

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: Projekte und Initiativen aus Bibliotheken und Archiven = Projets et initiatives dans le domaine archive et bibliothèque

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. 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. [Siehe Rechtliche Hinweise.](#)

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. [Voir Informations légales.](#)

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. [See Legal notice.](#)

Download PDF: 04.05.2025

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

II. Projekte und Initiativen aus Bibliotheken und Archiven / *Projets et initiatives dans le domaine archive et bibliothèque*

Die Bibliographic Framework Initiative (BIBFRAME) der Library of Congress – Offene, verknüpfte Datensätze

Bruno Wenk, HTW Chur

Die Bibliographic Framework Transition Initiative stellt ein zukunftsorientiertes Modell für den Übergang von MARC zu verknüpften offenen bibliografischen Daten vor. Das Modell orientiert sich sowohl an der Praxis als auch an bereits vorliegenden Modellen (FRBR) und neuen Katalogisierungsregeln (RDA).

Grenzen neuer Dienstleistungen für Bibliotheksnutzer/innen

Bibliotheken bemühen sich bei der Entwicklung neuer Angebote vermehrt um Zusammenarbeit, um ein grösseres Publikum zu erreichen und die Kosten für die Dienstleistungen teilen zu können. Ein Beispiel dafür ist der Velobuchkurier der fünf am «Bibliotheksplatz Chur»¹ beteiligten Bibliotheken: Die Kunden können sich Medien der teilnehmenden Bibliotheken kostenlos in die Lieferbibliothek ihrer Wahl bringen

lassen und Medien an der nächstgelegenen Bibliothek zurückgeben. Die innovative Dienstleistung hat allerdings einen Haken: Interessierte Nutzer/innen müssen in zwei verschiedenen Onlinekatalogen suchen, um den Gesamtbestand der beteiligten Bibliotheken abzufragen – wer ist heute noch bereit, diese Mühe auf sich zu nehmen? Der Aufwand für die Realisierung einer Webanwendung zur gleichzeitigen Suche in beiden Katalogen ist für die beteiligten Bibliotheken aber aus technischen und finanziellen Gründen nicht vertretbar.

Im Rahmen des Innovations- und Kooperationsprojekts der Schweizer Hochschulen «E-lib.ch: Elektronische Bibliothek Schweiz» wurde der Aufwand geleistet. Mit der Eingabe von Stichwörtern im Suchfeld von Swissbib² werden die Bestände von 900 Bibliotheken, Mediotheken und Archiven aus allen Landesteilen der Schweiz ab-

gefragt, darunter auch diejenigen der Churer Bibliotheken. Aber Swissbib wäre auch keine Lösung. Weil die Suche nur auf Bibliotheksverbünde und nicht auf einzelne Bibliotheken eingeschränkt werden kann, müssen zur Abfrage des Gesamtbestands des «Bibliotheksplatzes Chur» trotzdem zwei getrennte Suchen durchgeführt werden.

Verteilte geschlossene Datenbanken sind keine zukunftsorientierte Form der Datenhaltung für Gedächtnisinstitutionen. Um Bibliotheksnutzerinnen und -nutzern innovative Dienstleistungen anbieten zu können, ist Open Data eine (notwendige) Grundlage.

Open Data

Open Data ist eine Bewegung mit dem Ziel, Daten aus geschlossenen Datenbanken zu «befreien» und im Web öffentlich zugänglich und nutzbar zu machen. Wer eine auf offenen Daten beruhende Dienstleistung anbieten möchte, kopiert die Daten und programmiert eine passende Webapplikation. Ein Beispiel für offene geografischen Daten (Orte, Strassen, Gebäude usw.) bietet OpenStreetMap. Basierend auf den Daten von OpenStreetMap³ stellt eine an der HTW Chur realisierte Webanwendung die (auf OpenStreetMap erfassten) Schweizer Bibliotheken (und Museen) auf einer Karte dar (Bild 1)⁴.

Die Veröffentlichung von Daten bewegt momentan vor allem die Behörden

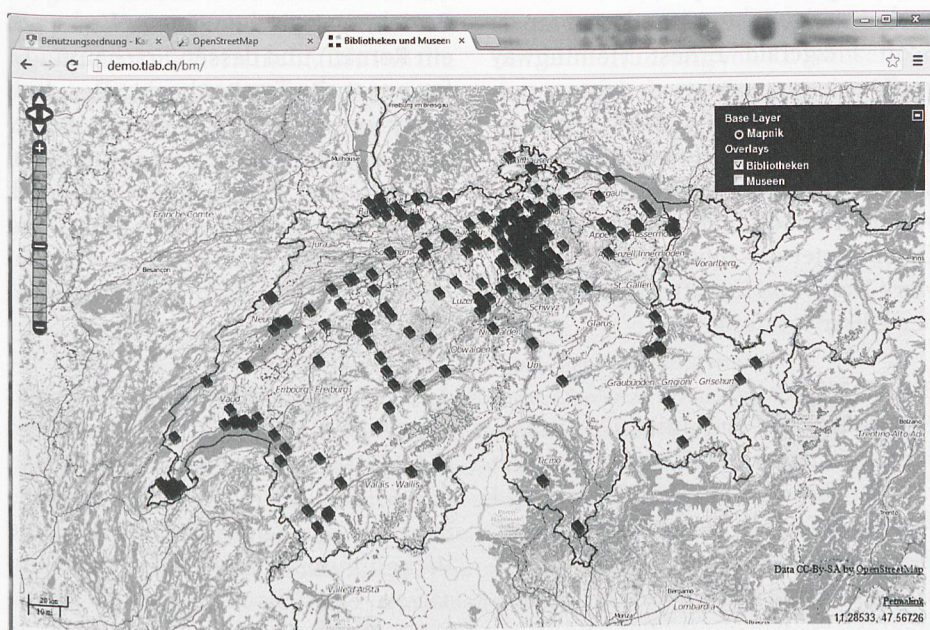


Bild 1: Karte mit den auf OpenStreetMap erfassten Schweizer Bibliotheken

- 1 www.bibliotheksplatz-chur.ch
- 2 www.swissbib.ch/
- 3 www.openstreetmap.org
- 4 <http://demo.tlab.ch/bm/>

1. The curious case of Benjamin Button das Hörbuch zum Sprachen lernen Verfasser / Beitragend: F. Scott Fitzgerald	2009	Ungekürzte Originalfassung	Medienkombination	Englisch	5 Bibliotheken ②
2. The Curious Case of Benjamin Button das Hörbuch zum Sprachen lernen Verfasser / Beitragend: F. Scott Fitzgerald ; David Williamson (Sprecher)	2009		Jf Ton	Deutsch	2 Bibliotheken ②

Bild 2: Wegen der unterschiedlichen Belegung der MARC-Felder ist für ein Programm nicht erkennbar, dass sich die beiden dargestellten Datensätze auf dasselbe Hörbuch beziehen.

(Open Government Data). Beispielsweise hat am 16. September 2013 die Bundesverwaltung ihr Pilotportal für offene Behördendaten der Schweiz⁵ eröffnet. Ein erster Schritt von Bibliotheken in Richtung Open Data wäre die Öffnung der OPAC und die Publikation ihrer bibliografischen Daten im Web. Vorreiterin diesbezüglich war die Bibliothek des Kernforschungszentrums CERN in Genf, welche im Dezember 2009 die Datensätze des CERN Document Server im Format MARCXML im Web zum Download zur Verfügung stellte⁶.

Die Open-Data-Bewegung legt das Format der offenen Daten nicht fest. Das Format MARCXML wäre demnach für die Veröffentlichung bibliografischer Daten durchaus geeignet. Aber das kryptische Feldformat ist nicht mehr zeitgemäss und die Belegung der Felder von Bibliothek zu Bibliothek zu unterschiedlich (vgl. Library of Congress, 2011). Die Programmierung von Metasuchfunktionen ist deshalb sehr aufwendig (Bild 2).

Linked Open Data

Das Konzept von Linked Open Data ist strikter als dasjenige von Open Data. Es legt nicht nur fest, dass die Daten im Web veröffentlicht und durch einen Unified Resource Locator⁷ (URL) eindeutig adressierbar sein müssen; das Konzept von Open Data legt auch fest,

dass die Datensätze in einer standardisierten Form (RDF⁸) maschinenlesbar formuliert sein und Verknüpfungen (Links) mit weiteren, verwandten Datensätzen enthalten müssen (Berners-Lee, 2006).

Auf bibliografische Daten übertragen ermöglichte das Konzept von Linked Open Data den Nutzerinnen und Nutzern, Werke, Autorinnen und Autoren, Institutionen und viele weitere Fakten zu entdecken. Im Linked-Open-Datensatz zu «The Great Gatsby» wäre anstelle des Namens des Autors (die Buchstabenfolge «Fitzgerald, F. Scott») eine Verknüpfung zum Datensatz mit Angaben zum Autor Francis Scott Fitzgerald zu finden und würden beispielsweise auch Verknüpfungen zu Datensätzen mit Angaben zu den Übersetzungen und Verfilmungen des Romans enthalten sein. Und weil diese Datensätze maschinenlesbar wären, würde eine entsprechende Webanwendung einer Person, die sich für den grossen Gatsby interessiert, durch automatisches Verfolgen der verknüpften Datensätze auch anzeigen können, dass Francis Scott Fitzgerald Ernest Hemingway kannte.

Einen Weg von offenen bibliografischen Daten im Format MARCXML zu verknüpften offenen bibliografischen Daten bietet die Bibliographic Framework Transition Initiative (BIBFRAME) der Library of Congress.

BIBFRAME

BIBFRAME⁹ ist eine Initiative der Library of Congress. In Zusammenarbeit mit der Firma Zepheira¹⁰, Library and Archives Canada, der British Library, der Deutschen Nationalbibliothek und anderen Nationalbibliotheken, zahlreichen MARC-Nutzergruppen und weiteren Communities wird ein Modell ent-

wickelt, das den Grundstein für die Zukunft verknüpfter offener bibliografischer Daten legt. Das praxisorientierte Modell berücksichtigt die Anforderungen an detaillierte bibliografische Beschreibungen, an Beschreibungen anderer kultureller Objekte und an weniger detaillierte Beschreibungen von Inhalten gleichermassen. Die Initiative befasst sich aber nicht nur mit einem zukunftsorientierten Ersatz für das Machine Readable Cataloging (MARC) Format, sondern untersucht auch Aspekte der Erfassung und des Austauschs bibliografischer Daten (z.B. Katalogisierungsregeln und Datenaustauschprotokolle).

Das Modell von BIBFRAME lehnt sich an dasjenige der Functional Requirements for Bibliographic Records (FRBR) an. Es unterscheidet klar zwischen einem Werk (work) und dessen physischen/digitalen Erscheinungsformen (instances), umfasst aber nur diese beiden Ebenen (anstelle der vier Ebenen Werk/Expression/Manifestation/Exemplar von FRBR). Werke und Instanzen können Beziehungen zueinander und untereinander sowie Beziehungen zu Autoritäten (authorities) und Notizen (annotations) aufweisen (Bild 3).

Zur Beschreibung der Beziehungen stellt BIBFRAME ein Vokabular zur Verfügung, das erweiterbar ist. Beispielsweise beschreiben die Ausdrücke «hasInstance» bzw. «instanceOf» die Beziehung zwischen einem Werk (z.B. ein Roman) und dessen Erscheinungsformen (z.B. Taschenbuch). Der Ausdruck «creator» beschreibt die Beziehung zwischen dem Roman und dessen Autor/in. Und mit dem Ausdruck «providerName» wird die Beziehung zwischen einer Instanz (z.B. das Taschenbuch) und dem publizierenden Verlag beschrieben.

Auf der Website von BIBFRAME¹¹ steht ein Onlinedienst zur Verfügung, um MARCXML-Datensätze automatisch in das Format BIBFRAME zu konvertieren. Bei der Konversion werden die Verknüpfungen mit Autoritätsdatensätzen ebenfalls automatisch eingefügt. Damit können auch Verbünde kleinerer (öffentlicher) Bibliotheken

5 <http://opendata.admin.ch>

6 <http://oldlibrary.web.cern.ch/oldlibrary/Library/bookdata.html>

7 Webadresse, z.B. <http://www.worldcat.org/oclc/727645936>

8 Resource Description Framework

9 www.loc.gov/bibframe/

10 <http://zepheira.com>

11 <http://bibframe.org>

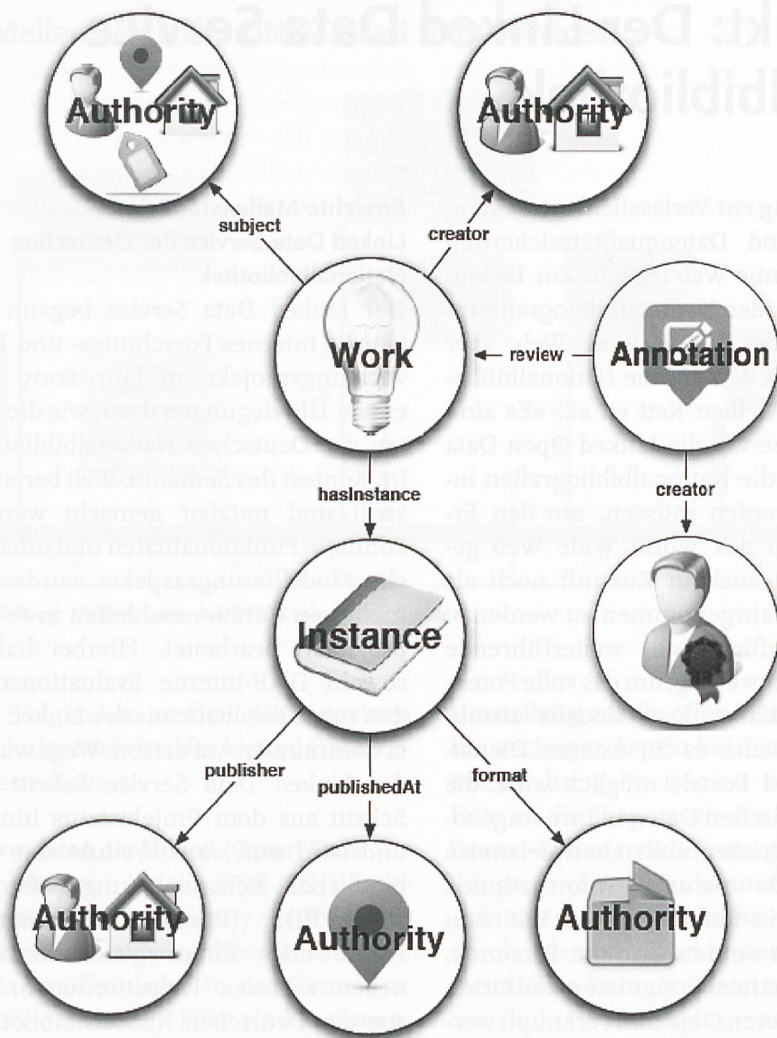


Bild 3: BIBFRAME-Modell (Quelle: Library of Congress, 2012; modifiziert)

Experimente mit Linked Open Library Data durchführen.

Die automatische Konversion löst allerdings das in Bild 2 illustrierte Problem der unterschiedlichen Belegung von MARC-Datenfeldern nicht. Die Inhalte der generierten BIBFRAME-Datensätze müssen auf die Einhaltung allgemein gültiger Katalogisierungsregeln

geprüft und nötigenfalls überarbeitet werden. Obwohl BIBFRAME keine bestimmten Katalogisierungsregeln voraussetzt, orientiert sich dessen Vokabular an RDA¹² (Resource Description and Access). Bibliotheken, die sich bereits mit der Einführung von RDA auseinandergesetzt haben, können beim Übergang von MARC zu BIBFRAME darauf aufbauen.

Um schliesslich den Nutzerinnen und Nutzern den im Zusammenhang mit Linked Open Library Data versprochenen Mehrwert zu bieten, müssen die konvertierten BIBFRAME-Datensätze noch mit anderen offenen Datensätzen verknüpft werden, beispielsweise mit Datensätzen der Linked Movie¹³ Database (LinkedMDB) und Datensätzen von Wikidata¹⁴, der freien Wissensbasis mit aktuell mehr als 13 Millionen Datensätzen.

Fazit

Die Bibliographic Framework Transition Initiative stellt ein zukunftsorientiertes Modell für den Übergang von MARC zu verknüpften offenen bibliografischen Daten vor. Das Modell orientiert sich sowohl an der Praxis als auch an bereits vorliegenden Modellen (FRBR) und neuen Katalogisierungsregeln (RDA). Die Arbeit an BIBFRAME ist noch im Gang, weshalb vorläufig nur Pilotprojekte in kleinerem Rahmen sinnvoll sind. Aber die Beschäftigung mit BIBFRAME und die aktive Beteiligung an der laufenden Diskussion über das BIBFRAME-Modell sind für die Entwicklung zukünftiger Informationsangebote von Bibliotheken auf jeden Fall nötig.

Kontakt: bruno.wenk@htwchur.ch

Quellen

- Berners-Lee, T., «Linked Data», 2006. (www.w3.org/DesignIssues/LinkedData.html)
- Library of Congress, «A Bibliographic Framework for the Digital Age», 2011. (www.loc.gov/bibframe/news/framework-103111.html)
- Library of Congress, «Web of Data: Linked Data Model and Supporting Services», 2012. (www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf)

ABSTRACT

Des données ouvertes et reliées au lieu de banques de données fermées

L'initiative Bibliographic Framework Transition présente un modèle d'avenir pour la transition de MARC à des données biographiques ouvertes et connectées. Le modèle s'oriente aussi bien vers la pratique que vers des modèles existants (FRBR) et de nouvelles règles de catalogage (RDA). Le travail sur BIBFRAME est encore en cours, raison pour laquelle seuls des projets-pilotes réalisés dans un cadre restreint sont judicieux pour l'instant. Cela étant, se pencher sur BIBFRAME et participer activement à la discussion en cours concernant le modèle BIBFRAME est utile et nécessaire pour le développement des futures offres d'informations des bibliothèques. (traduction: sg)

¹² www.rdatoolkit.org

¹³ <http://datahub.io/de/dataset/linkedmdb>

¹⁴ www.wikidata.org

Vom Projekt zum Produkt: Der Linked Data Service der Deutschen Nationalbibliothek

Julia Hauser,
Deutsche Nationalbibliothek

Das Umfeld, in dem Informationsspezialisten wie Bibliothekare, Archivare und Dokumentare arbeiten, hat sich mit der Verbreitung des World Wide Webs und den zugrunde liegenden Informationstechnologien grundlegend verändert. Damit ergaben sich neue Möglichkeiten der Datenbereitstellung, der Datennutzung und interdisziplinärer institutionenübergreifender Kooperationsmöglichkeiten.

Motivation und Hintergrund

Das Umfeld, in dem Informationsspezialisten wie Bibliothekare, Archivare und Dokumentare arbeiten, hat sich mit der Verbreitung des World Wide Web und den zugrunde liegenden Informationstechnologien grundlegend verändert. Damit ergaben sich neue Möglichkeiten der Datenbereitstellung, der Datennutzung und interdisziplinärer institutionenübergreifender Kooperationsmöglichkeiten. Die vollständige Überführung existierender Arbeitsprozesse und lang etablierter Strukturen hin zu einer modernen webbasierten Informationsinfrastruktur ist ein langer Weg – erste Schritte dahin sind bereits getan; weiteres Potenzial ist vorhanden und entsteht durch den fortschreitenden technologischen Entwicklungsprozess immer wieder neu.

Während Bibliotheksdaten viele Jahre lang in geschlossenen Katalogsystemen isoliert waren, sind sie heute dabei, zu einem stabilisierenden Element des World Wide Web zu werden. Kontrollierte Vokabulare, standardisierte Metadatenformate und ein persistenter Nachweis der Ressourcen sind Komponenten, die die Arbeit von Informationsspezialisten seit je begleiten. Bibliotheksdaten können nun ei-

nen Beitrag zur Verlässlichkeit, Beständigkeit und Datenqualitätssicherung des Semantic Web leisten. Zur Bedeutung etwa der Nationalbibliografie im Zeitalter des World Wide Web, aber auch des WWW für die Nationalbibliografie, schreiben Kett et. al.: «Es sind Datennetze wie die Linked Open Data Cloud, in die Nationalbibliografien integriert werden müssen, um den Erwartungen des World Wide Web gerecht und auch in Zukunft noch als relevant wahrgenommen zu werden.»¹

Zukünftig sind weiterführende Schritte notwendig, um das volle Potenzial des Linked-Data-Modells auszunutzen. Damit es für externe Dienstleister und Portale möglich wird, die bibliografischen Daten in ihre Angebote zu integrieren, sollen beispielsweise vorliegende textuelle Informationen (Literele) in Referenzen auf Entitäten umgesetzt werden, also mit Personen, Körperschaften, Ereignissen und Orten sowie anderen Objekten verknüpft werden. Dies ist ein weiterer Schritt in die Richtung, die Nationalbibliografie als einen Teilgraphen in der Linked Data Cloud zu verankern².

Viele Bibliotheken haben die darin liegende Chance erkannt, sich bibliotheksfremden Sparten und Kooperationspartnern aus der Forschung zu öffnen und die bibliografischen Daten für weitere Anwendungsfälle attraktiver zu vermarkten. Dies schafft eine bessere Nachnutzbarkeit in anderen Kultur- und Gedächtniseinrichtungen wie Museen und Archiven und ermöglicht eine bessere (technische) Zitierbarkeit im Bereich der Forschung und Wissenschaft – einem Bereich, der zunehmend webbasiert arbeitet und auf verlässliche und persistente Datenquellen im Web angewiesen ist.

Je stärker Bibliotheksdaten mit anderen Datenanbietern und Institutionen vernetzt sind, je mehr wird ihre Bedeutung in der heutigen Informationsgesellschaft gestärkt: der Wert eines Knotens im Geflecht des Semantic Web bestimmt sich danach, wie viele Fäden zu ihm hinführen.

Erreichte Meilensteine im Linked Data Service der Deutschen Nationalbibliothek

Der Linked Data Service begann als kleines internes Forschungs- und Entwicklungsprojekt im Jahr 2009 mit ersten Überlegungen dazu, wie die Daten der Deutschen Nationalbibliothek im Kontext des Semantic Web bereitgestellt und nutzbar gemacht werden könnten. Funktionalitäten und inhaltliche Modellierungsaspekte wurden in mehreren Iterationsschleifen in Folgeprojekten bearbeitet. Hierbei halfen sowohl DNB-interne Evaluationsrunden sowie Feedback aus der Linked Data Community. Auf diesem Wege wuchs der Linked Data Service Schritt für Schritt aus dem Projektstatus hinaus und wird nun zum Produkt mit verbindlicher Releaseplanung. Hiermit steht RDF (Resource Description Framework)³ fortan gleichberechtigt neben weiteren Exportformaten, die von der Deutschen Nationalbibliothek angeboten werden⁴.

Porträt

Julia Hauser ist Diplom-Bibliothekarin in der IT-Abteilung der Deutschen Nationalbibliothek (DNB). Hier arbeitet sie seit 2009 an Projekten zur Repräsentation der DNB-Daten im Kontext des Semantic Web mit. Julia Hauser verantwortet den Linked Data Service der DNB und koordiniert den DNB-Beitrag zur BIBFRAME-Initiative der Library of Congress.



1 Kett, Jürgen; Manecke, Mathias; Beyer, Sarah: Die Nationalbibliografie im Zeitalter des Internets. In: ZfBB 59(2012), 2, S. 70

2 vgl. ebd., S. 71 ff.

3 www.w3.org/TR/rdf-primer/

4 www.dnb.de/datendienst

Meilensteine des DNB-Linked Data Services

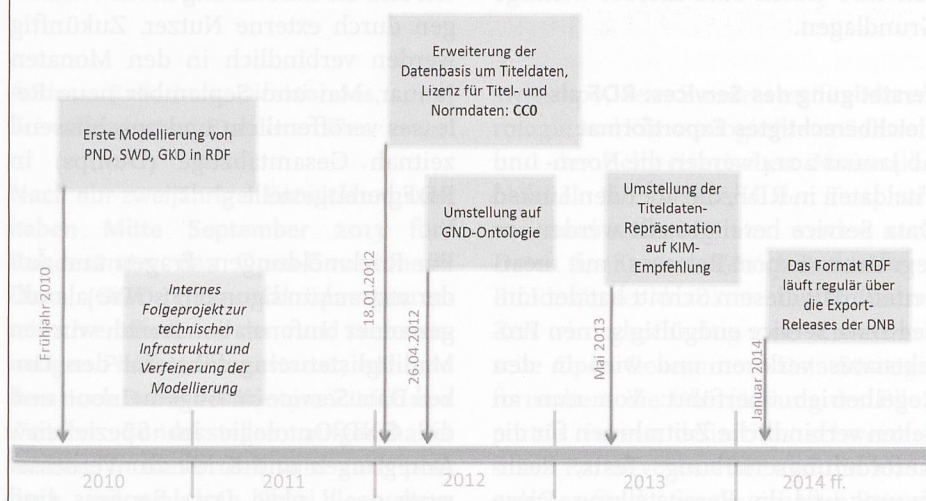


Abbildung: Meilensteine des DNB Linked Data Services

Im Folgenden werden die einzelnen Evolutionsschritte des Linked Data Services erläutert.

Normdaten als Linked (Open) Data

Im Jahr 2010 hat die DNB sich der Semantic Web Community geöffnet und erstmals ihre Normdaten als Linked Data veröffentlicht. In der Gemeinsamen Normdatei (GND)⁵ sind Personen, Familien, Körperschaften, Kongresse und Veranstaltungen, Geografika, Schlagwörter und Werke (z.B. die Einheitssachtitel des Deutschen Musikarchivs) enthalten. Die heutige GND bestand damals aus drei getrennten Normdateien: der Personennamendatei, der Schlagwortnormdatei und der Gemeinsamen Körperschaftsdatei. Die Modellierung erfolgte mittels bestehender Ontologien wie dem Friend-of-a-friend-Vocabulary (FOAF)⁶ und dem Elementset der Resource Description and Access (RDA)⁷. Diese Umsetzung konnte den Datenstrukturen nur zum Teil gerecht werden. Daher wurde im Zuge der Zusammenführung der drei früheren Normdateien zur GND auch die RDF-Repräsentation angepasst: Die DNB spezifizierte eine eigene GND-Ontologie⁸, die im April 2012 produktiv ging. Die GND-Ontologie wird sukzessive mit Elementen aus anderen, bereits existierenden Vokabularen verknüpft, im ersten Schritt mit dem RDA Element Set und FOAF⁹. Sie wird seit ihrer Erstveröffentlichung laufend aktualisiert und auch zukünftig an kommende Formatänderungen angepasst.

Umstellung auf die offene Lizenz CCO

Ein wichtiger Aspekt bei der Datenveröffentlichung im Kontext von Linked Data ist die Frage nach der Nutzungslizenz: kommerzielle externe Anwendungen und Services können nur dann auf einem Dataset aufbauen, wenn durch dessen Nachnutzung keine rechtlichen Konsequenzen zu befürchten sind, sie also unter einer offenen Lizenz stehen. Daher entschloss sich auch die DNB dazu, ab Januar 2012 sämtliche im Format RDF verfügbaren DNB-Daten unter die Nutzungslizenz Creative Commons Zero (CC0)¹⁰ zu stellen¹¹. Seit dem 1. Juli 2012 stehen darüber hinaus alle Normdaten und ein Grossteil der Titeldaten der DNB inklusive Kataloganreicherungsdaten auch in anderen Formaten kostenfrei zur Verfügung und werden unter CC0 zur freien Nachnutzung angeboten¹².

Die Entscheidung, die DNB-Daten schrittweise freizugeben und unter die Lizenz CC0 zu stellen, ist eine logische Konsequenz aus den Entwicklungen in der zunehmend globalisierten und vernetzten Informationswelt. Eine freie Lizenz ist die erste Voraussetzung für die Nachnutzung der Daten im Semantic Web.

Bereitstellung von Titeldaten

Anfang 2012 wurde der Linked Data Service um Titeldaten erweitert – die Voraussetzung, um auch im Kontext von Linked Data auf die Titel der Deutschen Nationalbibliografie zu referenzieren.

Der Umfang der konvertierten Titeldaten umfasst derzeit die Deutsche Nationalbibliografie inklusive der darin enthaltenen fortlaufenden Sammelwerke der Zeitschriftendatenbank (ZDB)¹³. Musikalien, Tonträger und die Bestände des Deutschen Exilarchivs und des Deutschen Buch- und Schriftmuseums sind aktuell noch nicht umgesetzt. Während für bibliografische Textdaten bereits ein Application Profile durch die Titeldaten der DINI-AG-KIM-Gruppe¹⁴ spezifiziert wurde, gibt es für Archivalien und Sonderbestände derzeit noch keinen Standard, der implementiert werden kann. Eine enge Zusammenarbeit mit Fachleuten aus dem Bereich Archiv und Dokumentation ist erforderlich, um eine RDF-Repräsentation für Sonderbestände zu erarbeiten. Aktuell ist eine Kooperation der DINI-AG-KIM-Gruppe-Titeldaten mit Musikbibliothekaren in Planung, um ein Anwendungsprofil für Musikalien und Tonträger zu erarbeiten. Langfristiges Ziel ist es, sämtliche Daten der Deutschen Nationalbibliothek inklusive kontrollierter Vokabulare als Linked Open Data bereitzustellen. Weitere Schritte sind geplant¹⁵.

5 www.dnb.de/gnd

6 <http://xmlns.com/foaf/spec>

7 <http://rdvocab.info>

8 <http://d-nb.info/standards/elementset/gnd#>

9 vgl. Haffner, Alexander: Internationalisierung der GND durch das Semantic Web, 2012, http://www.kim-forum.org/Subsites/kim/DE/Materialien/Dokumente/dokumente_no-de.html

10 <http://creativecommons.org/publicdomain/zero/1.0>

11 Svensson, Lars G.: Licensing Library and Authority Data Under CC0: The DNB Experience, 2013. http://www.w3.org/2013/04/odw/odw13_submission_57.pdf

12 www.dnb.de/geschaeftsmodell

13 www.zeitschriftendatenbank.de

14 <https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe>

15 Vgl. S. Hartmann, J. Hauser: Die Deutsche Nationalbibliografie in der Linked Data Cloud. In: Bibliotheken: Tore zur Welt des Wissens / 101. Bibliothekartag in Hamburg 2012. Hrsg. von Klaus-Rainer Brintzinger u.a. – Hildesheim; Zürich; New York, NY: Olms, 2013 (S. 57-71)

Geplante Schritte und Ausblick

Neben bereits erwähnten inhaltlichen Erweiterungen ist vor allem der Ausbau externer Verlinkungen für die Zukunft vorgesehen: Die Titeldaten sollen in kommenden Releases mit Titeln der British Library (BL)¹⁶ verknüpft werden. Auch die Normdaten sollen weiter mit externen Verknüpfungen angereichert werden, so etwa zu Geonames¹⁷, den LCSH¹⁸ und RAMEAU¹⁹.

Die Deutsche Nationalbibliothek arbeitet ausserdem in unterschiedlichen Projekten und Initiativen mit, die sich mit der Bereitstellung und Nutzung von Daten im Semantic Web beschäftigen, wie der Deutschen Digitalen Bibliothek²⁰ oder der europeana²¹. Darüber hinaus ist die DNB Mitglied der Early Experimenters Group der von der Library of Congress initiierten Bibliographic Framework Transition Initiative (BIBFRAME)²². Ziel dieser Initiative ist es, das MARC-21-Format durch ein nachhaltiges, RDF-basiertes Rahmenkonzept abzulösen. Die vorhandene DNB-Expertise hinsichtlich Konkordanzen, Formatumstiegen und

der Konvertierung sowie Bereitstellung von RDF-Daten sind hierbei wichtige Grundlagen.

Verstetigung des Services: RDF als gleichberechtigtes Exportformat

Ab Januar 2014 werden die Norm- und Titeldaten in RDF, die über den Linked Data Service bereitgestellt werden, in regulären Export-Releases mit bearbeitet²³. Mit diesem Schritt hat der Linked Data Service endgültig seinen Projektstatus verloren und wird in den Regelbetrieb überführt: Von nun an gelten verbindliche Zeitrahmen für die Anforderungserhebung, Tests, Realisierung und die Bereitstellung. Diese Abläufe werden für die Kunden transparent gemacht, sodass ggf. ausreichend Vorlauf für Anpassungen in ihren Systemen zur Verfügung steht. Die DNB schafft damit eine zuverlässige

Struktur für die Nachnutzung der Daten und die Einbindung in Anwendungen durch externe Nutzer. Zukünftig werden verbindlich in den Monaten Januar, Mai und September neue Releases veröffentlicht und anschliessend zeitnah Gesamtabzüge (Dumps) in RDF bereitgestellt.

Für Rückmeldungen, Fragen und Änderungsankündigungen sowie als allgemeiner Informationskanal wurden Mailinglisten eingeführt für den Linked Data Service im Allgemeinen²⁴ und die GND-Ontologie im Speziellen²⁵. Anregungen und Kritik zur Verbesserung des Linked Data Services sind willkommen über die Mailinglisten – weitere Kontaktmöglichkeiten sind auf der Website genannt²⁶.

Kontakt: j.hauser@dnb.de

ABSTRACT

Du projet au produit. Le service de Web de données de la Bibliothèque nationale allemande

Depuis 2010, la Bibliothèque nationale allemande (DNB) fournit ses données au format RDF, conforme aux besoins du Web sémantique, via un service de Web des données (linked data).

Par cette initiative, les utilisateurs et groupes d'utilisateurs ont eu, pour la première fois, la possibilité de réutiliser les données, sans avoir besoin d'une quelconque connaissance des formats propres aux bibliothèques.

Initialement, il s'agissait d'un prototype, lequel a été développé ensuite dans le cadre de projets. Après plusieurs versions, il a été optimisé sur les plans technique, fonctionnel et organisationnel. Parallèlement, RDF s'est imposé comme format d'exportation: le service de Web de données est ainsi passé du projet au produit.

La contribution donne un aperçu touchant les motifs de la démarche, évoque les problèmes de licence, les étapes de développement, rappelle l'évolution des contenus et fournit un aperçu des prochaines étapes.

(traduction: fs)

¹⁶ www.bl.uk/

¹⁷ www.geonames.org/

¹⁸ <http://id.loc.gov/authorities/subjects.html>

¹⁹ <http://rameau.bnf.fr/>

²⁰ www.deutsche-digitale-bibliothek.de/

²¹ www.europeana.eu

²² <http://bibframe.org/>

²³ <https://wiki.dnb.de/display/ILTIS/ILTIS-Change-Management>

²⁴ <http://lists.dnb.de/mailman/listinfo/lds>

²⁵ <http://lists.dnb.de/mailman/listinfo/gnd-ontology>

²⁶ www.dnb.de/lds

Andreas Kellerhals, Direktor des Schweizerischen Bundesarchivs

Nach nur zweijähriger Entwicklungszeit haben Mitte September 2013 fünf Bundesämter das Open-Government-Data- (OGD-)Pilotportal opendata.admin.ch eröffnet und den rund sechsmonatigen Pilotbetrieb gestartet, während dem noch Daten des Kantons Zürich und weiterer Bundesstellen aufgeschaltet werden. In dieser Zeit wird auch politisch und praktisch über die weitere OGD-Politik zu entscheiden sein.

Am 16. September 2013 hat der Bund anlässlich der Open Knowledge Conference in Genf sein Open-Government-Data-(OGD-)Pilotportal mit 1617 Datensätzen und vier Anwendungen der Öffentlichkeit vorgestellt¹. Die zentral verfügbar gemachten Daten stammen vom Schweizerischen Bundesarchiv BAR, dem Bundesamt für Statistik BFS, dem Bundesamt für Landestopografie

swisstopo, dem Bundesamt für Meteorologie und Klimatologie MeteoSchweiz und der Schweizerischen Nationalbibliothek NB. Während der Testphase von rund sechs Monaten werden noch Daten des Kantons Zürich und anderer Bundesämter dazukommen. In dieser Zeit wird gleichzeitig eine schweizerische Open-Government-Data-Strategie formuliert, die Überführung des Pilotportals in ein definitives schweizerisches Portal vorbereitet und die Massnahmen für dessen dauerhaften Betrieb beschlossen².

Die Schweiz ist kein OGD-Trendsetter, aber der Weg zum Pilotportal war doch ein erstaunlich kurzer: Von der ersten

Offene Behördendaten sind frei zugängliche und wiederverwertbare Datenbestände öffentlicher Verwaltungen, welche diese in Erfüllung ihrer gesetzlichen Aufgaben sammeln.

OGD-Konferenz im Bundesarchiv 2011 bis zum Launch in Genf hat es nur zwei Jahre gedauert. Der Weg ist gesäumt von verschiedenen Hackdays und wichtigen Publikationen zu OGD in der Schweiz sowie einer Abklärung zum Potenzial volkswirtschaftlicher Wertschöpfung³. Aktive Akteure waren der Verein opendata.ch und die parlamentarische Gruppe Digitale Nachhaltigkeit, die das Thema auf die politische Agenda gebracht hat⁴. Parallel dazu wurde unter der Leitung des Bundesarchivs pragmatisch dieses Pilotportal realisiert, welches sich als Tatbeweis für eine wirtschaftliche Machbarkeit in das E-Government-Vorhaben integriert.

Was sind offene Behördendaten?

Offene Behördendaten sind frei zugängliche und wiederverwertbare Datenbestände öffentlicher Verwaltungen, welche diese in Erfüllung ihrer gesetzlichen Aufgaben sammeln. Diese Datenbestände sind umfangreich, in der Regel von hervorragender Qualität und enthalten wertvolle Informationen zu unterschiedlichsten Themen auch

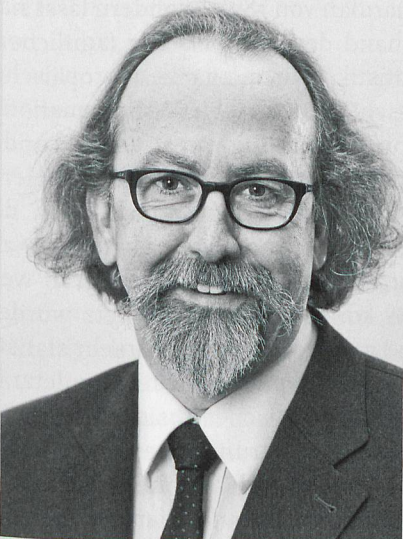
jenseits des ursprünglichen Verwendungszwecks⁵.

Damit interessierte Dritte diese Daten direkt und frei verarbeiten können, dürfen keine datenschutz-, informationsschutz- oder urheberrechtlichen Einschränkungen bestehen; am besten sollten sie unter einer einheitlichen Lizenz nutzbar sein⁶. Sie müssen zudem klar beschrieben sein und technisch in einer Form vorliegen, die einfache Weiterverarbeitung überhaupt ermöglicht. Gemäss einem Fünf-Sterne-Anforderungskatalog⁷ müssen sie

- 1 opendata.admin.ch. Kontrolle aller Links: 23.–27.9.2013.
- 2 Vgl. dazu den Bericht des Bundesrates zu Open Government Data vom 13. September 2013 (www.news.admin.ch/message/index.html?lang=de&msg-id=50250).
- 3 André Golliet et. al. (2012): Open Government Data Studie Schweiz, Bern (www.itopia.ch/repository/Publikationen/OGD_Studie_Schweiz_Juni_2012.pdf); Beat Estermann (2013): Schweizer Gedächtnisinstitutionen im Internet-Zeitalter. Ergebnisse einer Pilotbefragung zu den Themenbereichen Open Data und Crowdsourcing, Bern (www.wirtschaft.bfh.ch/de/forschung/institute/e_government_institut/tabs/publikationen.html). Adelheid Bürgi-Schmelz (2013): Wirtschaftliche Auswirkungen von Open Government Data. Verfasst im Auftrag des Bundesarchivs (www.bar.admin.ch/aktuell/index.html?lang=de).
- 4 OGD als priorisiertes E-Government-Vorhaben B2.12 (www.egovernment.ch/de/umsetzung/katalog_vorhaben.php) resp. in der Diskussion um das neue Meteorologiegesez (vgl. www.meteoschweiz.admin.ch/web/de/meteoschweiz/portrait/datenliberalisierung.html).
- 5 Vgl. dazu Golliet, wie Fussnote 3, S. 5. Die Studie weist auf die grosse Bedeutung unterschiedlichster Datenbestände etwa im Umwelt-, Gesundheits- oder Verkehrs- und Kommunikationsbereich hin.
- 6 Z.B. Creative Commons, die Praxis ist heterogener (vgl. andere Datenportale, wie data.gov.uk/ oder www.govdata.de/). Die konkrete rechtliche Ausgestaltung in der Schweiz ist noch in Arbeit.
- 7 Vgl. <http://5stardata.info/>.

Porträt

Andreas Kellerhals ist seit dem 1.11.2004 Direktor des Schweizerischen Bundesarchivs. Er studierte an der Universität Bern Geschichte, Architekturgeschichte und Staatsrecht. 1985 schloss er mit dem Lizentiat ab und war 6 Jahre als Wissenschaftlicher Assistent an der Universität Bern tätig. 1991 trat er als Direktionsadjunkt in den Dienst des Bundesarchivs, wo er ab 1995 als Vizedirektor amtierte.



demnach im Web auffindbar und strukturiert sein, in nicht proprietären Formaten vorliegen, eine URI haben und verlinkbar sein, d.h. beispielsweise im Resource Description Framework (RDF)⁸ vorliegen.

8 RDF ermöglicht «logische Aussagen über beliebige Dinge» und ist «ein grundlegender Baustein des Semantischen Webs» (vgl. dazu http://de.wikipedia.org/wiki/Resource_Description_Framework).

9 Das BAR ist beispielsweise daran, die bereits digitalisierten Staatsrechnungen seit 1848 zu datafizieren, damit interessierte Forschende diese nicht nur konsultieren, sondern auch direkt auswerten und bearbeiten können. Das Risiko von «Fehlinterpretationen» besteht, aber Fehlinterpretationen gab und gibt es auch bei nicht offenen und nichtverlinkten Daten.

10 EU-Schätzungen nennen rund 40 Milliarden, das Spektrum ist nach oben offen. Allerdings ist nicht einfach vergleichbar, was genau unter Public Sector Information verstanden wird und auf welchen Annahmen die Berechnungen beruhen. Vgl. dazu mit zurückhaltend-vorsichtigem Ansatz und trotzdem positivem Schlussergebnis Bürgi-Schmelz, wie Fussnote 3.

11 Beispielsweise: Artikel zum Beschaffungsfilz (Ex-Beamte profitieren vom Filz, Sonntagszeitung 25.11.2012, S. 7), Karte zum Hochwasser-Risiko in der Schweiz (www.sonntagszeitung.ch/fileadmin/dj/vis/flut_soz/flutsoz.html) in der Sonntagszeitung, die NZZ-Sommerserie mit verschiedenen Schweizerkarten (www.nzz.ch/aktuell/schweiz/ende-der-sommerserie-schweizer-karten-1.18130619) bzw. die Datenseite von Le Temps (www.letemps.ch/datas/) alle mit häufig starkem Echo in den Social Media.

12 Leitbilder in der Diskussion um offene Daten sind immer The Manchester Guardian, 5. Mai 1821, Publikation einer Liste von Schulen in Manchester und Salford mit Angaben zu Schüleranzahlen sowie den Jahresdurchschnitt an Ausgaben (vgl. z.B. www.nzz.ch/aktuell/digital/open-data-republi-ca-13-1.18078390#); anderes klassisches Beispiel sind die Mortalitätsstatistiken mit Diagrammen von Florence Nightingale aus dem Krimkrieg von 1858 (www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics#).

Was für eine Bedeutung haben offene Daten – speziell auch Linked Open Data – für Archive? Grundsätzlich unterliegen alle offenen Behördendaten einer Archivierungspflicht. Wenn sie öffentlich zugänglich bleiben, muss eine angemessene Archivierungslösung auf dem Konzept einer ununterbrochenen Nutzungsmöglichkeit aufbauen. Fragen des Zugangs und auch der Erschliessung stellen sich damit neu. Gleichzeitig eröffnet diese Entwicklung für Archive neben einem enormen Modernisierungspotenzial die Möglichkeit, ihr spezielles Fachwissen in den Dienst zeitgemäss standardisierter Katalogisierung und Sammlung von Daten über Daten einzubringen, um Orientierung zu schaffen und allen Interessierten Zugang zu Informationen zu eröffnen. Gleichzeitig kann Archivgut produktiv in einen offenen Datenraum integriert und an andere Informationen im oder ausserhalb des Archivs anschlussfähig gemacht werden. Archivgut zu digitalisieren und zu datafizieren⁹ braucht zwar einen nicht zu unterschätzenden Aufwand, verspricht aber erfreuliche Erfolgsaussichten.

Nicht zuletzt eröffnen Linked Open Data ebenfalls wesentlich einfachere Wege für archivische Kooperation in freier Form, ohne dass für jedes Projekt neuer Aufwand betrieben werden muss und jedes Update einer Datenbank zu einem ressourcenmässigen Flaschenhals wird. Archive gewinnen damit an Flexibilität und Archivgut kann wesentlich besser valorisiert werden – möglicherweise auch auf neue, aber kreative und spannende Art und Weise. Archivareinnen und Archivare müssen dabei allerdings auch einen Kontrollverlust in Kauf nehmen: Daten offen anbieten heisst, dass wir nicht mehr bestimmen können, was damit geschieht – aber das ist ohnehin nicht unsere Aufgabe. Umso wichtiger wird damit aber unser Beitrag zur Authentizitätsgarantie, welche nach wie vor nur die Archive leisten können.

Offene Behördendaten: ein Angebot und ...

Was ist der Nutzen frei verwendbarer Daten? Das legitimatorische Mantra ist kurz: Offene Behördendaten stärken die Transparenz sowie die demokratische Partizipation. Auch haben sie ein

grosses Potenzial für (volks-)wirtschaftliche Wertschöpfung, selbst wenn die Schätzungen hier weit auseinandergehen¹⁰. Neue Business-Modelle werden sich entwickeln und bereits jetzt profitieren beispielsweise die Medien vom sogenannten Datenjournalismus¹¹.

Ebenso gewinnen öffentliche Verwaltungen intern durch offene Behördendaten an Effizienz: Teilen der vorhandenen Daten dank Standardisierungen, bessere Qualität durch mehr Nutzung und intensivere Nutzung durch den Wegfall verwaltungsinterner Verrechnungen. Einnahmeausfälle für einzelne Amtsstellen können nicht al-

Die Forderung nach offenen Behördendaten schliesst sich dem Diskurs um das Öffentlichkeitsprinzip an, grenzt sich aber gleichzeitig klar von Wikileaks ab.

lein durch volkswirtschaftliche Wertschöpfung kompensiert werden, sondern müssen im Budget der Verwaltungsstellen ausgeglichen werden.

Ein staatspolitisch-demokratischer Nutzen ergibt sich aus der zusätzlichen Transparenz. Die Forderung nach offenen Behördendaten schliesst sich dem Diskurs um das Öffentlichkeitsprinzip an, grenzt sich aber gleichzeitig klar von Wikileaks ab. Auch steht und stand die Forderung nach mehr Transparenz immer in einem spannungsvollen Verhältnis zur Arkanpolitik. Seinen Anfang nahm dieses Spannungsfeld aber nicht mit der vielzitierten Publikation der Schulstatistik des Manchester Guardian von 1821¹², sondern lässt sich anhand des Beispiels der (amtlichen) Statistik über die ganze (europäische) Geschichte verfolgen. Informationsmonopolisierung von Daten und anderen Dokumenten ist stets ein Herrschaftsmittel, und so galt auch die Veröffentlichung von Statistiken als Verrat, der mit dem Tod bestraft werden konnte. Nichtsdestotrotz wurden im Laufe der Zeit immer mehr statistische Werke publiziert, nicht zuletzt in propagandistischer Absicht, und statistische Bevölkerungsdaten zirkulierten im 18. Jahrhundert ziemlich frei unter den Gelehrten und Pfarrherren über die Staatsgrenzen hinaus.

Die konkrete Wirkung offener Daten auf die politische Diskussion ist allerdings nicht einfach zu ermitteln; auch mag man bezweifeln, ob die Publikation von Wetterdaten oder der Fliessgewässertemperaturen – wie in der bei Berner Badenden beliebten Aare-App – politische Partizipation fördert. Wenn auch positive nichtökonomische Wirkungen und lebenspraktische Erleichterungen mehr dem Wohlergehen als dem Gemeinwohl (welfare effects) dienen, tragen sie doch zum «grössten Glücks der grössten Zahl» (Jeremy Bentham) bei.

... eine Herausforderung

Was sind mögliche Risiken und Nebenwirkungen? Bürgerinnen und Bürger mit qualitativ hochstehenden Daten und Informationen in die Lage zu versetzen, die Begründungen politischer Entscheidungen besser zu verstehen, sich selber eine eigene Meinung zu bilden oder überhaupt wohlbegründete Entscheidungen zu treffen – mit anderen Worten: durch intellektuelle Anstrengung aus möglicher Unmündigkeit herauszutreten –, ist ein hehres Anliegen. Offene Daten ersetzen allerdings das Denken nicht. Technisch frei kombinierbare Daten ergeben nicht automatisch inhaltlich zulässige Aussagen. Es muss immer ernsthaft geprüft werden, ob eine Kombination dekontextualisierter Daten zu logisch über-

zeugenden Ergebnissen führt. Jede Diskussion wird offen, der Umgang mit Daten überprüfbar und Resultate können fundiert bestritten, korrigiert

Offene Daten ersetzen allerdings das Denken nicht. Technisch frei kombinierbare Daten ergeben nicht automatisch inhaltlich zulässige Aussagen.

oder gar widerlegt werden. Damit geht die Pflicht einher, die neuen Möglichkeiten mit Sorgfalt und Respekt zu nutzen. Diskussionen werden sich von Fragen nach blossen Facts & Figures auf Methodenfragen verschieben: Was ist eine zulässige Argumentation? Was kann gesagt werden?¹³

Es kann als Aufgabe des Staats verstanden werden, Voraussetzungen für eine aktive Partizipation von Bürgerinnen und Bürgern zu schaffen: Daten sind nicht einfach Belege einer Meinung, sondern Rohstoff zur Meinungsbildung und müssen sich im Licht anderer Daten behaupten. OGD ist somit Teil einer Wissensinfrastruktur, die präventiv gegen Uninformiertheit und strukturell, nicht individuell begründetes Unwissen wirken soll¹⁴.

Unsere Gesellschaft braucht aber nicht allein offene Behörden-, Wissenschafts-

und anderen Daten, sondern auch intelligente Information. Das umfasst sowohl Inhalte aus verlässlichen Quellen als auch Instrumente mit transparenter Funktionsweise. Wir brauchen nicht bloss informationelle, sondern auch algorithmische Autonomie¹⁵. Ausserdem braucht es über die offenen (Behörden-)Daten hinaus auch einen offenen Geist. Im herrschenden Zwang zu argumentativer Auseinandersetzung kann der Staat dazu beitragen, dass alle gleich gute Voraussetzungen zur Partizipation haben. Dazu braucht es aber auch eine Haltung, wie sie Voltaire zugeschrieben wird: «Ich verachte Ihre Meinung, aber ich gäbe mein Leben dafür, dass Sie sie sagen dürfen.»¹⁶ Wenn nicht nur die Daten, sondern auch der Geist offen ist, dann haben wir etwas Positives erreicht.

Kontakt: andreas.kellerhals@bar.admin.ch

13 Berühmtes Beispiel ist die Reinhart-Rogoff-Debatte um die Studie *Growth in a Time of Debt*, in der sich weltwirtschaftliche Ratschläge als Ergebnis von falschen Formeln und unklar begründeter Datenauswahl bei der Modellbildung entpuppte – wie es die *New York Times* zugespitzt formulierte: in der Wirtschaftsexpertise von Excel-Expertise abhing. Vgl. dazu u.a. Paul Krugman, *The Excel Depression*, *New York Times*, 18. April 2013 (www.nytimes.com/2013/04/19/opinion/krugman-the-excel-depression.html?_r=0).

14 Willke, Helmut (1996): *Die Steuerungsfunktion des Staates aus systemtheoretischer Sicht. Schritte zur Legitimierung einer wissensbasierten Infrastruktur*. In Staatsaufgaben (Hg. von Dieter Grimm), S. 685–711. Frankfurt/M.

15 Rainer Kuhlen (1999): *Die Konsequenzen von Informationsassistenten. Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden?* Frankfurt/M (www.kuhlen.name/MATERIALIEN/Publikationen1995-2000/informationsassistenten.pdf).

16 Evelyn Beatrice Hall (Pseudonym Stephen G. Tallentyre) (1906): *The Friends of Voltaire* (vgl. <http://de.wikiquote.org/wiki/Voltaire>).

ABSTRACT

opendata.admin.ch

Après seulement deux années de développement, cinq offices fédéraux ont publié à la mi-septembre le portail-pilote Open Government Data (OGD) *opendata.admin.ch* et ont démarré ainsi un projet d'environ six mois durant lesquels seront ajoutées les données du canton de Zurich et d'autres offices fédéraux. Cette même période devrait permettre de décider de l'orientation politique et pratique de la stratégie de l'OGD.

L'OGD possède le potentiel – surtout sous sa forme en réseau, Linked Open Data – de transformer profondément l'archivage, aussi bien du point de vue de la permanence de la disponibilité des données que de celui de leur description et de leur organisation. Le patrimoine archivistique s'intègre dans un espace de données ouvertes. Ce potentiel de modernisation ne doit pas nous faire perdre de vue les risques et effets secondaires. Les données ouvertes représentent aussi un défi et une obligation pour leurs utilisateurs: il est également nécessaire de recourir, parallèlement aux données ouvertes, à des outils transparents, et de faire preuve d'un esprit ouvert. (traduction: fs)

Inscrire les données du Service de la culture dans le Web pour mieux valoriser le patrimoine

Alain Dubois, archiviste paléographe,
Archives de l'Etat du Valais

Les questions relatives à la mise en réseau et à l'interopérabilité entre les systèmes d'information et de description, ainsi qu'à l'échange et/ou à la réutilisation des données constituent aujourd'hui un enjeu professionnel majeur pour les institutions de conservation du patrimoine.

Elles sont actuellement au cœur des réflexions menées par le Service de la culture du canton du Valais, dans le sillage de la mise en place d'une plate-forme de pérennisation et d'archivage pour les collections numériques et numérisées des Archives de l'Etat du Valais, de la Médiathèque Valais et des Musées cantonaux conforme aux normes et standards internationaux en la matière. Elles aboutiront concrètement à la fin du premier semestre 2014 par la mise en ligne d'un outil de recherche fédérée dédié au patrimoine valaisan, l'ouverture d'un entrepôt OAI-PMH (open data culturel) et l'alignement des notices descriptives selon les principes du Web de données (linked open data).

Le présent article retrace en trois parties les étapes successives d'un projet qui a débuté en 2011 et qui a pour ambition d'offrir à toute personne intéressée un accès aux ressources patrimoniales conservées par les Archives de l'Etat du Valais, la Médiathèque Valais et les Musées cantonaux à partir d'un portail unique.

1. L'outil de recherche fédérée du Service de la culture

L'origine du projet de recherche fédérée se situe en fait dans le sillage de l'ouverture de la plate-forme de pérennisation et d'archivage en mai 2011. Il s'est agi, à cet instant, de mettre en réseau et de donner accès aux ressources numériques et numérisées disponibles sur cette plate-forme, indépendamment des outils métier institutionnels usuels – scopeArchiv pour les Archives de l'Etat du Valais et le catalogue collectif RERO pour la Médiathèque Valais. Rapidement, toutefois, s'est imposée l'idée de réaliser un outil qui, au-delà des seules ressources numériques et numérisées, donne accès à toutes les ressources en lien avec le patrimoine conservées par le Service de la culture. C'est ainsi qu'est né le projet d'outil de recherche fédérée. Actuellement en voie de finalisation, ce projet s'est déroulé en deux phases:

- une phase-pilote (avril-décembre 2011), qui a permis de démontrer la faisabilité d'une mise en réseau de ressources disparates;
- une phase de déploiement (janvier 2012-juin 2014), qui permet de consolider les acquis de la phase-pilote.

Sans revenir sur tous les éléments du projet, j'en mentionnerai ici les principaux.

Particularité de la structure de projet

Le projet est piloté par la cellule e-culture du Service de la culture. Cette structure originale, qui regroupe des représentants des Archives de l'Etat du Valais, de la Médiathèque Valais, des Musées cantonaux et de l'Encouragement des activités culturelles, a notamment pour objectif de mener des projets transversaux en lien avec les technologies numériques et relatives à la collecte, à la conservation, à la mise en valeur et à la promotion du patrimoine culturel valaisan et des ressources qui lui sont associées.

Particularité du projet

Dans sa philosophie, mais à une échelle nettement moindre, le projet de re-

cherche fédérée se rapproche par ailleurs de la bibliothèque numérique européenne lancée en novembre 2008 (Europeana). Il s'agit, en effet, de mettre à la disposition du public intéressé par le patrimoine valaisan un outil qui fédère des ressources descriptives qui ne le sont pas encore (fonds d'archives, collections audiovisuelles, bibliothèque numérique et collections muséales). Sur ce point, le défi principal porte sur la mise en commun et l'alignement de ressources décrites selon des normes et standards qui diffèrent fortement.

Choix de l'outil informatique

La particularité du projet a eu une incidence significative sur le choix de la solution informatique. Au terme d'une procédure sur invitation, c'est l'entreprise française Antidot SAS qui a été retenue pour réaliser l'outil de recherche fédérée¹. Ce choix s'est notamment justifié en raison des solides références dont disposait l'entreprise dans le domaine de la gestion de l'information, son expérience avérée dans la mise en œuvre d'outils de recherche fédérée et d'accès à l'information à travers de multiples portails et l'intégration des outils du web de données qu'elle proposait².

Choix du corpus

L'outil de recherche intègre actuellement 362 704 notices descriptives qui se répartissent comme suit:

- 68 947 notices de fonds d'archives conservés aux Archives de l'Etat du Valais (Etat du Valais, communes, bourgeoisies et paroisses, ainsi que fonds privés) et progressivement mis à la disposition du public sur le portail scopeQuery³;
- 279 182 notices de la bibliographie valaisanne disponibles sur le portail valaisan de RERO;
- 14 225 notices des ressources audiovisuelles conservées à la Médiathèque Valais – Martigny (photographies, enregistrements sonores et films) et mises à la disposition du public sur le portail «mémoire audiovisuelle du Valais»⁴;

1 www.antidot.net/ [consulté le 15 octobre 2013].

2 L'entreprise Antidot a notamment réalisé le portail du Centre national de la recherche scientifique pour les sciences humaines et sociales (www.rechercheisidore.fr/) [consulté le 15 octobre 2013].

3 <http://scopequery.vs.ch/suchinfo.aspx> [consulté le 15 octobre 2013].

4 <http://archives.memovs.ch/> [consulté le 15 octobre 2013].

– 350 notices issues des deux catalogues «Collectionner au cœur des Alpes» publiés par le Musée d'art et le Musée d'histoire du Valais⁵.

Livraison et traitement des données

L'un des principaux avantages du module Antidot Finder Suite, utilisé pour le traitement des données, réside dans le fait qu'il accepte en entrée des formats très variés, indépendamment de leur source. C'est ainsi que chaque office a pu déterminer son mode de livraison: les Archives de l'Etat du Valais exportent leurs unités de description en EAD (Encoded Archival Description), tandis que les notices de la bibliographie valaisanne sont livrées en Dublin Core et celles des collections audiovisuelles sont extraites de l'entrepôt OAI-PMH de la Médiathèque Valais – Martigny. Quant aux Musées cantonaux, ils fournissent leurs notices descriptives dans un fichier Excel.

Le module Antidot Finder Suite aligne ensuite les différents flux et les transforme en un format pivot dérivé du Dublin Core, avant d'appliquer sur le corpus ainsi constitué différents filtres d'indexation.

Editorialisation de l'outil de recherche fédérée

Le site dédié à la recherche fédérée est actuellement en cours de création⁶. A terme, il offrira au public un outil de recherche dont les résultats pourront être affinés à l'aide de facettes (provenance institutionnelle, lieux, périodes historiques et personnes). Il mettra également

en exergue un «document du mois» qui se fera l'écho de l'actualité d'un office (exposition, conférence, journée thématique ou encore nouvelle publication).

2. Vers un open data culturel

Dès les origines du projet, soucieux d'en assurer une diffusion aussi large que possible et dans le sillage des initiatives actuelles en la matière, le Service de la culture a décidé de rendre ses données interopérables et réutilisables. C'est ainsi que s'est très rapidement imposée l'idée de créer un portail qui rendent interopérables et moissonnables les données, dans la perspective de créer une plate-forme d'open data culturel au niveau valaisan. Sur ce point, le Service de la culture a pu bénéficier de l'expérience de la Médiathèque Valais, qui rend disponibles depuis 2009 déjà ses collections audiovisuelles à travers un entrepôt OAI-PMH⁷. Ce modèle a donc été repris et adapté en fonction des besoins. C'est ainsi que l'entrepôt OAI-PMH du Service de la culture, actuellement en cours de validation auprès de l'*Open Archive Initiative*⁸ et mis à disposition au début de l'année 2014 sur le site Internet de l'Etat du Valais⁹, proposera ses données dans deux formats: le format oai_dc, constitutif de tout entrepôt OAI-PMH, et le modèle de données d'Europeana (EDM).

Les sets seront ensuite rendus accessibles sous deux formes:

– par provenance institutionnelle (sets «Archives de l'Etat du Valais», «Bibliothèque numérique», «Mémoire audio-

visuelle du Valais» et «Musées cantonaux»);

– par type de ressources (sets «Photographies», «Enregistrements sonores», «Films», «Monographies», «Articles» et «Fonds d'archives»).

3. L'intégration du linked open data

Rendre interopérables ses données à travers un entrepôt OAI-PMH constitue certes une première étape importante, mais elle reste néanmoins insuffisante au regard des possibilités qu'offre désormais le Web de données en matière d'échange et de réutilisation des données. C'est ainsi que l'alignement des notices descriptives selon le modèle de

5 Ruedin, P., *Collectionner au cœur des Alpes: le Musée d'art du Valais, Sion, Sion – Paris 2007*; Elsig, P., Morand, M. C., *Collectionner au cœur des Alpes: le Musée d'histoire du Valais, Sion, Sion – [Paris] 2013*.

6 C'est l'entreprise Conchita-Plus qui a été choisie pour créer l'interface de consultation (www.conchita-plus.ch/ [consulté le 15 octobre 2013]).

7 Consultable à l'adresse <http://xml.memovs.ch/oai/>, l'entrepôt OAI-PMH de la Médiathèque Valais rend disponibles sous forme de sets ses collections de photographies, d'enregistrements sonores et de films. Les données sont disponibles dans les formats suivants: oai_dc, qdc (Dublin Core qualifié), ead, marcxml et ese (Europeana) [consulté le 15 octobre 2013].

8 www.openarchives.org/ [consulté le 15 octobre 2013].

9 www.vs.ch/oai.

Impressum

arbido print

arbido newsletter

arbido website

print: © arbido ISSN 1420-102X

website + newsletter: © arbido newsl. ISSN 1661-786X
www.arbido.ch

Offizielle Publikationsorgane – Organes officiels de publication

Bibliothek Information Schweiz BIS (www.bis.info)
Verein Schweizerischer Archivarinnen und Archivare VSA-AAS (www.vsa-aaa.org)
Bibliothèque Information Suisse BIS
Association des archivistes suisses AAS-VSA

Chefredaktor – Rédacteur en chef

Stéphane Gillioz (sg), E-Mail: stephane.gillioz@gmail.com

Redaktion BIS – Rédaction BIS

Stephan Holländer (sth), Katja Böspflug (kb), Daniela Rüegg (dr)

Redaktion VSA-AAS – Rédaction AAS-VSA

Frédéric Sartet (fs), Annkristin Schlichte (as)

Layout, Druck – Graphisme, Impression

Stämpfli Publikationen AG, Wölflistrasse 1, Postfach 8326, 3001 Bern, www.staempfli.com

Inserate – Annonces

Tel. 031 300 63 89, Fax 031 300 63 90
E-Mail: inserate@staempfli.com

Abonnemente – Abonnements

Tel. 031 300 63 41, Fax 031 300 63 90
E-Mail: abonnemente@staempfli.com

4 Hefte/Jahr: Fr. 115.– (im Mitgliederbeitrag inbegriffen).
Auslandsabonnemente: Fr. 135.– (inkl. Porto).

Einzelnummer: Fr. 30.– (plus Porto und Verpackung).

4 cahiers/année: Fr. 115.– (compris pour les adhérents).

Abonnements de l'étranger: Fr. 135.– (affranchissement compris).

Vente au numéro: Fr. 30.– (plus affranchissement et emballage).

Termine – Dates

Nächste Ausgabe erscheint am 20.2.2014
Le prochain numéro paraît le 20.2.2014

Thema – Thème

Alpenraum / Espace alpin / Spazio alpino

Redaktionsschluss – Délai de rédaction: 18.1.2014

Inserateschluss – Délai d'insertion: 14.1.2014

Newsletter

Pro Jahr erscheinen 8 bis 12 elektronische Newsletters.
Siehe www.arbido.ch

Parution de 8 à 12 newsletters électroniques par an.
Voir www.arbido.ch

Publiziert mit Unterstützung der Schweizerischen Akademie der Geistes- und Sozialwissenschaften SAGW
Publié avec le soutien de l'Académie suisse des sciences humaines et sociales ASSH

données d'Europeana permettra aux données du Service de la culture d'être présentées selon la structure du RDF (Ressource Description Framework).

Le référentiel des lieux, par exemple, sera aligné sur GeoNames, base de données géographique de plus de 6,5 millions de lieux existants¹⁰. Des discussions sont par ailleurs en cours à l'interne pour aligner le référentiel des personnes avec le fichier d'autorités VIAF¹¹. Par contre, l'abandon – temporaire toutefois – du référentiel des matières à ce stade du projet reporte l'alignement de ce dernier avec le thésaurus-matières pour l'indexation des archives locales proposé par le Service interministériel des Archives de France¹² ou la CDU abrégée¹³.

10 www.geonames.org/ [consulté le 15 octobre 2013].

11 <http://viaf.org/> [consulté le 15 octobre 2013].

12 <http://data.culture.fr/thesaurus/page/ark:/67717/Matiere> [consulté le 15 octobre 2013].

13 <http://udcdata.info/> [consulté le 15 octobre 2013].

Au final, les réflexions autour de la création d'un outil de recherche fédérée permettront non seulement d'offrir à toute personne intéressée un portail d'accès unique au patrimoine valaisan, qui pourrait du reste tout à fait être enrichi par des ressources en lien avec le Valais provenant d'autres institutions, mais surtout de l'inscrire dans le sillage des

projets d'open data et de linked open data actuellement en cours en matière de valorisation du patrimoine. En somme d'inscrire pleinement les données du Service de la culture dans le Web pour les valoriser au mieux.

Contact: alain.dubois@admin.vs.ch

ABSTRACT

Die Valorisierung des Kulturguts durch Onlinestellung von Daten der Dienststelle für Kultur des Kantons Wallis

Die Überlegungen zur Onlinestellung und Interoperabilität zwischen verschiedenen Informationssystemen, ebenso wie der Austausch und/oder die Wiederverwendung von Daten stehen im Zentrum der Überlegungen der Dienststelle für Kultur des Kantons Wallis, die im Gefolge der Inbetriebnahme der Speicher- und Archivierungsplattform für die digitalen und digitalisierten Bestände des Staatsarchivs des Kantons Wallis, der Mediathek und der kantonalen Museen angestellt wurden. Am Ende des ersten Semesters 2014 sollen diese Überlegungen ganz konkret zur Onlinestellung einer gemeinsamen Suche für das Walliser Kulturgut, zur Öffnung eines Datenlagers Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) und zum Abgleich der beschreibenden Daten nach den Grundsätzen des semantischen Webs führen.

Der vorliegende Artikel zeigt die wichtigsten Etappen des Projekts auf, das 2011 begonnen wurde und das zum Ziel hat, allen Interessierten den Zugang zu den kulturellen Ressourcen der Dienststelle für Kultur des Kantons Wallis über ein Guichet virtuel («One-Stop-Portal») zu ermöglichen.

(Übersetzung: as)

Linked Open Data in archivischen Fachinformationssystemen

Peter Sandner, Archivar,
Leiter der Abteilung Archivfachliche
IT/Digitales Archiv im Hessischen
Hauptstaatsarchiv in Wiesbaden

Die grosse Idee der vernetzten offenen Daten erreicht über den Umweg der übergreifenden Portalangebote auch die Archive und ihre Fachinformationssysteme. Bei der Erneuerung des hessischen Archivinformationssystems¹, das gemeinsam für das Hessische Landesarchiv und das Niedersächsische Landesarchiv entwickelt wird, ist von Anfang an erörtert worden, wie die Idee der Linked Open Data (LOD) in der Praxis realisiert werden kann.

URI für archivische Erschliessungsobjekte

Um Daten miteinander verknüpfen zu können, ist ein eindeutiger Identifikator unverzichtbar. Seit Langem werden Archivalien und Archivbestände, wie es ISAD (G)² beschreibt, mit Signaturen versehen, um sie eindeutig identifizieren zu können. Auch die Archive selbst haben in der Regel eindeutige Kürzel. In den meisten Fällen genügt also die Archivsignatur den Ansprüchen an einen einheitlichen Ressourcenidentifikator (URI – Uniform Resource Identifier), aber eben doch nicht immer. Probleme können sich ergeben bei Korrekturen, Umsignierungen oder bei

Mehrfachverzeichnungen eines Objekts. Daher ist die Verwendung maschinell generierter Identifikatoren der sicherere Weg zur eindeutigen Bezeichnung archivischer Erschliessungsobjekte.

1 Peter Sandner: Von der Findmitteldatenbank zum virtuellen Lesesaal im Netz. Rechtliche Fragen bei der Umgestaltung des Archivinformationssystems HADIS, in: Alles was Recht ist. Archivische Fragen – juristische Antworten. 81. Deutscher Archivtag in Bremen, Fulda 2012, S. 135–145

2 www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html

jekte. Dabei können beispielsweise Datenbank-ID verwendet werden. Sicherlich ist die längerfristige Stabilität solcher ID wünschenswert, noch wichtiger als ewige Geltung aber ist die Eindeutigkeit.

Verwendung von HTTP-Links für archivische Erschliessungsobjekte

Eine Vorschrift zur Verwendung von HTTP-Links für archivische Erschliessungsobjekte findet sich in den Archivgesetzen zwar nicht. Eine solche Forderung steht aber im Einklang mit dem Auftrag des neu gefassten Hessischen Archivgesetzes, Archivgut «unter Anwendung moderner Technologien für die öffentliche Nutzung zugänglich zu machen»³. Schon vergleichsweise früh haben die hessischen Staatsarchive ihre Erschliessungsdaten im Archivinformationssystem HADIS unter www.hadis.hessen.de/recherchierbar gemacht⁴. Allerdings waren dort die standardmäßig angebotenen HTTP-Adressen keine festen, sondern nur temporäre Weblinks, deren Gültigkeit mit Ablauf der Session verlief. Diese Einschränkung führte zu Unmut bei jenen Anwendern, die Links zu HADIS-Objekten setzen wollten. Seit 2006 werden von HADIS auch permanente Links für diverse Portale bereitgestellt. Infolgedessen sind hessische Erschliessungsdaten u.a. aus spartenübergreifenden Angeboten wie dem BAM-Portal oder der Deutschen Digitalen Bibliothek (DDB) heraus verlinkt worden⁵. Von der HADIS-Nachfolgeanwendung werden durchweg feste HTTP-Adressen so angeboten, dass eine Verlinkung zu den Erschliessungsobjekten für jeden leicht möglich ist. Dabei kommt das archivische Credo über die Bedeutung des Kontextes zur Geltung, indem jeder Erschliessungsdatensatz im Zusammenhang mit der Gliederung (Klassifikation, Tektonik) angezeigt wird; ein Navigationsbaum ermöglicht das Überwechseln zu benachbarten Objekten.

Eignung archivischer Erschliessungsdaten als Open Data

Manche hinterfragen, ob archivische Erschliessungsdaten überhaupt als Open Data geeignet seien. Diese Zweifel betreffen zum einen die – verglichen mit Bibliotheksdaten – heterogene Struktur der Archivdaten. Mehr noch

aber führen die hohen Anforderungen des Persönlichkeitsschutzes zu einer Zurückhaltung mancher Archive bei Open Data. Die öffentlichen Archive verstehen sich zu Recht auch als Datenschutzbehörden. Sie haben die gesetzliche Aufgabe, das Freiheitsrecht auf informationelle Selbstbestimmung⁶ auch nach Abgabe der Unterlagen ans Archiv zu gewährleisten. Dazu dienen vor allem die archivgesetzlichen Schutz- oder Sperrfristen. Aber deshalb

Die öffentlichen Archive verstehen sich zu Recht auch als Datenschutzbehörden. Sie haben die gesetzliche Aufgabe, das Freiheitsrecht auf informationelle Selbstbestimmung auch nach Abgabe der Unterlagen ans Archiv zu gewährleisten.

müssen die betreffenden archivischen Erschliessungsdaten keineswegs vollständig unter Verschluss gehalten werden. Vielmehr ist es Aufgabe der Archive, das Recht auf Persönlichkeitsschutz mit dem Grundrecht der Wissenschafts- und Forschungsfreiheit⁷ in Einklang zu bringen. Deshalb hat die bundesdeutsche Archivreferentenkonferenz 2007 eine Empfehlung zum Umgang mit sensiblen Erschliessungsdaten verabschiedet⁸. Dieses Papier arbeitet heraus, dass Erschliessungsdaten durchaus auch dann schon in öffentlich zugänglichen Netzen publiziert werden dürfen, wenn die Akteninhalte selbst noch Schutzfristen unterliegen. Entscheidend ist, dass der Erschließungstext die schutzwürdige Information (etwa einen Personennamen) nicht preisgibt. Archivinformationssysteme sollten diese Unterscheidung – so wie in Hessen seit vielen Jahren der Fall – abbilden und die beiden Merkmale «Zeitpunkt der Onlinepublikation der Erschließung» und «Ende der Schutzfrist» getrennt speichern. Dies sollte auf der Ebene der einzelnen Verzeichnungseinheit geschehen, und nicht pauschal für den ganzen Archivbestand, denn unter Onlinebedingungen entfällt die Notwendigkeit, die Fertigstellung eines kompletten Onlinefindbuchs abzuwarten. Dieses Verfahren erfordert unter den Vorzeichen von Linked Open Data eine erhöhte Wachsamkeit bei der Bearbeitung, denn ein versehentlich veröffentlichter Datensatz lässt sich kaum zurückholen.

Bereitstellung standardisierter Informationen zu den Objekten

Inzwischen hat sich EAD⁹ zum Standardaustauschformat für archivische Informationen entwickelt. In den letzten Jahren ist es im Förderprogramm der Deutschen Forschungsgemeinschaft (DFG) zur Retrokonversion von Findmitteln verwendet worden. Auch die Deutsche Digitale Bibliothek (DDB) nutzt EAD, um Daten von den Archiven in Empfang zu nehmen. Inzwischen haben die deutschen Archive sich über ein einheitliches EAD-Profil verständigt¹⁰. Die DDB reicht die Daten weiter an die Europeana, die diese in das klassische LOD-Format RDF verwandelt¹¹ und so öffentlich bereitstellt.

- 3 § 1 (1) Hessisches Archivgesetz (HArchivG). Vom 26. November 2012, Gesetz- und Verordnungsblatt für das Land Hessen Teil I, S. 458
- 4 Bernward Helfer: Das hessische Erschließungssystem HADIS 2000, in: Angelika Menne-Haritz (Hg.): Online-Findbücher, Suchmaschinen und Portale, Marburg 2002, S. 153–171, auch unter http://archivschule.de/uploads/Publikation/VOE35/Voe35_14HELPER.pdf; Peter Haberkorn: Das Fachinformationssystem HADIS der hessischen Staatsarchive, in: Gerald Maier, Thomas Fritz (Hg.): Archivische Informationssystem in der digitalen Welt. Aktuelle Perspektiven, Stuttgart 2010, S. 181–195
- 5 www.bam-portal.de; www.deutsche-digitale-bibliothek.de
- 6 In Deutschland abgeleitet aus der Menschenwürde und dem Recht auf freie Entfaltung der Persönlichkeit: Artikel 1 (1) bzw. Artikel 2 (1) Grundgesetz
- 7 Artikel 5 (3) Grundgesetz
- 8 www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/ark/20070320_veroeffentlichungsgrundsaeetze_ark.pdf
- 9 EAD = Encoded Archival Description; www.loc.gov/ead
- 10 www.landesarchiv-bw.de/ead; Ulrich Fischer, Sigrid Schieber, Wolfgang Krauth, Christina Wolf: Ein EAD-Profil für Deutschland. EAD(DDB) als Vorschlag für ein gemeinsames Austauschformat deutscher Archive, in: Archivar. Zeitschrift für Archivwesen 2/2012, S. 160–162, auch unter www.archive.nrw.de/archivar/hefte/2012/ausgabe2/ARCHIVAR_02-12_internet.pdf
- 11 <http://pro.europeana.eu/web/guest/in-brief>

Offenheit der Daten darf nicht auf Kosten der Datenqualität gehen. Daher widmete der LOD-Summit 2013 im kanadischen Montreal der Datenpflege («Curation») bei LOD eine eigene Sektion¹². Unter diesem Gesichtspunkt sind auch die Modalitäten der Datenaktualisierung zu klären. Die hessischen Staatsarchive hatten 2006 mit dem BAM-Portal eine tägliche Aktualisierung vereinbart, um mögliche Fehler – insbesondere solche, die den Persönlichkeitsschutz betrafen – ohne Zeitverzug bereinigen zu können. Derzeit werden die Daten bei der DDB noch nach Absprache im Einzelfall aktuali-

Offenheit der Daten darf nicht auf Kosten der Datenqualität gehen. Daher widmete der LOD-Summit 2013 im kanadischen Montreal der Datenpflege («Curation») bei LOD eine eigene Sektion.

siert, aber das ist nur eine Übergangslösung in der Startphase. Nach gegenwärtiger Planung werden die hessischen Archive ihre EAD-Dateien ohne Zugangsbeschränkung im Internet verfügbar machen. Über einen Zeitstempel lässt sich erkennen, wann eine Aktualisierung erfolgt ist. Ob darüber hinaus auch unmittelbar RDF-Daten publiziert werden, ist noch zu klären.

¹² <http://summit2013.lodlam.net/2013/08/18/curation-of-lod/>

¹³ <http://pro.europeana.eu/web/guest/linked-open-data>; <http://creativecommons.org/publicdomain/zero/1.0/deed.de>

¹⁴ <http://summit2013.lodlam.net/2013/08/18/curation-of-lod/>; <http://summit2013.lodlam.net/2013/07/08/notes-from-normalizing-licensing-and-data-models/>

¹⁵ <http://de.wikipedia.org/wiki/Wikipedia:Lizenzbestimmungen>; <http://creativecommons.org/licenses/by-sa/3.0/>

¹⁶ Richtlinie 2013/37/EU des Europäischen Parlaments und des Rates vom 26. Juni 2013, Amtsblatt Nr. L 175 vom 27/06/2013 S. 0001–0008, auch unter <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:01:DE:HTML>

¹⁷ www.lagis-hessen.de/de/subjects/index/sn/hadis

¹⁸ www.regesta-imperii.de; <http://monasterium.net>

Offenheit

Es wird diskutiert, wann Daten wirklich als «offen» zu bezeichnen sind. Die Europeana publiziert Erschließungsdaten nur, wenn sie unter der (Nicht-) Lizenz CCo verwendet werden können¹³. Dagegen wurden beim LOD Summit sachliche und rechtliche Zweifel an dieser umfassenden Lizenzierung artikuliert¹⁴. Ein Verzicht auf die Nennung des Datenurhebers widerspricht dem Ziel der Archive, ihre Archivalien besser auffindbar zu machen. Und bei Weiterverwendung der Daten zu rechtswidrigen Zwecken könnte die Reputation des Archivs leiden. Die deutschsprachige Wikipedia erlaubt eine Nachnutzung ihrer Artikel unter der Creative-Commons-Lizenz BY-SA, die die Nennung des Urhebers fordert und die Weitergabe von Bearbeitungen unter den gleichen Bedingungen zulässt¹⁵. Die möglichen Nachteile einer CCo-Lizenzierung sind abzuwägen gegenüber den Vorteilen, die im positiven Fall einen gesellschaftlichen Mehrwert für alle Beteiligten erzeugen. Sicherlich wird die 2013 neu gefasste EU-Richtlinie über die Weiterverwendung von Informationen des öffentlichen Sektors, die innerhalb von zwei Jahren in nationales Recht umzusetzen ist, die Entwicklung beflügeln: Erstmals werden nun nämlich auch Bibliotheken, Museen und Archive verpflichtet, die freie (wenn auch nicht kostenfreie) Weiterverwendung ihrer bereitgestellten öffentlichen Daten zu ermöglichen¹⁶.

Adressaten der verlinkten Daten

Wer mit einem gewissen Aufwand Links und weitergehende Informationen zu seinen Erschließungsobjekten im Internet bereitstellt, möchte auch, dass die Angebote genutzt werden. Bisher zeigen sich – neben den nationalen und europäischen Archivportalen – vor allem andere Archiv- oder Kulturinstitutionen daran interessiert, Links zu Verzeichnungsobjekten der regionalen Archivportale wie dem in Hessen zu setzen. Dadurch werden Quellen im Zusammenhang präsentiert oder Recherchemöglichkeiten durch Zusatzinformationen optimiert. Die landesgeschichtliche Anwendung LAGIS beispielsweise setzt Links, um die von den hessischen Staatsarchiven online gestellten digitalisierten Perso-

nenstandsregister gezielter recherchierbar zu machen¹⁷. Mehrfach haben Projektförderungen dazu geführt, dass mittelalterliche Urkunden der hessischen Staatsarchive im Verbund online präsentiert worden sind¹⁸.

In der nächsten Ausbaustufe der Datenverknüpfung könnten Archivdaten selbst als historische Forschungsdaten genutzt werden – für die Geschichtsforschung und für andere Disziplinen. Dazu dürften Erschließungsdaten nicht mehr nur als Verweise auf einzelne Akten und Urkunden verstanden werden, sondern als ein eigener Datenkorpus, der weiter gehende Auswertungen zulässt. Beispielsweise könnte eine systematische Auswertung eines grösseren Korpus von Erschließungsdaten Erkenntnisse darüber erbringen, welche Krankheiten im Zeitverlauf diagnostiziert wurden, welche Bodenschätze wann an welchen Orten gewonnen wurden; welche Personennamen über die Jahrhunderte hinweg verwendet wurden usw. Im Moment ist bei derartigen Forschungsthemen ein direkter Kontakt zum Archiv unumgänglich, und er wird auch in Zukunft sicherlich weiterhin angeraten sein, denn die archivischen Erschließungsdaten sind teilweise erklärungsbedürftig, und aufgrund archivischer Bewertung und Kassation haben statistische Auswertungen ihre Grenzen. Dennoch werden die Möglichkeiten der Forschung deutlich erweitert, wenn Daten in einer strukturierten Form frei zugänglich sind.

Open Government

Über die Onlinebereitstellung von Erschließungsdaten hinaus können staatliche und kommunale Archive – wie alle anderen öffentlichen Verwaltungen auch – ihren Beitrag zu einer offenen und transparenten Verwaltungskultur leisten. Im Sinne des Open Government lassen sich statistische Daten zur archivischen Tätigkeit in einer weiterverarbeitbaren Form publizieren. Bei diesem Aspekt der Datenbereitstellung stehen die meisten deutschen Archive noch am Beginn der Überlegungen. Zu denken wäre etwa an die jährliche Anzahl der Zugänge, der neu verzeichneten Archivalieneinheiten oder der Nutzerinnen und Nutzer im Archiv usw. – kurzum: an alle

Daten, die auch heute schon in Jahres- und Tätigkeitsberichten publiziert werden¹⁹, aber bislang, da sie nur in Papier- oder PDF-Dokumenten enthalten sind, nur schwer automatisiert ausgewertet werden können.

Links zu Ressourcen anderer Anbieter

Ein Kernpostulat der Linked-Open-Data-Idee ist die Verlinkung zu Ressourcen anderer Informationsanbieter, um die Daten zu Elementen des Semantic Web zu machen. Hierbei erweist es sich als Problem, dass die archivischen Erschliessungsdaten bei den meisten Archiven bislang nur bedingt strukturiert oder standardisiert sind. Zwar bieten EAD-Elemente eine Zuordnung, jedoch findet dabei oft kei-

ne Entitätenprüfung statt: Wenn im Titelfeld das Wort «Müller» vorkommt, geht häufig nur aus dem Kontext hervor, ob es sich um den Beruf oder um einen

Eine Anreicherung der Erschliessungsdaten um eindeutige Informationen ist dazu notwendig und auch erstrebenswert, jedoch mit den personellen Kapazitäten der Archive allein kaum zu bewältigen.

Personennamen handelt – und im zweiten Fall weiss man nicht, ob es sich um dieselbe Person namens Müller handelt wie in einer anderen Verzeichnung. Daher sind solche Datei-

en nur eingeschränkt für das Semantic Web geeignet. Eine Anreicherung der Erschliessungsdaten um eindeutige Informationen ist dazu notwendig und auch erstrebenswert, jedoch mit den personellen Kapazitäten der Archive allein kaum zu bewältigen. Ziel muss daher eine automatisierte Anreicherung um Normdaten sein, sodass der Aufwand der Archive sich auf eine Restmenge und auf die Qualitätskontrolle beschränkt. Hier sind Verbünde und Kooperationen unerlässlich. Vordringlich erscheint vor allem die Erstellung oder Optimierung von Personen- und Ortsregistern. Erschliessungsdatensätze sollten dazu mit Datensätzen grösserer Anbieter verlinkt werden. Dabei kommen etablierte Dienste wie GeoNames für ein Ortsregister und DBpedia für ein Personenregister infrage, die auch bei der Europeana hierfür verwendet werden²⁰.

Letztlich gilt es gerade für mittlere und kleinere Archivverwaltungen, sich verbreiteten Lösungen anzuschliessen, die im Verbund genutzt und weiter entwickelt werden. Nur dann kann die grosse Vision der Linked Open Data in der Praxis auch mit Leben gefüllt werden.

Kontakt: Poststelle@hhstaw.hessen.de

ABSTRACT

Linked Open Data dans les systèmes d'information professionnels d'archives

La grande idée des données ouvertes et reliées entre elles (Linked Open Data) intéresse également les archives et leurs systèmes d'information professionnels par le biais d'offres de portails transversaux. Dans le cadre de la rénovation du système d'information archivistique du Land de Hesse, développé conjointement pour les archives de Hesse et de Basse-Saxe, la question de savoir comment concrétiser l'idée des Linked Open Data (LOD) s'est posée dès le début. Les Archives du Land de Hesse permettent depuis des années de relier leurs données via des liens HTTP. L'EAD (encoded archival description; description archivistique encodée) a fait ses preuves comme standard d'échange pour les archives. La bibliothèque numérique européenne Europeana publie également des données archivistiques au format RDF. L'entretien («Curation») des open data représente un véritable défi pour le futur. La question porte également sur le degré d'ouverture des données. Jusqu'ici, ce sont surtout des institutions patrimoniales qui utilisent les liens des systèmes d'information archivistiques. Les données archivistiques devraient également être utilisées à l'avenir pour la recherche, et les archives devraient elles-mêmes se considérer comme des administrations transparentes au sens de l'open government.

(traduction: sg)

¹⁹ www.hauptstaatsarchiv.hessen.de/irj/HHStAW_Internet?cid=b95c44612cd6101b5bb41fbf2f40116b

²⁰ <http://pro.europeana.eu/web/guest/in-brief>