



Établir des ponts : des bibliothèques Europeana aux journaux Europeana

Susan K. Reilly
LIBER
The Hague, Netherlands
E-mail: susan.reilly[at]kb.nl

Session:

119 – Users and portals: digital newspapers, usability, and genealogy – Newspapers Section with Genealogy and Local History Section

Résumé :

Des études montrent que les personnes qui effectuent des recherches privilégient la facilité d'accès et particulièrement le guichet unique car ils constituent des moyens simples et efficaces d'accéder à des contenus numériques. En plus de la facilité d'accès, l'assurance de qualité du contenu est de première importance.

Le projet des bibliothèques Europeana¹ aborde ces deux questions par la sélection de 5,1 millions d'images, de livres, de vidéos et de thèses ainsi que d'articles provenant directement de dix-neuf des bibliothèques de recherche d'Europe de premier plan. La source des données indique permet la confiance dans les métadonnées et la qualité des images tandis que la facilité d'accès sera garantie par une recherche unique de tous les objets de la bibliothèque européenne² et les sites Web Europeana³.

Ce partenariat a posé le fondement d'une collaboration et d'innovations futures. Déjà, l'infrastructure d'agrégation durable et les capacités de recherche par texte entier conçues à travers le projet des bibliothèques Europeana sont configurées pour être appliquées à un nouveau contenu à travers le projet des journaux Europeana⁴. Le projet donnera accès à vingt-neuf million de pages de de journaux de toute l'Europe disponible par l'intermédiaire de la bibliothèque européenne et les plateformes Europeana.

¹ <http://www.europeana-libraries.eu>

² <http://www.theeuropeanlibrary.org>

³ <http://www.europeana.eu>

⁴ <http://www.europeana-newspapers.eu/>

Introduction :

Des études montrent que les personnes qui font des recherches privilégient la facilité d'accès et particulièrement le guichet unique car ils constituent des moyens simples et efficaces d'accéder à des contenus numériques. En plus de la facilité d'accès, l'assurance de qualité du contenu est de première importance.

Cet article décrit une partie du travail du projet des bibliothèques Europeana⁵, projet, qui permettra de faire des recherches dans des textes entiers des bibliothèques nationales et de recherche par le biais d'un portail unique. Il indique les motivations d'une collaboration et de ses avantages par l'intermédiaire d'organisations pour atteindre une vision commune. Il indique aussi les motivations du projet des journaux Europeana et ce qui y sera accompli. Il s'agit d'un projet qui bénéficie du portail et des capacités de recherche dans des textes entiers conçus par les bibliothèques Europeana.

La première pierre :

Le projet des bibliothèques Europeana est un projet de deux ans qui a commencé en janvier 2011. L'idée du projet vient de la prise de conscience de la nécessité d'avoir un agrégateur unique pour les bibliothèques de recherche européennes à la fois nationales et universitaires. Ces bibliothèques avaient déjà travaillées ensemble pour fournir un contenu thématique à Europeana, le portail du patrimoine culturel Europeana, à travers le projet de voyages Europeana⁵. Ce projet consistait à regrouper les contenus sur le thème des voyages et du tourisme à l'aide de deux agrégateurs différents, l'un pour les bibliothèques nationales et l'autre conçu spécialement pour regrouper le contenu des autres bibliothèques dans le projet. Ce projet a très bien réussi aussi bien en termes de fourniture d'un contenu numérique de grande qualité pour Europeana, qu'en termes de fondement d'une collaboration entre les bibliothèques nationales et celles de recherche, mais il a soulevé une interrogation sur le maintien de deux agrégateurs différents.

Le projet des bibliothèques Europeana traite de cette question du maintien par l'ouverture du service d'agrégation de la bibliothèque nationale, la Bibliothèque européenne, pour effectuer des recherches dans les bibliothèques. Il utilise ce service pour assembler une quantité énorme du contenu précieux des bibliothèques de recherche européennes. À la fin du projet, en décembre 2012, plus de 5,1 millions d'objets, notamment 1 200 films et clips vidéo, **850 000** images et **4,3 millions de textes** (livres, articles de presse, thèses, lettres) des bibliothèques de recherche auront été ingérés. Les textes entiers constituent une grande part de ce contenu et sont particulièrement précieux pour les personnes qui effectuent des recherches. Pour profiter au maximum du potentiel de ce contenu, le projet s'attelle aussi à développer des possibilités de recherche dans les textes entiers et un portail de recherche qui fournit des outils spécifiques aux recherches.

⁵ <http://www.europeanatravel.eu/>

Le partenariat :

Le rassemblement des contenus des bibliothèques nationales et de recherche a aussi facilité la création de réseaux clés de bibliothèques européennes, à savoir CENL, CERL et LIBER.

La Conférence européenne des directeurs de bibliothèques nationales (CENL) représente 48 bibliothèques nationales et possède actuellement l'assembleur de la bibliothèque européenne qui est le seul agrégateur de domaine de la bibliothèque européenne. Jusqu'à maintenant il n'agrégeait que le contenu des bibliothèques nationales.

L'intérêt du Consortium des bibliothèques de recherche européennes (CERL)(consortium) est d'améliorer l'accès et d'exploiter le patrimoine imprimé européen et comporte trente-trois membres à plein temps des bibliothèques nationales et de recherche. Le CERL a un intérêt et un savoir-faire particuliers pour l'indexation et les métadonnées.

LIBER, l'association des bibliothèques de recherche européennes comporte plus de quatre-cent-vingt membres (bibliothèques nationales, universitaires et d'autres de recherche) d'Europe et des alentours. Dix-neuf membres de LIBER fournissent le contenu pour le projet et, finalement, le service développé dans les bibliothèques Europeana sera étendu à tous les membres de LIBER.

Les bibliothèques Europeana sont la première occasion pour ces trois organisations de travailler ensemble à une communauté très forte, c'est à dire que les institutions membres ont toutes des contenus qui sont des éléments de recherche précieux et désirent toutes rendre leur contenu accessible aux personnes qui effectuent des recherches pour qu'ils puissent l'utiliser. La durabilité des résultats du projet sera assurée par la répartition dans cette communauté de précieuses propositions concrètes et acceptées et par la consolidation des relations entre les réseaux.

Définition de la valeur du partenariat :

La valeur des projets ne repose pas uniquement sur la création d'un seul service d'agrégation pour les bibliothèques, bien qu'il s'agisse d'un aspect important, elle repose aussi sur potentiel qu'elle offre pour apporter les contenus de recherche des bibliothèques aux personnes qui effectuent des recherches dans le monde entier. Elle étend potentiellement la recherche des collections à la fois des bibliothèques nationales et de recherche au-delà des frontières de leurs communautés et régions établies de personnes qui effectuent des recherches. Elle exploite la réputation collective de fournisseurs fiables de renseignements de qualité et de bonnes métadonnées, des collections des bibliothèques. C'est un fait établi qu'il existe une corrélation positive entre les bibliothèques et les livres⁶. La fourniture de textes entiers des collections de

⁶ De Rosa, Cathy et al. 2005. Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. Dublin, Ohio: OCLC.

<<http://www.oclc.org/reports/2005perceptions.htm>>.

livres numérisés avec d'autres contenus numériques comme des images, des vidéos et des fichiers audio sans parler des contenus universitaires comme des articles et des thèses, signifie que les personnes qui effectuent des recherches peuvent obtenir des résultats plus riches. En augmentant la visibilité de tels contenus par ce moyen, les bibliothèques peuvent accroître l'impact de l'investissement significatif de numérisation qu'elles font⁷. Pour les personnes qui effectuent des recherches, cette valeur réside dans la possibilité d'accéder à une masse conséquente de patrimoine culturel et de contenus se trouvant en un seul endroit et d'y effectuer des recherches. Des études montrent que les gens effectuent des recherches plus complexes et disposent aussi de moins de temps pour cela⁸, si bien que le guichet unique est une offre attrayante.

Conception du portail :

Une grande partie du contenu agrégé par l'intermédiaire du projet des bibliothèques Europeana sera disponible à travers le portail Europeana mais, de par la nature de ce contenu, le portail de la Bibliothèque européenne a été rénové en gardant à l'esprit utilisateurs qui effectuent des recherches en sciences humaines et sociales. Une telle offre est particulièrement adaptée à la communauté des personnes qui effectuent des recherches en sciences humaines et sociales pour laquelle la définition des données de recherche est complexe :

« La communauté des sciences humaines a besoin d'une masse importante de ressources numériques ainsi que d'outils pratiques, de services et de dépôts si elle veut aller au-delà de 'projets artisanaux' et constituer une fondation solide de théories et de méthodes⁹. »

Plusieurs séries d'ateliers et de tests par les utilisateurs ainsi que des recherches documentaires importantes ont contribué à la conception du nouveau portail qui comporte maintenant :

- la capacité de faire des recherches dans des textes entiers
- l'occasion d'examiner les archives des métadonnées d'origine d'objets individuels dans le but d'éventuellement donner accès à des groupes de données étendus pour des recherches.
- les titres de sujets CERIF de plus en plus utilisés. Cela rend les recherches possibles à travers un corpus d'objets qui sont liés par un thème commun ou une période commune.

⁷ European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. *The New Renaissance*. Brussels: European Commission, 2011.
http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_%20cds.pdf

⁸ Bulger, Monica, et al., *Reinventing Research? Information Practices in the Humanities*. London: Research Information Network, 2011.

⁹ Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" *Digital Humanities Quarterly* 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

- Le développement de la collection paneuropéenne en coopération avec les bibliothèques nationales, universitaires et de recherche d'Europe. C'est l'extension d'une exposition virtuelle ¹⁰, qui expose le contenu d'un éventail de sources de toute l'Europe. Des chronologies qui donnent les occurrences d'un terme de recherche précis à travers les siècles.
- Des API qui permettront au contenu de la base de données d'être analysé et affiché dans des contextes en-dehors du portail de la Bibliothèque européenne. Cela veut dire que les personnes qui effectuent des recherches peuvent apporter leur contenu dans leur propre environnement de recherche et explorer de nouvelles façons d'exploiter ce contenu. L'exportation directe des documents à des services populaires de traitement des références comme Mendeley et Zotero.

Une fois lancé, le portail sera constamment redéveloppé en ligne avec l'émergence de pratiques de recherche dans les sciences humaines et sociales et leur version numérique. D'autres études seront faites sur la façon dont les gens peuvent exploiter et utiliser ce contenu unique et une stratégie correspondante sera élaborée.

Établir des ponts :

Tout comme le service d'agrégation, le portail et les capacités de recherche dans les textes entiers développés dans le projet des bibliothèques Europeana doivent maintenant être utilisés pour exposer un nouveau type de contenu : les journaux. De récentes avancées dans la OCR (reconnaissance optique de caractère) à travers le projet IMPACT¹¹ vont maintenant être utilisées pour les journaux provenant de douze bibliothèques nationales et de recherche de toute l'Europe. Dix-huit millions de pages de journaux seront améliorées et mises à disposition à travers Europeana et le portail de la Bibliothèque européenne. Le projet des Journaux Europeana entreprend de traiter les difficultés très spécifiques à rendre les recherches possibles dans le texte entier des anciens journaux. Il utilisera des méthodes d'amélioration pour la OCR, la OLR/segmentation d'article, la reconnaissance des entités nommées (NER) et la reconnaissance de catégorie de page.

Une grande partie de ce qui a été développé et appris à travers les bibliothèques Europeana va maintenant être appliquée aux Journaux Europeana :

1. Agrégation :

Quatre types de collections de journaux numériques existantes peuvent être identifiés :

- a) Des images avec seulement des métadonnées structurales
- b) Des images avec des métadonnées structurales et des textes entiers à rechercher (OCR)
- c) Des images avec des métadonnées structurales, une reconnaissance d'articles (OLR/suivi des articles) et OCR

¹⁰ www.theeuropeanlibrary.org/exhibition

¹¹ <http://www.impact-project.eu/>

- d) Des images avec des métadonnées structurales, OLR, OCR et des enrichissements sémantiques.

Toutes les données disponibles seront récoltées par la Bibliothèque européenne. Les données seront transformées en EDM, le modèle de données d'Europeana et distribuées pour Europeana.

2. Normalisation des métadonnées.

De nombreux formats de métadonnées sont actuellement utilisés. Pour améliorer l'accès aux contenus numériques, il faut adopter des critères communs. Tous les formats de métadonnées existants seront étudiés et les solutions les plus pratiques seront fournies à la communauté.

3. Meilleures capacités d'affichage.

C'est actuellement un défi de rendre les recherches dans les journaux plus faciles et de les afficher de façon attrayante en ligne. Le projet de journaux en ligne examinera le travail accompli par les bibliothèques Europeana et l'utilisera pour préciser les exigences appropriées aux recherches et à l'affichage qui peuvent être utilisées par Europeana.

Des recherches en ligne pourront désormais être faites sur les textes entiers d'anciens journaux qui remontent au 18^e siècle. De plus, les utilisateurs pourront afficher ces journaux dans leur contexte, à côté d'images, de photos, de thèses correspondantes, de livres et d'articles. L'affichage du contenu des journaux de cette manière permet de nouvelles connections et l'ouverture de portes sur de nouveaux types de recherches et de collaborations.

Il y a d'autres avantages au travail de ce projet puisque les procédures par lesquelles le contenu des journaux sera amélioré comporte la OCR, le OLR/suivi des articles, NER et la reconnaissance de catégorie de page. Pour chacune de ces tâches techniques, les recommandations d'utilisation optimale seront identifiées et publiées. Cela constituera un immense avantage pour le réseau élargi des bibliothèques qui disposent des réseaux CERL, CENL et LIBER. Cela permettra de réduire le coût des projets de numérisation des journaux et augmentera l'accessibilité des collections de journaux numériques maintenant et à l'avenir.

Conclusion :

Les bibliothèques Europeana était un réseau d'utilisation optimale qui répondait au besoin très concret d'un agrégateur unique pour les bibliothèques européennes. Ce faisant il a aussi rassemblé des réseaux de bibliothèques clés. En œuvrant à une vision commune, les réseaux ont créé une ressource qui peut avoir une grande valeur potentielle pour les personnes qui effectuent des recherches. Il a aussi créé les conditions nécessaires pour que les bibliothèques nationales et de recherche travaillent ensemble de façon plus fluide en s'appuyant sur la vision d'un lien entre le contenu et les personnes qui effectuent des recherches. Le contenu des bibliothèques d'Europe est riche et varié. C'est particulièrement le cas des fonds de journaux. L'apport de ces fonds en ligne à un moment où des techniques d'amélioration se développent pour afficher des textes entiers avec clarté donne une occasion merveilleuse aux personnes qui effectuent des recherches de relier les documents du riche patrimoine culturel européen et d'établir des nouveaux liens entre eux. C'est aussi l'incarnation même de la

façon dont des organisations, des réseaux et des institutions qui travaillent ensemble peut produire des résultats innovants, améliorer leur efficacité et fournir une accessibilité. Un tel travail aura des effets à long terme, non seulement pour les bibliothèques ni même pour l'accessibilité du patrimoine culturel européen mais pour tous les pays du monde qui ont une masse de documents culturels imprimés.

Références :

Bulger, Monica, et al., *Reinventing Research? Information Practices in the Humanities*. London: Research Information Network, 2011.

Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" *Digital Humanities Quarterly* 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

De Rosa, Cathy et al. 2005. *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. Dublin, Ohio: OCLC.
<<http://www.oclc.org/reports/2005perceptions.htm>>.

European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. *The New Renaissance*. Brussels: European Commission, 2011.
http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_%20cds.pdf

Europeana (2012) Retrieved from <http://www.europeana.eu> on 19th April 2012.

Europeana Libraries (2012), Retrieved from <http://www.europeana-libraries.eu>, on 9th March 2012.

Europeana Travel (2012), Retrieved from <http://www.europeanatravel.eu/> on 19th April 2012.

IMPACT, Retrieved from <http://www.impact-project.eu/> on 19th April 2012.

The European Library(2012), Retrieved from <http://www.theeuropeanlibrary.org> on 9th March 2012.