

Draft recommendations page

Contents

[\[hide\]](#)

- [1 Recommendations](#)
 - [1.1 Assess](#)
 - [1.1.1 Identify sets of data as possible candidates for early exposure as LD](#)
 - [1.1.2 For each set of data, determine ROI of current practices, and costs and ROI of exposing as LD](#)
 - [1.1.3 Consider migration strategies](#)
 - [1.1.4 Foster a discussion about open data and rights](#)
 - [1.2 Facilitate](#)
 - [1.2.1 Cultivate an ethos of innovation](#)
 - [1.2.2 Identify Linked Data literacy needed for different staff roles in the library](#)
 - [1.2.3 Include metadata design in library and information science education](#)
 - [1.2.4 Increase library participation in Semantic Web standardization](#)
 - [1.3 Design](#)
 - [1.3.1 Translate library data, and data standards, into forms appropriate for Linked Data](#)
 - [1.3.2 Develop and disseminate best-practices design patterns tailored to LLD](#)
 - [1.3.3 Design user stories and exemplar user interfaces](#)
 - [1.4 Identify and link](#)
 - [1.4.1 Assign unique identifiers \(URIs\) for all significant things in library data](#)
 - [1.4.2 Create URIs for the items in library datasets](#)
 - [1.4.3 Create explicit links from library datasets to other well-used datasets](#)
 - [1.4.4 Directly use, or map to, commonly understood Linked Data vocabularies](#)
 - [1.5 Prepare](#)
 - [1.5.1 Develop best practices and design patterns for LLD](#)
 - [1.5.2 Commit to best-practice policies for managing and preserving RDF vocabularies](#)
 - [1.5.3 Identify tools that support the creation and use of LLD](#)
 - [1.6 Curate](#)
 - [1.6.1 Apply library experience in curation and long-term preservation to Linked Data datasets](#)
 - [1.6.2 Preserve Linked Data vocabularies](#)

Recommendations

| The general recommendation of the report is for libraries to embrace  the web of information,

both in terms of making their data available for use and in terms of making use of the web of data in library services. Ideally, library data should integrate fully with web resources, 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 around traditional library values of managing resources for permanence, application of rules-based data creation, and attention to the needs of information seekers.

The intended target audience of this section is not totally clear to me. I think we should keep in mind at least two different audiences: libraries that just want to participate in LOD and look for information how to do that and libraries/groups who are actively developing LOD and look for recommendations out of the LLD-XG how to proceed. It is clear that these groups can not always be distinguished, but for example some of the recommendations look like best-practice recommendations (1st audience), while others recommend the development of best-practices (2nd audience). As LLD is a new topic and best-practices are hardly available, I would primarily focus on the recommendations regarding the development of those and add a section of recommendations for libraries that want to start implement something right now. A note about the target audience and the scope of these recommendation section should be added in the introductory paragraph.



Assess

Add an introductory paragraph to each section that explains what you mean by Assess, Facilitate, ... This would also help to identify missing recommendations.-

Identify sets of data as possible candidates for early exposure as LD

A very early step should be the identification of high priority/low effort linked data projects. The very nature of linked data facilitates taking an incremental approach to making a set of data available for use on the Web. Libraries are in possession of a complex data environment and attempting to expose all of that complexity in the first steps to linked data would probably not be successful. At the same time, there are library resources that are highly conducive to being published as linked data without disrupting current library systems and services. Among these are authority files (which function as identifiers and have discrete values) and controlled lists. Identification of these "low hanging fruits" will allow libraries to enter the linked data cloud soon and without having to make changes elsewhere in their workflows.

For each set of data, determine ROI of current practices, and costs and ROI of exposing as LD

There must be some measurement of the relative costs of current library data practices and the potential of Linked Data to aid in making decisions about the future of library data. There are various areas of library metadata practices that could be studied, either separately or together. Among these are:

- The relative costs of the Record v. statement approach: for editing by humans, as full record replacement in systems, and the capabilities for sharing
- The use of text versus identifiers approach has costs: actual records must change when displays change (Cookery to Cooking); international cooperation requires extensive language mapping processes; some needed data elements must be extracted from textual field using algorithms, which also hinders sharing; and some library data formats require catalogs to duplicate information in the record, providing both textual fields and coded data for same information.

- Study ways to eliminate duplication of effort in metadata creation and in service development.

Consider migration strategies

A full migration to Linked Data for library and cultural heritage metadata will likely be a lengthy and highly distributed effort. The existence of large stores of already standardized data, however, makes possible economies of scale if the community can coordinate its activities.

Migration plans will need to recognize that there is a difference between "publish" and "migrate". Publishing existing data as library linked data will make limited use of linked data capabilities because the existing underlying data formats are built on prior data concepts. In particular, existing formats lack the ability to create many of the links that one would like. Migration is likely to be a multi-step process, perhaps publishing non-LD formats as RDF while encouraging libraries to include LD-friendly data elements in current data formats (e.g. MARC21 \$0 field for identifiers), then adding identifiers and relationships to that RDF. In addition, the data held in today's databases was designed to be coherent only within that database environment and does not interact with other data that might be found in the LD environment. The magnitude of this change will mean that it cannot be done as a single, one-time conversion; there will be many seemingly incomplete stages before the community arrives at a destination close to an idealized LD environment.

The length of time to perform the migration will be large because of the number of activities: emergence of best practices for LLD, creation and adoption of new software, consensus on global identifiers and deduplication strategies, and so forth. A plan must be drawn up that stages activities in ways that allow innovators to participate sooner while paving the path for the majority adopters to make use of the work later. Adoption of the plan would also reduce duplication of effort as the community moves from a self-contained, record-based approach to a worldwide graph approach for bibliographic data.

These tasks will require the cooperation of libraries and institutions in a broad coalition. The coalition will need to address some difficult questions. For example, not all institutions will be equally able to make this transition in a timely manner, but it will be important that progress not depend on the actions of a few institutions. The community must be allowed to move forward with new standards as a whole even where past practices have assigned development of standards to particular institutions.

Each of these possible paths have costs and benefits that should be studied and understood as part of the transition to linked data, taking into account the investment that libraries have in their current systems and economic factors. Concurrent with a plan to migrate data is the need for a plan to change data production processes to take advantage of linked data technologies.

Foster a discussion about open data and rights

Rights owners who are better informed of the issues associated with open data publishing will be able to make safer decisions. It makes sense for consortia with common views on the potential advantages and disadvantages of linked data to discuss rights and licensing issues and identify areas of agreement. A mixture of rights within linked data space will complicate re-use of metadata, so there is an incentive to have rights agreements on a national or international scale. For the perspective of UK higher education libraries, see the [Rights and licensing](#) section of the [Open bibliographic data guide](#).

[The principles at http://openbiblio.net/principles](http://openbiblio.net/principles) should be mentioned, if not recommended, also the [range of already existing endorsement: http://openbiblio.net/principles/endorse/](http://openbiblio.net/principles/endorse/).

Facilitate

Cultivate an ethos of innovation

Small-scale, independent research and development by innovators at individual library organization is particularly important, because small organizations have resources others don't, such as the freedom to make independent changes iteratively and close contact with internal and end-users. Sharing and reuse of these innovations is important, and it is particularly critical for innovators at small organizations, who may otherwise lack outlets for contact with their counterparts elsewhere. Communication of ideas and loose-knit collaboration across the community can save time and achieve common goals. Existing *ad hoc* communities such as [Code4Lib](#), [dev8D](#), and the [mashedUp series](#) provide support, networking, and information sharing for innovators. Developers and other innovators in these communities need to be further engaged and supported to grow libraries' capacity for problem-solving and innovation.

Research and development is also advanced at library and information-focused graduate schools, and through research-oriented organizations like ASIS&T and the Dublin Core Metadata Initiative, and in independent research groups like OCLC Research. Connections between such research organizations and individual libraries (especially small libraries, public libraries, and special libraries) could also be fruitful, both in translating research advances more quickly into production-level implementations and in directing research attention to new problems.

CREATION OF TOOLS [PEM]

Identify Linked Data literacy needed for different staff roles in the library

The linked data environment offers a very different perspective on metadata and its applications than traditional approaches. Obtaining best value from this environment requires orientation and education for professional staff interacting with metadata applications and vendors supplying metadata support infrastructures. This should be seen as an extension to existing knowledge and expertise, rather than a replacement of it. It is particularly important that decision-makers in libraries understand the technology environment well enough to make informed decisions.

Include metadata design in library and information science education

The principles and practice of Linked Data offer a fundamental shift in the way metadata is designed. To prepare future professionals in the creation of new services, metadata design should be included in professional degree programs. Topics could include evaluation of datasets and element sets with regard to quality, provenance, and trust, and Semantic Web modeling, with special attention to simple inference patterns and the semantics of data alignment.

Increase library participation in Semantic Web standardization

If Semantic Web standards do not support the translation of library data with sufficient expressivity, the standard 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 expressing concept coordination, LLD implementers should consider devising solutions within the idiom of Linked Data -- i.e., on the basis of RDF and OWL. In order to ensure that their structures will be understood by consumers of Linked Data, implementers should work in collaboration with the Semantic Web community both to ensure that the proposed solutions are compatible with Semantic Web 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 notions of named graphs and provenance, by joining working groups and participating in public review processes.

Design

Translate library data, and data standards, into forms appropriate for Linked Data

In the library environment, conformance to conceptual models and content rules has traditionally been tested at the level of metadata records, the syntactic conformance of which can be validated. As with natural language, there is more than one way to "translate" such models and constraints into the language of Linked Data. In an OWL ontology, for example, content rules may be expressed as "semantic" constraints on properties and classes, while an application profile (in the Dublin Core style) "uses" properties and classes, their semantics untouched, with "syntactic" constraints for validating metadata records. RDF data can also differentiate natural-language labels for things and identifiers (URIs) for the underlying things themselves -- a distinction relevant when translating authority data for subjects or personas, traditionally represented by text-string labels. In order to make informed choices between the design options, translators of library standards should involve Semantic Web experts who can verify whether the translations correctly convey the translators' intentions, and they can make the results of that process available for public comment and testing before widespread implementation is undertaken.

Develop and disseminate best-practices design patterns tailored to LLD

Design patterns allow implementers to build on the experience of predecessors. Traditional cataloging practices are documented with a rich array of patterns and examples, and best practices are starting to be documented for the Linked Data space as a whole (e.g., <http://linkeddatabook.com/editions/1.0/#htoc61>). What is needed are design patterns specifically tailored to LLD requirements. These patterns will meet the needs of people and developers who rely on patterns to understand new techniques and will increase the coherence of Library Linked Data overall.

Design user stories and exemplar user interfaces

Obviously the point of library linked data is to provide new and better services to users, as well as to allow anyone to create applications and services based on library data. Because the semantic web is new it isn't going to be possible to predict all of the types of services that can be developed for information discovery and use, but the design of some use cases and experimental user services are necessary to test library data in this environment and to determine fruitful directions for development activities.

Identify and link

Assign unique identifiers (URIs) for all significant things in library data

There are shared things... subject headings, data elements, controlled vocabs, that all need identifiers. The actual records in library catalogs that would be shared also need to be given identifiers, *although these may be local, not global, in their "range"*.

??? This is confusing, at best. There are no local identifiers in the semantic web. I would merge the section with the next about the creation of URIs. Creation and assignment can not be distinguished.

Create URIs for the items in library datasets

Library data cannot be used in a linked data environment if URIs for specific resources and the concepts of library standards are not available. 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. To avoid proliferation of URIs for the same thing and encourage re-use of URIs already assigned, 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 URI to others.

Some libraries or library organizations should play a leading role in curating the RDF representations of library metadata elements, including URIs, in a similar way to existing patterns of standards maintenance, where a specific organization acts on behalf of the community. Such roles should operate in a more cross-domain environment, to reflect the networking infrastructure of linked data. Agencies responsible for the creation of catalogue records and other metadata, such as national bibliographies, on behalf of national and international library communities should take a leading role in creating URIs for the resources described, as a priority over publishing entire records as linked data, to help local libraries avoid creating duplicate URIs for the same resource.

Namespace policies should be documented and published in a way that allows users of the URIs and namespace to make safe choices based on quality, stability, and persistence.

Create explicit links from library datasets to other well-used datasets

Libraries should also [assign URIs for create](#) relationships between their things, and between their things and other things in LD space. Without these relationships library data remains isolated, much as it is today. [Where no suitable vocabulary exist to specify these relationships, new relations have to be created.](#)

Directly use, or map to, commonly understood Linked Data vocabularies

In order to ensure linkability with other datasets in the cloud, library datasets must be described using commonly understood vocabularies. Library data is too important to be relegated to an "RDF silo" -- an island of information described formally with RDF, but using vocabularies not familiar to less specialized Linked Data consumers. If library data is described using library-specific vocabularies, then those vocabularies should, to the extent possible, be mapped to (aligned with) well-known RDF vocabularies such as Dublin Core, FOAF, BIO, and GeoNames ([TODO: Add Links](#)). Alternatively, the maintainers of library-specific vocabularies should promote the widespread use of their vocabularies in mainstream Linked Data applications.

Existing value vocabularies ([Maybe a reference to the section, where the distinction between vocabulary and value vocabulary is explained? I assume that some people read the recommendations first.](#)) for entities such as places, concepts, events, and persons should be considered for use in LLD. Where library-specific value vocabularies are created, either by translation from legacy standards or through new initiative, their widespread use and alignment with other vocabularies should be promoted.

Prepare

[Prepare should be before „Identify and Link“.](#)

Develop best practices and design patterns for LLD

Reusable design patterns for common data needs and data structures will be necessary to achieve some heterogeneity of metadata across the community. Design patterns will facilitate sharing, but will also create efficiencies for data creation by providing solutions to common problems and narrowing the range of options in development activities. These design patterns will also facilitate large-scale searching of these resources and dealing with duplication of data in the linked data space. Best practice documentation should emerge from this development to communicate LLD patterns within the community and among the broader users of LLD.

This exists two times, I would merge it with the best practice section under „Design“.

Commit to best-practice policies for managing and preserving RDF vocabularies

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. Policies encourage a consistent, coherent, and stable approach which improves effectiveness and efficiency. quality assurance for users of URIs and their namespaces. Policies might cover

- Use of patterns to design URIs, based on good practice guidelines.
- Persistence of URIs.
- Good practice guidelines and recipes for constructing ontologies and structured vocabularies.
- Version control for individual URIs and the namespace itself.
- Use of HTTP URIs (URLs) for the identification of library elements, vocabularies and bibliographic data, in keeping with the principles of the semantic web.
- Extensibility of use of the namespace by smaller organizations.
- Translations of labels and other annotations into other languages.'

Identify tools that support the creation and use of LLD

Easy-to-use, task-appropriate tools are needed to facilitate library use of linked data. Both generic linked data tools (e.g. a URI generator which facilitates the creation of URIs) and custom domain-oriented LLD tools are required (e.g. a MARC-to-RDF converter; tools incorporating LLD-related vocabularies in easy-to-use ways). Domain-oriented tools should be based on mainstream technologies, so should be adapted from generic tools and regularly maintained so that library approaches don't diverge too far from the mainstream. Developing appropriate tools will require identification of the necessary tasks (including programming, metadata development, metadata instance creation, and end-user browse and search), the needs and technical abilities of the tool users, as well as the resources available. Tools for non-programmers are especially important, and should adopt an appropriate technical level (i.e. that of a general computer user) and terminology (i.e. terms familiar to the user), including providing help and direction where decisions must be made.

Curate

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

Much of the content in today's Linked Data cloud is of questionable quality -- the result of ad-hoc, one-off conversions of publicly available datasets into RDF and not subject to regular accuracy checks or maintenance updates. With their ethos of quality control and long-term maintenance commitments, memory institutions have a significant opportunity to take a key role in the important

(and hitherto ignored) function of linked-data curation -- duties by which libraries and archives have proven their worth, extended to a new domain. By curating and maintaining Linked Data sets, memory institutions can reap the benefits of integrating value-added contributions from other communities. Data from biographers or genealogists, for example, would immensely enrich resource descriptions in areas to which librarians traditionally do not themselves attend, greatly improving the possibilities for discovering and navigating the collections of libraries, [archives, and museums, and archi-](#)

Preserve Linked Data vocabularies

Linked Data will remain usable twenty years from now only if its URIs persist and remain resolvable to documentation of their meaning. As keys to the correct interpretation of data, both now and in the future, element and value vocabularies are particularly important as objects of preservation. The vocabularies used in Linked Data today are developed and curated by maintainers ranging from private individuals to stable institutions. Many innovative developments occur in projects of limited duration. Such a diverse ecosystem of vocabularies is in principle healthier than a semantic monoculture, but most vocabularies reside on a single Web server, representing a single point of failure, with maintainers responsible individually for ensuring that domain-name fees are paid and that URIs are not re-purposed or simply forgotten.

This situation presents libraries with an important opportunity to assume a key role in supporting the Linked Data ecosystem. By mentoring active maintainers with agreements for caching vocabularies in the present and assuming ownership when projects end, maintainers retire, or institutions close, libraries could ensure their long-term preservation, much as they have preserved successive print copies of Dewey Decimal Classification since 1873. With help from automated mirrored-cache technology (such as the [LOCKSS system](#)) libraries could, in the short term, ensure uninterrupted access to vocabularies by maintaining versioned snapshots of a vocabulary's documentation at multiple sites, automatically coming online whenever a primary server should fail (see [Baker and Halpin](#)). In this role, libraries could provide protection against service outages and thus improve the robustness of Linked Data generally.