

Library Linked Data Incubator Group Final Report

W3C Incubator Group Report 25 October 2011

This Version:

<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>

Latest Published Version:

<http://www.w3.org/2005/Incubator/lld/XGR-lld/>

Authors

Thomas Baker, Dublin Core Metadata Initiative, US (W3C Invited Expert)
Emmanuelle Bermès, Centre Pompidou, France (W3C Invited Expert)
Karen Coyle, Consultant, US (W3C Invited Expert)
Gordon Dunsire, Consultant, UK (W3C Invited Expert)
Antoine Isaac, Europeana and Vrije Universiteit Amsterdam, Netherlands
Peter Murray, LYRASIS, US (W3C Invited Expert)
Michael Panzer, OCLC Online Computer Library Center, Inc., US
Jodi Schneider, DERI Galway at the National University of Ireland, Galway, Ireland
Ross Singer, Talis Group Ltd, UK
Ed Summers, Library of Congress, US
William Waites, University of Edinburgh (School of Informatics), UK
Jeff Young, OCLC Online Computer Library Center, Inc., US
Marcia Zeng, Kent State University, US (W3C Invited Expert)

See also [translations](#).

[Copyright](#) © 2011 W3C[®] (MIT, ERCIM, Keio), All Rights Reserved. W3C [liability](#), [trademark](#) and [document use](#) rules apply.

Abstract

The [mission](#) of the [W3C Library Linked Data Incubator Group](#), chartered from May 2010 through August 2011, has been "to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities — focusing on Linked Data — in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future." In Linked Data [\[LINKEDDATA\]](#), data is expressed using standards such as [Resource Description Framework](#) (RDF) [\[RDF\]](#), which specifies relationships between things, and [Uniform Resource Identifiers](#) (URIs, or "Web addresses") [\[URI\]](#). This final report of the Incubator

Group examines how Semantic Web standards and Linked Data principles can be used to make the valuable information assets that library create and curate — resources such as bibliographic data, authorities, and concept schemes — more visible and re-usable outside of their original library context on the wider Web.

The Incubator Group began by eliciting reports on relevant activities from parties ranging from small, independent projects to national library initiatives (see the separate report, [Library Linked Data Incubator Group: Use Cases](#)) [USECASE]. These use cases provided the starting point for the work summarized in the report: an analysis of the benefits of library Linked Data, a discussion of current issues with regard to traditional library data, existing library Linked Data initiatives, and legal rights over library data; and recommendations for next steps. The report also summarizes the results of a survey of current Linked Data technologies and an inventory of library Linked Data resources available today (see also the more detailed report, [Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets](#)) [VOCABDATASET].

Key recommendations of the report are:

- That **library leaders** identify sets of data as possible candidates for early exposure as Linked Data and foster a discussion about Open Data and rights;
- That **library standards bodies** increase library participation in Semantic Web standardization, develop library data standards that are compatible with Linked Data, and disseminate best-practice design patterns tailored to library Linked Data;
- That **data and systems designers** design enhanced user services based on Linked Data capabilities, create URIs for the items in library datasets, develop policies for managing RDF vocabularies and their URIs, and express library data by re-using or mapping to existing Linked Data vocabularies;
- That **librarians and archivists** preserve Linked Data element sets and value vocabularies and apply library experience in curation and long-term preservation to Linked Data datasets.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of [Final Incubator Group Reports](#) is available. See also the [W3C technical reports index](http://www.w3.org/TR/) at <http://www.w3.org/TR/>.

This document was developed by the [Library Linked Data Incubator Group](#).

Publication of this document by W3C as part of the [W3C Incubator Activity](#) indicates no endorsement of its content by W3C, nor that W3C has, is, or will be allocating any resources to the issues addressed by it. Participation in Incubator Groups and publication of Incubator Group Reports at the W3C site are benefits of [W3C Membership](#).

Incubator Groups have as a goal to produce work that can be implemented on a Royalty Free basis, as defined in the [W3C Patent Policy](#). Participants in this Incubator Group have agreed to offer licenses according to the licensing requirements of the W3C Patent Policy for portions of this Incubator Group Report that are subsequently incorporated in a W3C Recommendation.

Discussion on this document is welcome on the public mailing list public-ld@w3.org ([archive](#)).

Table of Contents

- [1 Scope of this report](#)
- [2 Benefits of the Linked Data Approach](#)
 - [2.1 Benefits to researchers, students, and patrons](#)
 - [2.2 Benefits to organizations](#)
 - [2.3 Benefits to librarians, archivists, and curators](#)
 - [2.4 Benefits to developers and vendors](#)
- [3 The Current Situation](#)
 - [3.1 Issues with traditional library data](#)
 - [3.1.1 Library data is not integrated with Web resources](#)
 - [3.1.2 Library standards are designed only for the library community](#)
 - [3.1.3 Library data is expressed primarily in natural-language text](#)
 - [3.1.4 The library community and Semantic Web community have different terminology for similar metadata concepts](#)
 - [3.1.5 Library technology changes depend on vendor systems development](#)
 - [3.2 Library Linked Data available today](#)
 - [3.2.1 Fewer bibliographic datasets have been published as Linked Data than value vocabularies and element sets](#)
 - [3.2.2 The quality of and support for available data varies greatly](#)
 - [3.2.3 Linking across datasets has begun but requires further effort and coordination](#)
 - [3.3 Rights issues](#)
 - [3.3.1 Rights ownership is complex](#)
 - [3.3.2 Data rights may be considered business assets](#)
- [4 Recommendations](#)
 - [4.1 For library leadership](#)
 - [4.1.1 Identify sets of data as possible candidates for early exposure as Linked Data](#)
 - [4.1.2 Foster a discussion about Open Data and rights](#)
 - [4.2 For standards bodies and participants](#)
 - [4.2.1 Increase library participation in Semantic Web standardization](#)
 - [4.2.2 Develop library data standards that are compatible with Linked Data](#)
 - [4.2.3 Develop and disseminate best-practice design patterns tailored to library Linked Data](#)
 - [4.3 For data and systems designers](#)

- [4.3.1 Design and test user services based on Linked Data capabilities](#)
 - [4.3.2 Create URIs for the items in library datasets](#)
 - [4.3.3 Develop policies for managing Linked Data vocabularies and their URIs](#)
 - [4.3.4 Express library data by re-using or mapping to existing Linked Data vocabularies](#)
 - [4.4 For librarians and archivists](#)
 - [4.4.1 Preserve Linked Data element sets and value vocabularies](#)
 - [4.4.2 Apply library experience in curation and long-term preservation to Linked Data datasets](#)
- [References](#)
- [Acknowledgments](#)
- [Appendix A: An inventory of existing library Linked Data resources](#)
- [Appendix B: Relevant Technologies](#)
 - [B.1 Using URIs to identify things not actually located on the Web](#)
 - [B.2 Discrete and bulk access to information](#)
 - [B.3 Front ends for mapping existing data stores to Linked Data and RDF](#)
 - [B.4 Tools for data designers](#)
 - [B.5 SKOS and related tools](#)
 - [B.6 Microformats, Microdata, and RDFa](#)
 - [B.7 Web Application Frameworks](#)
 - [B.8 Content Management Systems](#)
 - [B.9 Web Services for library Linked Data](#)
- [Appendix C: Semantic alignment](#)

1 Scope of this report

The scope of this report, "library Linked Data", can be understood as follows:

Library. The word "library" as used in this report comprises the full range of cultural heritage and memory institutions including libraries, museums, and archives. The term refers to three distinct but related concepts: a collection of physical or abstract (potentially including "digital") objects, a place where the collection is located, and an agent that curates the collection and administers the location. Collections may be public or private, large or small, and are not limited to any particular types of resources.

Library data. "Library data" refers to any type of digital information produced or curated by libraries that describes resources or aids their discovery. Data covered by library privacy policies is generally out of scope. This report pragmatically distinguishes three types of library data based on their typical use: **datasets**, **element sets**, and **value vocabularies** (see [Appendix A](#)).

Linked Data. "Linked Data" refers to data published in accordance with [principles](#) designed to facilitate linkages among datasets, element sets, and value vocabularies [[LINKEDDATA](#)]. Linked Data uses [Uniform Resource Identifiers \(URIs\)](#) as globally unique identifiers for any kind of resource, analogously to how identifiers are used for authority control in traditional librarianship [[URI](#)]. In Linked Data, URIs may be [Internationalized Resource](#)

[Identifiers \(IRIs\)](#), that is, [Web addresses](#) that use the extended set of natural-language scripts supported by [Unicode](#). Linked Data is expressed using standards such as [Resource Description Framework \(RDF\)](#), which specifies relationships between things; relationships that can be used for navigating between, or integrating, information from multiple sources [[RDF](#)].

Open Data. While "Linked Data" refers to the technical interoperability of data, "Open Data" focuses on its legal interoperability. According to the definition for [Open Bibliographic Data](#), Open Data is in essence freely usable, reusable, and redistributable — subject, at most, to the requirements to attribute and share alike. Note that Linked Data technology per se does not require data to be Open, though the potential of the technology is best realized when data is published as Linked Open Data.

Library Linked Data. "Library Linked Data" is any type of library data (as defined above) that is expressed as Linked Data.

2 Benefits of the Linked Data approach

The Linked Data approach offers significant advantages over current practices for creating and delivering library data while providing a natural extension to the collaborative sharing models historically employed by libraries. Linked Data and especially Linked Open Data is **sharable**, **extensible**, and easily **re-usable**. It supports multilingual functionality for data and user services, such as the labeling of concepts identified by language-agnostic URIs. These characteristics are inherent in the Linked Data standards and are supported by the use of Web-friendly identifiers for data and concepts. Resources can be described in collaboration with other libraries and linked to data contributed by other communities or even by individuals. Like the linking that takes place today between Web documents, Linked Data allows anyone to contribute unique expertise in a form that can be reused and recombined with the expertise of others. The use of identifiers allows diverse descriptions to refer to the same thing. Through rich linkages with complementary data from trusted sources, libraries can increase the value of their own data beyond the sum of their sources taken individually.

By using globally unique identifiers to designate works, places, people, events, subjects, and other objects or concepts of interest, libraries will allow resources to be cited across a broad range of data sources and thus make their metadata descriptions more richly accessible. The Internet's Domain Name System assures stability and trust by putting these identifiers into a regulated and well-understood ownership and maintenance context. This notion is fully compatible with the long-term mandate of libraries. Libraries, and memory institutions generally, are in a unique position to provide trusted metadata for resources of long-term cultural importance as data on the Web.

Another powerful outcome of the reuse of these unique identifiers is that it allows data providers to contribute portions of their data as statements. In our current document-based ecosystem, data is exchanged always in the form of entire records, each of which is presumed to be a complete description. Conversely, in a graph-based ecosystem an organization can supply individual statements about a resource, and all statements provided about a particular uniquely identified resource can be aggregated into a global graph. For

example, one library could contribute their country's national bibliography number for a resource, while another might supply a translated title. Library services could accept these statements from outside sources much as they do today when ingesting images of book covers. In a Linked Data ecosystem, there is literally no contribution too small — an attribute that makes it possible for important connections to come from previously unknown sources.

Library authority data for names and subjects will help reduce redundancy of bibliographic descriptions on the Web by clearly identifying key entities that are shared across Linked Data. This will also aid in the reduction of redundancy of metadata representing library holdings.

2.1 Benefits to researchers, students, and patrons

It may not be obvious to users of library and cultural institution services when Linked Data is being employed because the changes will lie "under the hood." As the underlying structured data becomes more richly linked, however, the user may notice improved capabilities for discovering and using data. Navigation across library and non-library information resources will become more sophisticated. Federated searches will improve through the use of links to expand indexes, and users will have a richer set of pathways for browsing.

Linked Data builds on the defining feature of the Web: browsable links (URIs) spanning a seamless information space. Just as the totality of Web pages and websites is available as a whole to users and applications, the totality of datasets using RDF and URIs presents itself as a global information graph that users and applications can seamlessly browse by resolving trails of URI links ("following one's nose") — a data-powered form of "toURism." The value of Linked Data for library users derives from these basic navigation principles. Links between libraries and non-library services such as Wikipedia, GeoNames, MusicBrainz, the BBC, and The New York Times will connect local collections into the larger universe of information on the Web.

Linked Data is not about creating a different Web, but rather about enhancing the Web through the addition of structured data. This structured data, expressed using technologies such as RDF in Attributes (RDFa) and microdata, plays a role in the crawling and relevancy algorithms of search engines and social networks, and will provide a way for libraries to enhance their visibility through search engine optimization (SEO). Structured data embedded in HTML pages will also facilitate the re-use of library data in services to information seekers: citation management can be made as simple as cutting and pasting URIs. Automating the retrieval of citations from Linked Data or creating links from Web resources to library resources will mean that library data is fully integrated into research documents and bibliographies. Linked Data will favor interdisciplinary research by enriching knowledge through linking among multiple domain-specific knowledge bases.

Migrating existing library data to Linked Data is only a first step; the datasets used for experiments reported in a paper and the model used by the authors to process that data can also be published as Linked Data. Representing a paper, dataset, and model using appropriate vocabularies and formalisms makes it easier for other researchers to replicate an experiment or to reuse its dataset with different models and purposes. If adopted, this practice could improve the

rigor of research and make the overall assessment of research reports outlined in research papers more transparent for easier validation by peers. (See, for instance, the [Enhanced Publications use case](#).)

2.2 Benefits to organizations

By promoting a bottom-up approach to publishing data, Linked Data creates an opportunity for libraries to improve the value proposition of describing their assets. The traditionally top-down approach of library data — i.e., producing catalog records as stand-alone descriptions for library material — has been enforced by budget limits: libraries do not have the resources needed to produce information at a higher level of granularity. With Linked Data, different kinds of data about the same asset can be produced in a decentralized way by different actors, then aggregated into a single graph.

Linked Data technology can help organizations improve their internal data curation processes and maintain better links between, for instance, digitized objects and their descriptions. It can improve data publishing processes within organizations even where data is not entirely open. Whereas today's library technology is specific to library data formats and provided by an Integrated Library System industry specific to libraries, libraries will be able to use mainstream solutions for managing Linked Data. Adoption of mainstream Linked Data technology could give libraries a wider choice of vendors, and the use of standard Linked Data formats will allow libraries to recruit from and interact with a larger pool of developers.

Linked Data may be a first step toward a "cloud-based" approach to managing cultural information, which could be more cost-effective than stand-alone systems in institutions. This approach could make it possible for small institutions or individual projects to make themselves more visible and connected while reducing infrastructure costs.

With Linked Open Data, libraries can increase their presence on the Web, where most information seekers can be found. The focus on identifiers allows descriptions to be tailored to specific communities such as museums, archives, galleries, and audiovisual archives. The openness of data is more an opportunity than a threat. Clarification of the licensing conditions of descriptive metadata facilitates its reuse and improves institutional visibility. Data thus exposed will be put to unexpected uses, as in the adage: "The coolest thing to do to your data will be thought of by someone else."

2.3 Benefits to librarians, archivists, and curators

The benefits to patrons and organizations will also have a direct impact on library professionals. By using Linked Open Data, libraries will create an open, global pool of shared data that can be used and re-used to describe resources, with a limited amount of redundant effort compared with current cataloging processes.

The use of the Web and Web-based identifiers will make up-to-date resource descriptions directly citable by catalogers. The use of shared identifiers will allow them to pull together descriptions for resources outside their domain

environment, across all cultural heritage datasets, and even from the Web at large. Catalogers will be able to concentrate their effort on their domain of local expertise, rather than having to re-create existing descriptions that have been already elaborated by others.

History shows that all technologies are transitory, and the history of information technology suggests that specific data formats are especially short-lived. Linked Data describes the meaning of data ("semantics") separately from specific data structures ("syntax" or "formats"), with the result that Linked Data retains its meaning across changes of format. In this sense, Linked Data is more durable and robust than metadata formats that depend on a particular data structure.

2.4 Benefits to developers and vendors

Library developers and vendors will directly benefit from not being tied to library-specific data formats. Linked Data methods support the retrieval and re-mixing of data in a way that is consistent across all metadata providers. Instead of requiring data to be accessed using library-centric protocols (e.g., [Z39.50](#) Information Retrieval Protocol), Linked Data uses well-known standard Web protocols such as the [Hypertext Transport Protocol \(HTTP\)](#).

Developers will also no longer have to work with library-specific data formats, such as [ISO 2709](#) and [MACHine-Readable Cataloging \(MARC\)](#), which require custom software tools and applications. Linked Data methods involve pushing data onto the Web in a form that is generically understandable. Library vendors that support Linked Data will be able to market their products outside of the library world, while vendors outside the library world may be able to adapt their more generic products to the specific requirements of libraries. By leveraging RDF and HTTP, library developers are freed from the need to use domain-specific software, opening a growing range of generic tools, many of which are open-source. They will find it easier to build new services on top of their data. This also opens up a much larger developer community to provide support to information technology professionals in libraries. **In a sea of RDF triples, no developer is an island.**

3 The Current Situation

3.1 Issues with traditional library data

3.1.1 Library data is not integrated with Web resources

Library data today resides in databases which, while they may have Web-facing search interfaces, are not deeply integrated with other data sources on the Web. There is a considerable amount of bibliographic data and other kinds of resources on the Web that share data points such as dates, geographic information, persons, and organizations. In a future Linked Data environment, all these dots could be connected.

3.1.2 Library standards are designed only for the library community

Many library standards, such as the MARC format or the information retrieval protocol Z39.50, have been (or continue to be) developed in a library-specific context. Standardization in the library world is often undertaken by bodies focused exclusively on the library domain, such as the [International Federation of Library Associations and Institutions](#) (IFLA) and the [Joint Steering Committee for Development of RDA](#) (JSC). By broadening their scope or liaising with Linked Data standardization initiatives, such bodies can expand the relevance and applicability of their standards to data created and used by other communities.

3.1.3 Library data is expressed primarily in natural-language text

Most information in library data is encoded as display-oriented, natural-language text. Some of the fields in MARC records use coded values, such as fixed-length strings representing languages, but there is no clear incentive to include these in all records, since most coded data fields are not used in library system functions. Some of the identifiers carried in MARC records, such as ISBNs for books, could in principle be used for linking, but only after being extracted from the text fields in which they are embedded, then normalized.

Some data fields, such as authority-controlled names and subjects, have related records in separate files, and these records have identifiers that could be used to represent those entities in library metadata. However, the data formats in current use do not always support inclusion of these identifiers in records, therefore many of today's library systems do not properly support their use. These identifiers also tend to be managed locally rather than globally, and hence are not expressed as URIs which would enable linking to them on the Web. The absence of links or insufficient support for them in library systems raises important issues. Changes to authoritative displays require that all related bibliographic records be retrieved in order to change their text strings — a disruptive and expensive process that often prevents libraries from implementing changes in a timely manner.

3.1.4 The library community and Semantic Web community have different terminology for similar metadata concepts

Work on library Linked Data can be hampered by the disparity in concepts and terminology between libraries and the Semantic Web community. Few librarians speak of metadata "statements," while the Semantic Web community lacks notions clearly equivalent to "headings" or "authority control." Each community has its own vocabulary, and these reflect differences in their points of view. Mutual understanding must be fostered, as both groups bring important expertise to the construction of a web of data.

3.1.5 Library technology changes depend on vendor systems development

Much of the technical expertise in the library community is concentrated in the small number of vendors who provide the systems and software that support both library management functions, such as acquisitions, user data, and circulation, as well as the user discovery service. This means that libraries must rely on these vendors and their technology development plans, rather than on their own initiative, when they want to adopt Linked Data at a production scale.

3.2 Library Linked Data available today

The success of library Linked Data will depend on the ability of practitioners to identify, re-use, or link to other available sources of Linked Data. However, it has hitherto been difficult to get an overview of the library datasets and vocabularies that are available as Linked Data. The Incubator Group undertook an inventory of available sources of library-related Linked Data (see [Appendix A](#)), leading to the following observations.

3.2.1 Fewer bibliographic datasets have been published as Linked Data than value vocabularies and element sets

Many metadata element sets and value vocabularies have been published as Linked Data over the past few years, including flagship vocabularies such as the [Library of Congress Subject Headings](#) and [Dewey Decimal Classification](#). Key element sets, such as [DCMI Metadata Terms](#), and reference frameworks such as [Functional Requirements for Bibliographic Records \(FRBR\)](#) have been published as Linked Data or in a Linked Data-compatible form.

Relatively few bibliographic datasets have been made available as Linked Data, and even less metadata has been produced for journal articles, citations, or circulation data — information which could be put to effective use in environments where data is integrated seamlessly across contexts. Pioneering initiatives such as [the release of the British National Bibliography](#) reveal the effort required to address challenges such as licensing, data modeling, the handling of legacy data, and collaboration with multiple user communities. However, these also demonstrate the considerable benefits of releasing bibliographic databases as Linked Data. As the community's experience increases, the number of datasets released as Linked Data is growing rapidly.

3.2.2 The quality of and support for available data varies greatly

The level of maturity or stability of available resources varies greatly. Many existing resources are the result of ongoing project work or the result of individual initiatives, and describe themselves as prototypes rather than mature offerings. Indeed, the abundance of such efforts is a sign of activity around and interest in library Linked Data, exemplifying the processes of rapid prototyping and "agile" development that Linked Data supports. At the same time, the need for such creative, dynamically evolving efforts is counterbalanced by a need for library Linked Data resources that are stable and available for the long term.

It is encouraging that established institutions are increasingly committing resources to Linked Data projects, from the national libraries of Sweden, Hungary, Germany, France, the Library of Congress, and the British Library, to the Food and Agriculture Organization of the United Nations and OCLC Online

Computer Library Center, Inc. Such institutions provide a stable foundation on which library Linked Data can grow over time.

3.2.3 Linking across datasets has begun but requires further effort and coordination

A major advantage of Linked Data technology is realized with the establishment of connections between and across datasets. Achieving these connections will be key to its success. Our inventory of available data (see [Appendix A](#)), shows that many semantic links have been created between published value vocabularies — a great achievement for the nascent library Linked Data community as a whole. More can — and should — be done to resolve the issue of redundancy among the various authority resources maintained by libraries. More links are also needed among datasets and among the metadata element sets used to structure Linked Data descriptions. Key bottlenecks are the comparatively low level of long-term support for vocabularies, the limited communication among vocabulary developers, and the lack of mature tools to lower the cost for data providers to produce the large amount of semantic links required. Efforts have begun to facilitate knowledge sharing among participants in this area, as well as the production and sharing of relevant links (see [Appendix C](#)).

3.3 Rights issues

3.3.1 Rights ownership is complex

Some library data has restricted usage based on local policies, contracts, and conditions. Data can therefore have unclear and untested rights issues that hinder their release as Open Data. Rights issues vary significantly from country to country, making it difficult to collaborate on Open Data publishing.

Ownership of legacy catalog records has been complicated by the degree of data sharing among libraries over the past fifty years. Records are frequently copied and the copies are modified or enhanced for use by local catalogers. These records may be subsequently re-aggregated into the catalogs of regional, national, and international consortia. Assigning legally sound intellectual property rights between relevant agents and agencies is difficult, and the lack of certainty hinders data sharing in a community that is necessarily cautious on legal matters.

3.3.2 Data rights may be considered business assets

Where library data has never been shared with another party, rights may be exclusively held by agencies who put a value on their past, present, and future investment in creating, maintaining, and collecting metadata. Some agencies treat records as assets in their business plans and may be reluctant to publish them as Linked Open Data. Others may only be willing to release their data in a stripped- or dumbed-down form with loss of semantic detail that affects the utility of the metadata.

4 Recommendations

Libraries should embrace the web of information, both by making their data available for use as Linked Data and by using the web of data in library services. Ideally, library data should integrate fully with other resources on the Web, creating greater visibility for libraries and bringing library services to information seekers. In engaging with the web of Linked Data, libraries can take on a leadership role grounded in their traditional activities: management of resources for current use and long term preservation; description of resources on the basis of agreed rules; and responding to the needs of information seekers.

4.1 For library leadership

4.1.1 Identify sets of data as possible candidates for early exposure as Linked Data

A very early step should be the identification of high-priority, low-effort Linked Data projects. By its very nature, Linked Data facilitates an incremental approach to making data available for use on the Web. The data environments of libraries are complex, and attempting to expose that complexity as Linked Data all at once could have limited success. However, some library resources lend themselves to publication as Linked Data without disrupting current systems and services. Among these are authority files (whose members identify things) and controlled term lists. Identification of such "low-hanging fruit" will allow libraries to quickly expand their presence in the Linked Data cloud without changing their workflows elsewhere.

4.1.2 Foster a discussion about Open Data and rights

In defining rights for data, rights owners must consider the impact of usage restrictions, as restrictions complicate the re-use of data in a Linked Data environment. It makes sense for library leaders to seek agreement with owners about rights and licensing at the level of library consortia or even on a national or international scale. (For an example, see the [Rights and Licensing](#) section of the [Open Bibliographic Data Guide](#) for UK higher-education libraries.)

4.2 For standards bodies and participants

4.2.1 Increase library participation in Semantic Web standardization

If [Semantic Web](#) standards do not support the translation of library data with sufficient expressivity, the standards can be extended. For example, if Simple Knowledge Organization System ([SKOS](#)), a standard used for publishing knowledge organization systems as Linked Data, does not include mechanisms for representing the components of pre-coordinated subject headings, implementers should consider devising solutions that extend its basic elements, e.g., using the [OWL Web Ontology Language](#). In order to ensure that these

new structures will be understood by consumers of Linked Data generally, implementers should collaborate with the Semantic Web community both to ensure that the proposed solutions are compatible with current best practice and to maximize the applicability of their work outside the library environment. Members of the library world should contribute in standardization efforts of relevance to libraries, such as the W3C efforts to extend RDF to encompass the concept of provenance, by joining technical working groups, or by participating in public review processes. A [W3C Community Group](#) could also play an important role in this area.

4.2.2 Develop library data standards that are compatible with Linked Data

Semantic Web technologies conceptualize data in a way that fundamentally differs from the conceptualization underlying the data formats of the twentieth century. Linked Data is primarily about meaning and meaningful relationships between things, while traditional library data formats combine the meaning of data and the structured encoding of data into a single package. The inseparability of meaning from encoding in data formats results in less flexibility for obtaining value from an investment in data. Since the introduction of MARC formats in the 1960s, digital data in libraries has been managed predominantly in the form of "records" that are bounded sets of information stored in files of a precisely specified structure. The Semantic Web and Linked Data, in contrast, structure data as graphs — constructs which, in principle, may be boundless. The difference between these two approaches means that the process of translating library standards and datasets into Linked Data is not trivial and must be undertaken with knowledge of new principles of data design. There is a need for best-practice documentation and recipes to guide participants in the construction of ontologies and structured vocabularies for library data.

4.2.3 Develop and disseminate best-practice design patterns tailored to library Linked Data

Design patterns allow implementers to build on the experience of predecessors. Traditional cataloging practices have been documented with a rich array of patterns and examples, and best practices are starting to be documented for the Linked Data space as well. Examples include publications on [Linked Data: Evolving the Web into a Global Data Space](#) and [Linked Data Patterns. Application profiles](#) provide a method for a community of practice to document and share patterns and constraints for using vocabularies to describe specific types of resources. What is needed are design patterns specifically tailored to the requirements of library Linked Data. Such design patterns could meet the needs of developers who are better able to understand new techniques through patterns and examples, as well as increase the coherence of library Linked Data overall.

4.3 For data and systems designers

4.3.1 Design and test user services based on Linked Data capabilities

Linked Data could ultimately lead to new and better services to users as well as enabling implementers outside of libraries to create applications and services based on library data. It is too early to predict what new types of services may be developed for information discovery and use. Experimental services using library Linked Data should be undertaken in order to explore potential use cases and inform the direction of larger development efforts.

4.3.2 Create URIs for the items in library datasets

Library data cannot be used in a Linked Data environment without having Uniform Resource Identifiers (URIs) both for specific resources and for library-standard concepts. The official owners of resource data and standards should assign URIs as soon as possible, since application developers and other users of such data will not delay their activities, but are more likely to assign URIs themselves, outside of the owning institution. When owners are not able to assign URIs in good time, they should seek partners for this work or delegate the assignment and maintenance of URIs to others in order to avoid the proliferation of URIs for the same thing and to encourage the re-use of URIs already assigned.

Agencies responsible for the creation of catalog records and other metadata, such as national bibliographies, are the logical organizations to take a leading role in creating URIs for their described resources.

4.3.3 Develop policies for managing Linked Data vocabularies and their URIs

Organizations and individuals who create and maintain URIs for resources and standards will benefit if they develop policies for the namespaces used to derive those URIs. Such "namespace policies" encourage a consistent, coherent, and stable approach which improves effectiveness and efficiency and provides quality assurance for users of URIs and their namespaces. Policies might cover:

- Patterns used to coin the URIs, preferably based on best-practice guidelines.
- Institutional commitments to the persistence of the URIs.
- Version control for a vocabulary and its terms.
- The use of "HTTP" URIs, which invoke the [Hypertext Transfer Protocol](#) supported universally by Web browsers, and their resolution to any Web pages or machine-readable representations which document the meaning of the URIs.
- Extensibility of the vocabulary by other organizations.
- Translations of labels and other annotations into other languages.

4.3.4 Express library data by re-using or mapping to existing Linked Data vocabularies

In order to maximize linkability with other datasets, library datasets must be expressed using Linked Data terms — properties, classes, and instances — that have well-defined relationships to those used in the wider Linked Data space. This can be done in two ways: by using Linked Data vocabularies based

on existing standards, and by defining explicit relationships ("alignments") between the Linked Data terms of the library world and those of other communities. (See further discussion in [Appendix C.](#))

4.4 For librarians and archivists

4.4.1 Preserve Linked Data element sets and value vocabularies

Many Linked Data vocabularies are essentially cultural reference works, giving authoritative information about people, places, events, and concepts within regional, national, or international contexts. As such, preservation of Linked Data vocabularies is a natural, and essential, extension of the activity of memory institutions. Linked Data will remain usable twenty years from now only if its URIs persist and can resolve to documentation of their meaning. As keys to the correct interpretation of data, both now and in the future, element sets and value vocabularies are particularly important as objects of preservation. This situation presents libraries with an opportunity to assume a key role in supporting the Linked Data ecosystem.

4.4.2 Apply library experience in curation and long-term preservation to Linked Data datasets

Much of the content in today's Linked Data [cloud](#) is the result of ad-hoc, one-off conversions of publicly available datasets into RDF and is not subject to regular accuracy checks or maintenance updates. With their ethos of quality control and commitment to long-term maintenance, libraries have a significant opportunity to take a key role in the important (and hitherto neglected) function of curating Linked Data as an extension of their existing mission. By curating and maintaining the resources described within datasets as truly linkable objects, libraries can reap the benefits of opening their data for value-added contributions from other communities. Adding links to data from biographers or genealogists, for example, could enrich library resource descriptions with data not usually provided by libraries, and could greatly improve the discovery and navigation of library collections.

References

[LINKEDDATA]

[Linked Data](#), Tim Berners-Lee, World Wide Web Consortium, accessed 18 October 2011. See <http://www.w3.org/DesignIssues/LinkedData.html>.

[RDF]

[Resource Description Framework \(RDF\)](#), World Wide Web Consortium, accessed 18 October 2011. See <http://www.w3.org/RDF/>.

[URI]

[RFC 3986 — Uniform Resource Identifier \(URI\): Generic Syntax](#), T. Berners-Lee, R. Fielding, L. Masinter, The Internet Society, January 2005, accessed 18 October 2011. See <http://tools.ietf.org/html/rfc3986>.

[USECASE]

[Library Linked Data Incubator Group: Use Cases](#), Daniel Vila Suero, Editor, W3C Incubator Group Report, 25 October 2011. See <http://www.w3.org/2005/Incubator/ldd/XGR-ldd-usecase-20111025/>. [Latest](#)

[version](http://www.w3.org/2005/Incubator/ld/XGR-ld-usecase/) available at <http://www.w3.org/2005/Incubator/ld/XGR-ld-usecase/>.

[VOCABDATASET]

[Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets](#), Antoine Isaac, William Waites, Jeff Young, and Marcia Zeng, W3C Incubator Group Report, 25 October 2011. See <http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/>. [Latest version](#) available at <http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset/>.

Acknowledgments

In addition to the editors, the Library Linked Data included the following participants, without whom this report would not exist: Alexander Haffner, Alexandru Constantin, András Micsik, Andrew Houghton, Anette Seiler, Asaf Bartov, Bernard Vatant, Brian Kelly, Carlo Meghini, Dan Brickley, Daniel Vila Suero, Dickson Lukose, Felix Sasaki, Fumihiko Kato, Glen Newton, Guenther Neher, Herbert Van De Sompel, Hideaki Takeda, Ikki Ohmukai, Joachim Neubert, Jon Phipps, Jonathan Rees, Kai Eckert, Kendall Clark, Kevin Ford, Kim Viljanen, Kosuke Tanabe, Lars Svensson, Laszlo Kovacs, Marcel Ruhl, Mark van Assem, Martin Malmsten, Michael Hausenblas, Mike Bergman, Monica Duke, Nicolas Delaforge, Oreste Signore, Ray Denenberg, Renato Iannella, Stu Weibel, Tod Matola, Uldis Bojars, Wolfgang Halb.

Reviews from the community also helped us shape this report. Special thanks go to: Adrian Pohl, Alan Danskin, Catherine Jones, Ed Chamberlain, J. McRee Elrod, James Weinheimer, Jennifer Bowen, Jody DeRidder, Juha Hakala, Laura Krier, Laura Smart, Lukas Koster, Nicolas Chauvat, Patrick Danowski, René van der Ark, Romain Wenz, Roy Tennant, Teague Allen.

Appendix A: An inventory of existing library Linked Data resources

The complexity and variety of available vocabularies, with their overlapping coverage, derivative relationships, and alignments, results in uncertainty for the re-use or linking efforts that are crucial to the success of library Linked Data. Many, especially among library professionals, are unfamiliar with the linked datasets and vocabularies that can be of use in the library domain because these have often been developed in the Semantic Web research community. A current and reliable bird's-eye view can help both novices seeking an overview of the library Linked Data domain and experts needing a quick look-up or refresher for a library Linked Data project.

The Incubator Group has therefore produced an inventory of useful resources for creating or consuming Linked Data in the library domain [[VOCABDATASET](#)]. This inventory, presented as a [separate document](#), shows that there are many areas where early adoption of Semantic Web and Linked Data principles and technology has led to the development of mature datasets and vocabularies. The inventory also points to areas where libraries and related organizations can still make key contributions. Finally, this document tries to provide the Linked Data community with an opportunity to understand the

specific viewpoint, resources, and terminology used by the library community for their data, while helping Library and Information Science professionals grasp the Linked Data notions corresponding to their own traditions.

Though Linked Data technology differs from traditional library data concepts, this report classifies available resources into three not mutually-exclusive categories that reflect library practices:

- [Datasets](#) describing library-related resources such as the British National Bibliography, the catalog of the Hungarian national library, the Open Library, CrossRef, and Europeana;
- [Value vocabularies](#) such as the Library of Congress Subject Headings, AGROVOC, the Virtual International Authority File (VIAF), Dewey Decimal Classification, and GeoNames;
- [Metadata element sets](#) such as DCMI Metadata Terms, the elements of RDA: Resource Description and Access, Simple Knowledge Organization System (SKOS), and the Friend of a Friend vocabulary (FOAF).

Specific datasets may re-use elements from various value vocabularies, and are structured according to the specifications for metadata element sets. For example, the British National Bibliography dataset re-uses terms from the Library of Congress Headings vocabulary and DCMI Metadata Terms (Dublin Core). Instances of these categories are listed in the inventory along with brief descriptions, links to their online locations, and to the [use cases](#) that our group has gathered from the community.

Our [inventory](#) is intended to provide a broad coverage of available data resources. However, we are well aware that this report cannot capture the full diversity of current datasets, especially given the dynamic nature of Linked Data: new resources are continuously made available, and existing ones are regularly updated. To get a representative overview, we intentionally based our work on the use cases we received. Additional coverage was provided by the experts who participated in the Incubator Group to ensure that key resources available at the time of writing were not overlooked.

To help make our report useful in the future we have included a number of links to tools or Web sites which we believe can provide up-to-date information after the Incubator Group has completed its work. In particular we have set up a [Library Linked Data group](#) as a site to collect information on relevant library linked datasets. This site is hosted by the [The Data Hub](#), a repository designed to be a central hub for descriptions of data packages with an emphasis on those that are published as Open Data. We hope that this Data Hub group will be actively maintained by the library Linked Data community after the Incubator Group has ended.

Appendix B: Relevant Technologies

Linked Data is an emerging technology, so most tools are still in development. The [principles of Linked Data](#) are not tied to any particular tool; rather, they are tied directly to Web standards. In many situations, the production and consumption of Linked Data can be layered or interwoven with existing applications without requiring massive redevelopment efforts. This list of tools and technologies is not exhaustive, but is intended to illustrate a few broad

categories. From a non-technical perspective, these technologies are relevant because they encourage the creation and discovery of reusable vocabularies and provide ways to combine those terms into reusable (syntactic) statements.

B.1 Using URIs to identify things not actually located on the Web

In the early days of the Web, it was unclear whether "HTTP URIs" (also known as "URLs") should be used to identify things that are not "located" on the Web. That concern was the basis for defining new URI schemes such as [URNs](#) and ["info" URIs](#). These uncertainties were eventually resolved by a report from the W3C Uniform Resource Identifier Interest Group ([RFC 3305](#)) and a [resolution](#) of the W3C Technical Advisory Group on the issue known as ["HTTPRange-14"](#). In the Linked Data paradigm, it is generally expected that HTTP URIs will also be used to identify "real world objects." Nevertheless, many applications have been built on the other identifier schemes. Using the [owl:sameAs property](#) is a good way to map these non-resolvable URI schemes to HTTP URI equivalents. Even if this mapping is not done, non-resolvable URIs are still useful in RDF and SPARQL.

B.2 Discrete and bulk access to information

The principles of Linked Data were introduced circa 2006, leading to a formalized notion of ["Cool URIs"](#) in 2008. What makes Linked Data identifiers special is the ability to help humans and machines understand, progress, and link information across a wide range of use cases; the [DBpedia](#) resource for [Jane Austen](#) is a good example. Resolvable URIs are great for casual use, for diagnosing data, and for serendipitous discovery, but discrete HTTP GET requests may be impractical for datasets with a large numbers of individuals. Fortunately, linked datasets are increasingly being published as [RDF dumps](#) and consistently described using the [Vocabulary of Interlinked Datasets \(VOID\)](#).

B.3 Front ends for mapping existing data stores to Linked Data and RDF

Related Use Case Cluster: [Vocabulary alignment cluster](#)

Unlike information represented hierarchically in typical XML documents, resources published as Linked Data allow information to be freed from use-case-specific hierarchies and thus available for unanticipated reuse. This not only makes the information easier to mash up, it also makes tools and services easier to [mash up](#). This is true for both producers and consumers of Linked Data. For example, an existing relational database can be mounted as Linked Data and [SPARQL](#) by using [D2R Server](#). The [W3C RDB2RDF Working Group](#) is currently working on standards for such mappings. Similarly, Linked Data can be produced from existing [SRU databases](#) with a few [rewrite rules](#). If the resources are already [described](#) from a SPARQL endpoint, then a Linked Data front end such as [Pubby](#) can be used to automate the content-negotiable [Cool URI](#) behavior for each individual. [Extensible Stylesheet Language Transformations](#) (XSLT) can be useful for converting generic XML into RDF/XML.

B.4 Tools for data designers

Related Use Case Cluster: [Vocabulary alignment cluster](#)

[Application profiles](#) provide a comprehensive way to document how a community of practice defines a domain model and a pattern for re-using vocabularies with particular constraints in describing particular types of resources. The current version of [OWL Web Ontology Language](#), which provides properties to represent alignments across vocabularies ([ontology mappings](#)), allows experts to describe their domain using community idioms while remaining interoperable with related or more common idioms. A variety of tools related to OWL can be found on the W3C's [RDF wiki](#) and [OWL wiki](#). [Unified Modeling Language \(UML\)](#) tools help designers represent and manipulate domain models visually. The [Ontology Definition Metamodel \(ODM\)](#) specification should help bridge some of the gaps between UML and OWL.

B.5 SKOS and related tools

Related Use Case Cluster: [Vocabulary alignment cluster](#)

Yet another key technology need is fulfilled by the [Simple Knowledge Organization System \(SKOS\)](#), which is an OWL ontology for expressing a broad range of concept schemes and thesauri, with support for broader and narrower relationships, and preferred and alternative labels. Many SKOS-related tools are listed on the W3C's [SKOS community wiki](#).

B.6 Microformats, Microdata, and RDFa

Related Use Case Cluster: [Social and new uses cluster](#)

[Microformats](#), [Microdata](#), and [RDFa](#) all provide ways to embed structured data into Web pages. As historically the emphasis on publishing information on the Web has meant publishing Web pages, these technologies provide ways to enhance what is already there rather than necessarily deploying additional infrastructure. RDFa supports the expression of RDF data embedded directly in Web pages; of the three, therefore, it is the most directly interoperable with other Linked Data infrastructure.

Microdata, which is defined in new [HTML5 specification](#) under development, provides another way of doing this. Microdata has notably gained prominence for Search Engine Optimization purposes with the announcement of [Schema.org](#) by Google, Microsoft, and Yahoo. This particular type of microdata does not appear to be intended to represent arbitrarily complex data, and the vocabulary that they have published places special emphasis on commerce and tourism. Although in principle they are extensible, microdata schemes would need to be heavily extended in order to express library information since most of the required vocabulary is lacking. There is some level of interoperability with Linked Data thanks to the efforts of [Schema.RDFS.org](#), but it currently seems like it would be difficult, using this approach, to cultivate the high level of interconnectedness between library and other datasets that is possible with Linked Data.

It should be noted that the Schema.org proponents do also support the harvesting of RDFa data and have pledged to continue doing so, so it does not appear to be the case that by publishing HTML pages marked up with RDFa one might somehow "miss out" on the opportunities afforded by microdata. Excluding bugs in the search engines' parsers, it should even be possible to use both metadata technologies in the same Web page. Ultimately, the conclusion is that some structured data is better than none.

B.7 Web Application Frameworks

Related Use Case Cluster: [Archives and heterogeneous data cluster](#)

As the Web has grown in popularity, the software development community has created a variety of software libraries that make it easier to create, maintain, and re-use Web applications. These libraries are often referred to as [Web application frameworks](#), and typically implement the [Model-View-Controller \(MVC\)](#) pattern in some fashion. In addition, Web application frameworks have typically encoded and encouraged best practices with respect to the [Representational State Transfer \(REST\) Architectural Style](#) and [Resource Oriented Architecture](#) which have informed much of the standardization around Web technologies.

A common component to Web application frameworks is a URI routing mechanism that allows software developers to define HTTP URI patterns and map them to controllers which, in turn, generate an HTTP response using the appropriate views and models. This activity encourages best practices with respect to [Cool URIs](#) and also forces developers to think about the resources that they are making available on the Web. Linked Data's focus on naming resources with HTTP URIs, and on delivering representations of those resources — in HTML for humans and RDF for machines — makes it a natural fit for Web application frameworks, which already provide some of the scaffolding for these activities. The [wide availability](#) of Web application frameworks in many different programming languages and operating system environments has led to their wide use in the cultural heritage sector.

Web developers are sometimes turned off by Semantic Web (Linked Data) technologies because they feel compelled to discard their current applications, swap their databases for triple stores, and their database query languages for SPARQL. This is simply not the case, as RDF serializations can be generated on-the-fly just as Web application frameworks do for HTML, XML, and JSON representations. The use of HTTP URIs to identify and link together resources using the RDF data model make it a natural choice for serializing and sharing entity state in a database-neutral way — a goal traditionally of great interest to cultural heritage organizations and the digital preservation community.

B.8 Content Management Systems

Related Use Case Clusters: [Social and new uses cluster](#), [Digital objects cluster](#), [Archives and heterogeneous data cluster](#)

Just as Web application frameworks have evolved with the spread of the Web, so has the class of Web applications known as Content Management Systems (CMS). CMSs are often built using a Web application framework but provide out-of-the-box functionality for easily creating, editing, and presenting content such as text, images, and video on the Web, and for managing workflows associated with the content. Since CMSs are typically built using Web frameworks, the same best practices for naming resources with HTTP URIs are naturally followed. The [wide availability](#) of Content Management Systems has led to their heavy use in the cultural heritage sector. Some content management systems such as [Drupal](#) are starting to expose structured database information to machine clients by seamlessly layering it into their HTML using RDFa. Data consumers such as Google Scholar, Google Maps, and Facebook are starting to leverage this structured metadata in their own service offerings. Conversely, Drupal is also starting to provide plug-ins for consuming RDF, such as [VARQL](#) and [SPARQL Views](#).

B.9 Web Services for library Linked Data

Related Use Case Clusters: [Bibliographic data cluster](#), [Authority data cluster](#)

In theory, most domain-specific Web Service API capabilities could be refactored as Linked Data URIs, OWL, SPARQL, and SPARQL/Update. But even though it should be possible to layer a Linked Data URI front end on an existing back-end datastore, it may not be so easy for the back end to support SPARQL and SPARQL/Update access. Security, robustness, and performance considerations could also preclude supporting SPARQL in production situations. SPARQL endpoints and bulk RDF downloads can facilitate discovery and re-use of the published Linked Data greatly. Most Web developers, however, face a steep learning curve before being able to exploit this, and for many application requirements this imposes too heavy a burden.

Web Services for the most common uses should be offered as an alternative. However, most Web Service APIs tend to be domain-specific, requiring custom-coded agents. This means they should be well-documented. More general approaches to Web Service interfaces include [OpenSearch](#) (which can be documented using a [Description Document](#)), the [Linked Data API](#) and ongoing work of the W3C [RDF Web Applications Working Group](#) on RDF and RDFa APIs. Some Linked Datasets could also benefit from syndicated access using the [Atom Syndication Format](#) or [RSS](#).

A few Linked Data implementations have endeavored to implement Web Services to enhance discovery and use of resources, often by providing some form of API. For example, [AGROVOC](#) and the [STW Thesurus for Economics](#) provide APIs for discovering resources based on relationships in the data. [VIAF](#), the [ID.LOC.GOV service of the Library of Congress](#), and [STW](#) offer [autosuggest](#) services for resources, delivering JSON responses ready for consumption in AJAX browser applications. (In principle, though, JSON responses could be content-negotiable via the Linked Data URI, as are responses in HTML and RDF.) AGROVOC and [STITCH/CATCH](#) include support for RDF responses. Some services provide full-fledged [SOAP](#) APIs, while others support a [RESTful](#) approach.

By focusing on request parameters and response formats to provide enhanced discovery, Linked Data Web Services diminish, if not eliminate, the requirement that data be stored in a triple store or be made searchable via SPARQL. And, because Web Service APIs are common, Web Services can lower the barrier to entry to adopting a Linked Data approach.

Appendix C: Semantic alignment

"Alignments" are links between semantically equivalent, similar, or related entities across different value vocabularies, metadata element sets, or datasets. Many semantic links across value vocabularies are already available, some of which obtained through high-quality manual work, as in the [MACS](#) or [CRISSCROSS](#) projects. Many value vocabulary publishers strive to establish and maintain links to resources semantically close to their own. [VIAF](#), for example, merges authority records from over a dozen national and regional agencies. [AGROVOC](#) has been published with links to six other major thesauri and subject heading lists. Though quantitative evaluation was outside the scope of our effort, we feel that many more such links should be created. Much work remains to be done to increase alignments among value vocabularies in the "library data cloud".

Alignments are likewise relevant for metadata element sets. As evidenced in the [Linked Open Vocabularies](#) inventory, practitioners generally follow the good practice of re-using existing element sets or building [application profiles](#) that re-use elements from multiple sets. Projects such as the [Vocabulary Mapping Framework](#) aim at supporting alignment.

The lack of institutional support for element sets can threaten the long-term persistence of their shared meanings. Moreover, some reference frameworks, notably Functional Requirements for Bibliographic Records (FRBR), have been expressed in a number of different ontologies, and these different expressions are not always explicitly aligned — a situation that limits the semantic interoperability of datasets in which their RDF vocabularies are used. The Library Linked Data community should promote the coordinated re-use or extension of existing element sets over the creation of new sets from scratch. Aligning already existing element sets when they overlap, typically using semantic relations from the [RDF Vocabulary Description Language \(RDF Schema\)](#) and [OWL Web Ontology Language](#), should also be encouraged. We hope that better communication among the creators and maintainers of these resources, as advocated by the [LOD-LAM initiative](#), the [Dublin Core Metadata Initiative and FOAF Project](#), and our own Incubator Group, will lead to more explicit conceptual connections between element sets.

Datasets may also be aligned. For example, Open Library attaches OCLC numbers to its bibliographic items. Re-use is arguably less central an issue for descriptions of individual books and other library-related resources than for metadata element sets and value vocabularies; union catalogs, for example, already realize a significant level of merging of book-level data. Yet it is crucial — indeed, one of the expected benefits of Linked Data applied in our domain — that library-related datasets be published and interconnected rather than continue to exist in their own silos. Because of past practices the community is already well aware of challenges such as "[deduplication](#)."

We also note that links are being built between library resources and resources originating in other organizations or domains. For example, VIAF aggregates authority records from various library agencies, identifies the primary entities involved and, where possible, links them to DBpedia, a Linked Data extraction of Wikipedia. The semantic alignment for Jane Austen in [VIAF](#), [Wikipedia](#), and [DBpedia](#), for example, illustrates one of the expected benefits of Linked Data, which is that data can be easily networked irrespective of its origins. In this way the library domain can benefit from re-using data from other fields, while library data can contribute to initiatives that did not originate in the library community.

The creation of alignments will benefit from the availability of better tools for linking. Much effort has been put into computer science research areas such as [Ontology Matching](#). This leads to implementations based, for example, on string matching and statistical techniques. These efforts have tended to focus on metadata element sets and typically are not ready to be applied more generally to the (often huge) datasets and value vocabularies of the library domain. Recent generic tools for linking data include [Silk - Link Discovery Framework](#), [Google Refine](#), and [Google Refine Reconciliation Service API](#). Nonetheless, the community still needs to gain experience in their use, to share results of this experience, and possibly to build tools better suited to library Linked Data.

One final caveat: data consumers should bear in mind that, in contrast to traditional, closed IT systems, Linked Data follows an **open-world assumption**: the assumption that data cannot generally be assumed to be complete and that, in principle, more data may become available for any given entity. We hope that more "data linking" will happen in the library domain in line with the projects mentioned here.