



The importance of open data to national libraries

Elisabeth Niggemann

Deutsche Nationalbibliothek
(German National Library)
Frankfurt am Main, Germany

Meeting:

181 — *National libraries and open data: new discovery and access services* — National Libraries

Abstract:

The German National Library (Deutsche Nationalbibliothek, DNB) publishes most of its bibliographic data as open data under a Creative Commons Zero License (CC-0), following the examples of the European cultural heritage portal Europeana and many national and other libraries, among them the British Library and the Spanish National Library. These cultural strongly believe that serving the public today means opening up as many services, holdings and data as (legally, technically, organizationally ...) possible, making them not only accessible via the World Wide Web but making their data an integral part of the World Wide Web! A very good example of a cooperative use of open data is Europeana and the paper will present the strategy and the process that led to the launch of Europeana's new Data Exchange Agreement.

An obvious next step is to publish metadata not only as open data but as Linked Open Data (LOD). By linking different representations of related concepts in the Semantic Web, libraries can draw from that Web and enrich their own metadata and benefit locally. What is even more important is that they are able to really and seriously reach out to many more users in typical internet "places" like social networks, portals, and search engines. This will not only help them fulfill their public mission but also increase traffic back to the library's site, helping them to monitor and prove their relevance in today's internet based information universe. By publishing open data as LOD and as significant knots in the Semantic Web, libraries will create benefits for themselves and, most important, for the internet users of today and tomorrow.

1. Introduction

On the 1st of July 2012, the German National Library (Deutsche Nationalbibliothek, DNB) published most of its bibliographic data as open data under a Creative Commons Zero License (CC-0).¹ In doing so the DNB followed the examples of the European cultural heritage portal Europeana and many national and other libraries, among them the British Library and the Spanish National Library, to name only two early movers. The DNB did so because we, the librarians of a publicly funded library, existing to serve the public, strongly believe that serving the public today means opening up as many services, holdings and data as (legally, technically, organizationally ...) possible, making them not only accessible via the World Wide Web but making our data an integral part of the World Wide Web! When asked by the organizers of this session to speak about "the importance of open data to national libraries", I therefore happily agreed. The aim of my presentation today is to convert those of you, who are not yet publishing your metadata as open data, into believers in open data, understanding the importance of open data to national (and other) libraries.

The term "data" can have many meanings and some combinations of "open" and "data" will lead to quite controversial discussion within our profession but even more so in a more general public. I will only talk about those data national libraries create themselves, our metadata. I will not talk about all the other types of data we curate, being it scientific data, governmental data or digital content – to name only the most obvious. But even so, open metadata is not a field without controversies. While I will on the one hand restrict the broad field of "data" to mean only "metadata", I will on the other hand use the term "open" in a very broad sense, meaning "without any restrictions", following in this respect the definition of Wikipedia²: "Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control." I would like to place emphasis on the essential technical implications of "being freely available" and "without any restrictions". It is not enough to set our data "free" in a legal or a financial sense, we also have to make them easily discoverable, reusable, sharable in a very concrete, technical, internet-based sense.

¹ Exceptions are the catalog data for the recent two years. All authority data are open data as are the bibliographic data for digital publications. Business model of the Deutsche Nationalbibliothek: <http://www.dnb.de/EN/Service/DigitaleDienste/Geschaeftsmodell/neuesGeschaeftsmodell.html>; Creative Commons Zero Licence: <http://wiki.creativecommons.org/CC0>

² http://en.wikipedia.org/wiki/Open_data; as of 11th June 2012

2. The openness of catalogs

The history of catalogs is probably as old as the history of libraries, and librarians have always been proud of their catalogs. As long as the use of a library's holdings was confined to the library's physical space, the catalog was where the holdings were. With easier, faster and more comfortable ways of traveling for the scholar – or the book – there was a need for published catalogs. Scholars were able to plan their journeys to use the books or they ordered the books they needed to come to their study from far away libraries. The first electronic catalogs only served the local users in local networks. Very soon CD-ROM copies were used for distribution to a geographically diverse public. But still, for another decade or two, libraries and a few specialized publishers held the monopoly for library catalogs. In retrospect I find it amazing how much of this development happened in a comparatively short time frame: When I was at school, libraries still had card catalogs that served both as internal, administrative tools and also as self-service reference tools for their readers. When I was at university, I came across bound volumes of famous libraries' catalogs, microfiche catalogs of single libraries or microfiche union catalogs of library consortia, and huge paper stacks that were the print-out catalogs of the first computer library systems. At library school, I learned of and later worked on OPACs, Online Public Access Catalogs. I welcomed and admired projects that fed these new catalogs with the data from the old card catalogs. All these catalogs I came across had one thing in common: they were open to the public, they were free to use for all the readers in the particular library, for the customers of the published volumes, the microfiches or the users of a certain network. There were restrictions, but these restrictions had to do with the physical nature of the catalog or the technological limitations of early data networks.

Today, technological restrictions, if there are any, are either intentional or a clear symptom of failure. Catalog data and library content are no longer a monopoly held by librarians and publishers; they are ubiquitous. There are many ways to find typical library material via search engines, retail systems like Amazon, social media applications and many more. Library catalogs are accessible in internet portals, big and small, commercial and non-commercial, free of charge and licensed data – offers are abundant. On the users' side, scholars are no longer the only heavy users of metadata. Many former bookshop customers order their books electronically or they use e-books in the first place. Music and movies are found and downloaded from commercial platforms. Readers and film or music lovers recommend the works they like to their friends in social networks or on distributors' platforms. They all need, use, create or share metadata that are freely available and floating around. Where are the libraries in this web scale

metadata business? To put it mildly: Internet users have to know where to find them. They do not find them where they regularly spend their time, in social media, on search engines or whatever is their latest virtual hangout.

Many librarians know that they are in competition with these new arrivals on the metadata scene and they realize that their pride of place, their leading the way or even their survival is at stake. Some hope, because they believe that library data is of better quality, that they will in the end win the competition and supply the readers, the music and film lovers, the scientists and scholars with high quality data. But hope is not enough. What can librarians do to win the competition? How can they bring users and high quality data together? Being in a competition also means having to monitor and count the scores. Companies that sell products, books, films, music, advertising space etc. can use their sales revenue as success rate. Libraries also have to prove their success. Some national libraries (among them my own library, the German National Library) have been selling their data for reuse. Giving up a reliable source of income is not easy psychologically speaking and even harder when funding or administrative bodies have to be convinced. In the case of the DNB, the decision to stop selling metadata, but to do it gradually, over a period of 5 years, was made in the end, because, after a long discussion, the conviction prevailed, that this market option would vanish anyway during that time. We now share the dilemma most cultural institutions find themselves in: How can we share metadata widely and still not lose track of the reuse?

Probably because of this dilemma, librarians have been reluctant to open up their catalog data completely. Sharing data is often still confined to cataloging network partners. As we all know, librarians, archivists, museum directors and curators are still discussing the best and most adequate licensing scheme for sharing metadata, and the discussion around the new Data Exchange Agreement (DEA) of Europeana is a good example for that kind of controversy. On the one hand, there are the pioneering cultural institutions, among them libraries like the British Library but also museums like the Amsterdam Museum, that now publish their data as open data. Driven by their rapidly changing and competitive environment, they offer them as linked open data for potential use in an emerging Semantic Web, or they cooperate with the big global search engines to improve the discoverability of their holdings via the content's metadata. On the other hand, there are colleagues I met in discussions especially with archivists and museum curators, who want to make sure, that their high quality data are not reused in an inadequate setting, in less than optimal form, mutilated, corrupted or reused by someone they do not want

to support. And they are quite right in that respect: If they really want to make sure that nothing of this kind happens, they have to keep their metadata strictly within tight business models. Once you free your data as open data you give up control entirely and you cannot regain it.

As an aside, I have to mention copyright in this context. In the early days of discussing open data, copyright was often mentioned as a barrier. I am not a copyright specialist but at least as far as I know, copyright is not an issue in most countries. It is definitely not a barrier according to German copyright law, where individual metadata are not protected³, nor are they protected by patents. So, to sum up: If librarians, archivists or museum curators do not open up their catalogs for any kind of reuse, they do so because they want to claim control over their metadata. They rarely do so because of financial issues. As far as I know, they could do it and still be compliant with their copyright law. There might be contractual issues with third parties, but contracts can be changed. Contracts are the result of a decision making process and therefore I would like to repeat my statement: Since the internet revolution and in the era of Google, Facebook, Amazon & Co, technological or physical restriction for metadata belong to the past or are the result of a conscious decision.

It is in this context that I want to show what can be accomplished with open data. A very good example of a cooperative use of open data is Europeana and I will talk about the strategy and the process that led to the launch of Europeana's new Data Exchange Agreement, offering its metadata for open reuse under a Creative Commons Zero license (CC-0). And I am proud to say that the Conference of European National Librarians (CENL) made a very early decision to supply Europeana with CC-0 metadata via their aggregator portal The European Library.

3. Open data and Europeana

Europeana is the portal to Europe's digital cultural heritage. It gives quick and easy access to over 23 million objects from more than 2.200 institutions from 33 countries (as of May 2012). The content on Europeana is cross domain, it is supplied from museums, archives, audiovisual archives and libraries. Europeana's Data Providers and Aggregators deliver high quality metadata that has been created by experts, usually to international

³ Only the catalogue database system can be protected.

standards, but in every domain, in every European country and language, with variant spellings and naming. Europeana works on the metadata and brings clarity to the users, helping them to access trusted, authentic cultural objects on their original sites without tedious individual searches, without getting frustrated because of ambiguity, duplication and uncertainty. Without this normalization of metadata and this aggregation of content, users would have to go and search every single cultural institution. Needless to say, this does not happen in real life, which, in consequence, already shows that Europeana is a valid service just by aggregating metadata linked to distributed content.

Europeana is an important resource not only for researchers but also to many users who are e.g. tourists, pupils or teachers – what you might call the average citizen. As a principle, Europeana holds only metadata that describe and link to digital content plus a thumbnail, if available. The actual digital content remains on the site of the metadata provider. When users search Europeana, the results link them back to the digital content on the original metadata provider's website. The metadata provider remains in total control of the content. The institution decides which content can be shown or shared in which way. This is not the case for the metadata: Since 2012 Europeana asks the content provider to share the metadata under a CC-0 license, because the Europeana Foundation, the governing body for Europeana, decided to be more than an aggregator and a search engine for Europe's cultural institutions. In its Strategic Plan 2011-2015⁴, Europeana outlined the approach it will take over the next years. The focus lies in four strategic tracks: Europeana will

- "Aggregate content to build the open, trusted source of European heritage;
- Facilitate knowledge transfer, innovation and advocacy in the cultural heritage sector;
- Distribute their heritage to users wherever they are, whenever they want it;
- Engage users in new ways of participating in their cultural heritage."

With the exception of the aggregator role, the strategic roles rely more or less on Europeana's ability to find new ways to open up the reuse of the metadata, provided by its Data Providers and Aggregators. The first Europeana Data Agreement in 2009 gave Europeana only the right to reuse the metadata in non-commercial environments. Since

⁴ http://pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602

then it became clear that the non-commercial clause makes it impossible for Europeana to reach the new objectives for the term 2011 to 2015. As Jill Cousins, Director of Europeana, points out in her presentation⁵ on this topic, under the first Europeana Data Agreement, Europeana metadata

- “cannot be distributed via APIs in order to put information into the user’s workflow, if the partner sites demonstrate some commercial activities,
- cannot be published as Open Linked Data (LOD), making full use of the semantic potential of the web,
- cannot be shared with Wikipedia, as all information posted there needs to be available also for commercial re-use,
- cannot be used by commercial companies that could help generate income for the content providers,
- cannot be used for the development of apps by commercial companies including for educational purposes.”

Europeana, its partners and many experts spent a lot of time and effort considering various licensing schemes. Europeana listened to arguments, answered questions, looked for solutions and helped cultural institutions that had to discuss the issue with their funding bodies, boards or stakeholders. In the end it became obvious, that only the Creative Commons Zero license provided the necessary freedom for the metadata to be used in all these cases.⁶ In 2011, the Europeana Foundation officially adopted a new Data Exchange Agreement (DEA). Under the DEA, Europeana is authorized by the data providers to publish all metadata under the terms of CC-0 1.0 Universal Public Domain Dedication, licensing them non-exclusively, unconditionally, free of charge for all types of use and for all territories to the public. The question of Intellectual Property Rights in metadata has been extensively dealt with and mentioned in detail in the DEA.⁷ There are Usage Guidelines that are non-binding, but they set out the responsibilities that people who want to reuse the data should be aware of. The DEA is part of the Europeana Licensing Framework. It was initiated during the Europeana Connect project. The University of Amsterdam Institute for Information Law, the National Library of Luxembourg and Nederland Kennisland worked on it together with the Europeana

⁵ Jill Cousin’s presentation on <http://pro.europeana.eu/web/guest/support-for-open-data>, there Strategic outlook

⁶ The process and the arguments are documented on the Europeana website: <http://pro.europeana.eu/consultation-process/>; <http://pro.europeana.eu/web/guest/support-for-open-data>; <http://pro.europeana.eu/consultation-process/>

⁷ For Germany, for example, there is an expert analysis on “The validity of the Creative Commons Zero 1.0 Universal Public Domain Dedication with regards to the German Copyright Law”: on <http://pro.europeana.eu/web/guest/support-for-open-data>, there Strategic outlook, An expert analysis.

Foundation. The Europeana Licensing Framework describes all the contractual elements that underpin the relationships between Europeana, its providers and its users.

Europeana has set out fundamental principles for its dealings with its partners' metadata:⁸

- "Europeana is committed to consultation with the network of data providers.
- Europeana does not intend to make direct commercial use of providers' metadata.
- The contribution of data to Europeana does not prevent you from selling metadata to a third party.
- Data Providers are not required to provide Europeana with complete metadata for digital objects.
- Providing metadata relating to some works in your collections does not create the obligation to provide metadata about complete collections.
- Thumbnails and previews will only be used by Europeana unless explicitly specified they cannot be reused by third parties."

There is an increased understanding that information that has been produced using taxpayers' money should be available for reuse by citizens for non-commercial and commercial purposes (PSI Directive, EU 2003)⁹. The new Data Exchange Agreement is in line with the role the European Council of Ministers referred to. Europeana was launched to encourage "new online services to emerge ... to democratize access and to develop the information society and knowledge-based economy."¹⁰ This position was reinforced by the Comité des Sages' report in January 2011, *The New Renaissance*,¹¹ which noted that: "In some cases cultural institutions charge for or impose other conditions for the reuse of metadata and they tend to be particularly wary of the commercial use of the data. This commercial use is broadly defined and includes the indexing by commercial search engines ... Metadata related to the digitized objects produced by the cultural institutions should be widely and freely available for reuse."

⁸ <http://pro.europeana.eu/web/guest/support-for-open-data>

⁹ PSI Directive EU 2003: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>

¹⁰ European Council of Ministers on the launch of the Europeana prototype, Brussels, 20 November 2008, quoted from "Europeana strategic plan 2011-2015"

¹¹ http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf

The DEA came into force on January 1, 2012, with a grace period of 6 months for those data providers that needed to resolve any issues or obtain permissions. By mid May 2012, the DEA had been signed for more than 17 million objects against a total of more than 23 million objects. By the end of June Europeana expects to reach 21 million, maybe even 22 million objects. By mid May, only two providers had asked for their content to be removed. CENL, the Conference of European National Librarians, was among the first who signed the new DEA, releasing all relevant metadata aggregated in their portal The European Library to Europeana under CC-0 and giving direct evidence that Europe's national librarians see the importance and the benefits of open data! And if you rephrase the "cannots" in Jill Cousins presentation into statements of what Europeana can do from now on under the new Data Exchange Agreement, you have a great list of examples of these benefits, showing how powerful open data can be.

4. Linked Open Data (LOD)

A presentation about open data needs to address Linked Open Data (LOD). LOD created by librarians and other information professionals could well be a reliable, trusted, consistent backbone for the Semantic Web, the structured "web of data", and connecting unstructured documents in the World Wide Web in a meaningful, machine-readable way. Admittedly, all in all only a limited amount of resources from the cultural heritage sector is published as LOD, there is still a lot of work to be done, but I think it is worth trying. The German National Library published authority file data from the German combined authority file (Gemeinsame Normdatei, GND) as LOD, GND being the authority file that unites the authority file datasets for names of persons, corporate bodies, geographic entities and subject index terms for concepts, time, etc from German libraries and their cataloguing networks. DNB is currently working on pilot applications and projects. But the best example for our LOD activities at present certainly is VIAF, the Virtual International Authority File.

Open Data is a first step towards and a requirement for Linked Open Data (LOD). And LOD is a first step towards and a requirement for the creation of the Semantic Web. Although Tim Berners-Lee's vision of a "web of data that can be processed directly and indirectly by machines", "remains largely unrealized"¹², there is much "to argue for a library data ecosystem built on linked data principles, published in the World Wide Web

¹² Quoted from Wikipedia's article on the Semantic Web: http://en.wikipedia.org/wiki/Semantic_Web as of 11th June 2012

under an open license.”¹³ There is a chance, that some of the often quoted challenges for the Semantic Web, “vastness, vagueness, uncertainty, inconsistency and deceit”, can be overcome at least for data from cultural institutions. But this is not a trivial task. Even if data from cultural institutions in the first place are not vague, uncertain or inconsistent, they easily tend to become so, once data from different geographies and different cultural sectors (libraries, archives, museums, audiovisual archives) with their different cataloguing rules and traditions, data formats, languages, scripts etc, get mixed. For example, many national libraries maintain controlled vocabulary lists of names for people, corporate bodies, conferences, geographic entities, works and other entities. These controlled vocabulary lists, or authority files, have been developed and maintained individually in many places around the world. Their differences become visible when data from many libraries are merged in shared catalogs and portals. LOD technology can be a way to improve the quality of the data, to remove much of the vagueness, uncertainty and inconsistency.

However, it is true that the LOD field is still under development, that it is still early days for a robust data model and technical infrastructure. But there is a certain sense of a great potential, there are many pilots showing what can be done, and there are some first movers. Again, Europeana is one example for cultural heritage institutions experimenting and participating in innovations, facilitating discovery for its users and counterbalancing less authoritative sources. Europeana experimented with a Linked Open Data pilot, which has been working with 3.1 million metadata records from 20 partners including the Austrian National Library, the Danish Film Institute and the Swedish National Heritage Board.¹⁴

Linked data is about the relationship of people, things, places, concepts, ideas etc. Authors, artists and other creators of intellectual and artistic works express their knowledge or feelings or beliefs about them in a way so they can share them with other people. In the context of cultural institutions, these expressions or works are collected, and the metadata describing the works are published. Individual metadata for an individual work are hard to find in those huge aggregations of metadata that the users find in the internet, if they do not know about the work before they start looking for it.

¹³ The Deutsche Nationalbibliografie as linked open data: Applications and opportunities. Jürgen Kett, Sarah Beyer, Mathias Manecke, Yvonne Jahns and Lars G. Svensson. Urn:nbn:de:101-2012052306, Presentation at IFLA 2012, section 215, Thursday 16.08.2012 (afternoon)

¹⁴ <http://pro.europeana.eu/web/guest/linked-open-data>

For a search that looks for the unknown work, a meaningful clustering of related metadata is necessary. What needs to be done is to connect or technologically speaking to link metadata that refer to the same topic or theme of the work, be it a person, a thing, etc. in a meaningful way, i.e. which can first be understood by machines and second be presented to human users. Linking to related data, the individual data then become part of a cluster that is like a knot in an intricate web. Linking individual data in the Semantic Web, i.e. in a web of meaningful relationships, makes the sum of the metadata, this huge amount of data all the cultural institutions have created and curate, fit for navigation and for discovery of the content they represent, for the users and their discovery tools, be they PCs, mobile devices or whatever comes next. How the user can then access the content is another important question, and libraries and other cultural institutions will have to find ways to make that part of the discovery easier, smoother, and more digital. But this is, as I mentioned at the beginning of my presentation, not my theme today. I mention it here only because we have to always bear it in mind and to make the necessary technological precautions so that the Semantic Web is not about metadata per se but a means to help users find the works, the content they need.

In the past, libraries developed elaborate ways for data exchange and cooperative cataloguing. Libraries and other cultural institutions often use specific interfaces and formats for their data. This makes data exchange and data aggregation very efficient within a defined community but makes it very difficult to work together outside that specific world. Working together with the commercial world, or across different cultural sectors, countries, languages, systems, is not part of our tradition. It can also be seen as a proof for the thesis that libraries and many other institutions and companies are not yet integral parts of the internet. To convert metadata to LOD is a way of making them an important part of the internet, making them usable by machines – machines of partner institutions and of unknown entities at the same time.

One interesting and successful example for a LOD application is the Virtual International Authority File (VIAF). VIAF was set up jointly by the Library of Congress, OCLC, the German National Library and the Bibliothèque nationale de France, in cooperation with an expanding number of other national libraries and other agencies. It is exactly what the name suggests, a virtual authority file in the sense that it explores the possibility of virtually combining the name authority files of participating institutions into a single name authority service. As of April 2012, it is made of 25 individual authority files,

supplied by 20 partners. There are 14.5 million clusters, generated by matching 18 million authority file entries, used in 80 million catalog records.

VIAF's goal is to make library authority files less expensive to maintain and more useful to libraries and their users. VIAF includes authoritative names from the partner libraries into a global Web-service, linking different names for the same person or organization. Linking the names means that national and regional variations, variations in preferred language, script and spelling in authorized form are allowed to coexist in a cluster of related records. For example, German users will be able to find a name displayed in the form established by the Deutsche Nationalbibliothek, while French users will see the same name as established by the Bibliothèque nationale de France, and English-speaking users will find the name as established by the Library of Congress NACO file. But users are also able to view all the different name records as established by the others, making the authorities truly international and facilitating research across languages anywhere in the world. VIAF provides a great platform for a wide community of libraries and other agencies to reuse bibliographic data produced by libraries serving different language communities. The VIAF dataset is currently available under the ODC-BY license¹⁵. It has a huge potential to play a role in the emerging Semantic Web, but to do that it should be released under a CC-0 license.

Right now VIAF is only clustering names of persons. Names for other entities like corporate names and geographical names are logical next steps. The most complex task then will be the clustering of subjects, themes or concepts. This is not such a straight forward affair as clustering entities. Themes tend to be fuzzy, vague; there can be endless discussions about slight differences of meaning; cultural differences come into play. A lot of work lies still ahead of us to link classification schemes, thesauri and subject heading lists to integrate them into VIAF. With LOD or Semantic Web technology this should be possible from the point of view of formats, interfaces and other technological issues. What remains as a challenge is the semantic issue. Linking concepts, ideas, themes, periods etc from different cultural backgrounds is not only time consuming but also a diplomatic nightmare. I nevertheless believe that it can be done!

¹⁵ <http://opendatacommons.org/licenses/by/>

Apart from cooperating within VIAF, the German National Library is in the process of developing a LOD system to share name authority data with the German Film Institute (Deutsches Filminstitut, DIF). DNB has books about movie actors, directors, composers etc as well as film-music recordings in its collections. The film institute owns copies of movies, posters, newspaper cuttings and other special collections. Both institutions maintain their name authority files. There are 4 million names in the combined authority file GND, hosted by the German National Library. Of these, 2 million personal name sets are individualized. There are 170.000 names in the German Film Institute's authority file. Both files show a big overlap. The two institutions want to reuse what is in each other's files, also on an ongoing basis. The way this was done in the past was to put up mirrored installations of the files in both institutions, to harmonize formats and rules so that merging data relevant to persons represented in both files becomes possible and to install a synchronization program to keep both systems updated. Because cataloguing rules differ, as do the formats and the database systems, this is not an easy task. The two institutions therefore decided to publish their data as LOD in RDF format, to match the datasets referring to the same person and to link them. The benefit lies not only in reduced costs when compared to the traditional method but also in the possible expansion of the system, should other institutions want to join the cooperation.

In another project, bibliographic data of one of the German union catalog systems and those of the German National Library are published as LOD and then get clustered (in the sense of FRBR clusters). All records describing the same entity, i.e. belonging to the same cluster, get a unique common persistent identifier for future use and for citation.¹⁶

5. Summing up

Stability, reliability and consistency are the important features of LOD and the web's clusters or knots. The Semantic Web is not only the web of facts and entities, but also the web of concepts and ideas. Librarians invented classification schemes and thesauri. If they are willing to publish their metadata as LOD, they can combine their traditional world of subject catalogs with the world of internet-based access to information, creating a trustworthy, reliable backbone of the Semantic Web. Cultural heritage institutions should take a lead in LOD, publish their data as LOD and make the vision of a Semantic Web happen – for the benefits of their users. On the other hand, cultural institutions will be able to enrich and upgrade their metadata by using the newly discovered information

¹⁶ Culturegraph.org

to describe their holdings. There will be a free flow of information on geo-locations, information coming from authority files, multilingual information and more, that can be used also locally by all participant providers and their users because LOD works both ways!

The importance of open data to national libraries lies in those two benefits: First of all, by publishing metadata not only as open data but as LOD and by linking different representations of related concepts in the Semantic Web, libraries can draw from that Web and enrich their own metadata and benefit locally. Combined with up-to-date search technologies like full text search and relevance ranking or by opening up their catalogs for tagging by external volunteers, libraries can improve their own institutional offer for their patrons' use. But the second benefit is even more important than this local advantage: libraries are able to really and seriously reach out to many more users in typical internet "places" like social networks, portals, and search engines. This will not only help them fulfill their public mission but also increase traffic back to the library's site, helping them to monitor and prove their relevance in today's internet based information universe. By publishing open data as LOD and as significant knots in the Semantic Web, libraries will create benefits for themselves and, most important, for the internet users of today and tomorrow.