

Linked Data im Kulturerbesektor

Linked Data for Cultural Heritage

Regine Stein

Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

Philipps-Universität

Biegenstraße 11, D-35037 Marburg

Tel.: +49 (0) 6421-28 23666, Fax: +49 (0) 6421-28 28931

Email: r.stein@fotomarburg.de, Internet: <http://www.fotomarburg.de>

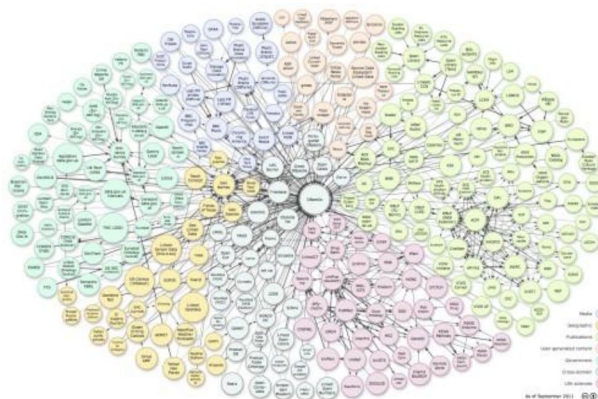
Zusammenfassung:

Das Konzept des Linked Data wird vielerorts als Königsweg zur Vernetzung von Informationen über Kulturgüter diskutiert. Im EU-geförderten Projekt "Linked Heritage" wird unter anderem der Frage nach einer Best Practice von Linked Data mit besonderem Augenmerk auf die relevanten Aktivitäten im Kulturerbesektor nachgegangen: Welcher Stand ist bislang erreicht, welche Aspekte sind besonders zu berücksichtigen?

Abstract:

The Linked Data concept is currently one of the hot topics in the area of knowledge management, being mentioned in many conference talks and papers on the future prospects for access to cultural heritage. Through the EU-funded project „Linked Heritage“ Linked Data best practice is to be explored with special focus on ongoing activities and specific aspects to be considered in the cultural sector.

“The Semantic Web done right!” lautet Tim Berners-Lees Beschreibung des “Linked Data” Konzepts, das die Interoperabilität, Nachnutzbarkeit und Wiederauffindbarkeit der Daten unserer Wissensgesellschaft im unendlichen Speicher des Internets verbessern kann. Die aussagekräftige Vernetzung der im Web verfügbaren Information gewinnt tatsächlich nicht nur angesichts der Fülle der Web-Angebote an Bedeutung, sondern erfährt durch die Aufbereitung als Linked Data eine entscheidende qualitative Anreicherung. Qualifizierte, formalisierte, automatisch auswertbare und damit neu kombinierbare Verknüpfungen nicht nur zwischen ganzen Webangeboten oder einzelnen Seiten, sondern zwischen kleinsten Informationseinheiten – Daten – sollen die Zukunft des Internet prägen, das „Web of Data“.



Linking Open Data cloud diagram as of September 2011
by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Die zugrunde liegenden Techniken können hier nicht im Detail vorgestellt werden, es seien aber die vier Grundregeln von Tim Berners-Lee für Linked Data aufgeführt:

1. Verwende zur Bezeichnung von Objekten URIs.

2. Verwende HTTP URIs, also URLs, so dass sich die Bezeichnungen nachschlagen lassen.
 3. Stelle zweckdienliche Informationen bereit, wenn jemand eine URI nachschlägt (mittels der Standards RDF und SPARQL).
 4. Zu diesen Informationen gehören insbesondere Links auf andere URIs, über die weitere Objekte entdeckt werden können.
- (Tim Berners-Lee, Linked Data – Design Issues, <http://www.w3.org/DesignIssues/LinkedData.html>)

Anstelle einer Vertiefung darf auf folgende Publikationen für die Lektüre zu Linked Data Techniken verwiesen werden:

Einführend:

Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136.

Morgan & Claypool.

Online frei verfügbar unter <http://linkeddatatoolkit.com/book>

Weiterführend:

Leigh Dodds and Ian Davis (2011), Linked Data Patterns, <http://patterns.dataincubator.org/>

(nur online publiziert).

Übertragen auf den Kulturerbesektor, in dem es inzwischen mit staatlich geförderten Projekten wie der Europeana oder der Deutschen Digitalen Bibliothek zur Selbstverständlichkeit geworden ist, Erschließungsinformationen zu den Kulturobjekten in Online-Datenbanken zu publizieren, lesen sich Tim Berners-Lees Grundregeln des Linked Data in etwa so: Man definiere für seine Institution eine Konvention zur eindeutigen Adressierung der Kulturobjekte im Netz und stelle unter dieser Adresse zweckdienliche Information – also die strukturierte Erschließungsinformation – zu diesem Objekt bereit. Wenn in der Erschließungsinformation Referenzen auf andere Entitäten enthalten sind – z.B. Personen- oder Ortsangaben, liefere man die Webadresse für diese Entität mit. Man verwende für die technische Kommunikation definierte Standards. Auf diese Weise kann die Information softwaregestützt ausgewertet und präsentiert werden.

Welche Anwendungen daraus entwickelt werden können, wurde beispielhaft von Europeana im Juni 2011 im Rahmen des Hackathons (zusammengesetzt aus „Hack“ und „Marathon“, eine zeitlich begrenzte Zusammenkunft von Programmierern zur kollaborativen Entwicklung von prototypischen Anwendungen) „Hack4Europe“ ausgetestet, die Ergebnisse können unter <http://www.version1.europeana.eu/web/api/hack4europe> konsultiert werden. So wurde als innovativste Anwendung die „TimeMash“-App ausgezeichnet, die auf Basis der aktuellen geographischen Position eines Nutzers historische Ansichten des selben Ortes in Europeana sucht. Weitere Hackathons können unter <http://www.version1.europeana.eu/web/api/hackathons> verfolgt werden.

Der potentielle Ertrag von als Linked Data publizierten Datenbeständen, insbesondere Normdateien und kontrollierten Vokabularen, liegt aber auch für die eigene Erschließungsarbeit auf der Hand: Wir alle machen für die Informationsgewinnung selbstverständlich ausgiebig Gebrauch von Webressourcen. Stehen diese nicht nur über den Browser recherchierbar zur Verfügung, sondern auch in einer standardisierten, maschinenlesbaren Form, müssen diese Informationen nicht in das eigene Informationssystem übernommen werden, sondern die Quellen können einfach referenziert und fortan für die Recherche und Präsentation genutzt werden. Allerdings ist eine solche Arbeitsweise im Umfeld der Kulturinstitutionen wohl nur denkbar, wenn die verwendeten Ressourcen als vertrauenswürdig eingestuft werden und auch dauerhaft verfügbar sind.

Mit diesen Stichworten – Vertrauenswürdigkeit und Nachhaltigkeit – sei hier übergeleitet zu einer allgemeinen Einschätzung zum Stand der Aktivitäten im Oktober 2011 im Kulturerbesektor. Diese resultiert aus Untersuchungen, die im Rahmen des EU-Projekts „Linked Heritage – Coordination of Standards and Technologies for the enrichment of Europeana“ (<http://www.linkedheritage.eu>) im Arbeitspaket „Linking Cultural Heritage Information“ durchgeführt wurden. In der ersten Arbeitsphase – das Projekt startete erst im April 2011 – wurde eine Bestandsaufnahme zur Linked Data Praxis durchgeführt, deren Ergebnis sich in folgenden Punkten zusammenfassen lässt:

- Viele Kultureinrichtungen, seien es Bibliotheken, Archive, Museen usw., haben die Bereitstellung und Nutzung von Linked Data bereits erwogen und sondiert.
- Es gibt jedoch noch keine konsolidierten Wege und Standardlösungen zur praktischen Umsetzung, es kann bislang eher von Experimenten gesprochen werden.
- Verschiedene Projekte wurden zwar bereits durchgeführt, jedoch haben sie bislang tendenziell zur Publikation von „Datensilos“ geführt: Die nach dem Linked Data Prinzip publizierten Daten sind in sich geschlossen und enthalten oft nur wenige oder gar keine Referenzen auf externe Quellen – zuweilen wird diese Situation als „Unlinked Data“ bezeichnet.

Darüber hinaus ist zu thematisieren, dass mit dem Anspruch der Vertrauenswürdigkeit und Nachhaltigkeit einhergeht, dass klare Aussagen zu den Nutzungsrechten der Daten getroffen werden. Während Europeana mit seinem kürzlich verabschiedeten, bis Sommer 2012 zu implementierenden „Data Exchange Agreement“ (s. <http://www.version1.europeana.eu/web/europeana-project/newagreement>) die gemeinfreie Nutzung aller in Europeana publizierten Metadaten ermöglicht, ist festzustellen, dass in der sogenannten Linked Open Data Cloud diesbezüglich eine erhebliche Lücke klafft. Das Grundprinzip des Open Data, das inzwischen in der Regel in einem Zug mit dem Prinzip des Linked Data benannt wird, ist allzu oft verletzt: Eine Analyse der Datenbestände im zugehörigen Verzeichnis unter <http://thedatahub.org> hat ergeben, dass für 55% der Datenpakete keine offene Lizenz ausgewiesen wird, so dass diese Pakete nicht seriös genutzt werden können.

Die Implementierung des Linked Data Prinzips auf Basis von qualifizierten Referenzen verspricht also zwar gerade im Kulturerbesektor nicht nur einen greifbaren Mehrwert durch neuartige Anwendungsszenarien, sondern auch für die eigenen Erschließungszusammenhänge. Die Ausbuchstabilisierung in die Praxis steht dabei aber noch weitgehend aus – die weiteren Entwicklungen im Umfeld der Europeana, die hier aktuell eine klare Vorreiterrolle einnimmt, sind in der nahen Zukunft mit besonderem Interesse zu verfolgen.