

Qu'est-ce qu'une bibliothèque numérique, au juste ? Au-delà des fonctions recherche et accès dans la National Science Digital Library

Carl Lagoze (1)
lagoze@cs.cornell.edu

Sandy Payette (1)
payette@cs.cornell.edu

Dean B. Krafft (1)
dean@cs.cornell.edu

Susan Jesuroga (2)
jesuroga@ucar.edu

(1) Computing and Information Science, Cornell University, Ithaca, NY

(2) UCAR-NSDL, Boulder, CO

Cet article est la traduction par Frédéric MARTIN (BnF) d'un article publié dans le numéro de novembre de la revue D-Lib Magazine. Il a été révisé par Catherine GUNET (INIST-CNRS), mis en ligne par l'équipe ARTIST et a bénéficié du soutien du RTP-DOC.

Il est paru sous la référence originale :

**What Is a Digital Library anyway, anymore ?
Beyond Search and Access in the NSDL**

Les bibliothèques numériques, bien que de tailles différentes¹, vivent à présent leur adolescence. Comme pour toute adolescence, il y a de quoi s'enthousiasmer et se préoccuper.

Les succès rencontrés pendant une décennie de recherche, de développement, de déploiement sont source d'enthousiasme. Toute liste en serait nécessairement incomplète, mais inclurait sans nul doute Google², le Handle System³, le Dublin Core⁴, l'OAI-PMH⁵ (protocole pour la collecte des métadonnées créé par l'Initiative des Archives Ouvertes), l'Open URL [40], arXiv⁶, DSpace [53] et LOCKSS [50]. Ces réalisations, parmi d'autres, sont à mettre en relation avec l'explosion généralisée du web lui-même, durant les quinze dernières années⁷. Elles tendent vers cette vision de la bibliothèque numérique comme « accès universel au savoir humain » exprimée dans le rapport du President information Technology Advisory Committee en 2001 (PAC)[45].

Les préoccupations proviennent en partie de problèmes soulevés lors des premiers ateliers de bibliothèque numérique [8, 34] et pour lesquels des solutions pratiques restent à mettre en oeuvre. Quelques exemples le montrent. Tandis que Handle et DOI⁸ ont été déployés avec succès dans les communautés des bibliothèques et de l'édition, l'objectif visant à disposer d'identifiants universels et pérennes n'est toujours pas atteint. L'adoption largement répandue du Dublin Core et de l'OAI-PMH semble répondre aux objectifs initiaux d'une description des ressources qui soit interopérable. Pourtant, des problèmes liés à la qualité des métadonnées [58] compromettent l'utilité des standards. Les logiciels médiateurs d'identité fédérée comme Shibboleth⁹ commencent à répondre aux questions d'autorisation et d'identification, mais l'infrastructure à clé publique, considérée comme « essentielle à l'émergence des bibliothèques numériques » [34], n'est pas encore développée. En dépit des efforts produits par l'initiative pour le Web sémantique du W3C [13], le Saint Graal de l'interopérabilité sémantique [42] reste hors de portée. Enfin, face aux volumes croissants d'information sous forme numérique native et stockée dans des entrepôts institutionnels, il nous manque toujours des techniques extensibles et normalisées pour préserver pleinement cette information.

¹ Bien que les catalogues de ressources numériques aient été introduits de bonne heure dans l'histoire de l'informatique [51] l'emploi généralisé du terme « bibliothèque numérique » remonte au début des années 90 [16] [20].

² <<http://www.google.com>>

³ <<http://www.handle.net>>

⁴ <<http://www.dublincore.org>>

⁵ <<http://openarchives.org>>

⁶ <<http://arxiv.org>>

⁷ Dont la richesse sera significativement accrue par des efforts de numérisation massifs tels que Google Print <<http://print.google.com>>

⁸ <<http://www.doi.org>>

⁹ <<http://shibboleth.internet2.edu/>>

Ces inconvénients techniques se situent dans un contexte institutionnel plus large et plus inquiétant, que certains¹⁰ ont caractérisé comme la « google-isation » des bibliothèques numériques et de l'information en général. Comme tout néologisme, « google-isation » a plusieurs sens. Ici, il réfère à l'idée fautive et agaçante selon laquelle Google représente l'apothéose de l'information numérique et que les problèmes restant dans ce domaine ont déjà été résolus - ou vont l'être (peut-être même par Yahoo !, MSN, etc.). Suite à des discussions informelles avec des collègues de la communauté des bibliothèques numériques de recherche, il ressort que la « google-isation » a contaminé les organismes de financement, à la fois publics et privés. Si l'absence de financement important pour un programme de bibliothèque numérique au sein de la National Science Foundation est imputable à de nombreuses causes, l'idée que « Google a résolu le problème » y est certainement pour quelque chose.

Les réalisations de Google sont certes frappantes. Mais cette vision réductrice d'une « Fin de l'Histoire »¹¹ apparaît comme le résultat d'une confusion sur « ce qu'est une bibliothèque (numérique) ». Peut-être sous l'influence de visions utopiques et trompeuses, comme les commentaires d'Al Gore sur « l'écolier de Carthage, Tennessee » [9], il existe semble-t-il une croyance selon laquelle une bibliothèque numérique ne concerne que la recherche d'information (« est-ce que je peux la trouver ? ») et l'accès (« est-ce que je peux l'obtenir ? »). Certes, ces fonctions sont essentielles (et demeurent des défis), mais elles ne sont que la partie d'un environnement informationnel. Les bibliothèques traditionnelles sont bien plus que des entrepôts bien organisés de livres, de cartes, de périodiques, etc. Elles sont par nature des lieux où des personnes se rencontrent pour accéder à un savoir qu'ils partagent et qu'ils échangent. Les ressources que les bibliothèques sélectionnent et les services qu'elles offrent devraient refléter l'identité des communautés qu'elles servent [31].

Comme le suggère Borgman [14-16], les bibliothèques numériques devraient non seulement ressembler aux bibliothèques traditionnelles mais encore aller beaucoup plus loin qu'elles. Elles ne doivent pas se limiter à de simples moteurs de recherche. Comme toutes les bibliothèques, il faut qu'elles procèdent à un haut niveau de sélection des ressources qui répondent aux critères de leur mission. Il est également nécessaire qu'elles fournissent des services, comme la recherche, qui facilitent l'utilisation des ressources par la communauté ciblée. Mais, libérées des contraintes physiques d'espace et de support, les bibliothèques numériques peuvent mieux s'adapter aux communautés qu'elles servent et mieux les refléter. Elles doivent être collaboratives, en permettant aux utilisateurs de contribuer et d'apporter du savoir, soit de façon active par des annotations, des comptes rendus de lecture, etc., soit de façon passive au travers de leurs profils d'utilisateurs. En outre, il faudrait qu'elles soient contextuelles, illustrant ainsi le réseau extensible des relations et des couches de savoir qui se tisse autour des ressources sélectionnées. De la sorte, le noyau de la bibliothèque numérique devrait être une base d'information évolutive, en entrelaçant ainsi dans un même tissu « sélection professionnelle » et « sagesse des peuples » [54].

Cette vision élargie du rôle des bibliothèques numériques implique de repenser les modèles informationnels sur lesquels elles reposent. Le poids de l'héritage légué par le catalogue collectif dans les bibliothèques traditionnelles, ajouté parfois à la disproportion prise par la fonction recherche, a conduit à l'utilisation répandue d'un modèle informationnel bâti sur un entrepôt de métadonnées. Même si souvent les bibliothèques numériques sont implémentées différemment, on constate généralement qu'à la base elles compilent, indexent et fournissent des requêtes sur un catalogue composé de notices de métadonnées. Comme nous le montrerons plus tard, ce modèle de catalogue simpliste est nettement insuffisant dans le cadre d'une vision plus étendue de ce qu'est une bibliothèque numérique.

Le présent article décrit un modèle informationnel pour les bibliothèques numériques qui va délibérément "au-delà de la recherche et de l'accès", sans pour autant ignorer ces fonctions de base et qui facilite la création d'environnements de savoir collaboratifs et contextuels. Ce modèle est un réseau d'information superposé (information network overlay) qui représente la bibliothèque numérique sous la forme d'un graphe. Ce graphe comporte des nœuds typés, qui correspondent aux unités d'information (documents, données, services, acteurs) au sein de la bibliothèque et des arêtes représentant les relations contextuelles qui se nouent entre ces unités. Ce modèle informationnel incorpore de l'information locale et distribuée intégrée aux web services, autorisant la création de documents enrichis (par ex., des objets d'apprentissage, des publications pour l'e-science, etc.). Il exprime les relations complexes entre les objets d'information, les acteurs, les services et la méta-information (comme les ontologies) et représente ainsi les ressources dans leur contexte, plutôt que comme le résultat d'un accès web isolé. Il facilite les pratiques collaboratives, fermant ainsi la boucle entre les utilisateurs-consommateurs et les utilisateurs-contributeurs.

¹⁰ Une recherche a montré qu'il existe plus de 13 000 occurrences de ce terme sur le web, dont une webémission de Clifford Lynch et Michael Keller sur le sujet [35].

¹¹ Il s'agit d'une référence à un livre de Francis Fukuyama publié en 1992 [21] remarquant une semblable myopie euphorique dans le domaine de l'économie politique.

Nous nous proposons de décrire comment ce modèle de données est implémenté dans Fedora, [27] logiciel libre de gestion d'entrepôt. Fedora est particulièrement adapté à cette tâche, grâce à sa manière unique de combiner un modèle d'objet flexible, l'intégration des web services, une gestion de l'accès permettant une fine granularité et l'incorporation de l'expressivité du web sémantique.

Ces travaux se situent dans le cadre du projet de la National Science Digital Library (NSDL) [61]. Les conditions posées par la NSDL, en termes d'échelle et de contraintes, nécessitent une approche aussi poussée. Frank Wattenberg en a décrit la vision originelle comme suit :

« A bien des égards, la NSDL pourrait aller bien plus loin que l'image traditionnelle de la bibliothèque. En plus de fournir un accès large et actualisé à des ressources à jour et de grande qualité destinées à la formation scientifique, la NSDL pourrait profiter de la connectivité apportée par internet et du potentiel des technologies interactives pour créer un lieu de travail riche et asynchrone : une salle de séminaire, une salle de lecture et un laboratoire où partager et construire la connaissance. Ainsi, la NSDL pourrait fournir un cadre qui, au travers d'un ensemble de ressources diversifié et puissant, faciliterait le travail des utilisateurs dans des environnements différents » [59].

Nous pensons que cette vision élargie de la bibliothèque numérique n'est pas propre à la NSDL. Bien que les communautés qui recherchent et partagent de l'information aient besoin de trouver des aiguilles dans des bottes de foin [28] - un créneau occupé par Google et ses concurrents - elles ont aussi besoin de fonctionnalité « au-delà des fonctions recherche et accès » , où les bibliothèques numériques « créent un lieu de travail riche et asynchrone » dans lequel l'information est partagée, agrégée, manipulée et affinée.

1 Construire une bibliothèque numérique avec un entrepôt de métadonnées : phase I de la NSDL

Les lecteurs de D-Lib Magazine et la communauté des bibliothèques numériques connaissent sans doute le projet de la NSDL. Aussi, cette section ne présente-t-elle que brièvement le contexte dans lequel s'inscrit le travail que nous décrivons dans le reste de l'article. Nous suggérons donc à ceux qui souhaitent davantage d'information de lire les articles déjà rédigés à ce sujet [6, 7, 25, 61] et de consulter la page « about NSDL » à l'adresse <<http://nsdl.org/about>>.

L'idée d'une NSDL est née en 1998 au cours d'un atelier [3] financé par la National Science Foundation (NSF). Cet atelier devait étudier les problèmes concernant l'état de l'enseignement en science, technologie, ingénierie et mathématiques (STEM en anglais) aux Etats-Unis et a mis en lumière les opportunités offertes par Internet et les technologies du web pour l'améliorer. En s'appuyant sur les résultats de cet atelier, la NSF commença en 2000 à subventionner des projets NSDL et à ce jour, elle a accordé plus de 180 bourses. Ces aides couvrent un grand nombre de domaines comme le développement des collections, les services et la recherche fondamentale. Les travaux décrits dans le présent article ont bénéficié de subventions de la part des universités Cornell et Columbia ainsi que l'université Corporation for Atmospheric Research (UCAR), pour la partie « noyau intégré », qui comporte la coordination de l'architecture, de l'organisation et de la stratégie pour la NSDL.

Les premiers travaux techniques de l'équipe « noyau intégré » (NI) ont abouti à une architecture et à un modèle de données ayant ces trois fonctions de base : sélectionner des ressources web en STEM, les interroger transversalement et en faciliter l'accès. Le paradigme architectural pour réaliser ces trois fonctions est essentiellement le catalogue collectif et un entrepôt de métadonnées [EM] en Dublin Core. Ce dernier correspond aux ressources développées et gérées par les projets de collections de la NSDL et par d'autres organismes participants. L'EM est implémenté sous la forme d'une base de données relationnelle Oracle™, dans laquelle les notices de métadonnées individuelles sont stockées dans des séries de tables.

Les notices de métadonnées en Dublin Core, qui contiennent des URL pointant vers les ressources numériques correspondantes, sont absorbées dans l'EM via l'OAI-PMH [29]. Au cours de ce processus d'alimentation, les dates et différents éléments de vocabulaire contrôlé dans ces notices sont normalisées. D'autres services, comme la recherche gérée par le "noyau intégré" et l'archivage des ressources, utilisent un serveur OAI-PMH¹² pour collecter ces notices normalisées et obtenir ainsi l'information nécessaire (par ex., pour construire des index de recherche à partir des métadonnées).

La fonction « recherche » utilise le système d'indexation en texte intégral Lucene¹³ pour indexer à la fois les métadonnées collectées décrivant la ressource et le contenu textuel de la première page HTML ainsi référencée.

¹² <<http://services.nsdlib.org:8080/nsdloai/OAI>>

¹³ <<http://jakarta.apache.org/lucene/docs/index.html>>

La fonction « archivage » utilise le Storage Resource Broker [10] développé par le San Diego Supercomputing Center. Elle parcourt chaque mois le web à la recherche de toutes les ressources numériques identifiées dans les notices de métadonnées collectées à partir des EM. La fonction « archivage » identifie une collection de pages reliées entre elles, considérée comme la plus représentative de la ressource elle-même et effectue une capture d'archive de ces pages.

Du point de vue de l'utilisateur, les ressources dans le catalogue de la NSDL et les services sous-jacents sont disponibles par le biais d'un portail central disponible à <<http://www.nsd.org>>. Celui-ci sera bientôt complété par des portails spécifiques à des communautés éducatives et soutenus par le programme NSDL Pathways [2].

Le portail central de la NSDL et son architecture fondée sur des entrepôts de métadonnées ont été déployés pour la première fois en décembre 2003. En deux ans, la collection s'est enrichie jusqu'à atteindre plus de 1,1 million de ressources, avec des notices de métadonnées collectées à partir de plus de 80 fournisseurs de données OAI-PMH.

2 Utilité de l'entrepôt de métadonnées en tant qu'architecture de bibliothèque numérique

En règle générale, l'usage à grande échelle du Dublin Core et de l'OAI-PMH dans l'EM de la NSDL a prouvé son utilité pour fournir les services de base d'une bibliothèque numérique, mais il a également révélé de nombreux problèmes d'implémentation. Le plus sérieux concerne la qualité des métadonnées [6] et la validité au regard de l'OAI-PMH, en particulier la conformité au schéma XML. Les coûts administratifs de maintenance de l'EM ont atteint ainsi des niveaux auxquels on ne s'attendait pas. Ces difficultés techniques feront l'objet d'un prochain article.

Cependant, notre sujet ici est d'examiner d'un point de vue plus large l'architecture de la NSDL existante et les bibliothèques numériques en général. Dans cette partie, nous passons en revue des travaux de recherche émanant de la communauté de l'enseignement qui étudient les conditions que doit remplir une bibliothèque numérique axée sur l'enseignement et les fonctionnalités nécessaires pour y répondre.

Les bibliothèques numériques ont une réelle valeur pour le monde de l'enseignement car elles offrent l'accès en ligne à des ressources primaires et des moyens de les utiliser. Mais, pour être vraiment efficaces en tant qu'outils didactiques, elles ne doivent pas se limiter au seul accès à des ressources de qualité. Selon Reeves, « les médias et les technologies ne réaliseront pleinement leur véritable pouvoir d'améliorer l'enseignement que lorsque les étudiants les utiliseront de façon active comme des outils cognitifs au lieu de ne voir en eux que de simples tuteurs ou stocks d'information avec lesquels ils peuvent interagir » [49]. Marshall constate aussi que les bibliothèques numériques doivent être plus que des entrepôts et accompagner la totalité du cycle de vie des données, de l'information, du savoir et de la construction du savoir en général [36].

Cette fonctionnalité plus large requiert un modèle informationnel pour les bibliothèques numériques qui soit plus riche qu'une collection de simples pages web ou de documents statiques. Wiley [60], entre autres, utilise la notion d'objets d'apprentissage pour indiquer une collection d'informations, qui comprend non seulement une ou plusieurs ressources primaires, mais aussi le contexte pédagogique dynamique de cette information. Ce contexte inclut des informations culturelles et sociales, les profils d'utilisation, les objectifs pédagogiques, la nature des systèmes éducatifs des apprenants, les capacités des apprenants, leurs profils individuels et leurs connaissances antérieures [37]. Le contexte informationnel peut être vraiment complexe, reflétant la diversité des publics desservis et les différences dans la façon qu'ont ces publics d'utiliser et de manipuler l'information.

Certains chercheurs ont examiné les différentes facettes de cette information contextuelle. Elles consistent notamment à :

- recueillir des opinions, des commentaires, des comptes rendus portant sur les ressources de la bibliothèque [39] et l'historique de leur utilisation [43],
- décrire la communauté des utilisateurs impliqués dans la création d'un objet d'apprentissage [48],
- cerner les interactions des utilisateurs et mettre en relation leurs profils avec les objets d'apprentissage [38],
- intégrer les recommandations des enseignants et les corrélations qui existent avec les programmes éducatifs [47],
- repérer et stocker des mots-clés utilisés pour l'interrogation qui conduisent à une utilisation éventuelle de la ressource [4].

Le modèle primaire de données et de métadonnées, orienté « notices », qu'utilisent la plupart des bibliothèques numériques (et traditionnelles), possède une capacité limitée à modéliser pleinement ce contexte informationnel multidimensionnel.

Premièrement, les notices de métadonnées et les entrepôts de métadonnées représentent principalement les propriétés d'un item individuel. Elles ne permettent souvent pas de modéliser complètement les relations contextuelles [43] qui entourent les ressources et n'opèrent aucune distinction entre les multiples entités - ressources, métadonnées, acteurs, ontologies - qui font partie de cette structure relationnelle. De plus, parce qu'elles reposent fréquemment sur des schémas ou des modèles figés, elles sont difficiles à adapter à des besoins en information évolutifs. L'entrepôt de métadonnées de la NSDL, par exemple, ne reconnaît que les collections et les items et ne représente que leur relation d'appartenance. Parce que l'EM est stocké dans une base de données relationnelle, chaque relation nouvelle nécessite une redéfinition du schéma. Ce manque de souplesse s'est avéré problématique à cause de l'évolution des contraintes au cours des activités de la NSDL.

Deuxièmement, la nature statique des notices de métadonnées, qui sont en général créées une fois pour toutes par les créateurs de ressources ou les catalogueurs, pose problème. Le contexte des ressources est dynamique, car il exprime les changements de profils d'utilisation, de préférences individuelles et d'environnement culturel. Selon Recker et Wiley, « un objet d'apprentissage appartient à un réseau complexe de relations sociales et de valeurs touchant l'apprentissage et la pratique. Ainsi la question se pose-telle de savoir si de telles notions contextuelles et mouvantes peuvent être représentées et regroupées dans une notice de métadonnées unique et figée » [48].

Troisièmement, un modèle informationnel centré sur les métadonnées se heurte inévitablement à la distinction floue entre « données et métadonnées »¹⁴ [19]. Par exemple, nous avons remarqué plus haut que les annotations sont une des formes utiles de l'information contextuelle. Les annotations sont-elles des métadonnées (portant sur quelque chose) ou des données à part entière ? Il n'y a pas de réponse univoque, mais une architecture qui marque nettement la distinction entre données et métadonnées rend difficile le traitement de telles ambiguïtés.

Enfin, nous avons aussi remarqué l'importance de la réutilisation de l'information - c'est-à-dire la capacité de prendre des ressources primaires et de les combiner dans des objets d'apprentissage agrégés ou des plans de cours [46], puis de les réutiliser pour fabriquer de nouveaux objets. Parce que les unités d'information sous forme physique, dans les bibliothèques traditionnelles, ne peuvent faire l'objet d'un tel réemploi, une approche centrée sur les métadonnées, avec des notices descriptives, était possible. Mais une bibliothèque numérique doit être centrée sur les ressources et fournir un cadre pour gérer, manipuler et traiter le contenu et les métadonnées tout comme la ligne continue qui les sépare.

3 Modélisation informationnelle pour gérer la complexité et le contexte

Quel est donc le modèle informationnel approprié pour dépasser les limites de l'approche "métadonnées" ? En cherchant une réponse à cette question, il faut veiller à ne pas se débarrasser trop vite des notices catalographiques ni ignorer la valeur des métadonnées uniformes, qui mettent de l'ordre dans une information hétérogène [10]. Il est nécessaire pourtant d'incorporer ces notices de catalogue dans une fondation plus riche, de nature à représenter des descriptions structurées et non-structurées, l'hétérogénéité et l'homogénéité, les métadonnées et le contenu, l'information statique et l'information dynamique, les relations complexes et toute une multitude d'autres complexités.

Cette partie décrit le cadre d'un modèle informationnel plus riche qui concilie la complexité et le contexte. Nous développons ce modèle en décrivant « le problème de l'item »¹⁵, avec comme point de départ les fonctions basiques de recherche et d'accès à des éléments homogènes (ressources), auxquelles nous ajoutons progressivement de la complexité. Nous soutenons que, bien que le contexte de cette description soit la NSDL, le problème envisagé ici peut être généralisé à toute une variété de bibliothèques numériques et d'environnements d'information.

¹⁴ Remarquons que c'est cette distinction problématique qui a été l'une des motivations premières [44] pour l'architecture Fedora, utilisée pour implémenter le modèle décrit plus loin.

¹⁵ Merci, pour cette expression, à un ancien collègue qui travaille maintenant chez Amazon (et qui restera anonyme). La coïncidence entre les problèmes de modélisation de l'information chez Amazon et dans les bibliothèques numériques n'est pas fortuite. Amazon est peut-être le meilleur exemple d'environnement informationnel qui offre aux utilisateurs une information riche et contextuelle construite à partir d'une couche basique de données (ses produits).

3.1 Représenter des matériaux numériques ¹⁶



Figure 1

Comme évoqué plus haut, le but initial de la NSDL était de proposer des fonctions de sélection, de recherche et d'accès portant sur des ressources en STEM accessibles par URL. Cet objectif limité était atteint par le modèle bien connu du catalogue collectif, où la bibliothèque est représentée comme un ensemble de métadonnées uniformes (Dublin Core) qui référencent les ressources via leurs URL. Il faut noter que dans ce modèle, la représentation des ressources passe au second plan. Elles ne sont pas représentées elles-mêmes mais n'existent qu'indirectement, par le biais de références (URLs) issues des métadonnées.

3.2 Décrire des matériaux numériques de plusieurs manières, structurées ou non-structurées



Figure 2

Si le Dublin Core permet une interopérabilité descriptive minimale, il lui faudra coexister avec d'autres formats, plus riches, spécifiques aux disciplines ou aux objectifs visés [24]. Par ailleurs, comme nous l'avons vu précédemment, des descriptions non structurées telles que des commentaires ou des annotations sont souvent aussi utiles que des notices de métadonnées structurées. Sans compter que ces descriptions, structurées ou non, proviennent de multiples contributeurs. Cette complexité supplémentaire met à mal les fondements du modèle de catalogue collectif, qui repose sur un ensemble unique de producteurs (les catalogueurs) qui créent et gèrent un ensemble uniforme de descriptions. Deux nouvelles difficultés de modélisation apparaissent. Tout d'abord, les ressources doivent être modélisées parallèlement aux notices descriptives, puisque les ressources constituent le point d'ancrage pour relier entre elles de multiples descriptions¹⁷. Ensuite, la modélisation des agents et des producteurs se révèle importante pour représenter le marquage (branding) des ressources (qui a sélectionné ou créé la ressource ?) et le distinguer du marquage des métadonnées (qui a fourni les métadonnées ?). Le marquage est un outil utile pour aider les utilisateurs à connaître la qualité des ressources numériques.

3.3 Ajouter d'autres types de matériaux numériques



Figure 3

Comme nous l'avons vu, le modèle nécessite déjà de représenter différents types de descriptions, les agents qui les produisent et les ressources qu'elles décrivent. Mais les ressources elles-mêmes ne sont pas homogènes. Les

¹⁶ L'emploi de ce terme est emprunté à Godfrey Rust [11]. "People Make Stuff, People Use Stuff, and People Do Deals About Stuff"

¹⁷Collecter des métadonnées provenant de plusieurs fournisseurs soulève d'intéressants problèmes d'équivalences. La capacité à déterminer que deux descriptions concernent la même ressource se fonde sur l'heuristique et la subjectivité.

bibliothèques numériques collectent une variété grandissante de ressources - images, fichiers audio, textes - et s'ouvrent à des types de ressources bien plus complexes comme les données, les simulations, les objets d'apprentissage multimédia et autres. C'est une source de complexité supplémentaire dans la modélisation - notamment en ce qui concerne la meilleure façon d'associer simultanément l'uniformité au niveau de l'interface utilisateur et la représentation des caractéristiques propres à chaque type de ressource. En plus des problèmes liés à la description (métadonnées), il existe des difficultés concernant l'accès et la présentation, puisque différents types d'informations peuvent requérir différents protocoles d'accès et applications d'aide, qui doivent tous être représentés dans le modèle informationnel.

3.4 Les matériaux numériques sont parfois durs à définir

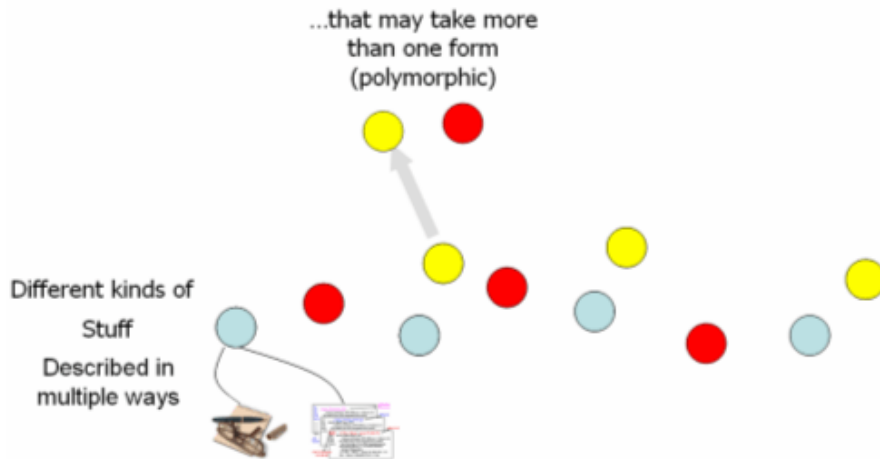


Figure 4

Les ressources dans une bibliothèque numérique ne sont pas toujours simples à caractériser. Par exemple, un livre électronique (e-book) est-il un livre ou un logiciel [33] ? Une information est-elle nécessairement soit une donnée soit une métadonnée ? Un acteur ne peut-il pas être aussi une ressource d'information ? Ce ne sont là que quelques exemples des difficultés que l'on rencontre lorsqu'on modélise l'information. Plutôt que de forcer les matériaux numériques à entrer dans des cases inadaptées, la structure des types du modèle informationnel doit être polymorphe. Toute entité doit pouvoir adopter différentes caractéristiques et différents comportements, en fonction du contexte d'accès ou d'utilisation.

3.5 Permettre aux utilisateurs de personnaliser les matériaux numériques

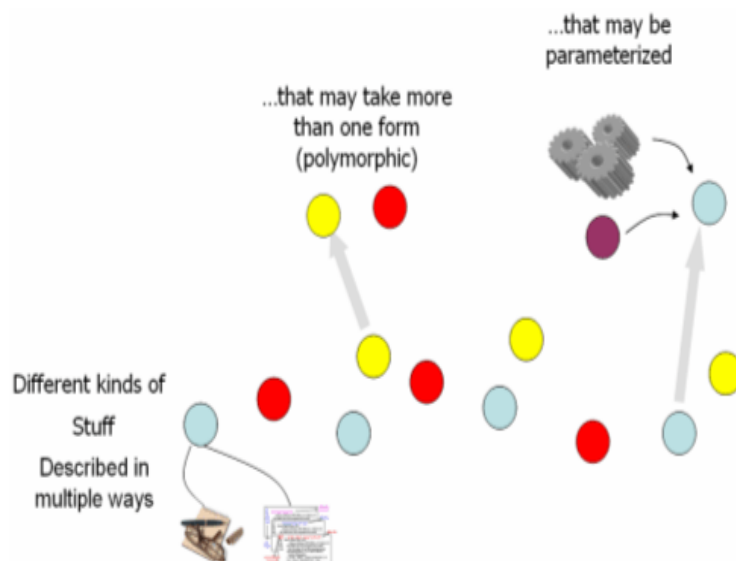


Figure 5

A l'origine, les bibliothèques numériques ont eu recours à la notion d'objets numériques, qui sont des paquets d'information avec de multiples diffusions disponibles par le biais de demandes de service [22, 26]. La plupart des systèmes de bibliothèques numériques modernes implémentent cette fonctionnalité en utilisant des standards comme les conteneurs [12, 32] d'objets complexes qui encapsulent les flux de données et de métadonnées associées à un objet numérique. Une demande de service peut alors inclure un paramètre qui spécifie la nature de la diffusion demandée - par exemple, une requête pour une diffusion en PDF ou en LaTeX d'un document scientifique.

Dans une architecture orientée "services", ces diffusions peuvent être produites aussi bien sous une forme dynamique que statique. Par exemple, plutôt que de stocker une image en plusieurs formats et résolutions, il est possible de répondre à la requête d'un utilisateur (par ex., 300 dpi, jpeg) en utilisant une seule forme d'archive (TIFF) qui sera traitée par un web service de manipulation d'images. Cette fonctionnalité est particulièrement attrayante dans une bibliothèque numérique à vocation éducative où la personnalisation du contenu, en fonction des divers besoins de l'utilisateur (par ex., la langue) est souhaitable.

C'est pourquoi le modèle informationnel doit modéliser les services parallèlement au texte, aux données, aux images et à toute autre information et doit caractériser les interactions de ces services avec les autres unités d'information dans la bibliothèque.

3.6 Exprimer les relations entre les matériaux numériques

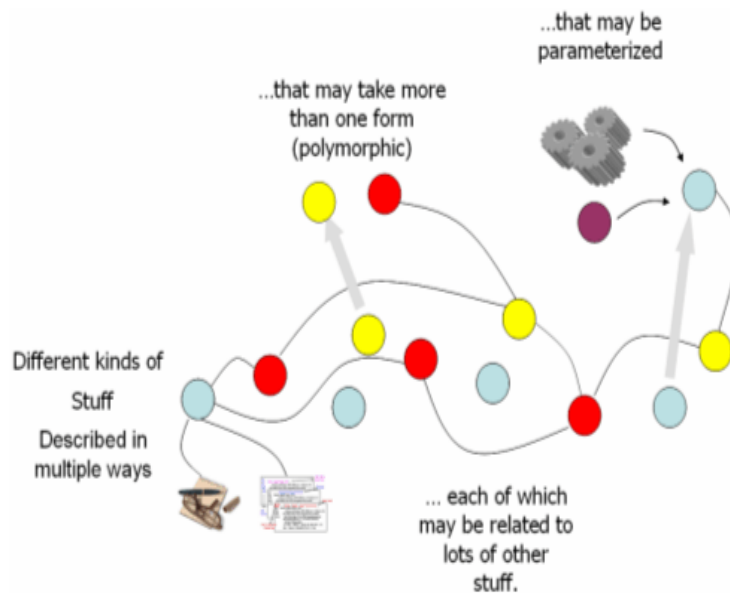


Figure 6

Dans le but de fournir une meilleure prise en compte de « l'objectif de collocation » [55], la communauté des bibliothèques a expérimenté plusieurs modèles informationnels pour modéliser les relations bibliographiques. Les Functional Requirements for Bibliographic Records (FRBR) [1] ou "Spécifications fonctionnelles des notices bibliographiques" en sont un exemple. Notre expérience au sein de la NSDL a montré que les relations bibliographiques ne sont qu'un aspect du problème. Il existe d'autres relations : entre les métadonnées et les producteurs, entre les ressources et leurs gestionnaires, entre les ressources et les taxonomies thématiques, entre les objets d'apprentissage et les programmes d'enseignement, etc. Au fur et à mesure de l'expansion de la bibliothèque, nous anticipons des besoins de modélisation pour d'autres relations spécifiques à des communautés. Le modèle informationnel doit représenter ce graphe de nœuds d'information interconnectés et les ontologies qui fournissent la méta-information sur ces relations. En outre, ces relations doivent pouvoir évoluer sans être contraintes par des schémas statiques.

4 Le Réseau d'Information Superposé (RIS)

La partie précédente proposait un modèle informationnel sous forme de graphe, avec des arêtes reliées sémantiquement et des nœuds qui soient typés de façon souple et compatible avec les web services. Nous utilisons la notion de réseau d'information superposé (RIS) [1] pour représenter ce modèle.

Le RIS emprunte à deux corpus préexistants. Les réseaux superposés ont été utilisés dans de nombreuses applications pour représenter un ensemble d'arêtes ou de connexions projeté au niveau supérieur d'un ensemble de nœuds qui existe dans d'autres environnements réseaux comme Internet, par exemple. Il existe deux domaines d'application particulièrement connus, le routage réseau [5] et la recherche sur les réseaux P2P [18]. Les graphes sémantiques, qui expriment les relations entre des ressources web, sont consubstantiels au web sémantique [13] et ont été utilisés dans des applications destinées au monde éducatif comme Edutella [41]. En fait, notre application de réseaux d'information superposés recourt aux technologies du web sémantique intégrées dans Fedora.

Les concepts qui sous-tendent le RIS sont illustrés par la figure ci-dessous, avec les couches suivantes :

- Les ressources primaires ou données brutes sélectionnées par la bibliothèque figurent au niveau inférieur. Dans la NSDL, ce sont les ressources en STEM accessibles par le web. Mais, comme nous l'avons vu, ces matériaux bruts consistent également en ensembles de données, d'agents et d'organismes qui contribuent à la bibliothèque et à ses services.
- Le réseau d'information superposé, qui se situe au niveau immédiatement supérieur, est la zone où sont modélisées les ressources de la bibliothèque, leurs descriptions et la toile d'informations tissée autour d'elles. Il est d'abord alimenté par les ressources primaires, ou les références à ces ressources via les métadonnées, qui sont représentées par des nœuds rouges. L'association et la dérivation de ces nœuds avec le niveau des ressources primaires sont matérialisées par des flèches rouges. Les tirets rouges dans le RIS indiquent les relations initiales entre ces nœuds, telles que les relations entre l'item et la collection dans la NSDL. Ici, l'alimentation du RIS est effectuée par le biais d'une collecte de métadonnées à partir des producteurs de collections, essentiellement en reprenant la fonctionnalité de l'entrepôt de métadonnées (phase I).
- L'API de contrôle d'accès, illustrée à un niveau supérieur, fournit l'accès programmatique total au RIS. Cela inclut l'accès en lecture et en écriture aux composants du modèle de données - documents, données, métadonnées, acteurs, relations, etc. - et la recherche au sein des relations (par ex., "trouver toutes les ressources impliquant une contribution de DLESE"¹⁸).
- L'API peut alors être utilisée par des contributeurs externes - par ex., des utilisateurs, des services, des fonctions de classification par ontologies, etc. - pour enrichir l'information dans le RIS. Ces requêtes effectuées à travers les API, représentées par des flèches vertes, ajoutent à la fois des nœuds supplémentaires (comme les objets d'apprentissage qui combinent des ressources existantes), qui apparaissent en vert dans la figure et de nouvelles relations entre ces nœuds, notées par des tirets verts.

Ce mouvement bidirectionnel (la représentation des ressources primaires à partir de la couche de données brutes / la représentation de l'information contextuelle à partir de la couche supérieure) permet au RIS d'évoluer à travers le temps vers un espace d'information de plus en plus riche. De la même façon que Amazon.com est une source d'information qui dépasse de loin le simple catalogue de produits, nous espérons que les bibliothèques numériques fondées sur le modèle du RIS reflèteront les communautés de savoir qui se construisent à partir des ressources de la bibliothèque.

¹⁸ <<http://dlese.org>>.

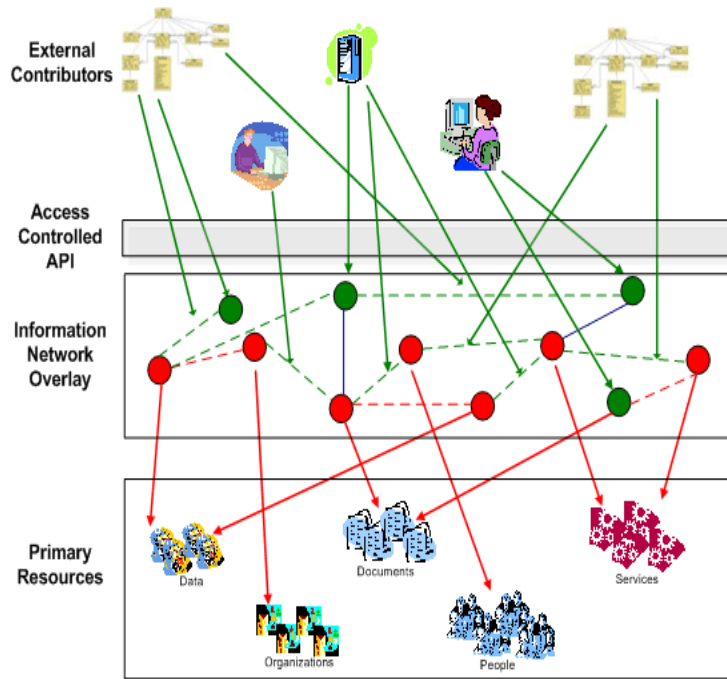


Figure 7

La plate-forme que nous utilisons pour implémenter le RIS est Fedora, un logiciel libre de gestion d'entrepôt¹⁹. Fedora a été déployé dans une variété d'applications incluant des entrepôts institutionnels, des archives, des musées, des projets de bibliothèques commerciales. Le modèle d'objets riche sur lequel repose Fedora et l'exploitation de ce modèle à travers une interface de service web font de ce logiciel un cadre idéal pour implémenter le RIS.

Chaque nœud dans le RIS correspond à un objet numérique dans Fedora. Le modèle d'objets numériques de Fedora offre des fonctionnalités considérables, combinant la gestion de contenu traditionnelle, des architectures orientées service et des technologies du web sémantique. Le modèle permet d'agréger des données locales ou distantes dans de multiples formats. Des services accessibles sur le web peuvent alors être associés aux données agrégées dans un objet numérique. Celui-ci devient alors accessible dans de multiples représentations, certaines étant des transcriptions directes des données agrégées et d'autres étant produites dynamiquement par les services web associés. Dans le contexte de la NSDL, c'est ainsi que se constitue le socle technique nécessaire au réemploi et à la construction d'objets d'apprentissage complexes [46] ; on mêle les ressources primaires et les commentaires de l'enseignant qui peuvent être présentés dynamiquement dans de multiples formats (par ex., comme des présentations Power Point ou Flash).

Chaque arête dans le RIS correspond à une relation sémantique exprimée à l'intérieur du modèle d'objets numériques de Fedora. On peut citer, comme exemples de relations entre les objets numériques dans le RIS, des relations de gestion bien connues (du type organisation des items dans une collection), des relations de structure (liens de la partie au tout entre des chapitres et un livre), des relations sémantiques utiles dans une organisation de bibliothèque numérique éducative comme la pertinence des sujets, des niveaux d'études, des programmes d'enseignement... Fedora définit une ontologie relationnelle de base en utilisant RDFS [17] et fournit un emplacement dans l'objet numérique pour exprimer des relations fondées sur cette ontologie. Des déclarations provenant d'autres ontologies peuvent aussi être incluses en complément des relations de base de Fedora. Toutes les relations exprimées dans les objets numériques sont converties dans le format Kowari [57], un triplet en RDF natif. L'interface de recherche RDQL [52] et ITQL [56] pour ce triplet est exposée en tant que service web. Comme tout service web, il peut être associé à un objet numérique, autorisant des diffusions à partir d'objets numériques qui sont paramétrés par leur contexte sémantique.

¹⁹ <<http://www.fedora.info>>.

5 L'entrepôt de données de la NSDL : NSDL phase 2

Pour distinguer notre travail de l'entrepôt de métadonnées (EM) de la première phase, nous appellerons notre implémentation du RIS "l'entrepôt de données NSDL" (EDN). La totalité des détails techniques du modèle de données implémenté dans l'EDN dépasse le cadre de cette publication. Les trois exemples de fonctionnalité suivants illustrent quelques unes des caractéristiques du modèle. L'EDN complet consiste en une multitude d'instances des éléments du modèle combinés à d'autres éléments. Par exemple, le modèle-type de métadonnées, décrit dans la [partie 5.1](#) est répété pour chacune des 1,1 million de ressources contenues dans la NSDL.

L'EDN est actuellement implémenté sous la forme d'un entrepôt Fedora unique géré par l'équipe « noyau intégré ». A l'avenir, nous pensons implémenter l'EDN sous la forme d'un ensemble d'entrepôts fédérés.

Chacun des exemples suivants s'accompagne d'une illustration, dans laquelle les cercles représentent des nœuds dans le réseau d'information, implémentés en tant qu'objets numériques Fedora. Chaque cercle a la couleur qui correspond au type d'information qu'il représente dans le contexte de l'exemple de modélisation. Les lignes représentent des relations sémantiquement chargées entre les unités d'information. Comme toutes les autres dans Fedora, ces relations sont stockées à l'intérieur de l'objet numérique et ensuite indexées dans le triplet Kowari. On peut alors rechercher ces relations à travers des requêtes de graphes.

5.1 Des métadonnées multi-sources et multi-formats grâce au marquage

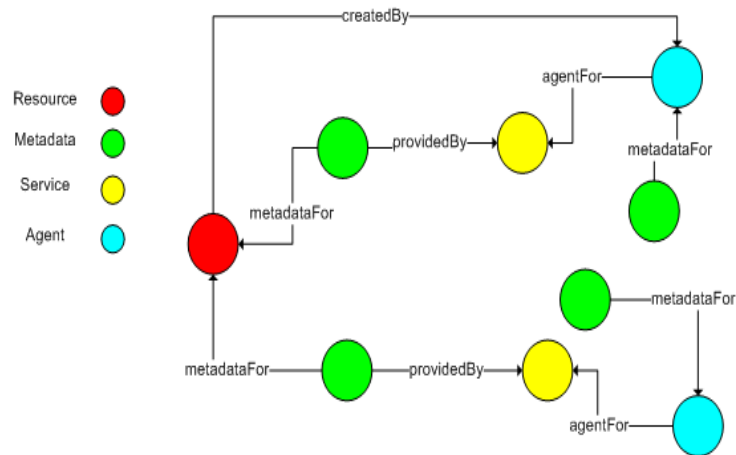


Figure 8

L'illustration ci-dessus montre le modèle EDN qui associe des métadonnées multi-sources et multi-formats à une ressource. Chaque objet numérique de métadonnées agrège plusieurs formats provenant d'un seul fournisseur de métadonnées. Grâce à la capacité de diffusion dynamique de Fedora, certains de ces formats sont générés de façon automatique à partir d'un format de base. Le lien entre l'objet numérique de métadonnées et son fournisseur et celui entre la ressource et son créateur ou sélectionneur fournit une information de marquage. Le marquage est important pour toute bibliothèque dont les données et les métadonnées proviennent de plusieurs sources. Relié à un référentiel de notoriété, le marquage permet de déterminer la qualité des ressources et de leurs descriptions.

5.2 Comptes rendus et annotations non structurés

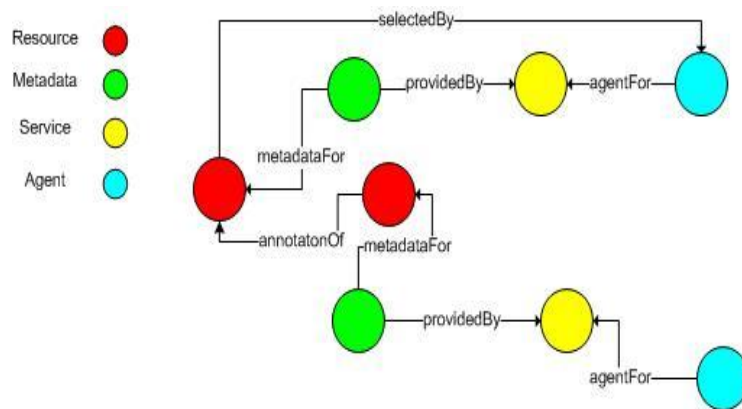


Figure 9

Bien que les métadonnées structurées soient utiles pour répondre à de multiples besoins, les annotations et les comptes rendus non-structurés ont eux aussi leur importance. Le modèle représente ces annotations et comptes rendus comme des ressources à part entière - leur statut d'annotation dépend de leur association à une ressource cible par le biais d'une relation "annotationFor". C'est un exemple de polymorphisme au sein du RIS, selon lequel un nœud peut endosser plusieurs caractéristiques. Autres exemples : une ressource peut aussi être un "agent", une "collection" peut aussi être un "item" qui peut être agrégé dans d'autres collections. L'objet numérique de Fedora rend tout cela possible, sans les contraintes des architectures orientées "objet" avec relation d'héritage unique. Par essence, un objet numérique peut revêtir toute combinaison d'identités types.

5.3 Collections et agrégations

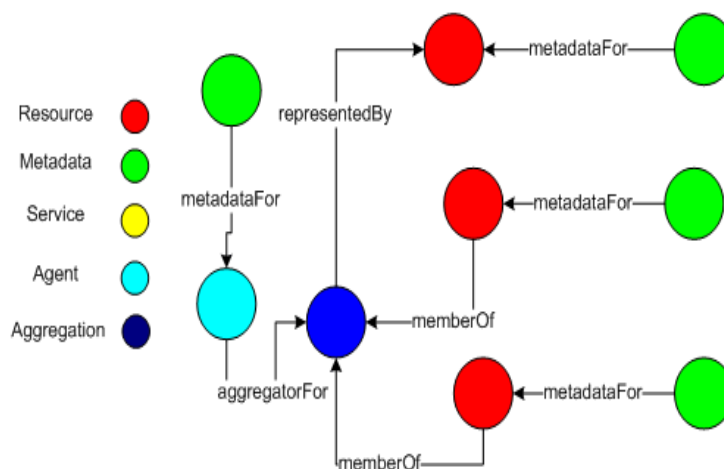


Figure 10

La phase I de l'architecture de la NSDL ne permettait qu'une seule forme d'agrégation, qui exprime les relations entre un fournisseur de métadonnées et l'ensemble des métadonnées collectées chez ce fournisseur. L'implémentation de l'EDN inclut un modèle d'agrégation fondé sur des ensembles, qui permet à un nombre quelconque de ressources d'être contenues dans un nombre quelconque d'ensembles. Comme le montre l'illustration, une agrégation est représentée par ("representedBy") une ressource. Cette ressource fournit la sémantique pour l'agrégation. Par exemple, une agrégation peut indiquer un ensemble de ressources qui se rapportent à un standard pédagogique officiel. Dans ce cas, la ressource au bout du lien "representedBy" exprime ce standard. Les agrégations sont elles-mêmes des "ressources", qui peuvent être imbriquées dans des agrégations additionnelles. Disposant d'une sémantique, ces agrégations constituent, dans le modèle de données, les fondations pour une contextualisation riche des ressources de la bibliothèque numérique.

5.4 Avancement de l'EDN

Au moment de la rédaction de cet article, le chargement initial de l'EDN, à partir des données de l'entrepôt préexistant, est presque terminé (plus de 1,1 million de notices). Le graphe RIS qui en résulte contient environ 1,5 million de nœuds et 10 millions d'arêtes explicites (d'autres arêtes implicites étant propres à Fedora). Dans un article à venir, nous rendrons compte de nos expériences concernant les triplets RDF et de la démarche d'élever le RIS à un niveau supérieur. En particulier, nous avons remarqué qu'un grand nombre d'indicateurs de performance sur les triplets s'appliquent à d'autres applications du web sémantique.

Une fois achevé le chargement de données, nous communiquerons la spécification de l'API de l'EDN à la communauté de la NSDL. Ce sera le début du processus de déploiement et d'approfondissement du RIS, qui s'appuiera sur l'information contextuelle ajoutée par la communauté de la NSDL. Des résultats intéressants concernant la nature du RIS sont à prévoir, au fur et à mesure qu'il se développera.

Conclusion

A l'heure de Google, qu'est-ce qu'une bibliothèque numérique, au juste ? Une telle question ne peut qu'enflammer les passions. Nous avons ardemment défendu les succès accomplis, en une décennie, par la communauté des bibliothèques numériques. Mais la stupéfiante réussite des moteurs de recherche commerciaux a changé la donne. Les fonctions de recherche et d'accès sur un ensemble de ressources, en dépit de leur importance, ne suffisent pas. Les bibliothèques numériques ont besoin de se distinguer des moteurs de recherche par la façon dont elles ajoutent de la valeur aux ressources internet. Cette valeur ajoutée consiste à mettre ces ressources en contexte, à les enrichir par de nouvelles informations et des relations qui expriment les modèles d'usage et le savoir de la communauté servie par la bibliothèque. La bibliothèque numérique devient alors un espace pour l'information collaborative et l'enrichissement - bien plus qu'un simple endroit où trouver de l'information et y accéder.

Le travail que nous avons mené au sein de la NSDL a démontré que le modèle centré sur les métadonnées, que tous connaissent, est insuffisant pour ce type de fonctionnalité. Nous avons conçu et implémenté un réseau d'information superposé au sein de Fedora, qui comporte toutes les fonctionnalités de l'entrepôt de métadonnées existant, mais qui modélise des relations, des services et de multiples types d'information à l'intérieur d'une application service web. Ce riche dépôt d'information fournira les bases de la prochaine étape de notre travail : implémenter une suite non limitée de services à l'utilisateur, à même de réaliser le "laboratoire pour le partage et la construction du savoir" imaginé dans le rapport initial du projet NSDL [59].

Remerciements : *Cet article reprend le travail de plusieurs personnes, en plus des auteurs. Le groupe Fedora, tout spécialement Chris Wilper et Eddie Shin, mérite un hommage pour le travail difficile qu'a nécessité l'implémentation de ces notions dans le logiciel libre Fedora. Les membres du groupe NSDL, en particulier Tim Cornwell, Elly Cramer et Naomi Dushay, ont joué un rôle majeur dans la formulation du modèle de données NSDL et son implémentation dans l'EDN. Le groupe NSDL dans son ensemble adresse ses plus vifs remerciements à Lee Zia, qui défend le projet auprès de la NSF depuis des années. Les réalisations décrites ici ont bénéficié de plusieurs subventions. Le travail concernant l'EDN de la NSDL a bénéficié des subventions n° 0227648, 0227656 et 0227888 de la National Science Foundation. Le travail concernant les réseaux d'information superposés a bénéficié de la subvention n° 0430906 de la National Science Foundation. Le travail sur Fedora est financé par la Andrew W. Mellon Foundation. Toutes les opinions, conclusions et recommandations contenues dans cet article sont celles de leurs auteurs et ne reflètent pas nécessairement les points de vue de la National Science Foundation ou de la Andrew W. Mellon Foundation. Un grand merci à Lucy Lagoze qui a montré à Carl Lagoze combien il est difficile pour un étudiant d'utiliser des moteurs de recherche et qui a livré quelques enseignements sur l'importance du contexte et des modèles d'usage.*

Merci, également, à Mike Keller et Vicky Reich de nous avoir autorisé à adopter et adapter un titre qu'ils ont utilisé dans un article antérieur [23].

Note du traducteur : [1] NDT : en anglais, Information network overlay (INO)

Bibliographie

[1] "Functional Requirements for Bibliographic Records," International Federation of Library Associations and Institutions March 1998. <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>>.

[2] New Pathways to the National Science Digital Library, 2004 <http://www.infosci.cornell.edu/news/NSDL_Pathways.pdf>.

- [3] "Report of the Science, Mathematics, Engineering, and Technology Education Library Workshop," National Science Foundation, Washington, DC, Workshop Report July 21-23 1998. <<http://www.dlib.org/smete/public/report.html>>.
- [4] J. Abbas, C. Norris, and E. Soloway, "Middle School Children's Use of the ARTEMIS Digital Library," presented at ACM/IEEE Joint Conference on Digital Libraries (JCDL '02), Portland, OR, 2002.
- [5] D. G. Andersen, H. Balakrishnan, and M. F. Kaashoek, "Resilient Overlay Networks," presented at 18th ACM SOSP, Banff, Canada, 2001.
- [6] W. Y. Arms, N. Dushay, D. W. Fulker, and C. Lagoze, "A Case Study in Metadata Harvesting : the NSDL," *Library Hi Tech*, 21 (2), 2003.
- [7] W. Y. Arms, D. Hillmann, C. Lagoze, D. Krafft, R. Marisa, J. Saylor, C. Terrizzi, and H. Van de Sompel, "A Spectrum of Interoperability : The Site for Science Prototype for the NSDL," *D-Lib Magazine*, 8 (1), 2002. <[doi:10.1045/january2002-arms](https://doi.org/10.1045/january2002-arms)>.
- [8] D. E. Atkins, Report of the Santa Fe Planning Workshop on Distributed Knowledge Work Environments : Digital Libraries, 1997 <<http://www.si.umich.edu/SantaFe/report.html>>.
- [9] K. Auletta, "Under the Wire," *New Yorker*, January 17, 1994.
- [10] C. Baru, R. Moore, A. Rajasekar, and M. Wan, "The SDSC Storage Resource Broker," presented at CASCON'98, Toronto, 1998.
- [11] D. Bearman, G. Rust, S. Weibel, E. Miller, and J. Trant, "A Common Model to Support Interoperable Metadata. Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities," *D-Lib Magazine*, 5 (January), 1999. <[doi:10.1045/january99-bearman](https://doi.org/10.1045/january99-bearman)>.
- [12] J. Bekaert, P. Hochstenbach, and H. Van de Sompel, "Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library," *D-Lib Magazine*, 9 (11), 2003. <[doi:10.1045/november2003-bekaert](https://doi.org/10.1045/november2003-bekaert)>.
- [13] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, (50), May, 2001.
- [14] C. L. Borgman, "Digital libraries and the continuum of scholarly communication," *Journal of Documentation*, 56 (4), pp. 412-430, 2000.
- [15] C. L. Borgman, "The invisible library : Paradox of the global information infrastructure," *Library Trends*, 51 (4), pp. 652, 2003.
- [16] C. L. Borgman, "What are digital libraries ? Competing visions," *Information Processing & Management*, 1999 (35), pp. 227-243, 1999.
- [17] D. Brickley and R. V. Guha, "RDF Vocabulary Description Language 1.0 : RDF Schema," W3C, Recommendation February 10 2004. <<http://www.w3.org/TR/rdf-schema>>.
- [18] A. Crespo and H. Garcia-Molina, "Semantic overlay networks for p2p systems," Stanford University, Palo Alto 2003.
- [19] R. Daniel Jr. and C. Lagoze, "Extending the Warwick Framework : From Metadata Containers to Active Digital Objects," *D-Lib Magazine* (November), 1997. <[doi:10.1045/november97-daniel](https://doi.org/10.1045/november97-daniel)>.
- [20] E. Fox, R. M. Akscyn, R. K. Furuta, and J. J. Leggett, "Digital libraries," *Communications of the ACM*, 38 (4), pp. 22-28, 1995.
- [21] F. Fukuyama, *The end of history and the last man*. New York, Toronto : Free Press, 1992.
- [22] R. Kahn and R. Wilensky, "A Framework for Distributed Digital Object Services," Corporation for National Research Initiatives, Reston, Working Paper cnri.dlib/tn95-01, 1995. <<http://www.cnri.reston.va.us/k-w.html>>.
- [23] M. A. Keller, V. Reich, and A. C. Herkovic, "What is a library anymore, anyway ?," *First Monday*, 8, May 5, 2003.
- [24] C. Lagoze, "The Warwick Framework : A Container Architecture for Diverse Sets of Metadata," *D-Lib Magazine*, 2 (7/8), 1996. <[doi:10.1045/july96-weibel](https://doi.org/10.1045/july96-weibel)>.
- [25] C. Lagoze, W. Arms, S. Gan, D. Hillmann, C. Ingram, D. Krafft, R. Marisa, J. Phipps, J. Saylor, C. Terrizzi, W. Hoehn, D. Millman, J. Allan, S. Guzman-Lara, and T. Kalt, "Core Services in the Architecture of the National Digital Library for Science Education (NSDL)," presented at Joint Conference on Digital Libraries, Portland, Oregon, 2002.
- [26] C. Lagoze and J. R. Davis, "Dienst - An Architecture for Distributed Document Libraries," *Communications of the ACM*, 38 (4), pp. 47, 1995.
- [27] C. Lagoze, S. Payette, E. Shin, and C. Wilper, *Fedora : An Architecture for Complex Objects and their Relationships*, 2005 <<http://arxiv.org/abs/cs.DL/0501012>>.
- [28] C. Lagoze and A. Singhal, "Information Discovery : Needles and Haystacks," *IEEE Internet Computing*, 2005 (May/June), 2005.
- [29] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, *The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0*, 2002 <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>.

- [30] D. Levy, "Cataloging in the Digital Order," presented at The Second Annual Conference on the Theory and Practice of Digital Libraries, 1995.
- [31] D. Levy, "Digital Libraries and the Problem of Purpose," *Bulletin of the American Society for Information Science*, 26 (6), 2000.
- [32] Library of Congress, METS : An Overview & Tutorial, 2004 <<http://www.loc.gov/standards/mets/METSOverview.v2.html>>.
- [33] C. Lynch, "The Battle to Define the Future of the Book in the Digital World," *First Monday*, 6 (6), June 4, 2001.
- [34] C. A. Lynch and H. Garcia-Molina, "Interoperability, Scaling, and the Digital Libraries Research Agenda," IITA Digital Libraries Workshop May 18-19 1995. <<http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>>.
- [35] C. A. Lynch and M. A. Keller, googlization, digital repositories, distance education, and privacy, 2005 <<http://www.learningtimes.net/acrlarchive.html>>.
- [36] B. Marshall, Y. Zhang, H. Chen, A. Lally, R. Shen, E. A. Fox, and L. Cassel, "Convergence of Knowledge Management and E-Learning : the GetSmart Experience," presented at ACM/IEEE Joint Conference on Digital Libraries (JCDL '03), Houston, TX, 2003.
- [37] K. Martin, "Learning in Context," *Issues of Teaching and Learning*, 4 (8), September, 1998.
- [38] G. McCalla, "The Ecological Approach to the Design of E-Learning Environments : Purpose-based Capture and Use of the Information about Learners," *Journal of Interactive Media in Education*, 7 (Special Issue on the Educational Semantic Web), 2004.
- [39] F. McMartin and Y. Terada, "Digital Library Services for Authors of Learning Materials," presented at ACM/IEEE Joint Conference on Digital Libraries (JCDL '02), Portland, OR, 2002.
- [40] National Information Standards Organization (U.S.), The OpenURL Framework for Context-Sensitive Services, 2003 <http://www.niso.org/standards/resources/Z39_88_2004.pdf>.
- [41] W. Nejdil, B. Wolf, and C. Qu, "EDUTELLA : A P2P Networking Infrastructure Based on RDF," presented at WWW2002, Honolulu, 2002.
- [42] A. M. Ouksel and A. Sheth, "Semantic Interoperability in Global Information Systems," *SIGMOD Record*, 28 (1), 1999.
- [43] P. Parrish, "The Trouble with Learning Objects," *Educational Technology Research and Development*, 52 (1), pp. 49-67, 2004.
- [44] S. Payette and C. Lagoze, "Flexible and Extensible Digital Object and Repository Architecture (FEDORA)," presented at Second European Conference on Research and Advanced Technology for Digital Libraries, Heraklion, Crete, 1998.
- [45] President's Information Technology Advisory Committee : Panel on Digital Libraries, "Digital Libraries : Universal Access to Human Knowledge," PITAC February 2001. <<http://www.itrd.gov/pubs/pitac/pitac-dl-9feb01.pdf>>.
- [46] M. Recker, Instructional Architect, 2004 <<http://ia.usu.edu/>>.
- [47] M. Recker, J. Dorward, and L. M. Nelson, "Discovery and Use of Online Learning Resources : Case Study Findings," *Educational Technology and Society*, 7 (2), pp. 93-104, 2004.
- [48] M. Recker and A. Walker, "Collaboratively filtering learning objects," in *Designing Instruction with Learning Objects*, D. A. Wiley, Ed., 2000.
- [49] T. C. Reeves, *The Impact of Media and Technology in Schools : A Research Report prepared for The Bertelsmann Foundation*, 1998. <<http://it.coe.uga.edu/treeves/edit6900/BertelsmannReeves98.pdf>>.
- [50] V. Reich, "LOCKSS : A Permanent Web Publishing and Access System," *D-Lib Magazine*, 7 (6), 2001. <[doi:10.1045/june2001-reich](http://doi.org/10.1045/june2001-reich)>.
- [51] G. Salton, *Dynamic information and library processing*. Englewood Cliffs, N.J. : Prentice-Hall, 1975.
- [52] A. Seaborne, RDQL - A Query Language for RDF, 2004. <<http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>>.
- [53] M. Smith, M. Bass, G. McClellan, R. Tansley, M. Barton, M. Branschovsky, D. Stuve et J. H. Walker, "DSpace : An Open Source Dynamic Digital Repository," *D-Lib Magazine*, 9 (1), 2003. <[doi:10.1045/january2003-smith](http://doi.org/10.1045/january2003-smith)>.
- [54] J. Surowiecki, *The wisdom of crowds : why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*, 1st ed. New York : Doubleday, 2004.
- [55] E. Svenonius, *The intellectual foundation of information organization*. Cambridge, Mass. : MIT Press, 2000.
- [56] Tucana Technologies, iTQL Commands, 2004 <<http://kowari.org/oldsite/271.htm>>.
- [57] Tucana Technologies, Kowari metastore, 2004 <<http://www.kowari.org/>>.

- [58] J. Ward, "A Quantitative Analysis of Unqualified Dublin Core Metadata Element Set Usage within Data Providers Registered with the Open Archives Initiative," presented at Joint Conference on Digital Libraries, Houston, 2003.
- [59] F. Wattenberg, "A National Digital Libraries for Science, Mathematics, Engineering, and Technology Education," D-Lib Magazine, 1998 (October), 1998. <[doi:10.1045/october98-wattenberg](https://doi.org/10.1045/october98-wattenberg)>.
- [60] D. A. Wiley, "Connecting learning objects to instructional design theory : A definition, a metaphor, and a taxonomy," in The Instructional Use of Learning Objects : Online Version, D. A. Wiley, Ed., 2000.
- [61] L. L. Zia, "The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program," D-Lib Magazine, 8 (11), 2002. <[doi:10.1045/november2002-zia](https://doi.org/10.1045/november2002-zia)>.