ducasse J.-P. , *Métadonnées pour une cyberinfrastructure de la recherche*, 2006.
33. Lévêque E. 2009, Nathalie Kosciusko-Morizet, *nouvelle secrétaire d'Etat à l'Economie numérique*, publié le 15 janvier 2009 <http://www.zdnet.fr/actualites/internet/0,39020774,39386604,00.htm?xtor=RSS-1>
34. Malleray E., *Meta données et analyses multidimensionnelles à travers les hypercubes*. 2008 [http://www.loria.fr/~peguiron/html/profile\\_fset.html](http://www.loria.fr/~peguiron/html/profile_fset.html).
35. Martre, H. *Intelligence économique et stratégie des entreprises*, Rapport du Commissariat Général au Plan, Paris, La Documentation Française, 1994.
36. Ministère de la Culture et de la Communication, Direction du Livre et de la Lecture. Nawrocki, François, *Le protocole OAI et ses usages en bibliothèque*, 15 février 2005 [[En ligne] <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> (consulté le 30 septembre 2008)
37. Ministère de la Culture et de la Communication, Mission de la Recherche et de la Technologie / projet européen Minerva. Foulonneau, Muriel. *Collaborer pour de nouveaux services culturels en ligne : Le protocole OAI*, janvier 2004 [En ligne] [http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/guide\\_oai.pdf](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/guide_oai.pdf) (consulté le 30 septembre 2008)
38. Muller P.-A., *Modélisation objet avec UML*, Eyrolles 1997.
39. Nancy-Université, 2008. *Projet de Développement Quadriennal 2009-2012 : Volet 4 Ressources documentaires*
40. Nataf J.-B., *Structure de l'entrepôt de données de pilotage*, 2001, <http://www.cpu.fr/Telecharger/NatafP6StructureEntrepotDeDonnees.pdf>.
41. Nicolas R. 2008, *Plan Besson : 154 mesures pour faire décoller la France numérique*. [http://www.svmlmag.fr/actu/03558/plan\\_besson\\_154\\_mesures\\_pour\\_faire\\_decoller\\_la\\_france\\_numerique](http://www.svmlmag.fr/actu/03558/plan_besson_154_mesures_pour_faire_decoller_la_france_numerique)
42. Nominé B., *ESUP portail : espace numérique de travail pour tous*, Nancy 2004.

43. Peguiron F., 2008. *L'intelligence économique au service des acteurs de l'université, La question du partage de l'information sur les campus*. Ed L'Harmattan
44. Peguiron F., 2009. *Politiques des bibliothèques publiques en France et en Europe : normes ou nouvelles tendances ?* SIEE2009. Hammamet Tunisie
45. Peguiron F., Thiery O. *Analyses multidimensionnelles de contenus documentaires dans un ENT au service de l'acteur enseignant chercheur*, Colloque International sur le Document Electronique (CIDE 10), Nancy INIST, Juillet 2007.
46. Peguiron F., Thiery O. *Un entrepôt de données vu comme une base de connaissance intégrant la modélisation des acteurs : application au système d'information universitaire*. ISKO 6ème colloque du chapitre français, Toulouse, Juin 2007.
47. Parrendt P., Frenot S., Hohn S., *Privacy-Aware Service Integration*, 2007.
48. Perenon P., *ProfilDoc : un Système de Recherche d'Information scientifique*, 2006.
49. Revelli C., *Intelligence stratégique sur Internet : comment développer des activités de veille et d'intelligence économique sur le web*, 2000.
50. Rochfeld A., Morejon J., *La Méthode Merise, Tome 3, Gamme opératoire*, Editions d'Organisation, 1989.
51. Rumbaugh J., et al., *Modélisation et conception orientées objet*, Masson, 1995.
52. Salençon J. et Moatti A. 2008, *Rapport du comité IST*
53. *Schéma directeur des espaces numériques de travail, Ministère de la jeunesse, de l'éducation nationale, et de la recherche*, 2004, <http://www.educnet.education.fr/chrgt/SDET-v1.doc>.
54. Teste O., *Modélisation et manipulation d'entrepôts de données complexes et historisées*, <http://www.irit.fr/recherches/IRI/SIG/personnes/teste/these/intro.pdf>.
55. Thiery O., David A., *Modélisation de l'utilisateur : systèmes d'informations stratégiques et intelligence économique*, Revue association pour le développement du logiciel (ADELI), 2002.
56. Welshons M. *Our Cultural Commonwealth*. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. [En ligne] <http://cnx.org/content/col10391/1.2/> (Consulté le 15 Décembre 2006).