

**Master SIB « Sciences de
l'Information et des Bibliothèques »
Option GISTE « Gestion de
l'information scientifique et technique
dans l'entreprise »**



Université Claude Bernard Lyon 1

Rapport de stage

Restructuration et mise à jour du registre de métadonnées de The European Library

Fesquet Guillaume

Sous la direction de

Jill Cousins

directrice de

**The European Library Office
c/o the Koninklijke Bibliotheek
National Library of the Netherlands
PO Box 90407
2509 LK The Hague
The Netherlands**

Stage de 4 mois

Année 2006 / 2007

Restructuration et mise à jour du registre de métadonnées de The European Library – Fesquet Guillaume

Résumé : au cours de mon stage au cœur du projet de portail d'interrogation des collections des bibliothèques nationales européennes The European Library, j'ai pu me rendre compte à quel point les projets de bibliothèques numériques foisonnaient de par le monde. Ce rapport tente de présenter de façon simple les bibliothèques numériques et le service The European Library. Cette présentation décrit l'architecture, les services proposés par ces projets et plus particulièrement le registre de métadonnées, composant clé et qui a été au cœur de mon travail durant le stage.

Descripteurs : The European Library, bibliothèque numérique, métadonnée, registre de métadonnées

Abstract : The European Library is a non-commercial organisation that allows visitors to search through the resources of the national libraries of Europe via its portal. My internship within technical department of The European Library Office gave me the opportunity to realize how broad the concept of digital libraries is. The current report provides an overview of digital libraries in general and of The European Library in particular. I will describe The European Library technical architecture and the added value behind its services. I will also highlight the Metadata Registry environment, key component of The European Library infrastructure and core of my contribution during this internship.

Keywords : The European Library, digital library, metadata, metadata registry

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.

Remerciements

Je tiens à remercier ici l'ensemble des membres de The European Library Office : Aubéry, Eric, Fleur, Georgia, Jill, Julie, Olaf, Sally et Sjoerd qui m'ont prodigué leurs conseils durant le stage et m'ont aussi fait vivre une expérience humaine très enrichissante.

Sommaire

Introduction	6
I. Les bibliothèques numériques.....	7
A. <i>Qu'est-ce qu'une bibliothèque numérique ?</i>	7
1. Définition	7
2. L'intérêt de créer une bibliothèque numérique	8
3. Les différents types de bibliothèques	9
B. <i>Les types de documents numériques</i>	9
1. Numérisé ou numérique pur	9
2. Vers un contenu multimédia	11
C. <i>Les services proposés aux utilisateurs</i>	11
D. <i>Des points à travailler</i>	12
1. Le contenu	12
2. La navigation et le visionnage.....	12
3. Les services supplémentaires	13
II. The European Library	15
A. <i>Le CENL</i>	15
B. <i>The European Library Office</i>	16
C. <i>Les groupes de travail</i>	16
D. <i>Architecture technique de The European Library</i>	18
1. Formats des données :	18
2. Processus d'échange et de rapatriement des données :	18
E. <i>Les services proposés</i>	20
III. Les métadonnées et les registres de métadonnées	22
A. <i>Origine</i>	22
B. <i>Présentation</i>	22
1. Le Dublin Core.....	22
2. Le registre de métadonnées de The European Library.....	23
C. <i>Evolution</i>	24
IV. Ma contribution à The European Library.....	26
A. <i>Le Metadata Registry</i>	26
1. Prise en main	26

2. Les choix effectués.....	27
3. Le développement	28
4. Les tests	29
<i>B. La réorganisation de la partie statique du portail.....</i>	<i>31</i>
1. Situation actuelle	31
2. Travail réalisé.....	31
<i>C. Remarques générales sur les tâches exécutées</i>	<i>32</i>
<i>D. Le cadre de travail</i>	<i>32</i>
1. Un projet international	32
2. Les outils de travail	33
3. La gestion du projet.....	34
Conclusion	35
Bibliographie.....	36
Annexes.....	38

Introduction

Le petit monde des bibliothèques numériques a connu une véritable explosion depuis 2005 et l'annonce de Google de numériser plus de 15 millions de livres afin de les mettre en ligne. Cette déclaration a eu pour conséquence d'accélérer les différents projets en cours, que se soit publics ou privés. On assiste donc depuis ce moment à une véritable course à l'innovation avec, pour les protagonistes, une multiplication des déclarations de presse et des partenariats.

Mais au-delà des spécialistes de la culture et de l'information et des bibliothécaires c'est l'ensemble de la population qui est concerné par les bouleversements récents dans ce domaine. Car il ne s'agit pas uniquement de livres, mais plus généralement de l'accès à la culture puisque les bibliothèques numériques semblent vouées à devenir un point central pour l'accès à un rassemblement de contenus culturels variés.

Cependant la question de l'évolution des bibliothèques numériques dans le futur ne peut pas seulement être contenue dans des aspects politiques et stratégiques d'accès à l'information. Il faut aussi prendre en compte l'apparition de nouvelles technologies et de nouveaux formats tels que l'OAI-PMH ou le Dublin Core. De plus, l'acceptation par le grand public des différentes plateformes développées ne pourra devenir effective que lorsque celui-ci aura à disposition un panel d'outils suffisamment intéressant et pratique. La confrontation entre un modèle privé et un modèle public se joue donc sur le contenu mais aussi sur les services à valeurs ajoutées qui seront proposés à l'utilisateur.

Le but de ce rapport est de présenter tout d'abord une définition générale des bibliothèques numériques ainsi que des différents modèles en construction, ceci sans perdre de vue les différentes lacunes actuelles.

Ensuite sera présenté le portail The European Library qui constitue une première étape vers la Bibliothèque Numérique Européenne.

Le point commun entre de nombreux projets fédérant plusieurs partenaires se trouve dans le registre de métadonnées. Je tenterais d'en expliquer le fonctionnement au cours d'une troisième partie.

Enfin, dans la dernière partie je développerai le travail que j'ai effectué et notamment ma contribution à la restructuration du registre de métadonnées de The European Library et à la mise à jour de l'interface web. Cette partie sera conclue par un commentaire général sur le déroulement de mon stage.

I. Les bibliothèques numériques

A. Qu'est-ce qu'une bibliothèque numérique ?

Les premières bibliothèques numériques ont été développées au milieu des années 90. Grâce à l'évolution rapide des technologies d'Internet durant cette période les bibliothécaires ont commencé à penser à proposer des ouvrages directement en ligne. Les technologies, les formats, les standards n'étant pas encore assez au point il a fallu attendre le début des années 2000 pour voir apparaître les premières bibliothèques numériques.

L'âge actuel des bibliothèques numériques pourrait être assimilé à l'adolescence, en effet l'évolution est loin d'être finie et de nombreux échelons restent encore à gravir avant de pouvoir vraiment affirmer l'âge adulte des bibliothèques numériques. Cette première partie va tout d'abord tenter de donner une définition aussi générale et universelle que possible de la bibliothèque numérique. Puis suivront des explications sur le contenu de ces bibliothèques tant au niveau des ressources que des services. Enfin les principaux défauts et les principales améliorations à prendre en compte seront présentés.

1. Définition

Une approche généraliste des bibliothèques numériques peut permettre de définir ces entités de la manière suivante : « Une bibliothèque numérique est un ensemble de ressources sélectionnées et organisées, au format électronique, et possédant des fonctionnalités de recherche avancées. Le contenu des ressources peut être accessible directement en ligne ». Pour simplifier, cette première définition donne aux bibliothèques numériques des fonctions de recherches et d'accès au contenu de ressources préalablement choisies. Il est important de souligner le rôle de la sélection et de l'organisation des ressources. Ce point particulier différencie une bibliothèque numérique d'un quelconque moteur de recherche basé sur l'indexation automatique et la recherche sur le web.

Une vision aussi générale mais plus technique de la bibliothèque numérique a été donnée par Oppenheim et Smithson en 1999¹. Il s'agit selon eux d'un « *système d'information dans lequel toutes les ressources d'information sont disponibles sous une forme traitable par ordinateur et dans lequel toutes les fonctions d'acquisition, de stockage, de conservation, de recherche, d'accès et de visualisation utilisent des techniques numériques* ».

Plus concrètement la bibliothèque numérique du futur doit donc permettre à un utilisateur situé n'importe où sur la planète de pouvoir rechercher puis accéder à un objet ou sa représentation numérique tel qu'un livre, une photographie, une vidéo ou même une pièce de collection (de musées par exemple). La notion de bibliothèque étant prise ici dans un sens très large, c'est-à-dire en tant qu'organisation répertoriant et classifiant des ressources afin de les mettre à disposition des usagers.

Cependant, il convient de nuancer ce propos. Une bibliothèque numérique peut en plus de cette fonction primaire proposer beaucoup d'autres services et, ne prendre en compte que les fonctions de recherches et de visualisations des documents numériques, serait ne considérer qu'une partie de toutes les nouvelles fonctions qui sont en train d'émerger grâce

¹ What is the hybrid library? Oppenheim and Smithson *Journal of Information Science*.1999; 25: 97-112 (payant)

aux développements de nouvelles technologies. Cette évolution sera développée plus tard dans la partie des services proposés.

2. L'intérêt de créer une bibliothèque numérique

Les raisons du développement et de la mise en place de bibliothèques numériques sont aussi variées que le nombre de bibliothèques numériques elles-mêmes. Cependant il est possible de distinguer trois grandes raisons et objectifs pour les organismes s'engageant dans de tels projets. Les deux premières, la préservation et la diffusion, ont toujours fait parties du rôle traditionnel des bibliothèques. La troisième raison, apparue avec l'émergence de bibliothèques numériques fondées par des sociétés privées, est l'aspect lucratif des bibliothèques numériques.

- Préservation : le rôle traditionnel des bibliothèques a toujours été depuis leur création de conserver et d'archiver les ouvrages et documents. C'est donc naturellement que les bibliothèques se tournent vers la numérisation de leur fond qui peut ensuite être intégré sur un site Internet. Cela leur permet de sauvegarder les ouvrages fragiles tout en permettant une visualisation par le grand public. D'autre part la préservation est facilitée par le rassemblement en un même endroit de toutes les données chose impossible pour une bibliothèque traditionnelle. Par exemple, la future bibliothèque numérique développée par la BNF sera un point d'accès unique à toutes les ressources collectées.
- Diffusion : après avoir collecté, traité, classifié ces documents une bibliothèque traditionnelle se doit de les mettre à disposition du public le plus large. Les bibliothèques numériques partagent aussi cet objectif, bien que développé de manière différente. Les possibilités des bibliothèques numériques sur ce point sont beaucoup plus avantageuses que pour les bibliothèques traditionnelles limitées par l'espace physique de consultation, les horaires d'ouverture et la disponibilité limitée des ouvrages (nombre fini de copies, impossibilité de rassembler plus d'un certain nombre d'ouvrages). Les bibliothèques numériques possèdent ce grand avantage de pouvoir rassembler les ressources offrant ainsi à l'utilisateur un nombre de documents sans équivalent avec les bibliothèques traditionnelles.
- Aspect Lucratif : l'engouement actuel pour les bibliothèques numériques est en lié à l'annonce faite par Google en 2005 d'un plan de numérisation à grande échelle de livres afin de les mettre en ligne. Cet engagement d'une société privée à but lucratif dans le domaine de la préservation et de la diffusion de la culture a eu pour effet de relancer les projets institutionnels afin de contrer ou du moins proposer une alternative publique à une bibliothèque uniquement basée sur le profit. Le but de Google est en effet de proposer à partir d'une interface unique le plus grand nombre possible de documents. Les revenus de Google et des grands moteurs de recherche proviennent majoritairement de la publicité et donc, plus il y a de contenu avec des formats différents, plus les utilisateurs viennent faire leur recherche sur le site et plus les recettes publicitaires sont importantes.

Finalement, cette évolution vers le numérique et les progrès technologique actuels changent la manière dont les bibliothèques travaillent, mais ne changent pas la raison de leur travail.

3. Les différents types de bibliothèques

Il faut cependant distinguer différents types de bibliothèque numérique. Bien que le but final soit de permettre à l'utilisateur de pouvoir visualiser en ligne des ressources électroniques, la nature et la forme des documents peuvent varier. De plus, pour chaque différent type le modèle économique varie et les objectifs à atteindre ne sont pas tout à fait les mêmes. Les bibliothèques peuvent être de nature :

- purement numérique : les ressources sont uniquement au format électronique et dont tout le contenu est visualisable en ligne.
- hybride² : les ressources sont à la fois numériques et physiques ce qui permet d'avoir accès aux notices descriptives ainsi qu'aux lieux de consultation. Ce modèle garde les avantages de la bibliothèque traditionnelle tout en offrant les fonctions avancées d'une bibliothèque numérique comme premier argument de consultation.
- éducatif : un grand nombre de bibliothèques numériques se sont développées dans le domaine de l'éducation. Elles offrent des cours et des outils pédagogiques tels que des forums pour les étudiants, en plus des ouvrages numérisés et des notices bibliographiques.

B. Les types de documents numériques

Le fond documentaire des bibliothèques numériques est composé de sources hétérogènes et, de plus en plus, les ressources proposées se composent de formats variés. C'est l'une des grandes forces des bibliothèques numériques que de proposer un contenu sous différents formats consultables à partir d'une même interface. A l'image des grands moteurs de recherche qui permettent une recherche sur un grand nombre de ressources différentes, les bibliothèques numériques sont appelées à offrir à l'utilisateur une grande variété de contenu. Ce regroupement de ressources électroniques demande cependant, avant la mise en place, une réflexion approfondie sur les différents types de documents et la manière de les présenter au public. En effet, il s'agit pour les bibliothèques numériques de proposer un contenu d'aussi bonne qualité que dans les bibliothèques physiques.

1. Numérisé ou numérique pur

On entend par documents numérisés les documents existants au format papier et, par le biais d'une numérisation pouvant être accessibles au format électronique. Le document numérisé est donc en fait une duplication de l'ouvrage d'origine. Ceci implique que le document a déjà été traité intellectuellement et qu'ainsi aucune description supplémentaire concernant l'ouvrage n'est à effectuer. Le résultat de la numérisation peut être de deux formes : image et/ou texte.

Dans le premier le cas il s'agit simplement d'une scannérisation du document. Le résultat est l'exacte copie du document d'origine. L'unique manière de faire une recherche sur un ensemble de documents de ce type passe par une interrogation des champs de la notice renseignés par le bibliothécaire.

² *La bibliothèque hybride*, Peter Brophy (<http://bbf.enssib.fr/sdx/BBF/pdf/bbf-2002-4/02-brophy.pdf>)

Le deuxième cas qui semble aujourd'hui le plus répandu et le plus demandé par les bibliothécaires ainsi que les usagers est la numérisation en mode texte. Dans ce mode, le document est scanné puis un logiciel de reconnaissance optique de caractère (ROC) génère le contenu sous format texte. Pour les utilisateurs cela permet de faire des interrogations en plein texte, interrogations qu'ils ont par ailleurs l'habitude d'effectuer sur les moteurs de recherche. En revanche, la qualité de l'océrisation³ est variable suivant les ouvrages. Les ouvrages récents ont une typographie assez régulière ce qui donne un pourcentage de reconnaissance d'environ 95%⁴ (rapport du nombre de caractères sur le nombre d'erreur). Les ouvrages plus anciens, souvent abîmés et avec une typographie moins régulière bénéficient en revanche d'une moins bonne reconnaissance. Un bon pourcentage de reconnaissance pour ces ouvrages s'élève à environ 85% actuellement.

L'océrisation des documents ne signifie pas forcément que la version image n'est plus utilisée. Les deux peuvent en effet être utilisées conjointement. La version océrisée servant pour l'interrogation et le format image pour le résultat rendu à l'utilisateur. Cela permet non seulement de charger les pages plus rapidement puisque le format image est moins lourd que le format texte, mais aussi de faciliter la sécurisation des documents afin d'éviter une reproduction ou une modification trop aisée des pages par les utilisateurs. De plus le format PDF étant très largement répandu la consultation en mode image semble donc tout à fait adaptée pour le visionnage. Ce mode de fonctionnement est actuellement celui qui semble se développer

Cependant la numérisation à grande échelle de documents ne peut aujourd'hui se faire qu'en ayant des moyens financiers importants. C'est pourquoi les bibliothèques de taille moyenne ne bénéficiant pas forcément de l'appui financier nécessaire font de plus en plus appel aux grands moteurs de recherche pour numériser leur fond. En contrepartie, ces derniers obtiennent l'autorisation d'utiliser les documents numérisés pour leur propre service de recherche de livre (Google Book Search, MSN Book Search, etc..). La Bibliothèque Municipale de Lyon a par exemple signé un accord avec Google pour la numérisation d'une partie de ses ressources. C'est également le cas entre la British Library et MSN (Microsoft) ce qui montre bien que ce ne sont pas seulement les bibliothèques de taille moyenne qui ont besoin de l'aide financière venant de sociétés privées pour la numérisation.

D'autre part, le stockage de toutes ces données numériques semble aussi poser problème. Si les bibliothèques deviennent le lieu d'archivage de la culture et du savoir alors la perte de ces données serait désastreux. Avant d'engager un processus de numérisation à grande échelle il faut donc prévoir comment vont être stockées les données. Ceci implique :

- le choix d'une politique globale concernant la façon dont les données doivent être stockées,
- le choix d'un format stable et répandu pour la description des données,
- la définition d'une politique globale pour le choix des ressources numériques qui seront sauvegardées ainsi que le niveau de sûreté accordé à chaque document,
- le choix des responsables et de la répartition des tâches.

Il semble probable qu'à terme les moyens financiers mis en œuvre pour l'archivage soient beaucoup plus importants que l'investissement initial pour la numérisation des documents.

³ Vient de OCR acronyme de Optical Character Recognition.

⁴ <http://www.culture.gouv.fr/culture/marche-public/dapa/cctp-mai2007.pdf>

2. Vers un contenu multimédia

Historiquement, la mission première des bibliothèques numériques était de proposer un catalogue en ligne de documents numérisés à partir d'ouvrages déjà existants au format papier dans les bibliothèques physiques. L'évolution rapide de l'édition électronique permet maintenant de mettre en ligne des documents créés directement au format électronique ce qui évite les frais dus à la numérisation. De plus les utilisateurs sont de plus en plus habitués à naviguer sur des sites offrant des contenus multimédias. L'augmentation très rapide du débit des connexions Internet autorise maintenant à pouvoir visualiser des films ou à écouter du contenu audio sans avoir à télécharger totalement la ressource (streaming). Cette amélioration des connexions est aussi un avantage pour les livres numérisés puisque cela permet d'avoir une interface plus agréable avec des temps de téléchargement des pages plus courts. La liste de tous les types de ressources intégrables est longue, mais ceux dont la mise en ligne est la plus évidente sont :

- textes : monographies, périodiques, publications électroniques scientifiques, archives, cours, thèses, ...
- vidéo : films d'archives, films documentaires, cours, conférences, ...
- fonds sonores : archives sonores, musique,...
- images : images d'archives, photographies de collections de musées, photographies d'art, gravures,...

Bien sûr, la mise en ligne de ces ressources demande au préalable l'autorisation du ou des ayants droits pour les œuvres n'étant pas tombées dans le domaine public.

A terme pourront ainsi être intégrées de nombreuses autres ressources notamment en profitant de la possibilité de partenariat avec d'autres institutions plutôt spécialisées dans la diffusion de contenus musicaux, vidéos ou d'images. Cela demande en revanche d'adapter l'interface à la diffusion de tels contenus (lecteur vidéo intégré, visualiseur d'image), d'adapter le format de description des données de manière à utiliser un schéma unique pour toutes les ressources. S'il est trop difficile d'implémenter un seul format pour tous les types de ressources il faudra alors prévoir des programmes de transfert automatique des données sources vers les données utilisées.

Enfin, comme cela l'a été évoqué dans la partie précédente, les livres non numérisés, les collections des musées ainsi que toutes autres ressources dont la numérisation n'est pas envisageable que ce soit pour des raisons juridiques, techniques ou économiques feront aussi partis des bibliothèques numériques. Seulement, les notices renverront vers le lieu de consultation ou d'exposition de la ressource, tout en fournissant l'information de base la concernant (créateur, date, couverture spatiale, résumé si possible, etc...).

C. Les services proposés aux utilisateurs

Toutefois avec l'avancement des projets il devient de plus en plus évident que cette définition uniquement basée sur l'accès à une ressource ne suffit plus. En effet, les utilisateurs d'aujourd'hui cherchent non seulement à avoir accès à l'information mais aussi à pouvoir partager leur connaissance, à avoir un accès personnalisé et personnalisable et s'attendent à ce que le « software » leur propose des réponses dynamiques en fonctions de leur profil. Pour cela il devient évident qu'une simple interface de recherche et d'accès aux documents n'est

pas suffisante. La maquette de future Bibliothèque Numérique Européenne⁵ développée par la Bibliothèque Nationale de France donne un aperçu de ce que pourrait être la bibliothèque numérique du futur. L'utilisateur peut notamment étiqueter les documents qui l'intéressent et ainsi se créer une mini bibliothèque personnalisée.

Comme le rappelle une étude de la BNF, la future Bibliothèque numérique européenne « doit répondre à quelques critères minimaux : ouverture sur le contemporain, attractivité du produit, rapidité de consultation. Mais, plus que la quantité de documents numérisés qui resteront toujours accessibles par le biais des moteurs existants, les plus-values nouvelles attendues se situent surtout au niveau de services conformes à l'essentiel de la demande »⁶.

D. Des points à travailler

Les bibliothèques numériques sont encore loin d'avoir terminé leur processus de développement, elles ne sont encore qu'à leur phase d' « adolescence ». Elles ne remplissent pour l'instant que leurs fonctions de recherche et de visualisation, mais sont appelées à court terme à fournir des services plus développés et des contenus plus actuels aux utilisateurs.

1. Le contenu

Les ressources présentées dans les bibliothèques numériques sont pour l'instant et dans leur grande majorité dans le domaine public. Les utilisateurs n'ont finalement pas accès à des livres récents car un livre entre dans le domaine public au bout de 70 ans (cas général). Les documents intéressants des bibliothèques numériques actuelles sont en général des documents d'archive ou historiques. Les autres étant des livres libres de droits que l'on peut en général se procurer à moindre coût au format papier. Ce format, même s'il est de mauvaise qualité reste tout de même plus agréable à lire qu'une image sur écran. Les ressources sont donc en grande partie destinées actuellement à un public éclairé représenté par les chercheurs, les enseignants ou les étudiants. Afin de palier à cet inconvénient majeur il semble nécessaire d'ouvrir un dialogue et d'initier des solutions avec les maisons d'éditions. La Fédération Européenne des Editeurs a fait savoir à la commission européenne que tout document sous droits consultables en ligne doit être soumis à rémunération pour les ayants droits. L'exemple le plus frappant de cette difficulté est la mise en veille du processus de numérisation de Google. Le groupe avait en effet entrepris de numériser toutes les œuvres sans prendre en considération si celles-ci étaient encore sous droits. La pression des syndicats d'auteurs et des maisons d'éditions a finalement poussé Google à suspendre son projet dans l'attente d'un accord. Cela pourrait amener à revoir le modèle économique des bibliothèques numériques actuelles et notamment remettre en question le libre accès pour l'ensemble des ressources. Une proportion de 80% de livres dans le domaine public et de 20% encore sous droits semble plausible selon le rapport de la BNF cité ci-dessus.

2. La navigation et le visionnage

La lecture sur écran, encore très difficile, est un autre défaut important qui pourrait freiner l'expansion des bibliothèques numériques et de tous types de projets proposant du contenu textuel. Bien que l'affichage ait beaucoup évolué durant la dernière décennie, le résultat final affiché à l'écran est encore largement en deçà de la qualité d'un livre. A cela

⁵ <http://www.europeana.eu/> maquette de la future bibliothèque numérique européenne proposée par la BNF.

⁶ <http://bbf.enssib.fr/sdx/BBF/pdf/bbf-2006-3/bbf-2006-03-0068-012.pdf> article sur le journal de la BNF

s'ajoute la qualité intrinsèque et le bon état du document s'il a été numérisé. La visualisation d'un livre en mode image laisse encore très largement à désirer.

D'autre part, les bibliothèques numériques actuelles hésitent souvent entre un modèle centralisé et un modèle décentralisé. Le premier ne donne à la bibliothèque que la fonction d'interface entre l'utilisateur et la ou les bibliothèques. L'utilisateur, en choisissant de voir un document est ensuite redirigé vers la bibliothèque possédant la ressource. Le problème de ce modèle est de proposer à l'utilisateur une interface de visionnage différente pour chaque bibliothèque et donc d'amoindrir la qualité de la navigation. Cependant cela donne aux bibliothèques partenaires l'occasion de mettre en valeur leurs fonds.

Le deuxième modèle consiste au contraire à rassembler toutes les ressources sur le site de la bibliothèque numérique et de proposer la visualisation des documents sur le site via une interface et un design unique. Pour l'utilisateur cette solution est la plus adaptée mais pour les responsables du projet elle est plus difficile à mettre en place car cela nécessite non seulement de rapatrier les notices des ouvrages mais aussi le contenu, puis de stocker et de maintenir l'ensemble.

3. Les services supplémentaires

Actuellement, les bibliothèques numériques proposent uniquement aux utilisateurs des fonctions de recherche et de visualisation des documents. Les résultats obtenus ne sont en fait que ceux qui correspondent à la requête formulée. Une avancée majeure pourrait être de proposer à l'utilisateur non seulement les résultats de sa recherche mais aussi des ressources connexes liées aux documents retournés par le sujet ou la concordance sémantique d'une part et par la concordance linguistique des ressources d'autre part.

Le premier point nécessite de créer des cartes ou des modèles de représentation des connaissances afin de pouvoir récupérer l'information connexe pertinente. Typiquement cela peut être envisagé, en partant du point de vue de l'utilisateur, en utilisant les connaissances que celui-ci peut apporter au système. Les sites communautaires actuels fourmillent d'applications de ce genre permettant de marquer (« tagger ») une ressource et de créer ainsi une cartographie de l'information, vue par l'utilisateur. Un autre moyen serait de développer des logiciels de traitement automatique de la langue travaillant sur un corpus de texte ou sur les métadonnées. Ainsi l'utilisateur bénéficierait de l'affichage de ces ressources connexes reliées entre elles par concordance de sens.

Le deuxième point correspond en fait à un grand chantier dans le monde de la diffusion de l'information multilingue. Il s'agit, pour résumer, de renvoyer comme résultats les ressources dans toutes les langues possibles correspondant à la requête formulée au départ en une langue unique. Mais, à l'instar des moteurs de recherches, les technologies ne sont pas encore au point dans ce domaine il est vrai très complexe. Un autre moyen pour parvenir à résoudre partiellement ce problème serait d'utiliser un identifiant unique pour chaque ressource numérique⁷. Ce travail de réflexion est actuellement en cours dans le Technical Working Group de The European Library. Cet identifiant pourrait définir chaque oeuvre de façon unique quelle que soit sa langue, sa forme ou sa bibliothèque d'origine. Ainsi, une oeuvre présente en plusieurs langues ou formats dans le catalogue numérique aurait pour chacune des versions le même identifiant, ce qui permettrait de les relier entre elles. D'autre

⁷ <http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8> , rapport du CERL sur les « persistent identifier ».

part cet identifiant donnerait aussi l'avantage de pouvoir éliminer les doublons dus à l'agrégation de plusieurs collections ou catalogues.

Il devrait aussi être possible, et de nombreux sites web actuels le montrent bien, pour l'utilisateur de pouvoir personnaliser l'interface de navigation. Cela voudrait dire sauvegarder les préférences dans le design, les caractéristiques des recherches à effectuer avec par exemple une possibilité de mémoriser le choix des collections, le format des documents, le domaine souhaité. Mais cela pourrait aussi permettre à l'utilisateur de faire partie d'une communauté d'utilisateurs ayant les mêmes centres d'intérêts. Cette dernière fonction serait fort utile dans le domaine scientifique dans lequel les chercheurs ou étudiants échangent des informations sur un même domaine. Dans ce même but de partage et d'enregistrement des préférences, il serait intéressant de développer des systèmes d'annotations des documents afin de faire profiter les autres membres du groupe des remarques et de se souvenir, pour la personne concernée, de la raison pour laquelle elle avait sélectionné ce document.

II. The European Library

Contrairement à ce que son nom pourrait laisser supposer le projet The European Library n'est pas à proprement parler une bibliothèque numérique. Bien qu'une partie des ressources soit des documents numérisés il ne s'agit en aucun cas d'un site permettant de les visualiser. The European Library peut se définir comme une interface permettant d'effectuer des recherches simultanées sur les catalogues et collections des bibliothèques nationales européennes. A cette interface s'ajoute des services proposés aux utilisateurs formant ainsi la valeur ajoutée du site.

Le projet The European Library s'inscrit dans une volonté européenne de favoriser la diffusion de la culture et de la connaissance européenne. Il est issu d'une concertation entre les bibliothèques nationales européenne. Il a fallu tout d'abord réfléchir à la conception d'un modèle viable économiquement et techniquement parlant.

De plus, le grand nombre de bibliothèques participant au projet a fait que cette concertation était indispensable afin de définir les rôles précisément et de mettre en place des processus de travail efficaces. Après trois années de travail et de réflexion sur le type de projet à développer, le bureau The European Library Office a été créé. Les objectifs donnés à ce bureau étaient de développer un portail d'interrogation des collections mises à disposition par les bibliothèques et de faire le lien entre les bibliothèques. Ce deuxième point est très important car pour un projet de cette envergure une gestion centralisée et efficace de la communication semble obligatoire. Après un an de développement et de rassemblement des premières collections la première version du portail web a finalement vu le jour au début de l'année 2005. Il convient ici de préciser quelque peu le vocabulaire utilisé. A partir du moment où le portail de The European Library a été fonctionnel il ne s'agissait plus d'un projet mais d'un produit ou services. Le terme projet n'est utilisé ici que pour désigner le mode de fonctionnement.

A. Le CENL

Le projet The European Library est la propriété de CENL (Conference of European National Library) mais est financé par les bibliothèques participantes. Comme son nom l'indique le CENL regroupe toutes les bibliothèques nationales européenne. Afin de contribuer au projet chaque bibliothèque fournit une contribution financière à hauteur de son budget. Cependant les bibliothèques peuvent choisir entre une participation totale au projet (full participant) et une participation partielle pour celles qui ne voudraient pas ou ne verraient pas l'utilité de s'engager plus avant (basic participant). Les frais de participation prennent en compte les moyens financiers dont disposent les bibliothèques. Ceux-ci étant très inégaux, ce fonctionnement permet l'adhésion des plus petites d'entre elles.

Le nombre de bibliothèques participantes est actuellement de 47 dont 23 participent complètement et donc les 24 restant sont participants partiels. Le 1^{er} juillet 2007 9 des participants partiels devraient devenir participants complets ce qui va donc porter leur nombre à 32.

Cependant actuellement l'accès aux collections et l'interrogation en ligne ne se fait que sur les catalogues et collections des bibliothèques participants pleinement au projet. Les autres ne pouvant que mettre en avant les « trésors » de leur collection dans une partie spécialement consacrée du site. Les collections de ces bibliothèques ne sont donc pas interrogeables.

L'ensemble des représentants des bibliothèques forme le comité de direction du projet. Ce comité décide des orientations du projet. Cependant afin de permettre un travail et une réflexion plus efficace, un comité exécutif plus restreint a été créé afin d'améliorer le processus décisionnel. Des réunions d'au moins 47 personnes ne sont en effet pas le meilleur moyen pour prendre des décisions rapidement (organisation des réunions, prise de décisions). Ce bureau travail restreint est composé de 4 membres.

B. The European Library Office

Le bureau de The European Library est hébergé par la Koninklijke Bibliotheek, la bibliothèque royale des Pays-Bas. The European Library Office bénéficie de toute l'infrastructure et de tous les services internes à l'institution. Notamment, les services informatiques sont hébergés et maintenus par le département informatique de la bibliothèque. C'est également le cas pour les services financier et logistique. Les membres de The European Library Office ont donc uniquement à travailler sur le ou les projets et n'ont pas à se soucier de maintenance ou de logistique. La seule tâche « externe aux projets » pour certaines personnes du bureau consiste à coordonner The European Library Office et les services de la bibliothèque.

The European Library office est composé actuellement de 9 personnes dont :

- 1 directrice, responsable des projets,
- 3 personnes pour la partie technique (dont une responsable),
- 2 personnes pour le management général des projets,
- 1 responsable du marketing et de la communication,
- 1 responsable éditorial,
- 1 responsable des finances et de l'interopérabilité.

Le bureau se compose donc à moitié de personnes chargées du développement et de l'édition du contenu de The European Library et l'autre moitié est attachée à la gestion/communication du projet. Cette égalité reflète bien le fait que The European Library est non seulement un service technique innovant mais aussi un projet mettant en œuvre un nouveau type de coopération internationale basée sur une mise en commun des ressources humaines ou matérielles à l'aide d'une gestion centralisée.

La présence d'une personne exclusivement réservée au marketing et à la communication semble un point intéressant à remarquer. Cela montre à quel point la nécessité de communiquer sur ce produit est importante pour faire face aux grands moteurs de recherche présents sur le marché de la diffusion de contenus culturels en ligne. Bien sûr il ne s'agit pas ici de vendre le portail, mais plutôt de présenter le produit et de profiter des canaux de diffusion du web tel que les groupes de discussion (Google Groups), les listes de diffusion ou des institutions spécialisées dans le domaine de l'information comme par exemple l'ABF ou l'ADBS en France. Une lettre d'information bimestrielle est par ailleurs diffusée sur abonnement.

C. Les groupes de travail

Les domaines de recherche étant très variés des groupes de travail spécifiques à un domaine ont été mis en place. Chaque groupe se compose de spécialistes du domaine ou de personnes ayant des connaissances dans le mode de fonctionnement en projet. Bien plus que

des groupes ayant simplement un avis consultatif sur les décisions prises par le comité exécutif et The European Library Office, ces groupes de travail ont une fonction de recherche et sont souvent en relation avec d'autres projets similaires. Ils ont aussi pour but de créer un lien informel entre les bibliothèques et The European Library Office.

Technical Working Group :

Groupe très important au sein du projet, son rôle est de définir et de proposer l'application de normes reconnues dans l'échange de données et sur l'architecture générale du système. Il a aussi pour rôle de promouvoir les technologies et standards utilisés dans le service afin que celui-ci soit interopérable avec le plus grand nombre d'autre projet. Il va sans dire que sur ce groupe repose l'ossature du projet et des futures décisions techniques qui feront ou non que The European Library soit reconnue. Le Metadata Working Group ainsi que d'autres groupes plus récents sont des sous groupes de ce groupe.

Metadata Working Group :

Le Metadata Working Group, groupe avec lequel j'ai collaboré, est responsable de la mise en place et de la mise à jour du « metadata registry ». Ce groupe est composé de 4 membres issus des bibliothèques et de deux membres venant de The European Library Office. Son petit nombre de participants permet des réunions assez fréquentes ainsi qu'une bonne efficacité. Ce mode d'organisation en comité restreint fonctionnant plutôt bien d'autres groupes de travail pensent à le reprendre.

Library Contacts Group:

Les personnes appartenant à ce groupe sont chargées de créer un lien entre leur bibliothèque respective et The European Library Office. Ils sont chargés de coordonner les activités pour leur bibliothèque, de surveiller le site en faisant participer les membres de leur bibliothèque et enfin de fournir les descriptions de collections. Du côté de The European Library Office ils sont chargés de faire des propositions concernant le fonctionnement général du projet mais aussi des propositions plus précises telles que des nouveaux services. Enfin ils doivent « surveiller » les décisions prises par les partenaires avec leur « œil » de bibliothécaire.

Editorial Working Group :

Ce groupe est responsable du maintien et de la mise à jour du contenu éditorial du site. Il peut faire des propositions concernant des nouvelles parties à mettre en place.

Languages Working Group :

Chargé de la traduction et de la vérification des pages éditoriales, de l'interface de recherche, des métadonnées et ce dans toutes les langues représentées dans le projet, ce groupe peut aussi faire des propositions concernant le traitement multilingue de l'information et notamment sur la recherche multilingue qui constitue un point important du projet.

The European Library étant en soi un service donnant accès à des services délocalisés, cette organisation en groupes de travail se prête particulièrement à son fonctionnement. Ainsi The European Library Office peut agir comme une plateforme mettant en lien entre eux les différents groupes de travail, les bibliothèques et les autres partenaires institutionnels. Cette fonction de gestion demande en revanche beaucoup de temps et c'est pourquoi, comme nous le verrons par la suite chaque personne du bureau est chargée d'établir des relations avec toutes ces organisations.

D. Architecture technique de The European Library

Les résultats obtenus sur le portail de The European Library sont la finalité d'un processus technique assez complexe. L'hétérogénéité des techniques et des formats mis en place dans les bibliothèques partenaires ne permet pas d'utiliser un protocole d'échange de données unique. The European Library Office a donc du développer des processus d'échanges bien définis en accord et avec le soutien des groupes de travail impliqués dans ces domaines. Ainsi afin d'obtenir un système viable, compatible et interopérable avec l'ensemble des participants les points suivants ont été particulièrement travaillés :

- formats des données
- protocoles pour l'échange des données
- traitement multilingue de l'information
- interface utilisateur

Lors du téléchargement de la page l'utilisateur récupère en fait de nombreuses données autres que la page html :

- la description des collections par défaut dans un fichier XML,
- le fichier des traductions.

Ces données sont ensuite utilisées avec JavaScript. La combinaison de ce langage et de fichiers XML s'appelle AJAX. L'application de cette technologie dans le portail sert à pouvoir traduire les pages sans utiliser de ressources côté serveur (intérêt de JavaScript) ainsi qu'à connaître les paramètres de connexion des bibliothèques cibles par défaut.

1. Formats des données :

Afin d'éviter des problèmes liés à une trop grande variété de formats The European Library demande à ses partenaires de fournir leurs descriptions en accord avec le registre de métadonnées qui a été développé et suivant le schéma du profil d'application. Ce registre de métadonnées est basé sur le Dublin Core qui est un format très utilisé. Je reviendrais plus loin sur le registre de métadonnées qui a été au cœur du travail que j'ai effectué durant mon stage.

Pour les bibliothèques il s'agit donc d'exporter leurs notices au format MARC (ou natif MARC) vers le format Dublin Core qualifié, au format XML du registre de métadonnées. Ces métadonnées sont ensuite moissonnées via le protocole OAI-PMH et stockées dans un index central.

Cependant, la conversion de toutes les notices vers un autre format ayant une architecture différente est un travail long et fastidieux et demandant des moyens techniques, humains et financiers assez important.

2. Processus d'échange et de rapatriement des données :

a) Les protocoles d'échange

La diversité des sources oblige The European Library à utiliser plusieurs protocoles différents.

Tout d'abord le protocole SRU qui est le seul compréhensible par le portail et par lequel toutes les requêtes partant de celui-ci sont effectuées. Ce protocole sert à interroger les bibliothèques sources via ou non un gateway (défini ci-après) et à récupérer des données dans

le central index. Les paramètres de la requête figurent dans l'URL demandée ce qui allège le poids de la recherche.

Le protocole Z39-50, traditionnellement utilisé dans les bibliothèques ne sert ici que pour faire un lien entre le gateway et les bibliothèques cibles. Ce protocole est encore utilisé car il est très répandu dans le monde des bibliothèques et est en général assez stable. Il faut cependant que le serveur cible soit disponible au moment de la demande et que le nombre de connexions ne soit pas trop important.

L'OAI-PMH est un protocole développé au départ pour la récupération de métadonnées dans les projets d'archivages du web. Grâce une syntaxe d'interrogation connue des 2 parties il est possible de récupérer les notices dans un format prédéfini (ici conforme au profil d'application) sur le serveur du partenaire. Le moissonnage des données est réalisé régulièrement et l'index est ensuite interrogeable par SRU lors de la requête de l'utilisateur.

b) Gateway (passerelle)

Le gateway est un des éléments clé du dialogue entre le portail et la bibliothèque cible si celle-ci ne dispose pas de serveur supportant le protocole SRU. Le gateway réalise la conversion de la requête SRU vers le protocole Z39-50 compréhensible par la plupart des systèmes. Ensuite les résultats au format MARC (ou dérivés) sont transformés au format XML tel que défini par le registre de métadonnées et conforme au profil d'application. Le gateway sert aussi à faire la conversion des formats d'encodage des caractères vers ceux supportés par le portail. Cette fonction peut parfois être compliqué à mettre en place car les catalogues utilisent quelques fois des encodages assez vieux ou plusieurs à la fois.

Le gateway peut-être mise en place et hébergé par l'organisation partenaire, mais c'est souvent celui de The European Library qui est utilisé pour réaliser les actions précédemment citées.

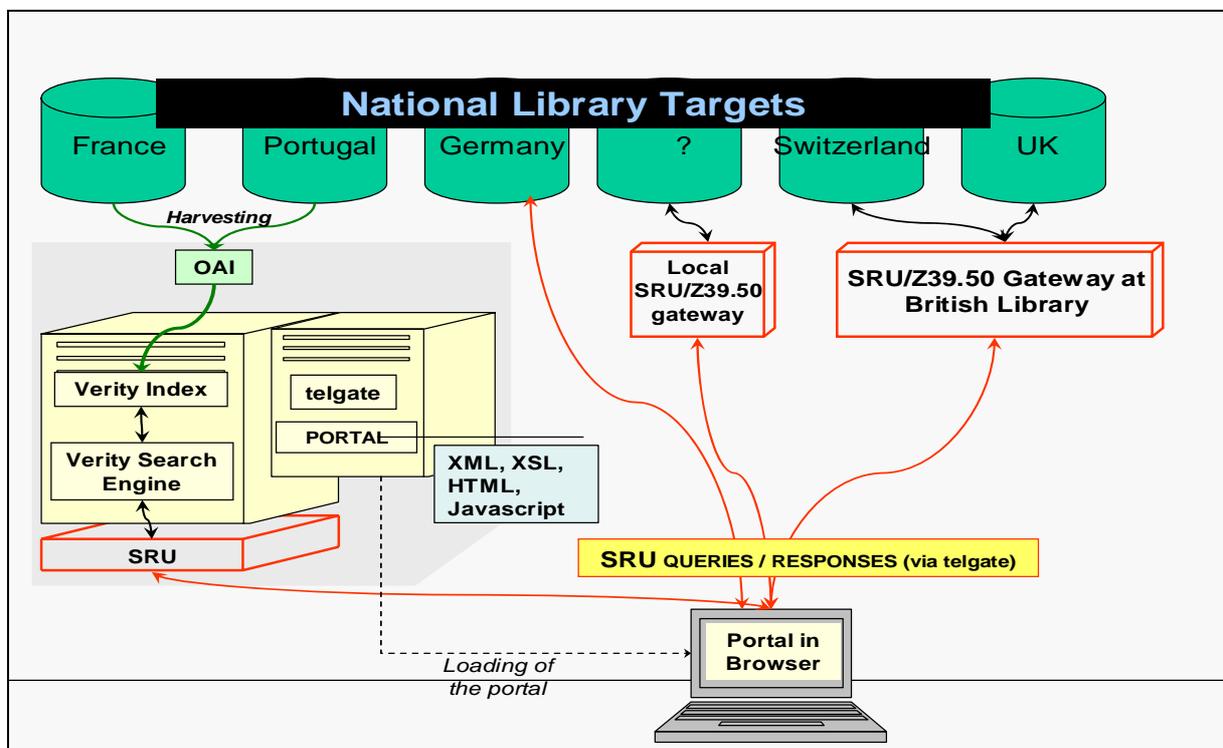


Figure 1: Architecture de The European Library

L'intérêt de l'architecture de The European Library est qu'elle repose sur des formats et des protocoles très répandus. De plus, il est plaisant d'imaginer qu'à partir de ces processus d'échanges bien définis il soit possible d'automatiser toute la chaîne de soumission des métadonnées descriptives des ressources. Ceci comprend la description par les partenaires, la validation de la description par The European Library puis l'indexation.

Mais cette automatisation semble difficile à mettre en place notamment à cause de la difficulté de validation des descriptions. La description doit être conforme au schéma utilisé, l'encodage des caractères doit être correct, les URL doivent être valides et bien encodées. De nombreux points sont encore difficiles à automatiser. Ce travail de validation est finalement réalisé par les développeurs de The European Library Office, ce qui constitue un coût en temps considérable. Pour chaque nouvelle collection soumise le développeur responsable du partenaire doit vérifier que les données sont bien structurées et que les informations transmises sont valides.

Le constat est le même pour le projet de National Science Digital Library aux Etats-Unis qui pensait pouvoir créer un processus automatique de moissonnage des métadonnées puis de leur validation. Mais les membres du projet se sont heurtés à la difficulté d'une telle mise en place, bien que leur architecture soit basée sur le Dublin Core et l'OAI-PMH qui sont des technologies avec un très haut niveau d'interopérabilité.

E. Les services proposés

La fonction première du portail est l'interrogation des collections fournies par les bibliothèques participantes. La recherche ne se fait actuellement que sur un nombre limité de collections car toutes ne sont pas encore interrogeables. D'autre part, la recherche par défaut ne se fait que sur un nombre limité de collections. Pour le choix de celles-ci, chaque bibliothèque « full participant » propose une et une seule collection de son catalogue pour être interrogée par défaut. A ces collections s'ajoute la propre base de données de The European Library appelée aussi « central index » et ne contenant que des items numérisés. Pour pouvoir définir plus précisément les collections à interroger, il est possible d'accéder à une interface de recherche avancée dans laquelle il est possible de choisir :

- les collections à interroger,
- le type de document : livres numérisés, périodiques, manuscrits, cartes, etc...,
- les champs à interroger avec l'utilisation de booléen.

Les résultats obtenus sont ensuite affichés par collections. Le croisement des collections n'est donc pas encore effectué ce qui pénalise l'utilisateur en recherche d'informations. En effet, toutes les collections doivent être visualisées avant de pouvoir être sûr de choisir l'information pertinente. Après avoir obtenu les résultats, l'utilisateur peut, si une version numérisée est disponible, visualiser le document sur le site de la bibliothèque proposant la ressource, sinon un lien vers la notice est donné.

A cette fonction de base s'ajoute des services supplémentaires tels que le lien vers le livre sur un site marchand, la gestion des notices récupérées afin de créer des références bibliographiques. Par ailleurs, depuis peu, une mini barre de recherche est proposée aux personnes en faisant la demande. Cet outil peut s'installer sur n'importe quelle page web et permet d'interroger le portail à partir d'un site extérieur.

Enfin, il convient aussi de voir comme un service la possibilité de pouvoir visualiser l'interface de recherche dans 21 langues différentes. Les autres pages sont en général

disponibles dans les 3 langues principales : allemand, anglais et français. Pour ce faire The European Library Office a demandé à tous les participants de fournir la traduction des pages dans la langue de leur pays. Ces pages sont divisées en parties ayant un contenu textuel et marquées par un identifiant permettant de remplacer la portion de texte dans la page par la traduction fournie (avec l'aide de JavaScript).

Les services présentés jusqu'ici sont uniquement présents sur le portail principal. Cependant, il est aussi possible de concevoir le Handbook comme un service puisqu'il constitue un guide pour les partenaires désirant proposer des collections. Le Handbook⁸ est un site contenant des explications sur la manière de préparer des métadonnées, soumettre des collections, gérer des collections soumises. The European Library met également à disposition des partenaires un exécutable pour installer automatiquement un serveur OAI-PMH.

Dans le même domaine des services connexes le « metadata registry⁹ » tient aussi une part importante, comme il va l'être démontré par la suite.

⁸ <http://www.theeuropeanlibrary.org/handbook> site en accès libre

⁹ <http://www.theeuropeanlibrary.org/metadataregistry> site en accès libre

III. Les métadonnées et les registres de métadonnées

A. Origine

Le registre de métadonnées ou « metadata registry » est une pièce centrale du service The European Library. Son utilité dépasse néanmoins le cadre de ce projet. En effet, que ce soit dans les bibliothèques « physiques » ou numériques le principal objectif est la diffusion de la connaissance. Or, face au danger possible d'un stockage hétérogène de l'information à partir de formats différents et sans interopérabilité entre eux il devient nécessaire de définir un socle commun pour décrire l'information.

C'est à partir de cette idée que s'est développé le registre de métadonnées de The European Library. La principale utilisation est donc pour le moment le projet actuel mais nul doute que si le registre s'avère suffisamment complet et puissant pour décrire le grand nombre de types de ressources présent dans le projet alors son utilisation pourra être étendue à un grand nombre de projet de bibliothèque numérique similaire.

La description des ressources en utilisant une même structure avec des vocabulaires prédéfinis permet à The European Library de récupérer des informations venant des toutes les bibliothèques européennes. Il devient ensuite aisé de traiter ou d'analyser ces données qui se retrouvent présentées de manière identique.

Cependant, la création d'un tel registre, utilisable par toutes les bibliothèques quel que soit leur pays d'origine, leur langue, leur culture n'est pas un projet évident. Bien que la plupart des notices bibliographiques existantes soient basées sur le format MARC, il est bien connu que pratiquement chaque pays a développé sa propre version. De plus, la grande hétérogénéité des ressources qui peuvent être des livres, des magazines, des photographies, des gravures, des vidéos ou autres complique encore la tâche. Le choix de métadonnées descriptives ne peut se faire qu'en prenant en compte toutes ces variables. Le registre de métadonnées doit donc être assez souple afin d'englober le plus grand nombre possible de ressources et en même temps assez strict pour que la description des ressources soit de qualité.

B. Présentation

A partir d'un format de description de ressources très répandu : le Dublin Core, le Metadata Working Group de The European Library a mis en place un registre de métadonnées pour décrire les ressources bibliographiques. Après avoir fait un rappel des origines et des principales caractéristiques du Dublin Core, le registre de métadonnées de The European Library sera abordé en détail.

1. Le Dublin Core

Le Dublin Core est un format de description de données dont les grandes lignes ont été dessinées en 1995 lors d'un workshop réunissant des experts dans la description de ressources. En constante évolution depuis sa création le Dublin Core est de plus en plus utilisé dans tous les projets mettant en jeu l'accès et l'échange de grandes quantités d'informations relatives à des ressources documentaires.

Le Dublin Core définit les éléments nécessaires pour décrire une ressource qu'elle soit numérique ou non. La grande hétérogénéité des objets à décrire implique que le format soit le plus large possible et permette de décrire celles-ci le plus simplement possible. Il est difficile d'imaginer un créateur lambda d'un site web décrire son site avec autant de finesse qu'un bibliothécaire décrirait un livre. Afin de répondre à cette problématique plusieurs variantes du Dublin Core ont été développées.

La plus simple, le Dublin Core permet de décrire une ressource par ses principaux attributs : le nom du créateur, le titre, l'année de création, l'état actuel des droits sur la ressource, etc. Au total 15 éléments principaux sont présents dans cette version simplifiée.

Ensuite pour des ressources plus complexes il est possible d'utiliser le Dublin Core qualifié. La base est la même que pour la version simple et se compose des 15 éléments de description principaux. Chaque élément peut ensuite être décrit plus finement grâce à des éléments de raffinement. Par exemple l'élément **title** peut être raffiné par l'élément de raffinement **alternative** qui permet de décrire un titre alternatif (parallèle) à la création. Ainsi pour chaque élément il est possible de décrire plus strictement la ressource et ses propriétés.

A cela s'ajoute la possibilité d'utiliser des vocabulaires contrôlés et des schémas d'encodages. Ceux-ci permettent, lors de l'exploitation de la description de la ressource de connaître par avance le format utilisé lors de description du contenu. Un exemple très parlant est l'élément **date**. Celui-ci peut être décrit de nombreuses manières : en chiffre, en lettre, dans des ordres différents. En général à chaque façon de décrire le contenu correspond une norme développée par un organisme officiel ou reconnu. En définissant le schéma d'encodage ou le vocabulaire contrôlé on permet à la « machine » qui va lire la description de la ressource de bien afficher le contenu dans l'interface avec l'utilisateur, et de permettre un traitement sûr de l'information.

2. Le registre de métadonnées de The European Library

Dans le registre de métadonnées de The European Library ces métadonnées sont reprises et regroupées en profils d'application. Un profil d'application permet de définir l'ensemble des métadonnées, le schéma d'application de ces métadonnées, les relations entre elles afin de décrire un certain type de ressource. Actuellement, dans le registre, sont utilisés deux profils d'application : un pour décrire les objets et un pour décrire les collections. Même si de nombreuses métadonnées sont communes aux deux, certaines ne peuvent être utilisables que dans un seul profil. Par exemple le terme **isPartof** qui permet de définir l'inclusion d'une sous collection dans une autre collection ne peut être utilisé que dans le profil d'application pour les collections. Ce système permet d'autre part de proposer à l'utilisateur des définitions et des guides d'utilisation différents suivant le profil ce qui autorise une description plus fine des métadonnées. A chaque profil d'application correspond donc un schéma de structure des données. L'exemple ci-dessous avec le terme **Collection** permet de mieux saisir l'utilité d'une telle séparation :

Profil	Profil d'application pour objet	Profil d'application pour collection
Exemples de champ		
Definition	A subject or topic associated with the object.	A subject or topic associated with the items in the collection.
Guidelines	The TEL Office recommends that you use the top-level classifications.	The subject index used by TEL when no encoding scheme is specified is the Dewey Decimal Classification.

Table 1: exemple d'utilisation des différents profils

En revanche l'utilisation de profils d'application multiplie le travail de description de chaque métadonnée descriptive par le nombre de profils d'application. Cette multiplication des descriptions ne facilite pas la tâche du ou des responsables du registre ni celle de l'utilisateur qui doit bien comprendre le registre avant de pouvoir l'utiliser correctement. D'autant plus qu'à l'avenir le MWG envisage de développer un troisième profil d'application pour la description des services des bibliothèques. Ce profil permettra de décrire tous les paramètres techniques nécessaires afin de pouvoir interagir avec bibliothèques partenaires lors de l'interrogation de celles-ci pour récupérer des notices. Cette description des services est actuellement incluse dans le profil d'application pour les collections. Pour une bibliothèque soumettant plusieurs collections, ce qui est généralement le cas, il y a donc actuellement redondance d'information puisque la description est réalisée pour chaque collection que soumet la bibliothèque.

C. Evolution

Le metadata registry est un élément assez complexe à prendre en main. Il nécessite d'avoir une connaissance approfondie des métadonnées utilisées et des relations entre elles. Afin de faciliter le travail de validation pour les partenaires il est évoqué de créer un schéma de validation en XSD (XML Schema Definition). Ce schéma de validation a le même but qu'une DTD (Document Type Definition) mais permet une description plus précise des relations entre les éléments de la structure XML. Par exemple il est possible de définir un nombre exact de sous éléments fils et non pas uniquement 0, 1 ou plus comme avec les DTD. D'autre part la syntaxe de XSD est similaire à XML contrairement à celle des DTD.

Actuellement les bibliothécaires réfléchissent à la mise en place de formats de description tel que les FRBR ou RDF permettant de décrire une ressource non pas en tant que simple item indépendant mais en tant qu'ensemble d'item reliés entre eux par le contenu. Il apparaît donc que certaines évolutions devraient s'opérer afin de faire évoluer les métadonnées descriptives. Comme expérimenté dans le projet National Science Digital Library¹⁰ aux Etats-Unis, un registre de métadonnées basé uniquement sur l'item ne permet pas d'exploiter toute la richesse possible.

D'autre part, en juin 2007 va démarrer le projet European Digital Library qui prendra en compte les archives, les collections de musées et tout le patrimoine culturel européen. Le

¹⁰ <http://arxiv.org/ftp/cs/papers/0601/0601125.pdf> , rétrospective sur le projet NSDL.

registre de métadonnées utilisé sera le même que pour The European Library. Le registre actuel va donc devoir étendre son domaine à d'autres types de ressources. Le Dublin Core est assez souple pour être utilisé dans des domaines aussi variés. En revanche, il faudra s'assurer que le registre de métadonnées développé par The European Library puisse donner à tous les partenaires les garanties d'une description assez fine tout en gardant un niveau de compréhension et d'utilisation le plus accessible possible. Etant donné que des langages de description existent déjà pour les archives (Encoded Archival Description), la question se pose de savoir s'il est possible de fusionner les deux ou du moins de reprendre des termes descripteurs.

Enfin, le registre de métadonnées a besoin d'une évolution importante concernant l'aspect multilingue de la description des termes utilisés. Suivant différents autres projets¹¹ il semble en effet qu'un registre traduit dans la langue de travail est plus facile à utiliser et est mieux utilisé. Cela vient en parti du fait que beaucoup de descriptions sont assez techniques et abstraites et donc difficilement manipulables. Actuellement seul le nom est traduit ce qui ne ne donne souvent pas beaucoup d'informations sur l'utilisation pratique du terme.

¹¹ http://www.idealliance.org/papers/dx_xmle03/papers/03-01-03/03-01-03.pdf article présentant l'intérêt d'utiliser des registres de métadonnées multilingues.

IV. Ma contribution à The European Library

A. Le Metadata Registry

Les origines, le contenu et l'intérêt de développer un registre de métadonnées ont été développés dans la troisième partie. Le principal objectif de mon stage était de permettre aux partenaires du projet de proposer de nouvelles métadonnées ou de faire des commentaires sur des métadonnées existantes. Ces actions sont typiquement réalisables par le biais de formulaires. Je vais présenter maintenant de quelle manière j'ai actualisé le site.

1. Prise en main

Les deux premières semaines de mon stage ont été consacrées à comprendre le fonctionnement général du projet The European Library, des différents projets européens avec lesquels il est en rapport. Au premier abord les projets européens consacrés à la diffusion du savoir et de la connaissance par le biais du numérique semblent tous se ressembler. Entre tous les acronymes qui se contenant souvent les mêmes lettres, les sous projets, les projets de préparation à des projets, la situation paraît très complexe. Par exemple, pour un néophyte il est difficile de différencier le projet EDL d'EDLnet ou d'EDLplus ou de bien faire la distinction entre The European Library et TEL-ME-MOR. Il m'a fallu aussi quelque temps avant de bien comprendre l'organisation du CENL et des différents groupes de travail présentés dans le deuxième chapitre.

D'autre part, j'ai du en même temps me documenter sur les métadonnées et les registres de métadonnées déjà développés et sur la structure de celui de The European Library. A chaque réunion du Metadata Working Group est rédigé un compte rendu contenant les actions effectuées, à effectuées ainsi que les différentes tâches pour chaque membre. Cette documentation m'a permis d'être rapidement au fait des évolutions du registre et de ses orientations. De plus, deux personnes du bureau faisant partie du MWG je pouvais toujours avoir des réponses rapidement.

Parallèlement à cela, j'ai commencé l'étude du site du registre de métadonnées mis en place avant mon arrivée. La précédente base de données contenant le registre était basée sur une structure XML et contenue dans un seul fichier. Chaque terme était décrit par un schéma XML dont la structure est présentée en annexe. L'affichage sur le site était ensuite réalisé avec l'utilisation de XSLT (XSL Transformation). La base du site avait été développée par un professionnel engagé pour deux mois. Le code est utilisé conjointement par le Metadata Registry et par le Handbook. Celui-ci étant très clair et très facile à utiliser la question ne s'est même pas posée de savoir s'il fallait continuer avec ou changer complètement le code de base du site.

En fait, les menus et sous menus du Handbook et du Metadata Registry sont gérés à partir d'un fichier XML puis automatiquement intégrés avec PHP. Les en-têtes et pieds de page sont automatiquement générés grâce à un code PHP. Le contenu est intégré à partir de pages html contenant simplement la partie centrale de la page (sans header, sans menu et sans pied de page). Les personnes chargées de la rédaction du contenu n'ont donc qu'à intégrer un petit morceau de XML dans le menu ou bien à modifier le contenu d'une page html pour mettre à jour le site. Le système mis en place s'apparente donc à un CMS (Content Management System) puisque seule la partie de la page contenant le contenu est à mettre à jour.

Bien sûr un tel code n'est pas simple à écrire...et donc à lire et à comprendre. Les deux premières semaines n'ont pas été de trop pour me familiariser avec les scripts. Mais cela m'a en revanche permis d'améliorer mes connaissances de PHP, en quelques sortes le code m'a servi de cours pour de nombreuses notions et m'a appris les bonnes pratiques de rédactions (clarté du code, commentaires, ne pas répéter les mêmes fonctions).

2. Les choix effectués

Suite à cette étape nécessaire avant la phase de développement j'ai donc entamé une réflexion avec Julie Verleyen (responsable technique du projet et membre du MWG) concernant la structure du nouveau registre et du site. La principale difficulté rencontrée a été de faire un choix entre un registre contenu dans un fichier XML comme précédemment ou dans une base de données relationnelles

Les deux ont des avantages et des inconvénients et le choix a été difficile. Cependant en faveur de la base de données on note une possibilité d'évolution plus facile. Or le MR est censé intégrer un nouveau profil d'application pour décrire les services à moyen terme. La structure relationnelle semble donc bien se prêter à cette organisation en profil. De plus, la génération de formulaire automatiquement est plus facile à réaliser avec une base de données et justement le site utilise de nombreux formulaires. Il a donc été choisi de réaliser le nouveau registre avec une base de données relationnelle.

La suite logique de ce choix a été de développer un modèle entité/association pour décrire le registre. Plusieurs modèles ont été envisagés mais c'est celui à mi-chemin entre une structure très abstraite avec beaucoup de relations et d'entités et une structure reprenant simplement le fichier XML dans la base de données. Le premier étant un peu trop compliqué pour être exploité et le deuxième ne profitant pas pleinement des avantages procurés par l'utilisation d'une base de données relationnelle.

Le choix d'utiliser le SGBDR PHPmyAdmin pour réaliser la base s'est imposé naturellement puisque celui-ci était déjà installé et utilisé pour d'autres applications. Il a en outre l'avantage d'être libre et de très bonne qualité.

En même temps il a fallu préparer le contenu des différents formulaires, c'est-à-dire définir les champs et leur type (liste de valeurs fermée, texte, entier) ainsi que le format des données saisies et réorganiser le menu en conséquence. La description des différents utilisateurs et des droits inhérents a aussi été réalisée durant cette phase. Le tableau ci-dessous a été très utile pour résumer les droits d'accès et pour la présentation de mon travail au MWG.

Nom d'utilisateur	Définition
Invité	Utilisateurs non enregistrés et ayant accès uniquement à la description des termes. ex. : visiteur lambda
Utilisateur inscrit	Utilisateur possédant un mot de passe et nom d'utilisateur unique et ayant accès à une partie des formulaires pour proposer ou modifier de nouveaux termes. ex. : Futurs partenaires dans le projet EDL (musées, archives) ou personne faisant la demande.

Nom d'utilisateur	Définition
Partenaire de The European Library	Partenaires du projet ayant accès aux mêmes contenus qu'un utilisateur inscrit (pour le Metadata Registry). ex. : BNF
Membre du MWG	Un membre du MWG ayant tous les droits concernant la gestion des termes : création, modification, suppression.
Administrateur	L'administrateur du site et de la BDD ayant les mêmes droits qu'un membre du MWG ainsi que la gestion des utilisateurs : création, suppression, changement du mot de passe. Cette personne doit être un membre de The European Library Office.

Table 2: Les utilisateurs du MR et leurs droits d'accès

D'autre part, au cours de l'implémentation de la nouvelle version du site de nouvelles fonctionnalités pour les utilisateurs se sont petit à petit intégrées. Notamment la possibilité de faire une recherche de termes¹² en utilisant certains champs de description (titre, définition du terme) ainsi que d'un historique des modifications pour chaque terme.

3. Le développement

La phase de développement a ensuite pu commencer. Il m'a fallu au total près de 3 mois pour achever le travail qui m'a été confié. Je me suis aidé de nombreuses fois de ressources Internet et notamment de forums qui constituent un formidable réservoir de connaissances et aide bien souvent à débloquer des problèmes.

Les développeurs de The European Library travaillent en général avec un logiciel de travail collaboratif spécialisé pour les projets informatiques appelé SVN Tortoise. Ce logiciel libre permet la gestion des versions et la mise à jour des modifications. En fait, un « repository¹³ » (dossier de base) est créé et ensuite chaque développeur travaille sur son propre dossier qui est lié au repository. Le développeur peut ensuite mettre à jour sa version avec les actions effectuées par les autres membres ou mettre à jour la version du repository.

Chaque décision importante que j'ai pris a été fait en accord avec Julie Verleyen et parfois avec les autres membres du site qui donnaient leur avis sur des points techniques. De plus, chaque évolution ou changement a été noté et figure dans au moins un document transmis aux membres du MWG. Cela permet d'être sûr que tout le monde est bien au courant des orientations choisies.

Mis à part le fait d'apprendre réellement à développer en PHP les principaux problèmes auxquels je me suis heurté sont les suivants :

Le passage des données d'une page à l'autre dans un formulaire :

Les formulaires développés font en général plus d'une page mais le script PHP est le même. Par exemple, le formulaire pour soumettre un nouveau terme contient quatre étapes mais ne s'exécute qu'à partir d'un seul script. Il faut donc à chaque étape transmettre les

¹² cf. annexe avec les copies d'écran

¹³ Je n'ai pas trouvé de traduction pertinente pour ce terme, j'ai donc gardé l'original.

données saisies à l'étape suivante. J'ai résolu ce problème en utilisant des champs html POST cachés contenant les valeurs des champs renseignés.

La validation et le nettoyage des données saisies par l'utilisateur :

Cette opération est très importante non seulement pour que les données saisies aient un sens, mais aussi afin d'éviter des problèmes de sécurité qui pourraient rendre le site instable.

La validation du formulaire contient deux volets. Le premier consiste à n'autoriser la soumission de l'étape que si tous les champs obligatoires ont bien été remplis. Le deuxième consiste à vérifier le format de certaines données entrées tel que la date, l'adresse mail ou une URL. Cette vérification se fait à l'aide d'expressions régulières. La validation de formulaire étant un problème récurrent chez les développeurs, il est facile de trouver les expressions régulières correspondantes sur des forums ou des sites. Il a été un moment envisagé de valider le formulaire en utilisant Java Script. Cela aurait été une très mauvaise solution car ce langage est exécuté côté client et est aisément modifiable par les utilisateurs en ayant une bonne connaissance.

Chaque valeur saisie est aussi nettoyé à l'aide de fonctions PHP ou « manuellement ». Les principales actions réalisées sont la suppression des espaces superflus, l'échappement des guillemets avant l'insertion dans la base car sinon ils seraient interprétés (plus connu sous le nom d'injection SQL), la suppression ou la conversion de caractères non autorisés tel que « < » et « > » pour éviter l'insertion de script dans la page html.

La gestion des authentifications :

Offrir aux différents utilisateurs des niveaux d'accès différents était une des grandes évolutions de cette nouvelle version du MR et du Handbook. Grâce à l'utilisation de session PHP et d'un formulaire d'identification les utilisateurs peuvent maintenant accéder aux données adaptées à leurs besoins. Cela permet aussi l'ouverture du site au grand public ce qui n'était pas le cas auparavant car les parties privées étaient inaccessibles.

Le fonctionnement est très simple, après authentification les droits d'accès de l'utilisateur sont stockés dans une variable de session située côté serveur. Ensuite, à chaque appel de page sécurisée une vérification est effectuée afin d'être sûr que le visiteur possède bien les droits nécessaires. Les noms des utilisateurs, les mots de passe et les droits d'accès sont stockés dans la base de données.

4. Les tests

Après 2 mois de travail une première ébauche de site a été réalisée. Le site a été développé sur un serveur accessible à distance ce qui a permis de faire une présentation de celui-ci lors du meeting du MWG fin mars à Londres. Le but était de faire une démonstration des nouvelles fonctionnalités ou de l'amélioration des fonctionnalités existantes et de valider les choix pris concernant la structure du site et de la description des métadonnées. Le retour a été plutôt bon même si des changements ont dû être effectués notamment concernant l'affichage et la navigation sur le site.

Après ce premier test j'ai amélioré l'affichage, les explications, la navigation sur certaines pages et fini les scripts qui ne possédaient pas toutes les fonctionnalités désirées. Par exemple les formulaires de soumission de nouvelles métadonnées ne permettaient pas d'afficher un résumé des informations saisies et de pouvoir rééditer les différentes étapes. J'ai donc dû développer un système donnant la possibilité à l'utilisateur de revenir en arrière et de modifier les valeurs.

Ensuite a eu lieu un deuxième test, début mai, réalisé par une société spécialisée dans le conseil aux entreprises ou institutions développant des sites Internet. Pour ce test il a fallu rédiger un document (en anglais bien sûr) comprenant toutes les spécifications fonctionnelles du site. Suite à ce rapport la société de test a réalisé un document présentant tous les cas de d'utilisation qu'elle allait effectuer. Ce document permet d'une part de fixer un identifiant à chaque cas pour les rapports d'erreur et d'autre part de vérifier que les deux parties sont d'accord sur les vérifications à effectuer et que les différentes fonctions ont bien été comprises. La société a testé :

- la structure du site,
- le design du site,
- la compréhension générale pour un nouvel utilisateur,
- l'affichage correct des résultats à l'aide de valeurs tests,
- l'affichage correct des messages d'erreur et leur compréhension,
- l'affichage de caractères spéciaux,
- la présence de toutes les fonctionnalités nécessaires (bouton de retour, bouton de remise à zéro,...).

Le test a été réalisé à la fois pour le portail The European Library et pour le Metadata Registry et en conséquence était prévu depuis quelques temps. J'ai donc du ajuster mon emploi du temps pour que le site soit « presque » terminé pour le test. Mais certaines pages n'étant pas encore tout à fait prêtes quelques problèmes sont apparus. C'est la difficulté de faire un test généralisé. D'autant plus que ne connaissant pas très bien PHP je ne pouvais estimer le temps nécessaire qu'il me fallait pour terminer le site. Cela étant, la vérification a été correctement réalisée et m'a permis de fixer de nombreux défauts ou failles, près d'une trentaine en tout, allant de la faute d'orthographe à des erreurs de programmation concernant l'insertion de données dans la base. La société possède un logiciel¹⁴ permettant aux testeurs de rapporter des erreurs avec des commentaires, des captures d'écran et aux développeurs de consulter les erreurs puis si elles sont fixées de les marquer comme résolues. Ce système très pratique permet d'effectuer, en même temps que le test se déroule, les corrections nécessaires et ainsi de pouvoir tester à nouveau la partie concernée.

Le 16 mai 2007 a eu lieu le lancement officiel du site. Le délai me semble assez court entre le test et la mise en ligne, la totalité des erreurs n'a pas pu être corrigée avant le lancement, d'autant plus que certaines pages ont dû être modifiées en profondeur.

Suite à la mise en ligne nous avons présenté le site au MWG afin qu'il donne son avis. La présentation a eu lieu au cours d'un meeting virtuel par le biais de Breeze, un logiciel de visioconférence. Ce logiciel permet d'afficher à l'ensemble des intervenants des documents tels que des présentations PowerPoint et même de partager son écran. Cette dernière fonctionnalité nous a permis de faire une démonstration du site en direct aux membres du MWG tout en faisant des commentaires. Mis à part quelques problèmes de son l'expérience a été très concluante. Nous avons ainsi pu recueillir les différents avis et mettre à jour quelques défauts.

Enfin à la suite de ce troisième test j'ai pu corriger les « derniers » défauts du site. Et la version définitive sera mise en ligne fin juin. Cependant dû au manque de temps de la personne devant mettre à jour le contenu du nouveau Metadata Registry les métadonnées ne

¹⁴ Logiciel Jira, pour plus d'informations : <http://www.atlassian.com/software/jira/>

sont pas encore complètement insérées à ce jour. Le contenu n'est pas totalement actualisé et les utilisateurs doivent utiliser la précédente base en XML qui a été intégrée au site pour le besoin.

B. La réorganisation de la partie statique du portail

1. Situation actuelle

Le site The European Library se compose de 5 parties principales¹⁵ :

- l'interface de recherche,
- la description des collections,
- la description des bibliothèques participantes,
- les trésors des bibliothèques,
- l'organisation du projet.

L'histoire du projet a fait que la partie « organisation » du site se trouve sur un serveur différent des autres pages. Il en résulte une grande confusion dans le développement ou la mise à jour des pages. Une conséquence importante est aussi la duplication importante d'un grand nombre de données ce qui entraîne forcément à moyen terme une perte de l'homogénéité du site. En outre les pages de la partie « organisation » n'utilisent pas l'une grande force de PHP qui est l'inclusion de code dans une page html. Ainsi il est possible de factoriser les en-têtes et pieds de page qui sont quasiment identiques sur chaque page. Par ce biais, si l'éditeur doit changer une partie de l'en-tête dans la page html, au lieu de devoir modifier à chaque fois les 139 fichiers que contient la partie « organisation » il suffit de changer une fois le fichier « header » et la modification s'effectue sur toutes les pages. Cette évolution était d'autant plus nécessaire à réaliser que la partie dynamique du site utilise déjà ce système. Enfin cette opération a eu le mérite de clarifier quelque peu la structure existante dans laquelle la cohérence entre les fichiers et leur position sur le site n'étaient pas forcément respectés.

2. Travail réalisé

Le travail a donc consisté à déplacer les pages d'un serveur à l'autre, puis de réorganiser la structure de la partie « organisation » et enfin de reconstruire les pages en enlevant l'en-tête et le pied de page pour chacune d'entre elles. La réorganisation a été effectuée par le responsable éditorial du site et je me suis chargé de la factorisation des parties communes.

Etant donné le nombre important de pages (139) j'ai décidé de tenter d'écrire un programme en PHP faisant automatiquement le travail avec en paramètre le nom des dossiers contenant les fichiers à modifier. Le programme « parse » les fichiers et enlève les parties en-tête et pied de page tout en ajoutant l'inclusion du code PHP. Ceci implique que les fichiers soient bien balisés et de façon identique car le programme reconnaît le début et la fin des parties grâce à certaines balises prédéfinies. Je dois dire que pour réaliser cette opération les cours de Perl de 1^{ère} année m'ont beaucoup servi ainsi que l'étude réalisée en deuxième année dans le cours de Veille et IST qui m'a remis les choses en mémoire.

¹⁵ <http://www.theeuropeanlibrary.org>

Entre le moment où nous avons commencé l'étude des opérations à faire et la production de la nouvelle structure s'est déroulé environ 3 semaines dans lesquelles 4 jours ont été passés pour les opérations techniques. Le temps passé à faire ces modifications aurait pu être économisé en organisant plus logiquement cette partie dès le début de sa construction.

C. Remarques générales sur les tâches exécutées

La totalité de mon stage a donc porté sur le développement d'interfaces web utilisant PHP. J'ai pu grâce à cette expérience avoir un regard critique par rapport à la formation en master sur ce sujet. Il me semble que l'accent n'a pas été assez mis en cours sur les possibilités de générer des pages web dynamiques grâce à ce langage, ainsi que le fait de pouvoir intégrer une ligne de PHP dans du code html. Mais il est vrai que les 3 cours dispensés sont trop peu nombreux en regard de la masse d'informations que nécessite une compréhension satisfaisante du langage. D'autre part, il me semble qu'un cours un peu plus approfondi sur des notions d'architecture d'un système d'information aurait été fort utile et ce pas seulement pour le développement (en première ou deuxième année). L'utilisation et la compréhension de l'architecture d'un intranet ou de la relation client/serveur est nécessaire dans de nombreux domaines (mise en place d'un logiciel de GED, d'outils de communication, ou d'un intranet).

J'aurais cependant bien aimé réaliser une autre tâche qui était importante à mes yeux. Celle-ci aurait consisté dans la réorganisation du système de gestion électronique des documents internes. Cette volonté est fondée sur la difficulté en tant que nouvel arrivant de retrouver une information ou un document au cœur de la masse de dossiers et de fichiers disponible sur plusieurs serveurs. Il m'a fallu presque 2 mois pour bien saisir où se situe l'information pertinente. Le principal disque dur sert à stocker les documents de travail, les archives, des dossiers personnels de travail ou encore des fichiers seuls et ce sans aucune nomenclature normalisée. Je pense donc qu'une réorganisation de ces espaces de travail de manière à obtenir une structure définie et respectée, des nommages de dossiers et fichiers pertinents pourrait être utile. La principale difficulté pour implémenter un nouveau système de gestion des documents est le temps. Il est en effet difficile pour The European Library Office de prendre du recul par rapport aux projets et à la maintenance de ces projets et de se pencher un peu plus sur le fonctionnement général du groupe. Le bureau est presque tout le temps entraîné dans des projets comportant des dates limites.

D. Le cadre de travail

1. Un projet international

Le projet dans lequel j'ai effectué ce stage comportait pas moins de 4 nationalités différentes. Au-delà de l'amélioration de mon anglais j'ai pu découvrir d'autres milieux, d'autres cultures et la vie aux Pays-Bas m'a immergé dans un mode de vie très différent de la France. Même si je trouve n'avoir pas assez progressé en anglais le séjour aura été bénéfique pour ma maîtrise de cette langue et notamment l'anglais parlé. Le fait d'être dans un pays qui parle anglais mais dont ce n'est pas la langue maternelle ne m'a pas aidé à pratiquer en permanence la langue.

Le cadre international et européen du projet m'a aussi permis d'appréhender de manière plus concrète les relations entre des partenaires ne parlant pas la même langue et ne partageant pas la même culture. Ainsi je comprends mieux maintenant l'utilité de développer

des systèmes multilingues en respect avec chaque langue engagée dans le projet. D'une part c'est une preuve de respect vis-à-vis des autres pays de ne pas tout développer en anglais allemand ou français et d'autre part, d'un point de vue plus technique cela permet une meilleure appropriation du système et de ses outils. Je pense notamment au registre de métadonnées qui à long terme devra offrir toute l'interface textuelle de description des termes dans toutes les langues représentées.

Cependant l'anglais reste la langue majeure dans la communication à l'intérieur du projet. Toute la documentation est rédigée en anglais et toutes les réunions se déroulent aussi dans cette langue. Il est donc nécessaire d'avoir un bon niveau afin de pouvoir participer au projet.

2. Les outils de travail

Il est important pour The European Library Office de disposer d'outils de travail efficaces et stables.

Concernant la communication le bureau utilise le logiciel Microsoft Outlook pour la messagerie et la gestion des plannings internes. Bien que parfois un peu difficile à paramétrer et ayant quelques défauts communs avec la suite MS Office (correction automatique, messages d'erreur intempestifs) ce logiciel se prête particulièrement à un travail en mode projet dans lequel de nombreux rendez-vous sont pris et dans lequel il est nécessaire de pouvoir regarder et mettre à jour rapidement son emploi du temps. Cet outil est de plus accessible à distance. Chaque membre dispose par ailleurs d'un téléphone personnel. La plupart des vidéoconférences sont effectuées avec Skype, un logiciel gratuit et très facile à manipuler. Pour les conférences regroupant un plus grand nombre de personnes le logiciel utilisé est Breeze¹⁶. La diffusion de documents peut se faire soit par l'envoi de pièce jointe par mail, soit pour certains groupes de travail via le site Google Docs sur lequel, après inscription, il est possible de gérer des documents de formats différents. Enfin, les groupes de travail le désirant peuvent créer une liste de diffusion réservée aux membres.

Chaque poste de travail est équipé de la suite Office 2003 et d'un éditeur de texte : UltraEdit, très facile à utiliser et offrant de très bonnes fonctionnalités avancées (gestion des expressions régulières, paramétrages avancés, macros).

Pour la partie marketing, The European Library utilise Google Analytics. La personne responsable dans ce domaine peut ainsi connaître le nombre d'utilisateurs, par page, par partie du site. Il est également possible de suivre le parcours des utilisateurs sur le site, leurs provenances et de nombreuses autres informations concernant le profil mais sans utiliser de données personnelles. En outre, ce logiciel génère automatiquement des graphiques de très bonne qualité.

Etant hébergé par la Koninklijke Bibliotheek, The European Library bénéficie de ses services. En particulier, le service informatique gère complètement tous les problèmes techniques pouvant apparaître et propose sur l'intranet de la bibliothèque un support en ligne. Sur ce même intranet sont également disponibles un annuaire et de nombreux formulaires comme par exemple le remboursement de frais ou toutes autres opérations administratives demandant souvent du temps.

¹⁶ Voir la partie sur la mise en place du Metadata Registry.

3. La gestion du projet

Il est possible de diviser la gestion du projet en deux grandes parties :

- les ressources humaines,
- les relations avec les partenaires et les groupes de travail.

Concernant le premier cas, toutes les deux semaines environ, l'équipe se réunit pour discuter des orientations générales des projets, mais aussi pour présenter le travail de chacun. Cette mise en commun du travail accompli permet à tous les membres d'être au courant des derniers changements et ainsi de pouvoir, lors des meetings ou pour communiquer, être au point sur l'avancement actuel des projets. Toutes les opérations étant reliées entre elles, une connaissance globale est nécessaire. D'autre part, durant ces réunions d'équipe, Jill Cousins, responsable de The European Library Office propose des pistes de réflexion sur des sujets tels que la communication en interne orale ou écrite, ou bien sur la manière dont chacun voit la gestion de projet. D'autres réunions ont lieu à l'extérieur du bureau durant une journée ou une après-midi. Cette gestion de projet rappelle beaucoup le fonctionnement d'une entreprise. Ainsi pour améliorer les performances globales de l'ensemble de l'équipe chaque membre a pu suivre une formation de gestion de projet dispensé par Prince2¹⁷, une société assez reconnue dans le domaine de la méthodologie de projet.

Dans le deuxième cas, c'est-à-dire la relation avec les partenaires et les groupes de travail, chaque rencontre est basée sur un modèle type :

- vérification technique des connexions et des outils si nécessaire,
- présentation des membres,
- rappel du précédent compte rendu et des points évoqués (minutes),
- présentation de l'état d'avancement de chaque action,
- définition des nouvelles actions à réaliser.

Suite à cette réunion, une personne est chargée d'éditer deux documents : les minutes de la rencontre et un compte rendu. La bonne gestion d'un projet d'une telle importance nécessite l'utilisation de règles strictes et normalisées pour que les informations ne soient pas perdues.

¹⁷ Le nom de la société se rapporte à la méthodologie de projet du même nom.

Conclusion

Les bibliothèques numériques ont suscité beaucoup d'engouement, engouement compréhensible par l'espoir d'une diffusion planétaire et pour une fois homogène (en théorie) de la culture. Pour les institutions porteuses de projet, qu'elles soient publiques, privées ou semi-publique toutes sont opposées aux mêmes problèmes de la mise en ligne de contenus sous droits et d'interopérabilité. Leur résolution demandera sûrement du temps et une certaine diplomatie mais nul doute que les différents acteurs finiront par trouver un terrain d'entente autour de ces services qui représentent pour tous une formidable avancée. Les institutions engagées ne devront cependant pas oublier que l'utilisateur doit être au centre de ces nouvelles entités et qu'au final lui seul sera juge de leur qualité. Enfin, les responsables auraient tout intérêt à rassembler leur savoir faire pour développer une bibliothèque numérique rassemblant ce qui se fait de mieux en tant que services et technologies

Sur un plan personnel, ces quatre mois de stages m'ont permis de me rendre compte à quel point la démarche de gestion de projet demande de la rigueur. Cela a été en quelque sorte une mise en pratique de tous les cours que nous avons eu sur ce sujet. Cette expérience m'a permis de devenir plus à l'aise dans les relations professionnelles, surtout grâce au fait que le stage se passait en anglais ce qui oblige à faire un effort supplémentaire, particulièrement au début, pour la communication. L'évolution de mes rapports avec le reste de l'équipe a aussi été marquée par ma progression en anglais.

Bien que n'ayant quasiment fait que du développement j'ai quand même beaucoup appris sur les autres domaines liés à un projet de cette taille, international et mettant en jeu de nombreuses connaissances multidisciplinaires.

Au final je retiens de ce stage une expérience vraiment très enrichissante, et ce à tous les niveaux.

Bibliographie

Ressources sur les bibliothèques numériques :

Sites :

D-Lib Magazine: <http://www.dlib.org/>

Bulletin des bibliothèques de France : <http://bbf.enssib.fr/>

Publications:

HAIGH S.; « A Glossary of Digital Library Standards, Protocols and Formats », <http://www.collectionscanada.ca/9/1/p1-253-e.html>

VEEN T. Van, OLDROYD B.; « Search and Retrieval in The European Library, A New Approach », *D-Lib Magazine*, February 2004, vol.10, n° 2

LAGOZE C., PAYETTE S., KRAFFT D., JESUROGA S.; « Qu'est-ce qu'une bibliothèque numérique au juste? » - http://artist.inist.fr/article.php3?id_article=245

BROPHY P., « La bibliothèque hybride », *BBF*, 2002, n° 4, p. 14-20

TESNIÈRE V., LESQUINS N. ; « La bibliothèque numérique européenne : Une stratégie culturelle de la Toile », *BBF*, 2006, n° 3, p. 68-80

Ressources sur les métadonnées et les registres de métadonnées :

Sites :

Dublin Core Metadata Initiative (DCMI), le site de l'initiative Dublin Core : <http://dublincore.org>

Guide d'utilisation du Dublin Core : <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>

The Open Metadata Registry, le registre de métadonnées du Dublin Core : <http://dublincore.org/dcregistry/navigateServlet>

Publications :

HILLMANN D., SUTTON S., PHIPPS J., LAUNDRY R.; « A Metadata Registry from Vocabularies Up: The NSDL Registry Project », <http://arxiv.org/abs/cs.DL/0605111>

LAGOZE C., KRAFFT D., SAYLOR J.; « Metadata aggregation and “automated digital libraries”: A retrospective on the NSDL experience », <http://arxiv.org/abs/cs.DL/0601125>

Ressources sur PHP, MySQL, et autres utilitaires pour le développement web :

PHP débutant :

<http://www.phpdebutant.org>

Manuel PHP - La référence en Cours et Manuels PHP / MYSQL / HTML en français, forums, scripts : <http://www.manuelphp.com>

PHP Index - La passerelle française des technologies PHP :

<http://www.phpindex.com>

PHP: Hypertext Preprocessor:

<http://fr.php.net>

PHPFrance:

<http://www.phpfrance.com>

PHP et MySQL - Club d'entraide des développeurs francophones :

<http://php.developpez.com>

MySQL - La Base de Données Open Source la plus Populaire au Monde :

<http://www-fr.mysql.com>

The W3C Markup Validation Service :

<http://validator.w3.org>

W3Schools Online Web Tutorials :

<http://www.w3schools.com>

Annexes

Structure XML de description d'un terme

```
<term_usage>
  <type></type>
  <name></name>
  <qualifiedName></qualifiedName>
  <sourceLabel> </sourceLabel>
  <label profile="TEL_AP_object"></label>
  <label profile="TEL_AP_collection"></label>
  <definedBy></definedBy>
  <identifier></identifier>
  <status profile="TEL_AP_object"></status>
  <status profile="TEL_AP_collection"></status>
  <sourceDefinition></sourceDefinition>
  <definition profile="TEL_AP_object"></definition>
  <definition profile="TEL_AP_collection"></definition>
  <guidelines profile="TEL_AP_object"></guidelines>
  <guidelines profile="TEL_AP_collection"></guidelines>
  <regexp></regexp>
  <refines profile="TEL_AP_object"></refines>
  <refines profile="TEL_AP_collection"></refines>
  <isRefinedBy profile="TEL_AP_object"></isRefinedBy>
  <isRefinedBy profile="TEL_AP_collection"></isRefinedBy>
  <encodes profile="TEL_AP_object"></encodes>
  <encodes profile="TEL_AP_collection"></encodes>
  <encodingScheme profile="TEL_AP_object"></encodingScheme>
  <encodingScheme profile="TEL_AP_collection"></encodingScheme>
  <obligation></obligation>
  <repetition></repetition>
  <translation></translation>
  <dataType></dataType>
  <function profile="TEL_AP_object"></function>
  <function profile="TEL_AP_collection"></function>
</term_usage>
```

Exemple de description avec les métadonnées de The European Library

<pre><term_usage> <type>property</type> <name>Title</name> <qualifiedName>dc:title</qualifiedName> <sourceLabel>Title</sourceLabel> <label profile="TEL_AP_object" xml:lang="en">Title</label> <label profile="TEL_AP_object" xml:lang="nl">Titel</label> <label profile="TEL_AP_object" xml:lang="fr">Titre</label> <label profile="TEL_AP_object" xml:lang="de">Titel</label> <label profile="TEL_AP_object" xml:lang="es">Título</label> <label profile="TEL_AP_object" xml:lang="it">Titolo</label> <label profile="TEL_AP_object" xml:lang="sl">Naslov</label> <label profile="TEL_AP_object" xml:lang="fi">Nimeke</label> <label profile="TEL_AP_collection" xml:lang="en">Title</label> <label profile="TEL_AP_collection" xml:lang="nl">Titel</label> <label profile="TEL_AP_collection" xml:lang="fr">Titre</label> <label profile="TEL_AP_collection" xml:lang="de">Titel</label> <label profile="TEL_AP_collection" xml:lang="es">Título</label> <label profile="TEL_AP_collection" xml:lang="it">Titolo</label> <label profile="TEL_AP_collection" xml:lang="sl">Naslov</label> <label profile="TEL_AP_collection" xml:lang="fi">Nimeke</label> <definedBy>The Dublin Core Metadata Element Set, v1.1 http://purl.org/dc/elements/1.1/</definedBy> <identifier>http://purl.org/dc/elements/1.1/title</identifier> <status profile="TEL_AP_object">in use</status> <status profile="TEL_AP_collection">in use</status> <sourceDefinition></sourceDefinition> <definition profile="TEL_AP_object">The name of the object.</definition> <definition profile="TEL_AP_collection">The name of the collection.</definition> <guidelines profile="TEL_AP_object">If not available the title should be constructed or "[No title]" should be the minimum title...</guidelines> <guidelines profile="TEL_AP_collection">This should be the known title of the collection. However it should mean something to the user and not be an internal library acronym such as: DB2 or BRSNi etc....</guidelines> <regexp></regexp> <refines></refines> <isRefinedBy>dcterms:alternative, Dublin Core Terms http://purl.org/dc/terms/alternative</isRefinedBy> <obligation>M</obligation> <repetition>yes</repetition> <translation>yes</translation> <dataType>text</dataType> <function>search</function> <function>openurl</function> <function>multilinguality</function> </term_usage></pre>

Copies d'écran de pages du nouveau registre de métadonnées

Step 4/5: Resume of your information

Personal information

Author : fe
Institution : fe
E-mail : fesq.fg@fd.fr

Edit step 1

Comments on this proposal

What is the content of your proposal?
fsdf

Why do you make this proposal?
fsdf

What will be the impact of your proposal?
fsdf

Edit step 2

Term description

Name	Identifier
Type	property
Qualified name	dc:identifier
Source label	Identifier
Label [TEL_AP_object] in english only	No value defined
Label [TEL_AP_collection] in english only	No value defined
Defined by	The Dublin Core Metadata Element Set, v1.1 http://purl.org/dc/elements/1.1/
Identifier	http://purl.org/dc/elements/1.1/title
Source definition	http://purl.org/dc/elements/1.1/identifier/
Definition [TEL_AP_object]	An unambiguous reference to the resource within a given context.

Figure 2: l'étape résumant les informations saisies avant l'envoi définitif

- Home: Metadata Registry
- Introduction
- Directory of terms
- Search
- View all
- View collection profile
- View object profile
- View history
- Controlled vocabularies
- Terms administration
- Crosswalks
- XML Validation Schemas
- Metadata Working Group
- Users administration

Author: TEL MWG
 Last update: May 16 2007

Search for a term

To search for a term (including properties, refinements and encoding schemes), type in the name of a term in the search box. The results display the brief term description. To see the full term description, click on the name of the term from the results screen.

Search for (exact phrase):

3 result(s) for **title**

Alternative (see more...)

Type	refinement
Qualified name	dcterms:alternative
Source definition	Any form of the title used as a substitute or alternative to the formal title of the resource.

Date (see more...)

Type	property
Qualified name	dcterms:alternative
Source definition	Any form of the title used as a substitute or alternative to the formal title of the resource.

Title (see more...)

Type	property
Qualified name	dc: title
Source definition	A name given to the resource.

Figure 3: la page de recherche de termes