

Tags: archives big data, numérisation,

Linked Open Data in archivischen Fachinformationssystemen

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 Archivinformationssystems, 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 Erschliessungs- Objekte

Um Daten miteinander verknüpfen zu können, ist ein eindeutiger Identifikator unverzichtbar. Seit Langem werden Archivalien und Archivbestände, wie es ISAD (G) www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html 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. 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» § 1 (1) Hessisches Archivgesetz (HArchivG). Vom 26. November 2012, Gesetz- und Verordnungsblatt für das Land Hessen Teil I, S. 458. Schon vergleichsweise früh haben die hessischen Staatsarchive ihre Erschließungsdaten im Archivinformationssystem HADIS unter www.hadis.hessen.de recherchierbar gemacht. 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_14HELFER.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. Allerdings waren dort die standardmässig angebotenen HTTP-Adressen keine festen, sondern nur temporäre Weblinks, deren Gültigkeit mit Ablauf der Session verfiel. 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. Infolge dessen sind hessische Erschließungsdaten u.a. aus spartenübergreifenden Angeboten wie dem BAM-Portal oder der Deutschen Digitalen Bibliothek (DDB) herausverlinkt worden¹. 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 In Deutschland abgeleitet aus der Menschenwürde und dem Recht auf freie Entfaltung der Persönlichkeit: Artikel 1 (1) bzw. Artikel 2 (1) Grundgesetz – auch nach Abgabe der Unterlagen ans Archiv zu gewährleisten. Dazu dienen vor allem die archivgesetzlichen Schutz oder Sperrfristen. Aber deshalb 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 EAD² 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 verständigt 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. Die DDB reicht die Daten weiter an die Europeana, die diese in das klassische LOD-Format RDF verwandelt³ und so öffentlich bereitstellt.

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 Sektion⁴. 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 aktualisiert, 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.

Offenheit

Es wird diskutiert, wann Daten wirklich als «offen» zu bezeichnen sind. Die Europeana publiziert Erschließungsdaten nur, wenn sie unter der (Nicht)Lizenz CC0 verwendet werden können⁵. Dagegen wurden beim LOD Summit sachliche und rechtliche Zweifel an dieser umfassenden Lizenzierung artikuliert⁶. Ein Verzicht auf die Nennung des Datenerhebers 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 BYSA, die die Nennung des Urhebers fordert und die Weitergabe von Bearbeitungen unter den gleichen Bedingungen zulässt⁷. Die möglichen Nachteile einer CC0 Lizenzierung sind abzuwägen gegen über 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 ermöglichen. 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>.

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 Archive 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 Personenstandsregister gezielter recherchierbar zu machen⁸. Mehrfach haben Projektförderungen dazu geführt, dass mittelalterliche Urkunden der hessischen Staatsarchive im Verbund online präsentiert worden sind⁹.

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 weitergehende 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 es wird auch in Zukunft sicherlich weiterhin angeraten sein, denn die archivischen Erschliessungsdaten sind teilweise erklärungsbedürftig, und aufgrund archivistischer 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 werden¹⁰, 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 keine 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 Personennamen handelt – und im zweiten Fall weiss man nicht, ob es sich um dieselbe Personnamens Müller handelt wie in einer anderen Verzeichnung. Daher sind solche Dateien 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 Geo-Names für ein Ortsregister und DB-pedia für ein Personenregister infrage, die auch bei der Europeana hierfür verwendet werden³.

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.

1 www.bam-portal.de; www.deutsche-digitale-bibliothek.de.

2 EAD = Encoded Archival Description; www.loc.gov/ead.

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

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

5 <http://pro.europeana.eu/web/guest/linked-open-data>;

<http://creativecommons.org/publicdomain/zero/1.0/deed.de>.

6 [>](http://summit2013.lodlam.net/2013/08/18/)

<http://summit2013.lodlam.net/2013/07/08/notes-from-normalizing-licensing-and-data-models/>.

7 <http://de.wikipedia.org/wiki/Wikipedia:Lizenzbestimmungen>;

<http://creativecommons.org/licenses/by-sa/3.0/>.

8 www.lagis-hessen.de/de/subject/index/sn/hadis.

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

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



Peter Sandner

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

Résumé

Français

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.