

**Zeitschrift:** Arbido  
**Herausgeber:** Verein Schweizerischer Archivarinnen und Archivare; Bibliothek Information Schweiz  
**Band:** - (2013)  
**Heft:** 4: Linked Open Data Big Data : alles vernetzt = tout en réseau = tutto in rete  
**Rubrik:** Grundlagen und Perspektiven

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 12.04.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# I. Grundlagen und Perspektiven

## Dimensionen und Zusammenhänge grosser, verknüpfter, offener und wissenschaftlicher Daten

René Schneider, Haute Ecole de  
Gestion de Genève

**Nachdem die Flut der Diskussionen, Kommentare und Prophezeiungen das Social Web betreffend nun verebbt ist, bricht bereits die neue Flutwelle über die Dämme und überspült uns mit dem, was im Social Web und anderswo in Massen ohne Unterlass und Überlegung produziert wird: Daten!**

### Einleitung

Man gewinnt nicht nur den Eindruck, dass die verhältnismässig alte Bezeichnung der elektronischen Datenverarbeitung eine ganz neue Bedeutung erhalten hat, sondern dass die Daten gleichsam selbst zu Information geworden sind, auch wenn dies auf den ersten Blick als widersinnig erscheinen mag. Dieser paradoxe Eindruck verstärkt sich jedoch, wenn in einflussreichen Kreisen nicht mehr vom Informationszeitalter gesprochen wird, sondern vom Zeitalter der Daten. So wird bereits ein neues und viertes wissenschaftliches Paradigma<sup>1</sup> heraufbeschworen und nach den wissenschaftstheoretischen Vorläufern der empirischen Beschreibung von Naturphänomenen (1. Paradigma), der theoretischen Generalisierung (2. Paradigma) und der computerbasierten Simulation (3. Paradigma) nun das Zeitalter der daten-

intensiven wissenschaftlichen Entdeckung verkündet.

Aber was heisst das eigentlich? Sind die übrigen Paradigmen der Wissenschaftsgeschichte damit bedeutungslos geworden? Und wie verhält es sich mit dem Kontext der Daten, durch den sie normalerweise zu Information werden? Oder sind Daten nicht mehr gleich Daten? Und was heisst dies für Bibliotheken, Archive und Dokumentationszentren? Sind sie oder ihre Daten betroffen? Ergeben sich neue Arbeitsaufträge für sie oder werden sie von einem weiteren Bedeutungsverlust betroffen sein?

Im Folgenden soll versucht werden, diese Fragen einzugrenzen und entsprechend dem derzeitigen Stand der Dinge zu beantworten. Dies soll anhand dreier unterschiedlicher Datentypen illustriert werden, die derzeit unterschiedlich stark diskutiert werden, sich trotz ihrer Unterschiede aber an einzelnen Punkten berühren bzw. in unterschiedlicher Beziehung zueinander stehen: Big Data, Linked Open Data und Research Data (d.h. grosse Datenmengen, verknüpfte und offene Datenmengen sowie Forschungsdaten). Diese dreifache Auffächerung von Datenverarbeitungsinitiativen beantwortet bereits unmittelbar eine der weiter oben formulierten Fragen und zeigt,

**Die Informationswissenschaft und alle damit verbundenen Institutionen stehen damit auch vor der Herausforderung, sich auf diese Daten und die damit verbundenen Aufgaben einzustellen.**

dass Daten wirklich nicht mehr gleich Daten sind. Die Informationswissenschaft und alle damit verbundenen Institutionen stehen damit auch vor der Herausforderung, sich auf diese Daten

und die damit verbundenen Aufgaben einzustellen.

### Big Data

Den grössten Block (im doppelten Sinn) machen dabei die gemeinhin unter Big Data subsumierten Datensätze aus, von deren Existenz und Konsequenz mittlerweile auch die Normalbürger erfahren haben, was insofern gut ist, da sie selbst zu den Hauptproduzenten, bedauerlicherweise aber nicht zu den Nutzern gehören. Big Data wurde dabei letztlich durch zwei Phänomene möglich: erstens die weiterhin exponentiell steigende Leistungsfähigkeit der Prozessoren sowie die damit parallel verlaufende Steigerung der Speicherkapazitäten der Rechner, zweitens durch den Anstieg der täglich bzw. sekundlich produzierten und aufgezeichneten Datenmengen, die allorts durch die Vernetzung der Endgeräte (vom Grossrechner bis zum Smartphone und den darauf befindlichen Apps) ins Web eingespeist und von dort weiterverarbeitet werden. Gespeichert wird dabei alles, was geht: Daten, mit oder ohne Metadaten, oder die Metadaten alleine, aus denen sich ja streng genommen der eigentliche Kontext der Daten bzw. ihr Verständnis ergibt. Entscheidend ist aber, dass aus diesen Daten ein neuer Kontext erzeugt wird und zwar mithilfe spezieller Algorithmen, genauer mit sogenannten Inferenzmechanismen, also Schlussfolgerungsregeln der Maschinen, die sie verwalten und bearbeiten. In den Schlussfolgerungsregeln des Big Data werden orthogonale, d.h. prinzipiell voneinander unabhängige Merkmale miteinander verknüpft, um neue oder bis dahin unbekannte Zusammenhänge aufzuzeigen<sup>2</sup>. Mit dieser an und für sich recht einfachen Vorgehensweise, die von Menschen täglich bewusst oder unbewusst ausgeführt wird, stellt sich

1 Jim Gray on e-Science: «A transformed Scientific Method», in: Hey, Tony, Tansley, Stewart, Tolle, Kristin: *The Fourth Paradigm: Data Intensive Scientific Discovery*. Microsoft Corporation. 2009, S. xviii [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_complete\\_lr.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf)

2 Dietmar Dath, Ranga Yogeshwar (2013): «Rechnen Sie damit, lebenslang ein Verdächtiger zu sein», FAZ vom 12.7.2013.



der einfache, und von daher sehr erfolgsversprechende Business Case von Big Data dar, der nun aber von den Maschinen ausgeführt wird. Damit sollte auch klar sein, dass hinter Big Data zu einem sehr grossen Teil wirtschaftliche oder andere Interessen stehen, die nicht primär im Interesse der Produzenten sind, auch wenn die ethischen Implikationen an dieser Stelle nicht ausgeführt werden sollen.

### Linked Open Data (LOD)

Eher unabhängig davon verläuft die Linked-Open-Data-Initiative<sup>3</sup>. Unabhängig in dem Sinne, dass die hier verknüpften (Linked) und veröffentlichten (Open) Datensätze nicht notwendigerweise eine Teilmenge der grossen (Big) Daten sind, im Gegenteil: Der Umstand, dass LOD die gemeinnützige Publikation sowie eine Verknüpfung zur *Conditio sine qua non* macht, schliesst den orthogonalen Charakter der grossen Datensätze aus und eröffnet eine ganz andere Sichtweise. Im Gegensatz zu Big Data beschränkt sich LOD auf Daten, die im Web zur Verfügung gestellt werden, und nicht in irgendwelchen nichtöffentlichen Datenbanken oder Clouds gespeichert werden. Des Weiteren ist ihr Umfang, trotz der an und für sich unvorstellbaren Menge der Daten, die die LOD-Cloud mittlerweile enthält, vergleichsweise klein.

Im Rahmen von Linked Open Data soll ein Grundproblem gelöst werden, das darin besteht, dass der grösste Teil des Wissens, das im Web publiziert wird, in relationalen Datenbanken gespeichert, aber nur einem sehr flachen Format (genauer: HTML, der Hypertext Markup Language) repräsentiert wird. Dies führt dazu, dass die ursprünglich gesetzten Relationen verloren gehen und eine direkte Kommunikation (im Sinne eines reziproken Verständnisses) zwischen den Datenbanken nicht möglich ist. Dieses Problem wird nun dadurch gelöst, dass eine neue Datenebene eingeführt wird, die Auskunft über die semantischen Relationen gibt. Der eigentliche Mehrwert von Linked Open Data besteht also darin, dass sie die Relationen aller Datenbanken in ein einheitliches Format überführen, was – genau wie im Fall von Big Data – dazu führen kann, dass die Maschine über

sie rasonieren kann. Wiederum geht es also nicht um kognitive Prozesse von Menschen, sondern darum, diese in Rechnern zu repräsentieren und zu simulieren und den Maschinen zu erlauben, selbstständig Rückschlüsse aus den verknüpften Daten zu ziehen.

### Forschungsdaten

Etwas im Abseits dieser grossen Bewegungen findet – auf den Bereich der Wissenschaften beschränkt – seit einigen Jahren eine weitere Diskussion statt, die die sogenannten Forschungsdaten zum Hauptgegenstand des Interesses gemacht hat.

Die gesamten Anstrengungen lassen sich sehr gut an zwei grundlegenden Modellen illustrieren: dem Modell des Datenlebenszyklus<sup>4</sup>, das von ersten Projektideen über die Ablage und dauerhafte Speicherung von Datensätzen bis hin zu deren Integration in Publikationen und deren Zitation reicht, und dem Modell für das Datenkontinuum<sup>5</sup>, das eine kontinuierliche Weitergabe und den Austausch der Daten garantieren soll und zwar unabhängig davon, ob es sich um Rohdaten, strukturierte Daten, Metadaten oder die in späteren Publikationen eingebundenen Daten handelt. Der Mehrwert für den Forscher und die Öffentlichkeit ergibt sich daraus, dass beide quasi uneingeschränkt und über alle Soft- und Hardwarewechsel Zugang zu diesen Daten

**Der eigentliche Mehrwert von Linked Open Data besteht also darin, dass sie die Relationen aller Datenbanken in ein einheitliches Format überführen, was dazu führen kann, dass die Maschine über sie rasonieren kann.**

für eine Nachnutzung oder Überprüfung haben. Auch wenn man davon ausgehen kann, dass sich das Interesse dafür sowohl bei den meisten Forschern als auch der Gesellschaft, die diese Forschung finanziert, in Grenzen hält, wird eine Sicherung der Forschungsdaten mittlerweile zu den guten und festen Regeln für eine transparente Forschung gezählt. Der grösste Vorteil für den Forscher ergibt sich daraus, dass die Datensätze selbst, d.h. die Primärdaten, zitier- und publizier-

bar werden. In einigen Disziplinen gibt es parallel dazu bereits «Peer Reviewed Data Paper», die allein die Datensätze enthalten.

Was das spezielle Verhältnis von Forschungsdaten zu Big und Linked Open Data betrifft, lässt sich sagen, dass sie – etwa im Fall des Large Hadron Colliders – sehr gross sein können. Sie können auch mit LOD verknüpft und angereichert werden, müssen es aber nicht.

Dabei sollte nicht ausser Acht gelassen werden, dass bei den Forschungsdaten eine ganz andere grundsätzliche Frage im Vordergrund steht: wie bewahren wir (Roh-)Daten langfristig auf, sodass sie uns zu einem späteren Zeitpunkt zur Nachnutzung wieder

**Der grösste Vorteil für den Forscher ergibt sich daraus, dass die Datensätze selbst, d.h. die Primärdaten, zitier- und publizierbar werden.**

problemlos zur Verfügung stehen? Es geht also in erster Linie um ganz grundsätzliche Fragen der Archivierung, der Kompatibilität und der Interoperabilität, im Grunde also um Fragen, die auch die Bereiche der grossen und der verknüpften Datenmengen irgendwann betreffen könnten. Man kann sich aber auch nicht des Eindrucks erwehren, dass gerade in der Debatte um die Forschungsdaten so prinzipielle Fragen angegangen werden, dass den Beteiligten noch gar nicht genau klar ist, wie diese gelöst werden sollen, auch wenn es zumindest in einzelnen Disziplinen bereits Lösungsansätze bzw. respektable Lösungsvorschläge gibt.

3 Bizer, Christian, Tom Heath, and Tim Berners-Lee (2009). «Linked data-the story so far.» *International Journal on Semantic Web and Information Systems (IJSWIS)* 5:3, 1–22.

4 Sarah Higgins (2008): «The DCC Curation Lifecycle Model». *International Journal of Digital Curation* 3(1), 134–148.

5 Andrew Treloar (2011): *Private Research, Shared Research, Publication, and the Boundary Transitions*. [http://andrew.treloar.net/research/diagrams/data\\_curation\\_continuum.pdf](http://andrew.treloar.net/research/diagrams/data_curation_continuum.pdf)



## Bibliotheken und Daten: eine neue und notwendige Herausforderung

Bleibt die Frage nach der Rolle der Institutionen, die traditionell mit diesen Fragen beschäftigt waren. Zuvorderst handelt es sich dabei um die Bibliotheken, auch wenn die Archive und Dokumentationszentren in einzelnen Bereichen ein Wort mitsprechen könnten, sofern ihre Expertise noch gefragt ist.

Die grossen Datenmengen (Big Data) können die Bibliotheken, Archive und Dokumentationszentren schnell ausser Acht lassen, auch wenn sie traditionell damit beauftragt waren, grössere Informationsmengen aufzubewahren. An Big Data zeigt sich vielmehr, wie schnelllebig die Zeitläufte geworden sind. Wähten sich die Bibliotheken (zumindest einige erlauchte unter ihnen) bis vor Kurzem noch an vorderster Front in der grössten Digitalisierungsinitiative und durften sie sich noch als strategisch wichtige Datenlieferanten der erfolgreichsten Suchmaschine ansehen, müssen sie nunmehr erkennen, dass sie vom Gehilfen des globalen Players zum staunenden Betrachter noch weitaus grösserer Datenverwalter geworden sind.

Anders verhält es sich bei der Linked-Open-Data-Initiative. Sollten die Bibliotheken und ihre Katalogisierungsabteilungen sowie die Verbundzentren dazu bereit sein, würde sich Linked Open Data als der notwendige, wenn nicht sogar zwingende kommende Schritt darstellen, um die Kataloge in ein Format zu bringen, von dem deren Anbieter selbst, aber auch viele andere profitieren können. Schlussendlich könnten damit – neben vielen anderen positiven Seiteneffekten und neuen Nutzungsformen der Katalogdaten – vor allen Dingen die Probleme der Konvergenz unterschiedlicher Verbunddaten sowie die komplizierten Bemühungen für die Erstellung standardisierter Metakataloge ein Ende haben.

Ähnlich verhält es sich mit den Forschungsdaten, auch wenn hierbei weniger die Katalogisierenden als vielmehr die Kollegen aus dem Bereich des E-Publishing sowie die Betreiber der Open-Access-Plattformen eine Rolle spielen werden. Ihre Aufgabe wird es sein, neben den Langzeitarchivaren und im Verbund mit den Datenzentren die Datensätze, so gross oder (eher) klein sie auch sein mögen, mit einer persistenten Adressierbarkeit für eine Nachnutzung, Nanopublikation oder -zitation zur Verfügung zu stellen.

### Erinnerung, Betrachtung, Erwartung

Festzuhalten bleibt, dass es insgesamt sehr unterschiedliche Vorstellungen vom Umgang mit Daten gibt, die derzeit in drei unterschiedlich grossen Initiativen angegangen und diskutiert werden, wobei durchaus Berührungspunkte bestehen. Das gegenwärtige Bild der Dateninitiativen zeigt sich so dynamisch, dass derzeit nicht antizi-

**Aus dieser Perspektive betrachtet, lassen sich kurioserweise Parallelen zu den drei Dimensionen der Zeit erkennen, die Augustinus im elften Buch der Confessiones in seiner Reflektion über das Wesen der Zeit entwickelte.**

pierbar ist, ob die Dateninitiativen weiter zersplittern, konvergieren oder etwa zu einem Schichtenmodell zueinander finden können.

Die reine Betrachtung der Gegenwart liefert aber noch keine Antworten auf die zwei grundlegenden Fragen: Warum heben wir die Daten eigentlich auf? Und: Warum interessieren wir uns so sehr für sie? Diese zwei Fragen lassen sich letztlich mit ganz grundsätzlichen menschlichen Bedürfnissen beantworten: erstens einem eher rückwärtsgewandten Bedürfnis, nämlich dem Sammeln und dem daraus resultierenden Bedürfnis nach Erinnerung, zweitens mit der Organisation bzw. der Repräsentation von Wissen, um es immer, d.h. gegenwärtig, zur Verfügung zur haben, und drittens dem eher vorwärtsgewandten Wunsch, Zukünftiges vorhersagen zu können. Aus dieser Perspektive betrachtet, lassen sich kurio-

serweise Parallelen zu den drei Dimensionen der Zeit erkennen, die Augustinus im elften Buch der Confessiones in seiner Reflektion über das Wesen der Zeit entwickelte. Interessanterweise sprach er nicht allein von Praeteritum, Praesens und Futurum, sondern von «praesens de praeteritis memoria, praesens de praesentibus contuitus, praesens de futuris expectatio»<sup>6</sup>, d.h., er assoziierte die Erfahrung der Zeit mit drei zentralen Begriffen: Memoria (Erinnerung), Contuitus (Betrachtung) und Expectatio (Erwartung).

So mag es für ein grundlegendes Verständnis hilfreich sein, die unterschiedlichen Konzepte zum Umgang mit Daten schlussendlich mithilfe dieser drei Dimensionen zu begreifen: der Memoria entsprechen die Bemühungen um die Archivierung und Nachnutzung von Daten; die LOD-Initiative zeigt das Bemühen, eine Datenschicht für die gegenwärtige Betrachtung, also den Contuitus, aufzubauen; die Bestrebungen von Big Data entsprechen schliesslich dem Wunsch, die Handlungs- und Gedankenwege anderer vorherzusagen.

Nicht umsonst begann die Wissenschaft der Daten mit Euklids Geometrie, in denen zum ersten Mal von «Dedomena» (dem Gegebenen)<sup>7</sup> die Rede ist, die in der ersten Übersetzung ins Lateinische eben mit dem Wort Data wiedergegeben wurden. Euklid sah als Dedomena jenes Gegebene an, aus dem sich gesuchte, d.h. nicht bekannt geometrische Zusammenhänge erschliessen lassen. Nichts anderes ist – wenn wir den geometrischen Zusammenhang beiseitelassen – bei Big Data und Linked Open Data der Fall. Hier ist der Zusammenhang so einfach wie offensichtlich.

Wir sollten dabei auch nicht vergessen, dass seit Anbeginn der Wissenschaftsgeschichte praktisch alle namhaften Forscher darum bemüht waren, die Zukunft vorherzusagen, mit welchen Mitteln auch immer. Auch wenn wir es heute nicht mehr wahrhaben wollen, schrieb Newton, der gerne als Vater der modernen Wissenschaften angesehen wird, und darauf verwies «auf den Schultern von Giganten» zu stehen, mehrere Bücher zur Prophetie und der Alchimie<sup>8</sup>.

6 Augustinus, *Confessiones*, XI, 20, 26, <http://www9.georgetown.edu/faculty/jod/latinconf/11.html>

7 Jakob Voss (2013): «Was sind eigentlich Daten?» *Libreas. Library Ideas*, 23, <http://libreas.eu/ausgabe23/o2voss>

8 John Freely (2009): *Aladdin's Lamp*, Alfred E. Knopf, New York.



Diese beiden alles überragenden Interessen (Prophetie und Alchimie) führten letztlich zum Entstehen der ungleichen Schwestern Astronomie und Astrologie sowie der Ablösung der Alchimie durch die Chemie. Wissenschaftstheoretisch würde sich die ganze Diskussion um Daten also als nichts Weiteres herausstellen als die Suche nach dem Beweis, dass Prophetie nichts anderes als ein ganz rationales und berechenbares Unterfangen ist.

Kontakt: [rene.schneider@hesge.ch](mailto:rene.schneider@hesge.ch)

## ABSTRACT

*Les grandes quantités de données ouvertes, scientifiques et reliées entre elles: dimensions et contextes*

Mais que peuvent donc signifier ces énormes quantités de données ouvertes et reliées entre elles? Les autres paradigmes de l'histoire de la science sont-ils donc devenus obsolètes? OU bien les données ne sont-elles plus des données? Et qu'est-ce que cela signifie pour les bibliothèques, les archives et les centres de documentation? Sont-ils concernés ou est-ce leurs données qui le sont? En résulte-t-il de nouvelles tâches pour ces institutions ou leur importance s'en trouvera-t-il amoindrie?

Dans cet article, l'auteur tente de cerner ces questions et d'y répondre. Il le fait en discutant de trois types de données qui, malgré leurs différences, ont quelques points communs qui les relient d'une certaine manière, à savoir: big data, linked open data et research data (c.à.d. de grandes quantités de données ouvertes et reliées entre elles ainsi que des données de recherche). Ce triptyque d'initiatives de traitement des données répond d'ores et déjà à l'une des questions formulées ci-dessus et montre que les données ne sont plus tout à fait des données. Les sciences de l'information et toutes les institutions qui y sont liées se trouvent donc face à un défi: maîtriser ces données et les tâches qu'elles induisent.

(traduction: sg)

# Linked Open Data: quelles nouvelles compétences pour les professionnels de l'information?

Jasmin Hügi et Nicolas Prongué, spécialistes en information documentaire

**Les LOD ne représentent pas la première thématique qui remet en question les compétences voire la formation des professionnels de l'information. Ces dernières années, la littérature scientifique a proposé des réflexions par rapport à des programmes d'études en lien avec les bibliothèques numériques<sup>1</sup>, la «Digital Curation» (gestion des actifs numériques)<sup>2</sup> ainsi qu'avec la gestion de données de recherche<sup>3</sup>.**

## Des études concernant les nouvelles compétences des bibliothécaires

Les LOD ne représentent pas la première thématique qui remet en question les compétences voire la formation des professionnels de l'information. Ces dernières années, la littérature scientifique a proposé des réflexions par rapport à des programmes d'études en lien avec les bibliothèques numériques, la «Digital Curation» (gestion des actifs numériques) ainsi qu'avec la gestion de données de recherche.

Concernant les LOD en bibliothèque, encore très peu d'études ont été réalisées pour identifier les compétences qui devraient être enseignées aux futurs professionnels de l'information. Deux projets méritent d'être mentionnés dans ce contexte-là. D'une part, un curriculum concernant les LOD en général a été développé pour le projet EUCLID<sup>4</sup>, sans considérer les institutions culturelles. Sur la base de ce curriculum, des matériels d'apprentissage sous forme de webinars, de tutoriels et d'un cahier d'accompagnement ont été élaborés et mis à disposition en ligne. Afin de définir les connaissances nécessaires pour travailler quotidiennement avec les LOD, des experts de différentes industries ont été interrogés. D'autre part, le projet Learning Linked Data<sup>5</sup> s'est focalisé spécifiquement sur les besoins des professionnels de l'information. Ce projet, qui s'est terminé en septembre 2012, avait pour but de faire un répertoire des sujets d'apprentissage afin de développer une plate-forme soutenant les enseignants et les étudiants dans l'interprétation et la création des données.

## Offres d'emploi et avis d'experts

Le Web des données et les LOD représentent un sujet très technique, avec lequel les frontières entre le métier d'informaticien et celui de professionnel de l'information s'estompent. L'étude qui est présentée dans cet article vise à identifier les compétences prioritaires à acquérir pour qu'un professionnel de l'information puisse travailler

- 1 POMERANTZ, Jeffrey, et al. 2006. Curriculum development for digital libraries. Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries. New York, NY, USA: ACM. DOI: 10.1145/1141753.1141787.
- 2 YAKEL, Elizabeth, et al. 2011. Digital Curation for Digital Natives. Journal of Education for Library & Information Science. Vol. 52, no. 1, p. 23–31.
- 3 STANTON, Jeffrey M., et al. 2011. Education for eScience Professionals: Job Analysis, Curriculum Guidance, and Program Considerations. Journal of Education for Library and Information Science. Vol. 52, no. 2, p. 79.
- 4 [www.euclid-project.eu/](http://www.euclid-project.eu/)
- 5 <http://lld.ischool.uw.edu/wp/>



avec des LOD en bibliothèque. Pour ceci, des offres d'emploi trouvées en Suisse, en Allemagne, en Autriche et en France ainsi que des avis d'experts internationaux ont été analysés afin de dresser un premier constat. Seules les offres d'emploi exigeant un diplôme en science de l'information, en plus des compétences liées aux LOD, ont été retenues. Quant aux experts, ils devaient, pour être sélectionnés, avoir étudié les sciences de l'information ou avoir travaillé en tant que bibliothécaire, en plus de leur expérience en LOD<sup>6</sup>. Au final, huit offres d'emploi et douze réponses d'experts ont été prises en compte.

### Huit compétences-clés

La nature technique des LOD avait une certaine influence sur les réponses. Ainsi, ces dernières mentionnent beaucoup de détails techniques, tels que des langages, des formats ou encore des logiciels. L'analyse du contenu des offres d'emploi et des réponses a permis d'identifier plus de 40 éléments de compétence différents. La figure 1 présente les compétences qui apparaissent dans le nombre de cas le plus élevé<sup>7</sup>.

#### Ontologies et métadonnées

Dans cette première catégorie, il s'agit de connaître des schémas de métadonnées tels que Dublin Core, savoir évaluer et choisir un élément de métadonnées pour l'appliquer à ses propres données, savoir créer des vocabulaires contrôlés ainsi que des ontologies, et savoir effectuer un alignement de deux vocabulaires différents.

#### Développement web

Cet élément englobe des connaissances concernant le fonctionnement des protocoles web et des serveurs, ainsi que les principes du Web design. La maîtrise de

ces connaissances est nécessaire afin de savoir créer et attribuer des URIs aux triplets, les publier en ligne ainsi que créer des interfaces web pour que les utilisateurs puissent interagir avec.

#### Triplets

En troisième position arrivent les compétences en lien avec les triplets, c'est-à-dire la connaissance des spécifications RDF émises par le W3C, qui représentent la base pour la création de triplets. Il est également important de savoir développer et modéliser un graphe qui peut être appliqué aux données. Puisque les triplets peuvent être exprimés à travers différentes syntaxes (RDF/XML, Notation 3, etc.), il est utile de comprendre ces dernières ou de connaître des outils qui peuvent les produire.

#### Web sémantique et Linked Data

Il s'agit de connaître les principes émis par le W3C et par Tim Berners-Lee, ainsi que leurs objectifs. Cette catégorie inclut également la compréhension basique de l'inférence ou du raisonnement, par exemple.

#### Standards de catalogage

Les compétences concernant les standards de catalogage sont essentielles dans un projet LOD, car pour transformer des données en triplets, il faut bien connaître les données de base. Il s'agit donc de comprendre les règles de saisie des données (telles que les AACR2), le format d'enregistrement (tel que MARC), ainsi que le fonctionnement du SIGB utilisé (tel que Virtua).

#### Programmation

Certains experts interrogés ont évoqué la connaissance basique d'au moins un langage de programmation (Java, Ruby, PERL, etc.) comme compétence nécessaire. Cela servirait avant tout de moyen de communication avec les informaticiens. Ces connaissances peuvent également être utiles pour la programmation de codes simples qui permettent la conversion en masse de données en LOD.

#### Technologies du Web sémantique

Les données organisées en triplets nécessitent une base de données spécifique (un «triple store»), un langage d'interrogation spécifique (SPARQL) et une interface d'interrogation spécifique (SPARQL Endpoint). La maîtrise de ces technologies est fortement souhaitée pour travailler avec les LOD.

#### XML

Enfin, dernier élément mais non des moindres, des connaissances de XML et des langages annexes comme XSLT ont été mentionnées dans les cas analysés. XML est à la base de nombreux standards du Web sémantique, et il est donc indispensable de comprendre et de savoir exploiter des données structurées en XML.

#### Un monde changeant

Le travail avec des LOD en bibliothèque nécessite visiblement des compétences très spécifiques et au premier abord, cela peut sembler intimidant. Néanmoins, il est important de noter que les compétences déjà existantes parmi les

- 6 La méthodologie sera expliquée en détail dans un rapport qui paraîtra à la fin de cette année.
- 7 Pour donner un exemple: si une compétence est mentionnée dans chacun des 20 cas, la compétence a une haute priorité. Si une compétence n'apparaît que dans un seul cas, sa priorité est basse (même si la compétence est mentionnée très souvent à l'intérieur du cas).

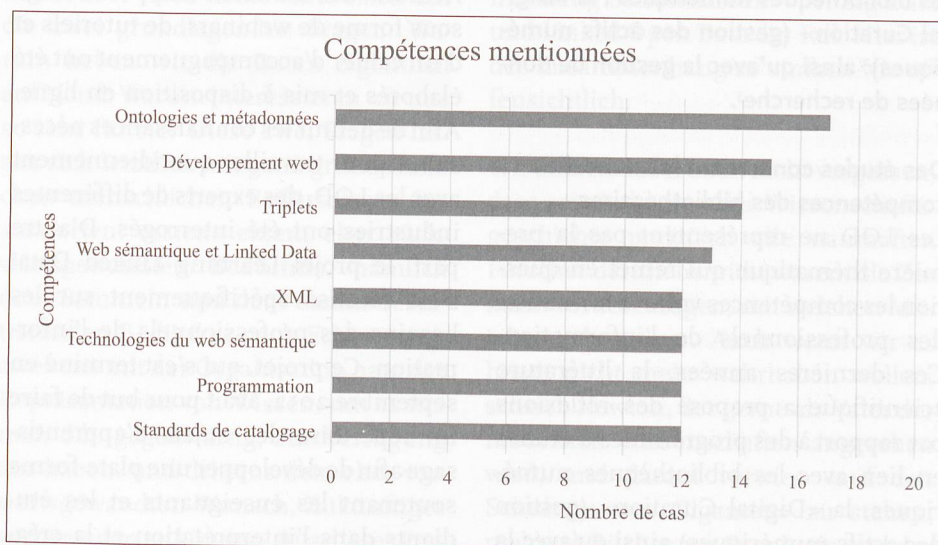


Figure: Compétences les plus mentionnées dans les 20 cas



professionnels de l'information se trouvent tout à fait dans les éléments cités dans le diagramme. Ainsi, la première compétence touche aux ontologies et aux métadonnées, qui sont utilisées pour la description des ressources, et ceci a toujours fait partie des connaissances de base des professionnels de l'information. Concernant les standards de catalogage, il est inutile d'évoquer que cela appartient à la bibliothéconomie. En outre, les formations de bachelors en information documentaire en Suisse contiennent des modules d'enseignement de développement web et de XML. Par rapport à la programmation, la frontière entre informaticien et professionnel de l'information se montre fluctuante, et l'attribution de cette compétence à l'un des deux métiers représente un sujet de discussion.

Ce qui manque réellement en ce moment, ce sont les compétences strictement

liées au Web sémantique (principes, standards et technologies). Selon nous, l'enseignement de ces compétences devrait être inclus dans le curriculum des professionnels de l'information. Une introduction obligatoire à la thématique ainsi que des cours à choix devraient être dispensés au niveau bachelors, alors que les étudiants au niveau master devraient atteindre une compréhension globale afin d'être capables de prendre des décisions stratégiques.

Des cours de formation continue sont par ailleurs à mettre en place pour que les professionnels sur le terrain puissent eux aussi se former aux LOD. En effet, le Web sémantique est bien plus qu'une tendance, et le monde des bibliothèques ne peut se permettre d'attendre.

Contact: nicolas.prongue@etu.hesge.ch  
jasmin.hugi@hesge.ch

## ABSTRACT

*Linked Open Data – Welche neuen Kenntnisse brauchen I&D-Fachleute auf diesem Gebiet. Die Linked Open Data (LOD) verbreiten sich progressiv in der kulturellen Domäne. Eine wachsende Anzahl an Projekten entstehen in Institutionen und LOD-Applikationen halten Einzug in Bibliotheken. Dieser Einzug stellt gewisse Kompetenzen der I&D Fachleute infrage, stärkt andere und erfordert neue Kenntnisse und neues Know-how. Dieser Artikel präsentiert die Resultate einer Studie, die zum Ziel hat, diese neuen Kompetenzen, die sich I&D-Fachleute angesichts LOD aneignen müssen, zu bestimmen.*

# «Daten öffnen, weltweit kooperieren» Grosse Fragen, noch keine abschliessenden Antworten

Vera Münch,  
freie Journalistin und PR-Beraterin

Die neuen Möglichkeiten der nahezu unbegrenzten Verlinkung von Daten, die Aggregation und Analyse riesiger Datenmengen – diskutiert unter den Überbegriffen Big Data, Linked Data und Linked Open Data (LOD) – bieten Bibliotheken völlig neue Chancen, ihre Bestände und besonderen Angebote international sichtbar zu machen. Die neue Technik des Cloud Computings, also der Bereitstellung und Nutzung von Software, Rechen- und Speicherkapazitäten als Onlinedienstleistung, ermöglicht es zudem, Routinearbeiten im eigenen Haus zu reduzieren und sie stattdessen durch Automatisierung, Vernetzung und Abgleich von z.B. Katalogen und Metadaten auf viele Schultern in der Bibliothekswelt zu verteilen. Auch die Informationsauslieferung auf mobile Computer und Smartphones lässt sich über Cloud Services relativ problemlos organisieren.

Soll man seine Daten wirklich für die ganze Welt öffnen? Wie sichert man in einer völlig offenen Welt die nationalen Interessen, wie die des eigenen Hauses? Wer sorgt für eine vernünftige, nachhaltige Ordnung in der Datenablage, der Datenverknüpfung und der Informationspräsentation? Wie kann man es als Bibliothek schaffen, mit dem atemberaubenden Forschungsschritt mitzuhalten? Viele grosse Fragen, auf die es heute noch keine abschliessenden Antworten gibt.

## Globale Clouds: Problemlösung oder Herstellerabhängigkeit?

Als eine der ersten Bibliotheken in Europa hat die Universität Tilburg Services der cloudbasierten Softwareinfrastruktur WorldShare™ Management Services (WMS) von OCLC in Betrieb genommen: «Wir betreiben, historisch gewachsen, seit Jahren mehrere Bibliotheksinformationssysteme (LIS) paral-



### Portrait

Vera Münch (55), Hildesheim (D) begleitet als freie Journalistin und PR-Beraterin für wissenschaftliche Einrichtungen die Entwicklungen in der Fachinformation und im Bibliothekswesen. Sie berichtet für die deutsche Bibliotheksfachzeitschrift b.i.t.online über Konferenzen, Messen und Trends.



lel. Mit der Cloud ist jetzt die Technologie da, hier etwas zu ändern», so ein Mitarbeiter.

Durch die Nutzung einer Cloud-Plattform bindet man sich technisch eng an einen Anbieter. Entsprechende Befürchtungen (in Fachkreisen unter dem Begriff «Vendor Lock-In») diskutiert, begleiten die neue Technologie. Anja Smit, Direktorin der niederländischen Universitätsbibliothek Utrecht und amtierende Vorsitzende des OCLC EMEA Regional Council hielt den Bedenken in Strassburg das Verbundargument entgegen: «Alle Macht geht von den Mitgliedern aus».

### Die Schweizerische Nationalbibliothek hat eine Strategie

Die Direktorin der Schweizerischen Nationalbibliothek (NB), Marie-Christine Doffey<sup>1</sup>, hat mit ihrem Team für ihr Haus eine klare Entscheidung getroffen. In ihrem Konferenzvortrag<sup>2</sup> brachte sie diese auf den Punkt: «Daten öffnen, weltweit kooperieren». Das bedeutet, die NB kooperiert nicht nur mit den Schweizer Bibliotheken wie zum Beispiel im Rahmen des Metakatalogs Swissbib und «den anderen etwa 900 Bibliotheksdiensten, die im Metakatalog sind», sondern auch mit der Europeana, mit dem WorldCat «und allen anderen Möglichkeiten», so Doffey. Sie begründete die Entscheidung vor allem mit der Überzeugung, dass alle Metadaten und Informationsbestände der Schweizerischen Nationalbibliothek

mit öffentlichen Geldern aufgebaut sind und deshalb auch der Öffentlichkeit zur Verfügung stehen müssten; sei es vor Ort in der Bibliothek oder virtuell. «Durch die Öffnung können andere Bibliotheken und Interessenten die Metadaten für eigene Zwecke und Innovationen nachnutzen.»

Die NB hat für die Europeana den Schweizer Bibliothekskatalog Helveticat vollumfänglich geöffnet. Alle Metadaten sind unter der Corporate-Commons-Lizenz CC-O für alle Interessenten verfügbar und können angebunden werden. Doffey erklärte: «Wir müssen dort sein, wo andere Partner, andere Bibliotheken auch dabei sind, damit wir an der Diskussion teilnehmen können. Die Schweiz ist so klein, dass wir diese Öffnung auch weltweit nutzen wollen und können.» Genevieve Clavel-Merrin, bei der NB verantwortlich für den Bereich «Nationale und Internationale Kooperation» ergänzte, durch den Big-Data-Austausch auf Basis vernetzter Metadaten würde ein Bestandsabgleich mit grossen anderen Metadatenansammlungen möglich. Zur Veranstaltung sagte Direktorin Doffey, dass auch «die ganze Diskussion über die Cloud» sehr wertvoll sei. «Als einzelne Institution haben wir zwar nicht die gleichen Probleme wie Konsortien, aber ich denke, es ist auch einmal ganz interessant, Leute zu hören wie Raymond Bérard, zu erfahren, was es da für Probleme gibt und welche Fragen gestellt werden.» Bérard, Direktor des französischen staatlichen bibliografischen Instituts ABES berichtete u.a., dass ABES für die französischen akademischen Bibliotheken eine Studie zur Ablösung der bisherigen LIS durch ein neues, geteiltes «shared LIS» durchgeführt hat. Auf Basis der gewonnenen Erkenntnisse hat ABES sein Bibliothekssystem neu ausgeschrieben.

### Frankreich setzt stark auf Wikipedia für Bibliotheken

Wie Wikipedia in Bibliotheken eingebunden und von ihnen genutzt werden kann, ist in Bezug auf Big Data und Linked Data ebenfalls ein grosses Thema. Die Wissenschaftlerin Titia van der Werf, die vom niederländischen Leiden aus für OCLC die F&E-Aktivitäten mit europäischen Bibliotheken koordiniert,

griff die Diskussion mit Rémi Mathis als geladenem Gast und Max Klein von OCLC Research<sup>3</sup> in einem Workshop auf. Mathis ist Bibliothekskurator und Präsident der Wikimedia France. Er berichtete, dass das französische Kultur- und Kommunikationsministerium mit Wikimedia France kooperiert, um die französischsprachigen Ressourcen in Wikipedia semantisch zu erschliessen. Klein beschäftigt sich als «Wikipedian in Residence» bei OCLC Research mit allen Fragen rund um die Weiterentwicklung von Wikipedia als Quelle für verlinkbare Daten und Informationen. Er versuchte, in 45 Minuten die Idee der freien, von Maschinen und Menschen gleichermaßen lesbaren Datenbank Wikidata zu erklären – und wie die Forscher dieser Datenbank Multilingualität beibringen wollen.

### Culturomics: Kulturgeschichte aus Digitalisaten herausrechnen

Auf diesem hohen Niveau internationaler Forschung ging es nahtlos weiter. Jean-Baptiste Michel erforscht in Harvard «die Nutzung von Millionen digitalisierter Bücher und Dokumenten, um durch quantitative Analysen daraus z.B. Rückschlüsse zur Geschichte und Kultur zu ziehen». «Culturomics»<sup>4</sup> haben er und sein Kollege Erez Lieberman Aiden dieses neue, auf Digitalisaten aufsetzende Forschungsgebiet getauft. Ihre Arbeiten führten zu einer Software, die aus grossen Datenbeständen in den Daten verborgene Trends berechnet und grafisch als Kurven in Diagrammen sichtbar macht. Die Forschung brachte ihnen im letzten Jahr Titelseiten auf «Nature and Science» und auf der «New York Times». Google hat das Programm gekauft.

### In Deutschland entsteht ein «Culturegraph»

Die Deutsche Nationalbibliothek (DNB) baut an einer Plattform, über die Dienste und Projekte zu Datenvernetzung, Persistent Identifier und Linked Open Data für kulturelle Entitäten zusammengeführt werden sollen. Markus Geipel, Softwareingenieur bei der DNB, stellte die «Culturegraph» getaufte Plattform vor. Sie soll zu einem globalen Kulturschatz-Netzwerk weit über die Grenzen einzelner Unternehmen und Einrichtungen hinaus werden.

1 [www.nb.admin.ch/org/amtsvorsteherin/index.html?lang=de](http://www.nb.admin.ch/org/amtsvorsteherin/index.html?lang=de)

2 Aufzeichnungen der Konferenzbeiträge sind, soweit sie von den Referenten freigegeben wurde, auf der Website von OCLC bereitgestellt. [www.oclc.org/en-europe/events/2013/emearc-annual-meeting-feb-26.html](http://www.oclc.org/en-europe/events/2013/emearc-annual-meeting-feb-26.html). Von einigen Präsentationen gibt es zudem Videoaufzeichnungen auf YouTube, u.a. vom Vortrag von M.-C. Doffey <https://www.youtube.com/watch?v=aONFHqMbdiq> <http://www.oclc.org/en-europe/events/2013/emearc-annual-meeting-feb-26.html>

3 OCLC-Research betreibt eine eigene Website, auf der sehr interessante Forschungsprojekte und -ergebnisse stehen [www.oclc.org/research.html](http://www.oclc.org/research.html)

4 [www.culturomics.org/](http://www.culturomics.org/)



## Neue Informationsdienste mit offenen Daten

Wie auf Basis offener Daten mit neuen IKT entwickelte Dienste aussehen können, zeigten nach Jean-Baptiste Michel zwei weitere Referenten auf sehr beeindruckende Weise: Silver Oliver, Informationsarchitekt mit Spezialgebiet strategische Ausrichtung der Datennutzung für Onlinepublishing bei der britischen Unternehmensberatung Ontoba<sup>5</sup>, und Dr. Klaus Ceynowa, stellvertretender Direktor der Bayerischen Staatsbibliothek (BSB). Oliver referierte über das «Zeitalter des Hinweisens auf Dinge, die in Beziehung zum persönlichen Interesse stehen» («The age of pointing at things»). Gemeint ist das automatische Anbieten von weiterführenden Informationen direkt in den Informationsraum hinein, in dem sich der Nutzer gerade bewegt.

Dr. Klaus Ceynowa zeigte innovative Nutzungsszenarien für digitalisierte Inhalte aus Bibliotheksbeständen, die von der BSB mithilfe von Technologien wie GPS-gestützte Lokalisierung (Location based Services), durch virtuelle Darstellung angereicherte Informationsbereitstellung (Augmented Reality) sowie gestengesteuerte Präsentation und Informationssuche (Gesture based Computing and Search) entwickelt wurden.

## OCLC Research informiert umfassend über Datennutzung

Nicht nur die eingeladenen Redner, auch OCLC Research hatte Spannendes aus den Forschungslaboren zu berichten. Roy Tennant, Senior Programm Officer, führte in die «Entfesselung des Katalogisierens»<sup>6</sup> durch Datamining in LOD ein. Die über 290 Millionen WorldCat-Einträge, die auf rund 1,95

Milliarden Holdings hinweisen, liefern den Rohstoff für schier unendliche Nutzungsmöglichkeiten. Van der Werf zeigte in ihrem Vortragsthema «Metadaten ausser Kontrolle: Metadatenaggregation auf Netzwerkebene», welche Aufgabe gute Metadaten erfüllen und wo die Grenzen der Technik liegen.

## W3C Group für LOD-Erweiterung von Schema.org eingerichtet

Last, but not least sprach OCLC Technology Evangelist Richard Wallis im letzten Vortrag über «Linked Data und OCLC»<sup>7</sup>. Dabei berichtete er unter anderem, dass das World Wide Web Consortium (W3C) an der Weiterentwicklung des Metadatenbeschreibungsstandards Schema.org in Richtung LOD-Nutzung arbeitet. Dazu wurde eine «Schema Bib Extend Community Group»<sup>8</sup> eingerichtet. Wallis leitet sie als Vorsitzender.

Kontakt: vera-muench@kabelmail.de

## ABSTRACT

### *Ouvrir les données et coopérer au niveau mondial*

Les nouvelles possibilités que présentent la mise en réseau pratiquement illimitée des données, l'agrégation et l'analyse d'énormes quantités de données (Big Data, Linked Data et Linked Open Data) offrent aux bibliothèques de nouvelles chances de valoriser leurs fonds à l'échelle internationale et de proposer des offres spécifiques. La nouvelle technique du Cloud Computing, autrement dit la mise à disposition et l'utilisation de logiciels, de capacités de calcul et de stockage comme prestation en ligne, permet en outre de réduire les travaux de routine et de répartir ces derniers via l'automatisation et la mise en réseau de catalogues et de métadonnées. La fourniture d'informations sur des ordinateurs portables et des téléphones intelligents peut également être organisée sans trop de problèmes via les Cloud-Services. *(traduction: sg)*

OCLC hat sein 4. EMEA Regional Council Meeting 2013 unter das Thema «Dynamic Data: a World of possibilities» gestellt.

5 [www.ontoba.com/blog/bbc-news-labs](http://www.ontoba.com/blog/bbc-news-labs)

6 [www.oclc.org/research/news/2013/02-27.html](http://www.oclc.org/research/news/2013/02-27.html)

7 Ein Interview mit Richard Wallis zu LOD erschien in B.I.T.-Online 15 (2012), Ausgabe 6, S. 566

8 [www.w3.org/community/schemabibex/](http://www.w3.org/community/schemabibex/)

[ a [ r [ b | i | d ] o ]

ANNONCES/INSERATE

– arbido newsletter

– arbido print

– arbido website

[inserate@staempfli.com](mailto:inserate@staempfli.com)

031 300 63 89



# Les institutions patrimoniales suisses sont favorables à l'open data

Beat Estermann, collaborateur scientifique, Institut pour le gouvernement électronique de la Haute école spécialisée bernoise

Depuis l'avènement de l'Internet, les institutions patrimoniales connaissent des changements successifs nombreux. On par exemple la numérisation systématique des objets patrimoniaux, la mise en réseau accrue des organisations, la coproduction de contenus par les utilisatrices et les utilisateurs ainsi que la mise à disposition de données et de contenus sur l'Internet sous des licences «libres». Une étude de la Haute école spécialisée bernoise<sup>1</sup> montre où les institutions patrimoniales suisses se situent dans ce contexte.

## Les institutions patrimoniales en pleine mutation

Le Musée de Derby mise sur des codes QR pour enrichir via des articles de Wikipédia les explications concernant les objets d'exposition, et ce en plusieurs langues bien entendu. La Bibliothèque nationale australienne propose

à des bénévoles des tâches de correction et de transcription sur l'Internet, offre qui rencontre un véritable succès. Le Tropenmuseum à Amsterdam met des centaines de milliers de photos haute résolution sur Wikimedia Commons, se félicite que ses fonds soient désormais plus faciles d'accès et en espère en tirer profit par l'augmentation du nombre de visiteurs. Le Brooklyn Museum s'en remet au jugement des visiteurs de son site web pour sélectionner des photographies destinées à une exposition. Un consortium de bibliothèques européennes utilise le crowdfunding pour financer la numérisation des œuvres tombant dans le domaine public. A Vienne, des chercheurs misent sur les données libres d'accès pour visualiser d'une nouvelle façon d'appréhender l'histoire de l'art.

Ces exemples montrent les profonds changements qui traversent le champ des institutions patrimoniales depuis l'avènement de l'Internet dans les années 1990 et les innovations qu'il a induites.

On peut, concrètement, identifier plusieurs tendances successives:

Depuis le début des années 2000, les institutions patrimoniales en Europe conjuguent leurs efforts afin de numériser de manière coordonnée et la plus complète possible les objets patrimoniaux et leurs métadonnées. Outre une mise en réseau renforcée qui permet aux institutions d'échanger leur savoir-faire et d'éviter les doublons, sont également proposées dans la foulée des offres de «single-point-of-access» sous la forme de systèmes de catalogues inter-organisations et de bibliothèques virtuelles.

Dès 2005, on assiste à l'avènement du Web social qui offre encore plus de possibilités d'interaction. Cette tendance se caractérise par de nouvelles formes de personnalisation et par une implication plus forte des utilisatrices et des utilisateurs, ou des visiteuses et des visiteurs respectivement, qui sont appelés à coproduire.

Le développement spectaculaire de Wikipédia, Flickr et d'offres analogues basées sur le principe de la communauté virtuelle focalise de plus en plus l'attention sur la constitution et la collecte collaborative de contenus sur l'Internet. De premières collaborations s'installent entre des institutions patrimoniales et des communautés en ligne, tandis que certaines institutions lancent des projets spécifiques en faisant appel au crowdsourcing<sup>2</sup>.

Les communautés Wikipédia et Flickr découvrent de leur côté que les institutions patrimoniales sont des partenaires importants et cherchent depuis quelques années à collaborer activement avec elles. A partir de 2009, c'est le mouvement pour l'ouverture des données publiques qui, depuis les

**Le développement spectaculaire de Wikipédia, Flickr et d'offres analogues basées sur le principe de la communauté virtuelle focalise de plus en plus l'attention sur la constitution et la collecte collaborative de contenus sur l'Internet.**

Etats-Unis et la Grande-Bretagne, se propage rapidement pour toucher quelque 50 Etats de par le monde. Son exigence: les données des autorités publiques doivent être rendues accessibles sous une forme lisible par ordinateur afin qu'elles puissent être utilisées par des tiers. Si le mouvement pour l'ouverture des données publiques trouve son origine dans une vision de la politique étatique que l'on veut plus transparente et plus participative, il est également motivé par une vision technico-économique: la mise en réseau du plus grand nombre possible de données «ouvertes» doit permettre l'émergence de ce qu'il est convenu d'appeler le «Web sémantique». Les données des autorités publiques ainsi que, le cas échéant, celles d'entreprises privées, deviennent ainsi des ressources infrastructurelles libres d'accès que des tiers

1 En allemand: Estermann Beat (2013): Schweizer Gedächtnisinstitutionen im Internet-Zeitalter. Ergebnisse einer Pilotbefragung zu den Themenbereichen Open Data und Crowdsourcing, Berner Fachhochschule, E-Government-Institut. / En anglais: Estermann Beat (2013): Swiss Heritage Institutions in the Internet Era. Results of a pilot survey on open data and crowdsourcing, Bern University of Applied Sciences, E-Government Institute.

2 Le concept de «crowdsourcing» désigne des activités participatives en ligne dans le cadre desquelles une personne ou une organisation demande publiquement à un groupe (en général non défini de manière précise) de personnes d'horizons différents d'accomplir une tâche à titre bénévole. On parle également de «coproduction» ou, en particulier dans le contexte de Wikipédia, de création collaborative de contenus.



peuvent utiliser pour offrir des services à valeur ajoutée: pour la valorisation des données, la création et la visualisation de combinaisons de données d'un nouveau genre ou pour d'autres prestations basées sur des données. L'utilisation de licences standardisées «libres», qui permettent également de modifier et de diffuser plus loin les œuvres à des fins commerciales, est primordiale dans ce contexte, étant donné qu'elles réduisent sensiblement les coûts de transaction pour l'utilisation d'œuvres par des tiers<sup>3</sup>.

### Où se situent aujourd'hui les institutions patrimoniales suisses par rapport à l'open data/open content et au crowdsourcing?

Afin de montrer où se situent actuellement les institutions patrimoniales suisses par rapport à l'open data / open content et au crowdsourcing, comment elles évaluent les chances et les risques de ces tendances et où elles voient leurs bénéfices potentiels, la Haute école spécialisée bernoise a réalisé une enquête-pilote à la fin 2012. Quelque 200 institutions de Suisse alémanique d'importance nationale ont été sollicitées dans ce contexte<sup>4</sup>, dont 72 ont répondu au questionnaire.

Environ 60% des institutions interrogées mettent, à tout le moins en partie, à disposition sur l'Internet des objets patrimoniaux et des métadonnées. Quelque 40% échangent des métadonnées dans le cadre d'alliances. Ces deux pratiques, qui représentent des conditions importantes du point de vue de l'open data / open content et du crowdsourcing, sont donc déjà très répandues. Les institutions qui ne sont pas encore parvenues à numériser leurs fonds les plus importants, pour des raisons de manque de moyens ou de savoir-faire, peuvent toutefois causer quelque souci.

On constate deux dynamiques différentes en ce qui concerne l'open data / open content et le crowdsourcing. Pour ce dernier, on constate un pourcentage certes relativement élevé (10%) des institutions interrogées qui disposent déjà de collaborateurs qui apportent leur contribution à Wikipédia dans le cadre de leur travail, ainsi qu'un pourcentage

tout aussi élevé de participants à l'enquête qui estiment que le travail bénévole sur l'Internet joue un rôle plutôt important pour leur organisation. Cela étant, la plupart des institutions interrogées estiment que le crowdsourcing recèle davantage de risques que de chances. Il n'en va pas de même pour ce qui est de l'open data/open content: à la fin 2012, il n'y avait certes que très peu d'institutions qui étaient disposées à mettre des contenus disponibles en ligne libres de droit, mais les institutions interrogées estimaient néanmoins que les chances qu'offre une telle approche étaient bien réelles. Le feu, pour l'open data, est donc pour ainsi dire au vert! Comme l'on s'y attendait, les linked data constituaient également déjà un thème pour quelques-unes des institutions interrogées, mais aucune d'elles n'avait toutefois de projet dans ce contexte à la fin 2012.

### Open data/open content: besoin d'information concernant les licences «libres»

Les résultats de l'enquête permettent de conclure qu'à ce jour moins de 10% des institutions interrogées mettent «librement» à disposition sur l'Internet des illustrations de leurs objets patrimoniaux, mise à disposition qui permet, autrement dit, à des tiers de les modifier et de les utiliser gratuitement. Plus de la moitié mettent certes des illustrations à disposition sur le Net, mais en limitent toutefois leur utilisation. Le plus souvent, les institutions sont prêtes à mettre gratuitement à disposition des contenus pour la formation et la recherche, pour une utilisation privée ou des projets d'intérêt public. Elles sont nettement plus réticentes lorsqu'il s'agit d'autoriser la mise en ligne de contenus qui peuvent être utilisés à des fins commerciales ou faire l'objet de modifications.

Par rapport aux licences «libres», il existe en outre un besoin d'information: près de la moitié des institutions qui rendent des objets patrimoniaux accessibles en ligne ne semblent pas être conscientes du fait qu'il n'est pas possible de libérer des œuvres pour leur utilisation dans Wikipédia et en même temps d'interdire à des tiers de modifier des images ou de les utiliser à des fins commerciales.

Plus de la moitié des institutions interrogées considèrent l'open data comme important. Les groupes cibles les plus importants dans ce contexte sont les institutions de recherche et de formation, les personnes privées et d'autres institutions culturelles. Les institutions interrogées estiment que les plus grandes chances de l'open data/open content résident dans la meilleure visibilité et accessibilité des fonds, dans la meilleure visibilité des institutions elles-mêmes ainsi que dans leur meilleure mise en réseau. Concernant les risques de l'open data, les institutions interrogées craignent surtout des dépenses supplémentaires en ressources et une certaine perte de contrôle sur les données, en particulier pour ce qui est des violations des droits d'auteur, de la protection des données et des infractions contre les règles de confidentialité.

### Crowdsourcing: important mais porteur de risques

Près de 40% des institutions interrogées considèrent le crowdsourcing comme étant important. La mise en œuvre du crowdsourcing, c'est-à-dire la possibilité de créer ou de compléter des contenus sur l'Internet par des bénévoles qui collaborent entre eux, est considérée par ces mêmes institutions comme envisageable surtout dans les domaines de la classification et de la complétion des métadonnées ainsi que pour des tâches de correction et de transcription. Près de la moitié des institutions constatent par ailleurs des besoins d'amélioration dans le domaine des métadonnées, en particulier pour ce qui est de l'intégralité, de la disponibilité et de la numérisation de ces dernières.

En revanche, les institutions interrogées estiment que le crowdsourcing présente lui aussi une série de risques. Parmi ces derniers figurent la charge

3 P. ex. les licences de droits d'auteur de Creative Commons, telles que CC-by ou CC-by-sa (pour plus d'informations à ce sujet, voir: [www.creativecommons.org](http://www.creativecommons.org))

4 L'inventaire suisse des biens culturels d'importance nationale a servi de base pour la sélection des institutions.



importante qui y est liée au niveau de la préparation et du suivi. Autres risques: la continuité incertaine de l'entretien des données, de la difficulté d'influer sur les résultats et d'estimer correctement la charge temporelle ainsi que la faible sécurité en matière de planification.

### Conclusions

Les résultats de l'enquête montrent que miser sur des approches de type crowdsourcing implique une certaine charge et des impondérables. Cela s'explique également par le fait que les institutions doivent participer activement aux communautés de bénévoles et de partenaires impliquées, ce qui ne manque souvent pas de poser des problèmes compte tenu des différentes cultures organisationnelles. Les processus d'intégration qui y sont liés peuvent certes, en fin de compte, déboucher sur un bénéfice pour les deux parties, mais cela ne va en général pas sans un investissement temporel considérable.

En contrepartie, opter pour une stratégie open data/open content est nettement plus simple pour les institutions. Il est donc conseillé de miser, à court terme et à grande échelle, surtout sur des approches qui permettent aux institutions la « libre » mise à disposition des données et des contenus, sans qu'elles aient à assumer elles-mêmes les dépenses qu'impliquent de longs processus de coopération. Néanmoins, un certain niveau de standardisation des métadonnées, de la qualité des données et des interfaces techniques pourrait malgré tout être utile pour réduire au minimum les coûts de transaction. Il s'agit également de mettre l'offre en rapport avec la demande, car si tel n'est pas le cas, on court le risque que les données et les contenus mis à disposition restent inutilisés.

Compte tenu du fait que de nombreuses institutions considèrent la thématique

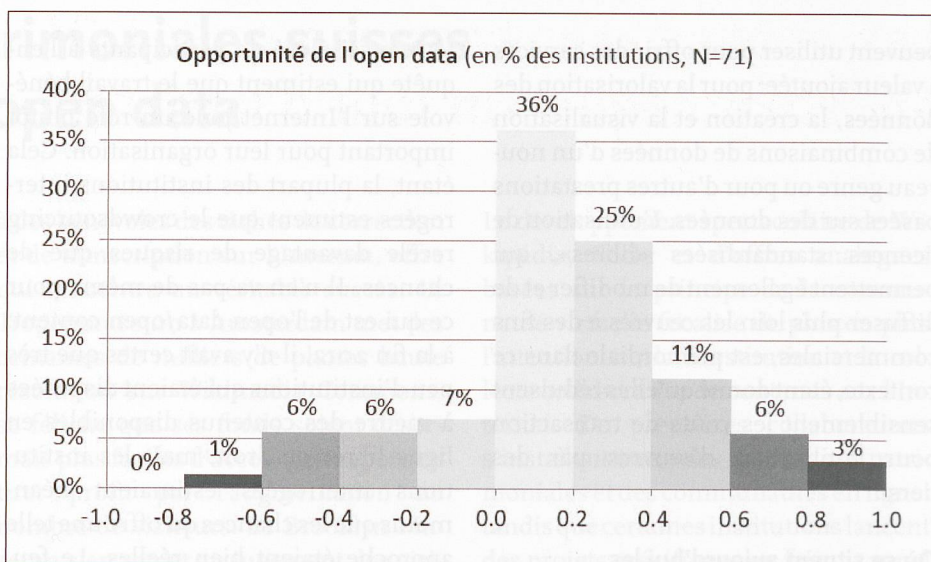


Illustration 1: Opportunité de l'open data (chances vs risques)

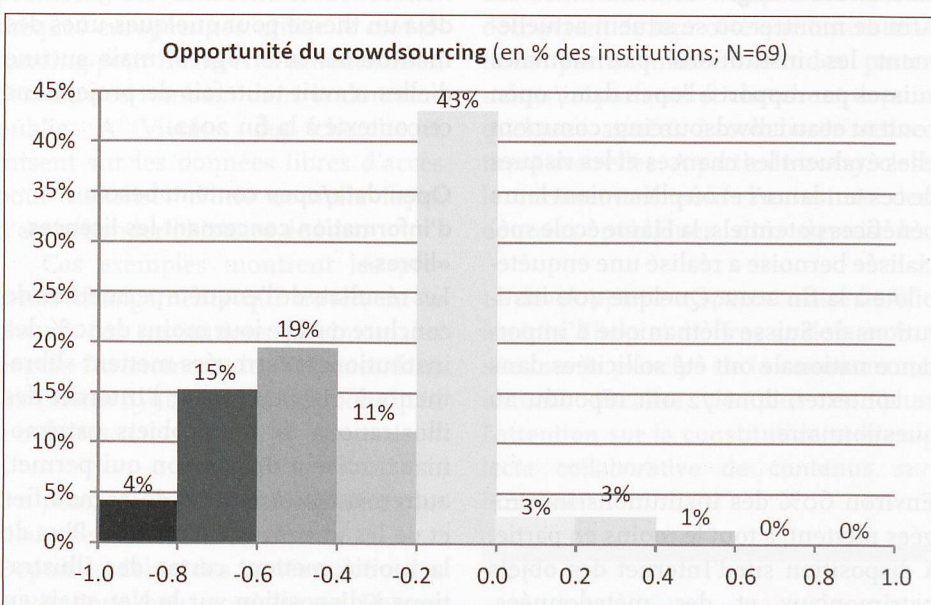


Illustration 2: Opportunité du crowdsourcing (chances vs risques)

du crowdsourcing et de la création collaborative de contenus comme étant aussi importante que l'open data/open content, il serait judicieux de mener quelques projets-pilotes dans ce domaine et d'entamer un processus d'apprentissage collectif parmi les institutions patrimoniales et les communautés en ligne concernées. Les conditions

pour ce faire sont une culture positive de l'erreur, la documentation et l'évaluation systématique des expériences faites ainsi que la création de plateformes d'échange appropriées, tels que conférences, ateliers ou offres de formation continue.

Contact: [beat.ester mann@bfh.ch](mailto:beat.ester mann@bfh.ch)



# Les applications basées sur les LOD en bibliothèque: un tour d'horizon

Nicolas Prongué et Jasmin Hügi,  
spécialistes en information documentaire

**Les Linked Open Data (LOD) sont devenus aujourd'hui un thème récurrent en bibliothèque, mais la plus-value qu'ils peuvent apporter est encore très difficile à percevoir. Quels résultats concrets ont déjà été atteints? Quelles sont, en bibliothèque, les applications LOD existantes à ce jour? Qu'en est-il de la situation en Suisse?**

## Tour d'horizon

De nombreux projets de mise à disposition de données bibliographiques en LOD sont en cours de réalisation ou ont abouti. En 2008 déjà, la Library of Congress (États-Unis) publiait le LCSH<sup>1</sup> en RDF. Mais le phénomène touche en particulier les bibliothèques européennes. Par exemple, le réseau suédois Libris<sup>2</sup> fut, la même année, le premier à publier l'intégralité de son catalogue en RDF.

D'autres institutions ont suivi. Plusieurs bibliothèques nationales (Allemagne, France, Grande-Bretagne, Hongrie, etc.) ainsi que des bibliothèques et réseaux universitaires, principalement allemands et français, ont publié certaines de leurs données.

Les données bibliographiques incluent également les fichiers d'autorités, qui sont particulièrement visés par les projets LOD. Ainsi, les fichiers d'autorités auteurs de plus de 30 bibliothèques ont été convertis en RDF et reliés (ils sont accessibles sur le site du VIAF<sup>3</sup>), tout comme les fichiers d'autorités matières RAMEAU en français, LCSH en anglais et GND en allemand<sup>4</sup>. Du fait qu'ils représentent des concepts du monde réel, comme des personnes ou des lieux, ces fichiers ont un potentiel de liaison bien plus grand que les données décrivant des documents: ils peuvent être liés à des données qui ne sont pas d'ordre bibliographique.

Néanmoins, la plupart de ces initiatives, après avoir converti et lié les don-

nées, se contentent de les mettre à disposition. Mais au-delà de la simple exposition des données, la possibilité de les exploiter, au moyen d'applications, est encore rarement considérée par les bibliothèques. Très peu d'applications LOD existent à ce jour. Et pourtant, c'est là que réside le réel potentiel de plus-value pour l'utilisateur.

## Des données aux applications

Une application consiste en un logiciel traitant des données en vue de fournir un service à l'utilisateur. Dans cet article, nous avons donc considéré comme applications LOD en bibliothèque uniquement les services web destinés directement aux utilisateurs et apportant une plus-value basée sur les LOD. Ceci exclut notamment certaines interfaces dont l'utilisation nécessite des connaissances spécifiques, telles que le langage SPARQL par exemple.

Quelques bibliothèques disposent aujourd'hui de services web rendant accessibles les données bibliographiques en LOD. Néanmoins, la plus-value est bien souvent peu perceptible pour l'utilisateur direct, voire inexistante. De tels services sont entre autres proposés par le Sudoc<sup>5</sup> (France), Libris (Suède), le réseau HBZ<sup>6</sup> (Allemagne), la bibliothèque de l'Université de Winchester<sup>7</sup> (Angleterre) ou encore les bibliothèques nationales allemande<sup>8</sup>, hongroise<sup>9</sup> et britannique<sup>10</sup>.

Il existe néanmoins de réelles applications destinées directement à l'utilisateur lambda. En voici trois des plus abouties:

- Le Virtual International Authority File (VIAF). Maintenu par OCLC, ce portail fournit un accès centralisé aux notices d'autorités auteurs de différents réseaux et grandes bibliothèques, reliées entre elles.
- Le Centre Pompidou virtuel<sup>11</sup>. Le nouveau site du Centre Pompidou à Paris, inauguré en octobre 2012, propose une interface moderne donnant accès à toutes les ressources de ses diverses institutions. On y trouve notamment

des notices de documents, des œuvres d'art numérisées, mais également toutes les ressources produites par le Centre lui-même: affiches, vidéos et enregistrements d'événements, dossiers pédagogiques, etc. Cependant, les données ne sont pas disponibles sous une licence ouverte, donc il ne s'agit pas d'une application strictement LOD.

- Data.bnf.fr. Cette application propose un accès unique à trois bases de données de la Bibliothèque nationale de France: le catalogue général, Gallica ainsi qu'Archives et manuscrits. Les données ont été FRBRisées pour être regroupées au sein de pages d'œuvres, d'auteurs et de sujets. De plus, des liens renvoient vers des notices d'autorités externes et des articles Wikipédia. Dans ses pages «ateliers», data.bnf.fr propose également de nouvelles manières d'exploiter les données, par exemple en les présentant sous forme de cartes et de frises chronologiques générées automatiquement.

## La plus-value des applications

L'exploitation des LOD au moyen d'applications est encore très peu courante dans les bibliothèques. Pourtant, par rapport à un catalogue en ligne, cela

- 1 Library of Congress Subject Headings, fichier d'autorités matières en anglais, disponible sur <http://id.loc.gov/authorities/subjects.html>
- 2 <http://libris.kb.se/>
- 3 Virtual International Authority File: <http://viaf.org/>
- 4 ISAAC, Antoine. RAMEAU subject headings as SKOS linked data. Department of Computer Sciences, University of Amsterdam (en ligne). 2012. [Consulté le 14 octobre 2013]. Disponible à l'adresse: [www.cs.vu.nl/STITCH/rameau/](http://www.cs.vu.nl/STITCH/rameau/)
- 5 [www.sudoc.abes.fr/](http://www.sudoc.abes.fr/)
- 6 Hochschulbibliothekszentrum (Rhénanie-du-Nord-Westphalie): [www.lobid.org/](http://www.lobid.org/)
- 7 <http://prism.talis.com/winchester-ac/>
- 8 <https://portal.dnb.de/>
- 9 <http://nektar2.oszk.hu/>
- 10 <http://bnb.data.bl.uk/>
- 11 [www.centrepompidou.fr/](http://www.centrepompidou.fr/)



peut apporter plusieurs avantages à l'utilisateur final, parmi lesquels:

#### La recherche fédérée

Dans data.bnf.fr par exemple, les diverses données, initialement encodées en MARC, Dublin Core et EAD, ont été converties dans un même format du modèle RDF<sup>12</sup>. Grâce à cela, une seule requête permet d'interroger trois réservoirs de manière optimisée. La page d'un auteur fournit ainsi un accès unique à toutes les ressources liées à cet auteur, quels qu'en soient leurs réservoirs de base. Cela permet également, de pair avec la FRBRisation, d'éviter les doublons.

#### L'enrichissement des catalogues

Le principe de base des LOD réside dans la liaison des données. Les applications peuvent donc proposer des liens vers d'autres ressources, internes ou externes, pour favoriser la découverte et la sérendipité lors de la recherche. Le Centre Pompidou propose par exemple d'augmenter l'expérience de l'utilisateur par de nombreux liens internes, notamment vers des vidéos et des dossiers pédagogiques créés par ses institutions.

#### Les recherches multilingues

Les LOD sont prometteurs dans ce domaine. Par exemple, lors d'une requête dans The European Library<sup>13</sup>, le système propose des sujets connexes au terme recherché, en français, allemand et anglais, grâce à l'alignement des fichiers d'autorités matières RAMEAU, LCSH et GND en format RDF.

#### La visibilité des catalogues sur le Web

Un format compatible avec les standards du Web permet aux données d'apparaître dans les résultats des moteurs de recherche.

#### Et en Suisse?

Pour connaître l'état de la situation en Suisse, nous avons contacté les principaux acteurs susceptibles d'être inté-

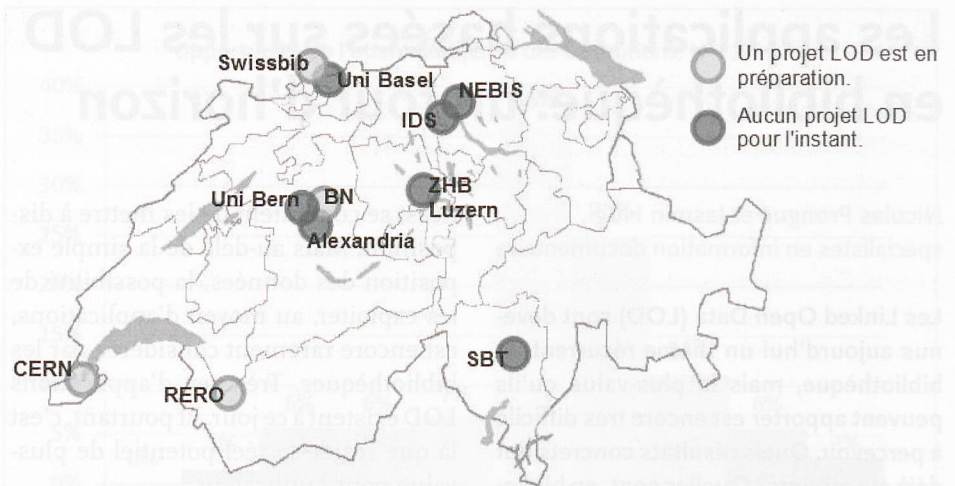


Figure 1: Situation des LOD en Suisse

ressés par le développement des LOD, soit les grandes bibliothèques et les réseaux. Selon les résultats obtenus (figure 1), un certain retard peut être constaté. En effet, seules quatre institutions sur les onze ayant répondu ont un projet LOD en cours de préparation: RERO, la Bibliothèque nationale, Swissbib et la bibliothèque du CERN. Cependant, rien n'a encore abouti à un résultat concret. RERO, par exemple, a déjà préparé le terrain pour le passage aux LOD, notamment en participant au VIAF, en intégrant RAMEAU, en encourageant la création de notices d'autorités auteurs au sein du réseau ainsi qu'en attribuant des URI à ses données.

Ce retard peut être dû à différents facteurs. D'une part, les institutions suisses sont très petites en comparaison à des géants tels que la Bibliothèque nationale de France, et cela représente un frein. D'autre part, les institutions suisses n'ont pas encore toutes déterminé à qui revient la responsabilité du développement des LOD. Est-ce la tâche des réseaux ou de chaque bibliothèque?

La Bibliothèque nationale doit-elle montrer l'exemple? Faut-il organiser un projet au niveau national? Cette problématique est essentiellement d'ordre stratégique.

#### Conclusion

Les LOD reposent sur une technologie très jeune et encore en pleine phase de recherche. Très peu d'applications ont déjà été créées par les bibliothèques directement pour leurs utilisateurs. La plus-value des LOD est encore difficile à percevoir: aucune «killer application», révolutionnant la recherche et l'exploitation des données, n'a pour l'instant été inventée. Les bibliothèques suisses, en retard dans ce domaine, devraient se fixer des objectifs et prendre position par rapport aux LOD. Si elles ne souhaitent ou ne peuvent pas encore investir dans un tel projet, elles peuvent du moins déjà soigner et enrichir leurs fichiers d'autorités, qui sont la clé de voûte des LOD en bibliothèque.

Contact: nicolas.prongue@etu.hesge.ch  
jasmin.hugi@hesge.ch

## ABSTRACT

### Applikationen, die mit LOD arbeiten, in Bibliotheken: eine Übersicht

LOD beruht auf einer sehr jungen Technologie, die sich noch im Forschungsstadium befindet. Es gibt erst sehr wenige Anwendungssoftware, die von Bibliotheken für ihre Nutzer entwickelt wurde. Der Wert der LOD kann noch nicht beziffert werden: Es wurde bis jetzt keine eigentliche «Killer-Applikation» erfunden, die die Suche und die Verwaltung von Daten revolutioniert hätte. Die Schweizer Bibliotheken, die diesbezüglich der Entwicklung etwas hinterher hinken, müssen in Bezug auf LOD Stellung beziehen und sich Ziele setzen. Wenn die Bibliotheken sich noch nicht in solche Projekte einbringen können oder wollen, so können sie zumindest ihre Autoritätsdaten, die den Zugangsschlüssel zu LOD in Bibliotheken darstellen, anreichern und pflegen. (Uebersetzung: sh)

12 Bibliothèque Nationale de France. Web sémantique et modèle de données. data.bnf.fr (en ligne). [Consulté le 14 octobre 2013]. Disponible à l'adresse: <http://data.bnf.fr/semanticweb>

13 [www.theeuropeanlibrary.org/tel4/](http://www.theeuropeanlibrary.org/tel4/)



# Alle Bibliotheken können die Zukunft der Katalogisierung gestalten

Eliane Blumer, Karsten Schuldt\*

**Linked Open Data ist Bestandteil einer grösseren Bewegung hin zur datengetriebenen Bibliotheksarbeit. Alle Bibliotheken haben die Chance, bei der Gestaltung der zukünftigen Katalogisierungsregeln und -formate aktiv mitzuwirken. Dazu ist es notwendig, sich mit den Möglichkeiten und aktuellen Diskussionen in diesem Feld auseinanderzusetzen und sich in diese einzubringen.**

Einige wichtige Stichworte sind dabei FRBR, RDA, Bibframe, Kataloganreicherung und Metadatenmanagement. Die Functional Requirements for Bibliographic Records (FRBR) stellen ein Framework für die bibliografische Beschreibung von Bibliotheksmedien aus Sicht der Nutzerinnen und Nutzer dar, welches als Grundlage für die Formulierung zukünftiger Regelwerke dienen soll. (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). FRBR unterteilt die bibliografischen Daten eines Mediums in vier Entitäten: (Work, Expression, Manifestation und Item). Die Entität Item stellt dabei das konkrete, in einer Bibliothek vorhandene Medium dar. Von dieser Ebene wird der Inhalt aufsteigend abstrahiert. Wichtig sind die Beziehungen zwischen den Ebenen. Resource Description and Access (RDA) stellt das erste auf FRBR aufbauende Regelwerk dar. RDA betont die Verfügbarmachung eines Mediums als Aufgabe des Katalogs. Es wird seit 2013 in der Library of Congress und ab 2014 in der Deutschen Nationalbibliothek genutzt. Der Entschluss der Schweizerischen Nationalbibliothek, RDA einzuführen, ist ebenfalls prinzipiell gefallen (Schweizerische Nationalbibliothek, 2013). Bibframe stellt den Versuch eines Austauschformats dar, welches die Bibliothekskataloge an das Semantic Web anschliesst. Es soll MARC und seine Dialekte ablösen. (Library of Congress, 2012).

Gleichzeitig gibt es in Bibliotheken die Tendenz, Bibliotheksdaten möglichst automatisch mit anderen Daten anzureichern. Die Anreicherung mit

Coverabbildungen ist heute Standard, genannt werden in der internationalen Diskussion weitergehende Anreicherungen, zum Beispiel Reviews, Inter-

**Gleichzeitig gibt es in Bibliotheken die Tendenz, Bibliotheksdaten möglichst automatisch mit anderen Daten anzureichern.**

pretationen, oder Querverbindungen zu anderen Medien. Begründet werden die Anreicherungen regelmässig damit, dass die Nutzerinnen und Nutzer diese wünschen, obgleich das nicht immer nachgewiesen wird.

Grundsätzlich kann gesagt werden, dass Fragestellungen im Bezug auf bibliografische Arbeit und Bibliothekskataloge international einer intensiven Diskussion unterliegen. Sicher ist, dass die Katalogarbeit in zehn Jahren radikal anders aussehen wird. Dabei beziehen einige Einrichtungen oder Einzelpersonen klare Positionen und bestimmen so die zukünftige Gestaltung der Katalogarbeit mit.

## Trends und Szenarien

Wer sich in Diskussionen um die Zukunft der Katalogarbeit einbringen will, sollte den heutigen Zustand kennen und gleichzeitig Vorstellungen von möglichen und anstrengenswerten Zukünften haben. Dies ermöglicht, Positionen dazu zu beziehen, was möglich sein soll, was notwendig ist und was verhindert werden muss. In diesem Abschnitt wollen wir als Diskussionsgrundlage drei erkennbare Trends als Zukunftsszenarien darstellen.

### *Kollaborative Katalogisierung*

Die aktuell diskutierten Regelwerke und Austauschformate zeigen eine klare Tendenz: Es werden Netzwerke von Wissen generiert, in denen einzelne Medien Endpunkte darstellen. Diese Netzwerke entstehen nicht, indem bibliografische Daten jedes einzelnen Bibliotheksmediums aufgenommen und anschliessend untereinander verbun-

den werden. Vielmehr implizieren die Diskussionen, dass bei Integration in einen Bibliotheksbestand die einzelnen Medien als Item in ein Wissensnetzwerk eingebunden werden.

Die Pflege dieses Wissensnetzwerks ist als Ergebnis kollaborativer Arbeit möglich: Alle Bibliotheken übernehmen gemeinsam die Weiterentwicklung des Netzwerks, ergänzen und updaten die notwendigen Entitäten und vertrauen gleichzeitig darauf, dass die Kolleginnen und Kollegen in den

**Die aktuell diskutierten Regelwerke und Austauschformate zeigen eine klare Tendenz: Es werden Netzwerke von Wissen generiert, in denen einzelne Medien Endpunkte darstellen.**

anderen Bibliotheken, im Idealfall auf internationaler Ebene, das Gleiche mit dem gleichen qualitätsbewussten Impetus tun.

### Szenario

2021, der Bibliothekar L. der juristischen Bibliothek im Centovalli hat soeben die 5. Auflage eines Buches erhalten und stellt während seiner Suche innerhalb des Wissensnetzwerks fest, dass er der erste Besitzer der besagten Auflage ist. Mit einem Klick fügt er die Auflage unter die Entität «Manifestation» zur betreffenden Expression hinzu. In einem weiteren Schritt fügt er das Buch – als Entität «Item» – hinzu. Früher hätte er zeitraubend den bereits katalogisierten Datensatz des Buches als Fremdtiteldaten übernommen, überprüft und seinen Hausregeln angepasst.

\* Eliane Blumer ist Lehr- und Forschungsassistentin im Bereich «Information documentaire» an der Haute Ecole de Gestion, Genève.

Dr. Karsten Schuldt (Chur/Berlin) ist Wissenschaftlicher Mitarbeiter am Schweizerischen Institut für Informationswissenschaft, HTW Chur.



### Zentrale Katalogisierung

Die aktuellen Diskussionen beinhalten auch die Möglichkeit, dass ein Grossteil der Katalogarbeit nicht mehr in den einzelnen Bibliotheken, sondern zentral stattfindet, bspw. bei den Verbundzentralen, Nationalbibliotheken oder zentralen Katalogisierungsstellen. Wenn konkrete Bibliotheksmedien als Item an ein bestehendes Wissensnetzwerk angehängt werden, kann es sich als sinnvoll herausstellen, wenn dieses Wissensnetzwerk von einigen Expertinnen und Experten gepflegt wird. Die Bibliotheken hätten weniger Personalzeit in die Katalogisierung zu investieren und könnten sich anderen Aufgaben widmen. Gleichzeitig würden sich die Expertinnen und Experten, welche

**Die Bibliotheken hätten weniger Personalzeit in die Katalogisierung zu investieren und könnten sich anderen Aufgaben widmen.**

das bibliografische Wissensnetzwerk pflegen, intensiver mit den Konsequenzen von Entscheidungen in Regelwerken auseinandersetzen und sich aktiv in Diskussionen über diese einbringen. Für solche Systeme gibt es heute schon in öffentlichen Bibliothekssystemen Vorbilder.

### Szenario

2018, die Mitarbeiterin M. der Verbundzentrale Gesamtschweiz verbringt ihren Morgen mit der Metadatenerstellung auf allen Entitätsebenen eines Pakets von Büchern zum rätoromanischen Schulunterricht und schickt die Bücher anschliessend an die Bibliotheken, welche die Bestellung in Auftrag gegeben hatten. Diese können die Medien direkt zur Ausleihe verwenden.

Da sie von einer Kollegin des Verbunds Ostdeutschland vernommen hat, dass das Sorbische sich ähnlich entwickelt, es aber im Wissensnetzwerk noch keine Verbindung der zwei Kleinsprachen gibt, beschliesst sie, dies auf der nächsten bibliografischen Konferenz zu thematisieren.

### Metadata Librarian

Die erkennbare Tendenz, Katalogdaten innerhalb der Bibliothekssysteme mit

weiteren Daten anzureichern, wird sich verstärken. Die Frage ist nur, welche Anreicherungen sinnvoll sind. Dies hängt von den verfügbaren Daten, den Möglichkeiten der zukünftigen Bibliothekssoftware und den Einschätzungen der Bibliotheken, was für die Sammlungen sowie die Nutzerinnen und Nutzer sinnvoll ist, ab. Grundsätzlich wird mit einer steigenden Anzahl der Anreicherungen die Aufgabe an die Bibliotheken übergehen, aus diesen Möglichkeiten zu wählen. Dies könnte die heutige Katalogisierung ersetzen.

### Szenario

2035, der Musikbibliothekar D. des Instituts für Populärmusik in Versoix erhält eine Sammlung von Stücken des Schweizerischen Kammermusikorchesters. Für die gespielten Stücke wählt er innerhalb internationaler und regionaler Anreicherungen, z.B. Lexikonartikel zu Musikstilen, Webportale von Musikfans oder musikwissenschaftliche Artikel, aus. Dadurch werden die Stücke in einen erweiterten Kontext gesetzt, und es wird so ermöglicht, weit über das einzelne Musikstück hinweg Informationen direkt im Katalog zu finden. Der alte Katalog hingegen beantwortete vor allem die Frage, ob ein Medium vorhanden war.

### Fazit

Alle Bibliotheken können sich aktiv in die Gestaltung der Zukunft des Katalogs und der bibliothekarischen Regelwerke einbringen. Bislang tun dies aber nur wenige. So brachte sich die Deutsche Nationalbibliothek engagiert in die Diskussion um FRBR und RDA ein und hinterliess in diesen Spuren europäischer Katalogtheorie und -praxis. Die Schweizerische und Österrei-

che Nationalbibliothek arbeiten nun mit der Deutschen Nationalbibliothek bei der Umsetzung von RDA zusammen.

Die Mailinglisten für die Diskussionen um Bibframe (bibframe@listserv.loc.gov), FRBR (frbr@infoserv.inist.fr) oder die RAK-Liste (rak-list@lists.dnb.de) zeigen, dass sich Bibliothekarinnen und Bibliothekare aus unterschiedlichsten Einrichtungen tatkräftig daran beteiligen können, die bibliothekarischen Regelwerke, Frameworks und Werkzeuge zu gestalten. Dazu bedarf es eines Wissens darüber, was möglich wäre und was sein sollte. Wir haben drei Szenarien als Input für Diskussionen entworfen. Andere Meinungen sind möglich und erwünscht.

Kontakt: eliane.blumer@hesge.ch

karsten.schuldt@htwchur.ch

### Literatur

- IFLA Study Group on the Functional Requirements for Bibliographic Records/Functional Requirements for Bibliographic Records: Final Report. [ohne Ort] : International Federation of Library Associations and Institutions, 2009
- Library of Congress/Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Washington, D.C.: Library of Congress, 2012
- Schweizerische Nationalbibliothek / Schweizerische Nationalbibliothek 99. Jahresbericht 2012. Bern: Schweizerische Nationalbibliothek, 2013

## ABSTRACT

### *Bibliothèques et catalogage ou comment maîtriser le futur*

Dans un proche avenir, le travail de catalogage changera dans les bibliothèques, notamment en raison de la place qu'occuperont les Linked Open Data. Toutes les bibliothèques peuvent s'insérer dans ce processus de changement et y participer activement. Les auteurs décrivent dans cet article trois tendances qui leur semblent particulièrement pertinentes à ce stade, à savoir le catalogage collaboratif, le catalogage central et le *metadata librarian*, et élaborent pour chacune d'elle un scénario envisageable pour le futur travail des bibliothèques. Ces scénarios devraient surtout permettre de susciter le débat sur les futurs possibles et souhaitables pour les bibliothèques. (traduction: sg)