

Access Beyond Borders: Linked Open Data Applications on Cultural Heritage

Sümeyye Akça

To cite this article: Sümeyye Akça (2022): Access Beyond Borders: Linked Open Data Applications on Cultural Heritage, New Review of Information Networking, DOI: [10.1080/13614576.2022.2145334](https://doi.org/10.1080/13614576.2022.2145334)

To link to this article: <https://doi.org/10.1080/13614576.2022.2145334>



Published online: 17 Nov 2022.



Submit your article to this journal [↗](#)



Article views: 65



View related articles [↗](#)



View Crossmark data [↗](#)



Access Beyond Borders: Linked Open Data Applications on Cultural Heritage

Sümeyye Akça

Department of Information and Records Management, Faculty of Arts and Sciences, Marmara University, Istanbul, Turkey

ABSTRACT

The main mission of cultural memory institutions is to protect their collections and to provide access to a wider audience. World Wide Web technology offers different tools to fulfill this mission. Among a range of applications, the most effective is the search engines' search for metadata of collections. However, search engines are not always able to see all sources caused by various problems. Linked data is a publishing technique that uses standard web technologies to associate data with other data and makes them available on the Web. In available literature, it is seen that many memory institutions use Linked Open Data as a solution to the search engine problem with the main motivation of making their collections more accessible. However, these applications are still not at proficient level. In this study, Linked Open Data studies in the available literature were scanned and a process map was created based on these applications. These applications, which especially require technical knowledge and infrastructure, can be confusing for memory institutions that have not yet participated in the process. In this context, it is clear that the process map created as a result of this study will be enlightening for memory institutions. The most important point on the map is that an extensive decision-making mechanism should be formed with the stakeholders before starting the process and the implementation plan should be developed in detail over the problems of existing projects.

KEYWORDS

Cultural heritage; metadata; metadata schemas; linked data; linked open data

Introduction

The rapid changes in the means of access to information caused by the emergence of the World Wide Web have created opportunities to access information beyond traditional boundaries (Duval et al.; Eito-Brun 441). Technologies such as Hyper Text Transfer Protocol -HTTP, Extensible Markup Language – XML have facilitated access to previously inaccessible data and information universe. However, information access, protection, and data processing processes depend on the identification of metadata. Metadata refers to the data that we can store about an object and manage it later, as well as the relationship data that this object maintains with other entities and the

CONTACT Sümeyye Akça ✉ sumeyyesakca@gmail.com; sumeyye.akca@marmara.edu.tr 📧 Department of Information and Records Management, Faculty of Arts and Sciences, Marmara University, Istanbul, 34722, Turkey

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

subject to which it is related (Eito-Brun 441–442). In a way, metadata plays an important role in expanding the original target audience of accessing content through tools available on the Internet (Foulonneau and Riley 127).

As there are large search engines and digital libraries on a global scale, there are more and more subject or genre-oriented different portals, digital libraries and information systems available on the Web. Moreover, metadata schemes have developed in line with the needs and purposes of these areas on the Web, each of which has a different background in terms of discipline, that is, branch of knowledge (Sicilia 21).

These metadata schemes applied on cultural and visual objects have been developed to describe information sources from Memory Institutions (Galleries, Archives, Libraries and Museums – GLAM Labs). GLAM Labs should be recognized as significant centers in a society to represent an open, transparent, collaborative, ethical and especially accessible culture for researchers, users, the public and all other stakeholders. However, these centers have varied and complex data in their collection and integrating that data, especially across institutions, has always been a challenge (Alexiev). Along with the burst of digital library, the complexity of the data in this sector and the problems encountered. Digital library is a concept that includes digital collections, digital repositories, or digital information services focusing on cultural and scientific digital content (Sicilia 22). In the context of this wide-ranging knowledge, many metadata standards have been developed to organize and describe information, especially from libraries, archives or museums, as well as from the fields of art, archeology, and architecture. In most cases, cultural collections contain heterogeneous characteristics with unique features such as written texts, photographs, physical objects, sound recordings, maps and even born-digital materials (Schlötterer et al.). At the same time, a single object can have rich meanings and complex structure. For example, while the Tortoise Trainer painting has the characteristics of a picture, it can also be said that many objects related to the painting, such as books and brochures, exist at the same time.

Because of that features of the cultural heritage objects there are many addressed access problems in the literature (Kando and Adachi; Kanhabua et al.; Signore) regarding the inability to access cultural heritage via the Internet by wider domains. First is about low quality of metadata content. It is seen as a big problem that the accessed metadata is not sufficient to meet the needs of the User (Agosti et al.; Berardi et al.; de Boer et al.) and the inadequacy of metadata in terms of quality also causes access problems. Based on this problem, Seifert et al. worked on a dataset consisting of 400 objects from the European Digital Library (Europeana) to enrich the digital metadata content. They connected each object in this dataset with the content of Wikipedia and evaluated the previous and the next system for performance.

The second problem is addressed as difficulties in accessing metadata. Accordingly, the fact that the users do not have information about the collection causes the search queries to be incomplete. The occurrence of wrong queries as a result of the lack of knowledge about specific and commanding words related to the cultural heritage terminology prevents the matching of the sources (Parikh et al.). Kollia et al. proposed a system to solve this problem that has been addressed in the available literature. Designed to provide information to the non-expert users more effectively, the system consists of three components: personal metadata, semantic enrichment, and query response techniques. In addition, Tonkin and Tourte proposed an open data source to understand and interpret metadata content. The heterogeneity of metadata also emerges as a separate problem (Candela et al.). To address this problem, suggestions have been made to ensure semantic interoperability of heterogeneous metadata (Candela et al.; Liao et al.; Xu et al.).

Today, there are many studies to enrich the metadata content and make it more accessible. Rich metadata content and interoperability of metadata emerge as an active research area for cultural heritage collections. Attempts to use Linked Open Data – LOD along with metadata also enable users to reach more satisfactory information about the cultural heritage products they access. This information also creates access of the users.

The aim of this study is to examine the existing approaches in the application of metadata, which is one of the important factors that increase access to cultural heritage, and to evaluate the role of these approaches in the context of the available literature in order to expand the target audience of cultural heritage collections. In line with this purpose, the concept and principles of metadata have been examined comprehensively in this study, first of all. Then, standards and problems in accessing cultural and heritage collections via metadata have been discussed within the framework of metadata components. After giving examples from contemporary applications, the use of LOD – a semantic-based application designed to enable collaboration and the sharing of a wide range of information among cultural memory institutions- on metadata and its access benefit have been analyzed. Finally, a sample process map has been created on how, in the context of metadata, memory institutions benefit from linked open data applications to make their collections more visible.

The target audience of this study is defined as the managers of cultural memory institutions, their employees, policymakers and academicians conducting research on this subject.

It is one of the goals of cultural heritage management to discover rich resources in cultural memory institutions and to make them accessible to a wide range of people. With the widespread use of the web, it has become essential to produce policies according to this goal. It is known that publishing metadata as LOD creates an important traffic in order to make collections in

cultural memory institutions more visible at the point of discovery of resources.

Method and research questions

In this study, the descriptive method, which tries to describe and explain “what” events, entities, institutions, groups and various fields are, and to reveal their current situations, conditions and characteristics in a precise manner (Johnson 241).

In the first stage, the LOD applications constructed upon libraries or other memory institution collections of the world have been examined. Then the objectives of how memory institutions can make the existing cultural collections more visible in the internet environment have been determined in the context of existing access problems and the wider user audience brought by the up to date applications. Also, the steps to be taken have been determined while forming a process map.

In this context, the questions of the study are as follows:

- To what extent do cultural memory institutions use linked open data through the metadata of the collections?
- Does the linked open data applied on metadata of cultural collections create a difference in the access of collections?
- How are the metadata of cultural memory institutions prepared for the linked open data application?
- What should be the processes in the LOD applications of cultural memory institutions?

In order to answer these questions, previous research results, reports and applications related to the subject in the available literature will be examined and an easy process map will be created in order to use the linked open data in the metadata of the collections of cultural memory institutions in light of the data obtained in previous studies. This map will be particularly enlightening for memory institutions that have not opened metadata as linked data previously.

Metadata definition and its principles 2010

Along with the storage of the data, information about the data also had to be stored. Metadata entered the literature as data about data in the 1990s (Hay 1). This definition is a definition from computer science, and according to Caplan 1), the first usage of the word as data related to data can be found in the first edition of NASA’s Directory Interchange Format Manual published in 1988. In the early 1990s, the term metadata was used to mean the information

required to make computer files related to science, social sciences, and geospatial data sets useful for humans. With the rise of the Internet and the Web, this term has come to be used more in the context of describing information objects on the network. IFLA defines metadata as any data used to help identify and locate networked electronic resources (International Association of Library Associations and Institutions).

It is necessary to accept metadata as the descriptor of any information source, including printed publications as many collections consist of both digital and non-digital objects and there are many resources in the same collection that are later digitized. In fact, the World Wide Web Consortium – W3C has explained the current function of metadata in the narrowest sense. In this description, metadata is defined as machine-understandable information for the web (Metadata and Resource Description).

It is also seen that the definition of metadata is made differently depending on the community and content in which it is used. As the sum of all these usages, metadata is actually used to mean structured information about an information source in any medium or form (Baca; Sicilia 17). Librarians defined metadata as data that eliminates the need for a user (human or machine) to have fully developed knowledge of the existence or characteristics of potentially interesting things in the environment (Dempsey; UKOLN, t.y). It is the variety of information a user contains about an object, whether it is a human or a machine. It can be used to describe many different types or levels of entities ranging from abstract concepts to physical objects. Metadata enables various functions to be performed on digital resources such as discovery, interpretation, protection, management, representation, and reuse of objects (Foulonneau and Riley 4–6).

However, there are two criteria for which information can be considered as metadata. First, the information is expected to be structured, that is, a set of randomly collected data items cannot be considered as metadata. This information should be recorded in accordance with standardized metadata schemes. The other is that metadata should define a source of information (Caplan 3).

The National Information Standards Organization – NISO sees the functions of metadata as a general method used to classify metadata. For example, it states that metadata are used functionally in discovery of resources, organizing electronic resources, ensuring interoperability between institutions and resources, digital identification and archiving, and protection of resources (NISO). In general, it is seen that metadata are used to discover and use resources (Campbell). But today wide access to cultural heritage objects in a semantic dimension in the Web environment as can be seen from the current problems. At this point, the metadata used should have some features such as modularity, expandability, multilingualism, and

curability. Metadata should be created in a flexible structure to meet any specific (local need/site-specific need) requirement that arises over time. Although many schemes and standards have been developed for accurate and precise recording of metadata in time (Ullah et al.), the innovations in technology allows metadata schema designers to create new assemblies (attachments) based on built-in metadata schemas and leverage observed best practice, rather than reinventing elements. In addition, this feature of metadata schemes is important to realize the principle of interoperability (Duval et al.).

Interoperability, at its most basic level, is the conversation or dialogue between different systems. Linking content and applications over networks on the basis of collaboration and collections between multiple institutions requires definition (Caplan 33; Eito-Brun 443; Foulonneau and Riley 119). In order for metadata to work together, it must be created in a certain standard that is semantic and syntactic. While semantics can be expressed as a definition of what metadata items mean, the syntax of a standard is about how things are encoded in a machine-readable format (Duval et al.; Eito-Brun 447–449). There are some crosswalks for ensuring the ability to exchange information between the standards. One of them, Resource Description Framework – RDF creates an architectural shape for Web metadata and has been developed to provide the infrastructure for semantic web activity in W3C (Resource Description Framework). Interoperability is the most effective steps to communicate networks between systems. In order to increase the discovery of resources, cultural memory institutions should transmit their metadata to any collective medium or allow these collective media to extract their own metadata (Foulonneau and Riley 19).

Exploration of resources

Cultural memory institutions have developed numerous online services to highlight the value of their content and allow users to interact with their collections. These services generally remain local. It has to ensure that metadata of the collections of cultural memory institutions can be seen by the spiders of the search engines. However, search engines cannot always explore all resources. The spiders discover new resources to be indexed by usually following links between web pages. This mechanism is useful for static websites where all pages of the website are linked on one page because software robots follow links to discover new pages and index them. However, many institutions such as museums, libraries, and archives use Web resources stored in repositories. In such repositories, resources are discovered through internal directories rather than links. This means that search engine robots cannot index a document without an explicit link. The set of resources found on the web that has been dynamically created and not indexed by search engines is

called the “Deep Web” (Foulonneau and Riley 131). Cultural memory institutions are also included in this definition.

There are some common approaches that content providers can take to make their collections more search engine friendly. The use of Title, Description and Keyword tags can support efficient search engine processing of the resource. Creating a sitemap or an OAI-PMH data provider for resources is also recommended as methods to ensure that a search engine can find all relevant content of an institution (Foulonneau and Riley 134).

Another practice of the cultural memory institutions is to send metadata to third-party search engines in order to make their content more visible for example Europeana. Set out as a European digital library, this application shares cultural heritage items from thousands of European countries’ archives, libraries and museums (Europeana). One of the main motivations for institutions to publish their content on the Web is to increase the discoverability of their resources by both existing users and new users (Foulonneau and Riley 127). Metadata sent to third-party search engines attracts attention in terms of adding value to local content by increasing discoverability. Institutions’ sharing of their collections on third-party search engines provides a broader perspective on the policies and goals of cultural memory institutions. For example, the Australian National Library has created a common database of Australian government records to satisfy curiosity about the past, present, and future of Australia’s history, social life and many important issues with the contribution of many institutions (National Archives of Australia).

In recent years, the system that provides better availability and interoperability of systems by search engines is accepted as Linked Data (Heath and Bizer). In this context, Cultural Memory Institutions convert their metadata to LOD for creating a data network and providing more discoverability on their collection.

Linked data and linked open data applications

Data aggregation technologies in cultural memory organizations are not the same as any other internet data used to access information. In the cultural heritage sector, institutions are undertaking more sustainable data gathering initiatives for their collections. For example, institutions increase the number of potential users by sending metadata of their collections to organizations such as Europeana, Digital Public Library of America – DPLA, DigitalNZ, Trove, and Digital Library of India. However, this data collection is a costly task for cultural memory institutions. In order to reduce the cost, there are actually applications found in the literature. One of them is Linked Data – LD. The process of sharing their data with cultural heritage data collectors will be extremely simple for those implementing LD in digital libraries. For those who will be new to the LD application on their collection, institutions will gain

extra benefit from this situation as this application provides interoperability with areas other than cultural heritage. For example, a significant number of linked data sources from many cultural heritage sources were successfully collected in the Research and Education Space – RES project (van der Steen). It can be said that the most popular one of this initiatives is Wikidata. Wikidata has been a pioneering app in the implementation of linked data and removes many of the challenges posed by other connected data environments. It encouraged a wider user base by supporting the data model that anyone can edit. It also allows the creation of connected data projects that are not limited to the availability of technical skills such as programming and data management, and enables wider participation of people with greater expertise in the content (Allison-Cassin et al.).

LD is a publishing technique that uses standard web technologies to link data and make them available on the Web (Hallo et al. 2). Cultural memory institutions should develop LOD applications and enable researchers to access embedded data through cross-collections. Accessing old information is an important benefit of LOD applications (Mayer 10). Essentially, LOD allows users to interrelate communication about artifacts without requiring a curator's interpretation (Heath et al.).

In addition, as the traditional goals of cultural memory institutions have been to share resources and advance human knowledge the goals overlap with applications of LOD. Thanks to this technology, data transfer can be realized by ensuring interoperability of metadata between institutions (Davis and Heravi). LOD applications are considered important for the reason that they allow memory institutions to share the metadata they create and also for the enrichment of this data by users. Otherwise, the standards that memory institutions use to store and later access metadata create difficulty in retrieving data. On the other hand, most of the metadata standards used by memory institutions are not used outside of memory institutions. LOD is a standards and technologies framework that can complement existing vocabulary and ontologies in memory institutions. In other words, libraries can continue to use their existing standards while creating and publishing LD (Alemu et al.). There is also a connected data vocabulary (see <http://rdvocab.info/>) that helps to highlight the library-specific metadata.

Berners-Lee, talked about four principles for applying this LOD that the existence of open data regardless of the format registered with an open license on the web, the data being structured machine-readable, the use of open standards of W3C (RDF and SPARQL) to describe the data, and linking your data to other institutions' data for context and semantic integrity. In addition, Hendler and Berners-Lee's five-star model is also taken as the basis for LOD applications. American Art Collaborative proposes a series of steps to make the data suitable for LOD. Preparing a complete data set, associating it with an existing ontology, mapping the data to an open machine-readable

standard, preferably RDF, and linking these data from external data repositories to headers whenever possible (Folsom et al.).

Leading memory institutions in the world benefit from LOD facilities to open their metadata to mass access. The Library of Congress have an initiative to transfer library catalogs to LD (The Library of Congress Bibliographic Framework Initiative). The work that started as an attempt to export MARC21 (Machine-Readable Cataloging21) records to linked data was defined as the BIBFRAME model. According to this model, MARC21 records are divided into four basic classes: study, example, authorization, and explanation. The BIBFRAME dictionary is published as a single namespace and the model is defined in RDF. After the conversion of the MARC21 format to the linked data model, a long testing period was conducted. Work continues to develop tools and support services (Hallo et al. 5; Library of Congress). The Library of Congress Linked Data Service, since 2009, has published named authorities, topics and other LC vocabulary and standards in LC cataloging as LOD. URIs have been published for the vocabulary and the singular values they contain. The subjects of the Library of Congress, which has published 260,000 authority records so far, link to Wikipedia article titles (Hallo et al. 5; Library of Congress; Marden et al.; Mayer 6).

Other national libraries are also active in the field of LOD. The British National Library has published a version of the British National Bibliography as LOD including both monographic and serial publications. Unlike previous attempts, they have modeled interesting topics such as people, places, and events about any book without converting MARC collections to RDF/XML (Bizer et al.). The library mainly uses linked data such as Virtual International Authority File – VIAF, Library of Congress Subject Headings – LCSH, Lexvo, GeoNames (for country of publication), MARC (for country code and language), and RDF Book Mashup (books, authors, reviews and semantic Web, and an app to learn about online bookstores), Dewey.info (for linked Dewey Decimal Classification data) (Deliot 174). BBC televisions have also published the archive as LOD to make the large archive they have more usable (van der Steen). The German National Library partners with many other institutions in LOD projects and has established its own Linked Data service, which includes both bibliographic and authority data (DNB). There is an ongoing work in the National Library of France (Bibliothèque Nationale de France) to make the resources more usable. Aiming to collect resources from different catalogs in a single hand using the semantic web, the institution has combined metadata in a single RDF format, such as the INTERMARC bibliographic format for the main catalog, XML-EAD (Encoded Archival Description) for the inventory files, and Dublin Core for the digital library. These resources were assigned a continuous marker (URI), converted to RDF triplets into a single format and published as LOD (BNF). URIs are created based on Archival Resource Key identifiers. Most of the LD in the National Library of France is based on the

Functional Requirements for Bibliographic Records – FRBR model. This model website creates web pages that provide standard information, references, and links about authors, works and themes. In addition, the vocabulary used is Rameau and it has been transformed into a version of the Simple Knowledge Organization System – SKOS vocabulary that reveals approximately 160,000 concepts (Wenz, “Linked Data and Librarie,”; “Linked open data for new library services”). CubicWeb has been used as a software program for this. This open-source software collects the resources defined in any standard and creates pages in the form of different file types (BNF).

The technique used in these initiatives is common. To make their collections more accessible, they convert their metadata first to the RDF triplets and then as LOD. Together with leading cultural memory institutions, applications on a smaller scale are also encountered in the literature aimed at to see the effect of LOD. Westrum et al. took a subset of MARC records translated into FRBR representing the work of two designated authors and converted them into RDF. From there, the metadata were linked to the data in the LOD cloud, allowing the library to get benefitted from this new enriched metadata for its users and produce new applications (<https://lod-cloud.net/>). In parallel with information technologies, the contents used in the initiatives can also differ. In one study, user comments and tags were associated with LOD also added when user comments are linked to metadata (Holgersen et al.). In another study, Johnson used JSON-LD, BibJSON and Elasticsearch to create a robust index combining cross-linked open data sets. There are also wide spread project incorporating linked authority registers for named entities from different regions and countries, creating a unified supervisory record set like Virtual International Authority File – VIAF, the Social Networks and Archival Context – SNAC, DBpedia, Bibliographica, The Linked Open Copac and Archives Hub Project (DBpedia; VIAF: The Virtual International Authority File; Yoose and Perkins).

It is also applied in museums as one of the cultural memory institutions. The Smithsonian American Art Museum (SAAM) project is also a joint study with the Information Sciences Institute and the Department of Computer Science at the University of Southern California. In the study, the museum aims to make its entire database of 40,000 records available as a five-star Linked Open Data. It is envisaged that the five stages, which are intended to be a model project for other museums, consist of five stages: data preparation, ontology development, RDF mapping, connecting to central data sets and publishing data. Ontology is taken from Europeana Data Model, SKOS, Dublin Core, RDA and schema.org. Mapping was done with KARMA, an open source data integration tool developed by ISI. Links have been made to Union List of Artists Names, GeoNames, social media sites and other museums using external centers such as DBpedia (2,194) and New York Times (70) (; Mayer 9; Szekely et al.). In the study that collected data from

43 National Museums in Thailand, the Thai Open Museum system was created by correlating these data with the LOD (Chanhom and Anutariya).

The Getty Research Institute, National Gallery of Art, and Rijksmuseum have both opened and linked the content as open data. Although these institutions publish open data on their websites, they have also encouraged their users to improve the data (Getty Vocabularies as LOD; The J. Paul Getty Trust; The Rijksmuseum). In particular, the Getty Institute has published controlled open data lists such as The Thesaurus of Geographic Names (TGN), the Art and Architecture Thesaurus (AAT).

Step by step LOD

The importance of LOD cannot be denied for the metadata of cultural memory institutions to be more discoverable on the Web. Therefore, the steps to be taken by memory institutions in order to open their collections have been emphasized in this section. Transferring the metadata to LOD in sustainable way comprise multiple components, such as software and hardware, staff to build and operate them, expertise and work practices, tools, and methods. The process to be applied when cultural memory institutions decide to open their metadata requires business and sustainability plans. How to generate repositories and minimize costs to maintain the repository long - term are the cardinal to archive the mission they idealized. One approach is grant-funded projects in which cultural memory institution collaborate with stakeholders to address their project needs.

The first step is the planning phase, in which the preparations for the work must be completed. This study should be planned as a multi-element study that should be carried out by cultural memory institutions with appropriate funding or support programs. Linking metadata to other sources involves uncertainty and involves various levels of reference or cooperation (Berners-Lee). The inclusion of the Ministry of Culture and Tourism, to which cultural memory institutions are affiliated, as stakeholders, is important in terms of the dissemination of the study and its adoption as a policy. A working group should be established by identifying all these stakeholders. It is necessary to have a computer programmer or engineer in the team who will carry out the technical infrastructure and operations. It is also important to put all the processes on paper and to plan the timeline. It is necessary to explain to the stakeholders how the process will be implemented and what will be its purposes and benefits.

While planning the process, ROI assessments should also be made. The money to be spent should be calculated considering the expected return at the end of the job as it has not been proven yet whether open data applications turn into additional budgets as revenue. It is necessary to decide on the material to be opened, especially considering the issues such as copyright,

and also be mindful of which memory institutions to be included in the work. It would be more meaningful to work on a data set that will benefit in terms of impact. It may be problematic for a small-scale memory institution with modern art and design objects to open its collection both due to lack of material and copyrights issue (Mayer 13). The determination of the collection should be done together with the stakeholders. It is necessary to conduct analyses such as who is currently using the data, which user audience is targeted, and a decision-making process with broad participation should be established by sharing the results with all stakeholders.

Before the creation a linked data model that will represent the data objects, data clean up and normalize process should be taken into account. The suitability for use of the data to be converted into LOD indicates the quality of the data. At this stage, problems such as data consistency, accuracy and conflicting data problems are encountered. For example, problems related to accuracy can appear in two ways, syntactically and semantically. Syntactic errors are frequently encountered, as well as semantic problems such as the representation of data values in the real world (Ahmed). The data model to be applied according to the current metadata standard should be selected while deciding on this model. The standards of the W3C should be taken into consideration in the selection. The expectations of the users should also be taken into account in the creation of the model strategy. In addition, the choice of licenses for open data should be determined at this time (Creative Commons, Open Foundation Standards). An important part of the LOD implementation is the inclusion of Cross-Origin Resource Sharing (CORS) headers in websites and APIs to be linked (Van Kesteren).

After all this preparation phase, the metadata of the decided collection is determined and necessary transformation is planned. The steps to be taken before converting to RDF – XML format are shown in the model in [Figure 2](#). These circuit steps may differ depending on the currently used standard. This model is mainly designed for MARC21 metadata. However, if the available data are defined in a different standard, it will be necessary to determine an action accordingly. For Linked Open Data transformation, all metadata must be converted to RDF standard. RDF is a language created to display information about resources in the [WWW](#). Metadata in RDF consists of triplets containing subject, object, and predicate (see in [Figure 1](#).) (Manola et al.; Resource Description Framework – RDF). Among the RDF triplets, the subject refers to the source provided with metadata, the predicate refers to a property (metadata element) of this resource, and object refers to the value attributed to this property (Eito-Brun 471).

Creating a URI strategy for linking data should also be considered at this stage. Creating a plan suitable for HTTP URIs and providing sustainable URIs is the basic step point. Whether standard databases will be used to link

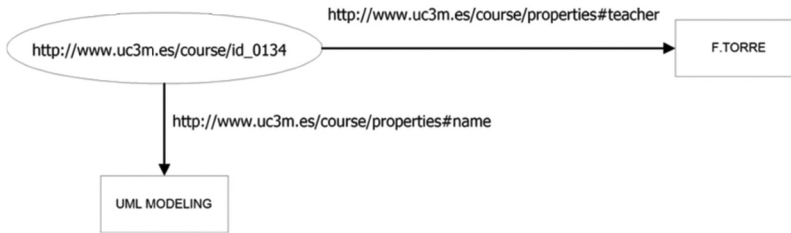


Figure 1. A simple RDF trio (Source: Eito-Brun 471).

metadata to open data, or will new words be created, these should also be carefully selected.

There are various standards and protocols that can be used to retrieve RDF data. The most widely used standard for retrieving RDF data is Simple Protocol and RDF Query Language (SPARQL), a language that can be used to search and retrieve RDF results via HTTP or SOAP. (RDF, 2021). SPARQL is similar to other query languages, such as SQL, but is designed specifically on RDF trios. The results obtained from the SPARQL query can be presented in three formats as XML, CSV/TSV and JSON. This choice in result formatting provides flexibility in application development while processing data from a specific query (Sicilia 7).

To publish metadata as triple RDF, Dunsire suggested nine basic steps: Take a record and determine its local identifier. Parse into single expressions in subject-predicate-object syntax. Create a URI for the record – that is, define it by appending its identifier by the record as a local part to a base domain. Replace the record ID with the URI as the subject of parsed expressions. Find URIs for attributes – that is, find properties that match the attributes. Replace attributes with URIs as predicates of parsed expressions. Find URIs for values – that is, identify what strings are labels. Replace values with URIs as objects of parsed expressions. Finally, publish the triplets as linked data.

While converting data into RDF trio, open-source applications can be used. Open Refine, formerly known as Freebase Gridworks and Google Refine, is an open-source application that can clean and standardize scattered data, connect data to other public databases and export it in your preferred structure. For example, if you have a spreadsheet metadata table exported from an old database, Open Refine can take most of the steps required to apply that spreadsheet to open linked data. In addition, it has an extension that exports data as RDF and a structure that transfers the words created by users to the RDF data set with its vocabulary (OpenRefine).

Another standard applied in linked open data studies is Web Ontology Language-OWL. OWL is a standard used in the semantic web using ontology to explain the knowledge and relationship between objects. An OWL standard uses axioms, which are the basic expressions expressed by an OWL ontology,

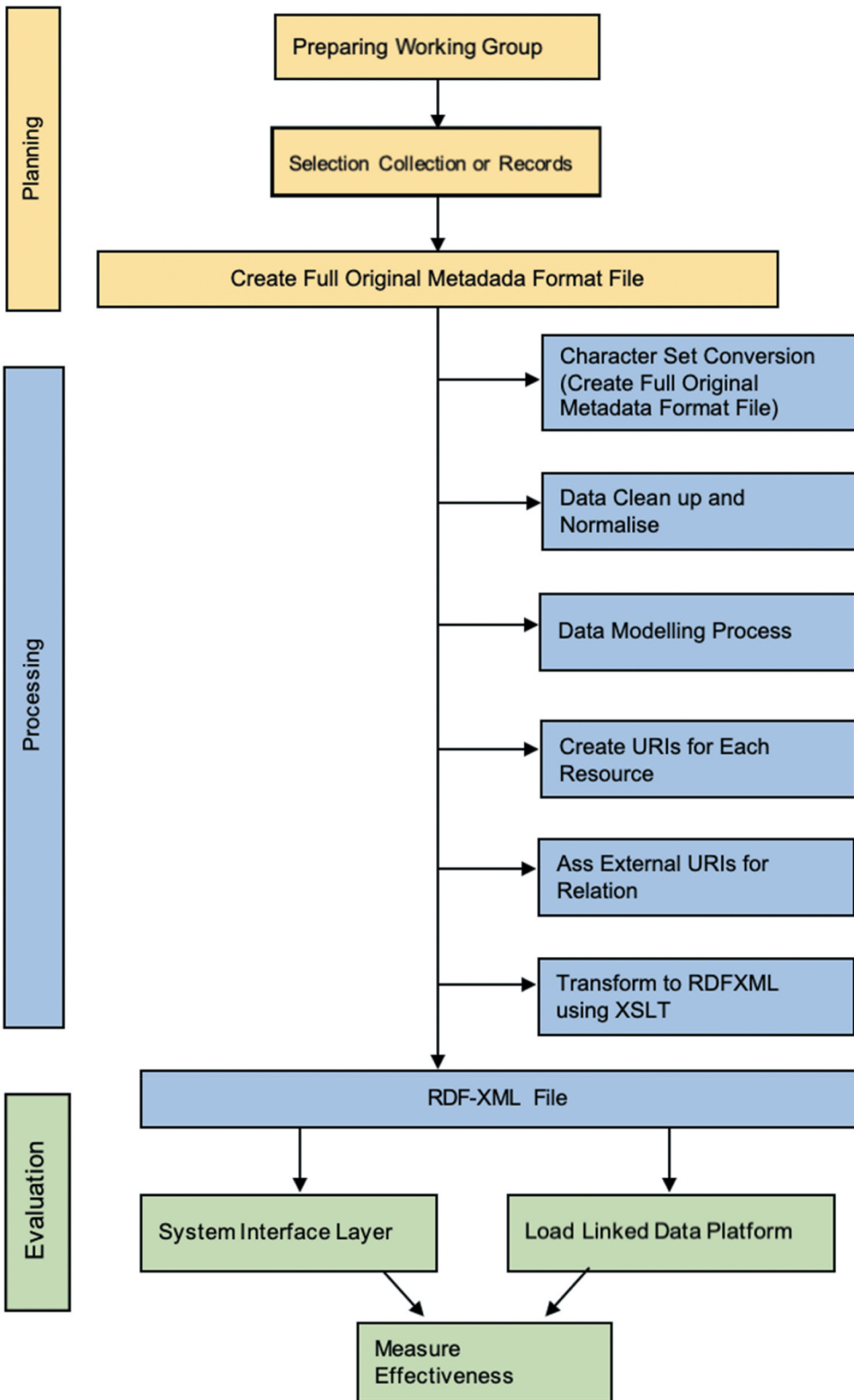


Figure 2. The process of adapting the metadata of cultural memory institutions to LOD.

entities used to refer to real-world objects, and combinations of entities to create complex descriptions from the underlying knowledge (Hitzler et al.). Both RDF and XML lack a solid vocabulary for expressing linked relationships. However, OWL provides a more detailed expression of the relationships between classes and properties. An OWL ontology can be created more serially with standards such as RDF triplets and XML (Yoose and Perkins).

There are many websites that present open-source data to users for linked data application. The Data Hub (<http://datahub.io/>), Open Metadata Registry (<http://metadatarregistry.org/>), Linked Open Vocabularies (<http://lov.okfn.org/dataset/lov/>) and DBpedia (<http://www.wiki.dbpedia.org>) are some of them. The disclosed data can be associated using the vocabulary here.

At this stage, the reports of libraries or memory institutions that have implemented Linked Open Data applications in the world should be examined in detail. Considering all these points, the processes of a simple LOD application for cultural memory institutions can be seen in [Figure 2](#).

Discussion-conclusion

The accessibility of cultural heritage collections for the development and promotion of creative thinking and the advancement of science and art should be seen as the most fundamental democratic right. In order to achieve this, cultural memory institutions have to play the game according to its rule in this internet age that we are passing through. Especially in the Covid-19 pandemic process, digital resources and access to them have become more important. Cultural memory institutions, whose social responsibilities form the basis of their basic policies, are also responsible for fulfilling their duties.

The collections defined for various reasons cannot be discovered on the web. Although many approaches have been developed to solve this problem, it is quite useful to benefit from the semantic side of the web. Singer stated that the use of linked data should become widespread in the field of librarianship. Singer has argued that metadata generated and maintained by libraries will be useful to both the libraries themselves and wider audiences on the Web. Although it has progressed slowly since then, linked data has been applied to library catalogs (Alemu et al.). Metadata produced by librarians is only available in databases that are dynamically formatted pages on the web when searched (Caplan 48).

However, search engines do not index these dynamically formatted pages. Coyle states that this means that the library data will not take place in the information environment on the Web and this limit the visibility of libraries on the Web. In other words, library metadata does not appear in search results submitted to users by search engines. Associating metadata of libraries or other memory institutions with linked data means that this metadata can be

scanned by search engine bots and included in search results presented to users along with titles or articles on platforms such as Wikipedia (Coyle).

It is known that serious studies have been carried out on this subject in memory institutions since 2009. However, it is a fact that the studies are still local and the semantic feature of the web cannot be used in full access and deeper understanding of the content. Despite the serious studies of the leading memory institutions of the world, the developments are still not at the desired level. Indeed, memory institutions' motivations to publish Linked Data include the difficulties of cataloging an increasing number of sources without increasing staff, but if metadata is published as Linked Data, it can be expanded and enriched from external sources and its reuse is also expected to increase (Hallo et al. 4). In addition to all these measurable effects, Linked Open Data applications gain importance in terms of data sustainability. It provides the user with the opportunity to easily extract information with other connections and interact with deeper and more effective information without the need to browse separately.

For cultural memory institutions, integration with Linked Open Data applications and initiatives such as Shema.org means more traffic in more quantity. It is a factor that improves the quality of definition and metadata. In addition, users' feedback about the resource is also important. A wider participation in the definition of the source provides a qualified definition. Analyzing the resources used also gains value. Memory institutions are now required to have expert staff in the Linked Open Data field and form a team that can carry out the work, prepare their policies, and start implementation as soon as possible, even if it is a pilot. Currently, the lack of personnel with technical infrastructure suitable for this study in memory institutions is among the factors that make this work difficult.

In this study, a process map has been created for beginners by examining the Linked Open Data applications of memory institutions around the world. Memory institutions can take meaningful steps to announce their metadata and institutions to a wider audience in line with this process. In future studies, general and understandable schemes should be developed to overcome problems with some specific aspects about the LOD applications beyond this general process.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Works cited

Agosti, M., et al. "Interacting with Digital Cultural Heritage Collections via Annotations: The CULTURA Approach." Proceedings Of the 2013 ACM Symposium on Document Engineering, 2013, New York, USA: ACM, pp 13–22.

- Ahmed, H. H. “Data Quality Assessment in the Integration Process of Linked Open Data (LOD).” 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). Hammamet, Tunisia, Oct. 2017, IEEE, pp. 1–6.
- Alemu, G., et al. “Linked Data for Libraries: Benefits of a Conceptual Shift from library-specific Record Structures to RDF-based Data Models.” 78th IFLA General Conference and Assembly, 24 May 2012, <http://conference.ifla.org/past/ifla78/92-alemu-en.pdf>.
- Alexiev, V. “Museum Linked Open Data: Ontologies, Datasets, Projects.” *Digital Presentation and Preservation of Cultural and Scientific Heritage*, vol. 8, 2018, pp. 19–50. doi:10.55630/dipp.2018.8.1.
- Allison-Cassin, S., et al. “ARL White Paper on Wikidata: Opportunities and Recommendations.” 2019, <https://scholarworks.iupui.edu/bitstream/handle/1805/18902/ARL-white-paper-on-Wikidata.pdf>.
- Baca, M. (Ed.). *Introduction to Metadata: Pathways to Digital Information*. 2002. <http://www.getty.edu/research/institute/standards/intrometadata>.
- Berardi, G., et al. “Metadata Enrichment Services for the Europeana Digital Library.” Proceeding Of Second International Conference on Theory and Practice of Digital Libraries, 2012, Berlin, 508–11.
- Berners-Lee, T. “Linked Data-Design Issues.” 2006, <http://www.w3.org/DesignIssues/LinkedData>.
- Berners-Lee, T. “Linked Data - Design Issues.” 18 June 2009, <http://www.w3.org/DesignIssues/LinkedData.html>.
- Bizer, C., et al. “The RDF Book Mashup: From Web APIs to a Web of Data.” Proc. of the 3rd International Workshop on Scripting for the Semantic Web (SFSW2007), 6 June 2007, Austria, [CEUR-WS.org/vol-248/paper4.pdf](http://www.ceur-ws.org/vol-248/paper4.pdf).
- BNF. 31 Jan. 2021, <https://data.bnf.fr/en/semanticweb>.
- Campbell, D. G. “Metadata, Metaphor, and Metonymy.” *Cataloging & Classification Quarterly*, vol. 40, no. 3–4, 2005, pp. 57–73. doi:10.1300/J104v40n03_04.
- Candela, L., et al. “The DELOS Digital Library Reference model—foundations for Digital Libraries.” 2008, http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.
- Caplan, P. *Metadata Fundamentals for All Librarians*. American Library Association, 2003.
- Chanhom, W., and C. Anutariya. “TOMS: A Linked Open Data System for Collaboration and Distribution of Cultural Heritage Artifact Collections of National Museums in Thailand.” *New Generation Computing*, vol. 37, no. 4, 2019, pp. 479–98. doi:10.1007/s00354-019-00063-1.
- Coyle, K. “RDA Vocabularies for a twenty-first-century Data Environment.” *Library Technology Reports*, vol. 46, no. 2, 2010, pp. 1–39.
- Davis, E., and B. Heravi. “Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation.” *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 14, no. 2, 2021, pp. 1–18. doi:10.1145/3429458.
- DBpedia. “About DBpedia.” 30 Jan. 2021, <https://www.dbpedia.org/about/>.
- de Boer, V., et al. “Linking the Kingdom: Enriched Access to a Historiographical Text.” Proceedings of the 7th International Conference on Knowledge Capture (KCAP 2013), 2013, New York, USA: ACM, pp. 17–24.
- Deliot, C. “Publishing the British National Bibliography as Linked Open Data.” *Catalog & Index*, vol. 174, 2014, pp. 13–18. http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf.
- Dempsey, L. “Registries: The Intelligence in the Network [Blog post].” 2006, <http://orweblog.oclc.org/archives/001105.html>.

- DNB. Jan. 2021, http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html.
- Dunsire, G. "Interoperability and Semantics in RDF Representations of FRBR, FRAD and FRISAD." *Concepts in Context: Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization*, Cologne, Germany, July 2010, p. 113.
- Duval, E., et al. "Metadata Principles and Practicalities." *D-lib Magazine*, vol. 8, no. 4, 2002, pp. 1082–9873. doi:10.1045/april2002-weibel.
- Eito-Brun, R. "Technologies for Metadata Integration and Interoperability." *Handbook of Metadata, Semantics and Ontologies*. By Miguel-Angel Sicilia, World Scientific, 2014, pp. 441–85.
- Europeana. "About Us." 16 Nov. 2020, <https://www.europeana.eu/en>.
- Folsom, D., et al. "The Blossoming of the Semantic Web: Linked Open Data and the American Art Collaborative." Presentation at the Museum Computer Network 2013 Annual Conference, 22 Nov. 2013, Montreal, Quebec, <http://americanartcollaborative.org/info/slides-the-blossoming-of-the-semantic-web/>.
- Foulonneau, M., and J. Riley. *Metadata for Digital Resources: Implementation, Systems Design and Interoperability*. Elsevier, 2014.
- Getty Research Institute. "Getty Vocabularies as LOD." 2014, <http://www.getty.edu/research/tools/vocabularies/lod/>.
- Hallo, M., et al. "Current State of Linked Data in Digital Libraries." *Journal of Information Science*, vol. 42, no. 2, 2016, pp. 117–27. doi:10.1177/0165551515594729.
- Hay, David C. *Data model patterns: A metadata map*. Elsevier Science, 2010.
- Heath, T., et al. "How to Publish Linked Data on the Web." Tutorial in the 7th International Semantic Web Conference, 2008, Karlsruhe, Germany.
- Heath, T., and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space, ser. Synthesis Lectures on the Semantic Web: Theory and Technology*. Morgan & Claypool, 2011.
- Hendler, J., and T. Berners-Lee. "From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web." *Artificial Intelligence*, vol. 174, no. 2, 2010, pp. 156–61. doi:10.1016/j.artint.2009.11.010.
- Hitzler, P., et al. (eds.). *OWL 2 Web Ontology Language Primer*. 2nd ed., 2012. <http://www.w3.org/TR/owl2-primer/>.
- Holgersen, R., et al. "Using Semantic Web Technologies to Collaboratively Collect and Share user-generated Content in order to Enrich the Presentation of Bibliographic records- Development of a Prototype Based on RDF, D2RQ, Jena, SPARQL and WorldCat's FRBRization Web Service." *Code4lib Journal*, vol. 17, 2012, pp. 1–17. <http://journal.code4lib.org/articles/6695>.
- International Association of Library Associations and Institutions. "Digital Libraries: Metadata Resources." 2002, <http://www.ifla.org/II/metadata.htm>.
- Johnson, L. "Limitations of the Descriptive Method." *The Phi Delta Kappan*, vol. 34, no. 6, 1953, pp. 241–45. <http://www.jstor.org/stable/20332351>.
- Johnson, T. "Indexing Linked Bibliographic Data with JSON-LD, Bib-JSON and Elasticsearch." *Code4lib Journal*, vol. 19, 2013, pp. 1–11. <http://journal.code4lib.org/articles/7949>.
- Kando, N., and J. Adachi. "Cultural Heritage on Line: Information Access across Heterogeneous Cultural Heritage in Japan." Proceedings of international symposium on digital libraries and knowledge communities in networked information society (DLKC'04) Tsukuba, Ibaraki, Japan, 2004, p. 136.
- Kanhabua, N., et al. "How to Search the Internet Archive without Indexing It." Proceedings of 20th International Conference on Theory and Practice of Digital Libraries. Hannover, Germany, 2016, Springer.
- Kollia, I., et al. "A Systemic Approach for Effective Semantic Access to Cultural Content." *Semantic Web*, vol. 3, no. 1, 2012, pp. 65–83. doi:10.3233/SW-2012-0051.

- Liao, S-H., et al. "A Semantic Web Approach to Heterogeneous Metadata Integration." *Proceeding of Second International Conference ICCCI Computational Collective Intelligence*, 2010, Taiwan, pp. 205–14.
- Liao, S-H. "Bibliographic Framework Initiative." 22 Jan. 2021, <https://www.loc.gov/bibframe/>. Linked Data Service. "Library of Congress." 2014, <http://id.loc.gov/>.
- Manola, F., et al. "RDF Primer." *W3C Recommendation, 1-107*. 2004, <https://www.uazuay.edu.ec/bibliotecas/mbaTI/pdf/RDF%20Primer.pdf>.
- Marden, J., et al. "Linked Open Data for Cultural Heritage: Evolution of an Information Technology." *Proceedings of the 31st ACM International Conference on Design of Communication ACM* (2013), 2013, pp. 107–12, <https://dl.acm.org/doi/abs/10.1145/2507065.2507103>.
- Mayer, A. "Linked Open Data for Artistic and Cultural Resources." *Art Documentation: Journal of the Art Libraries Society of North America*, vol. 34, no. 1, 2015, pp. 2–14.
- Metadata and Resource Description. 2002, <http://www.w3.org/Metadata>.
- National Archives of Australia. "Annual Report 2001." 2001, http://www.naa.gov.au/publications/corporate_publications/ar2001/outcomes_outputs_reports2.html.
- National Gallery of Art. "NGA Images." 2014, <https://images.nga.gov/>.
- NISO. "Understanding Metadata [online]." 2004, Bethesda: NISO P, <http://www.niso.org/standards/resources/UnderstandingMetadata.Pdf>.
- OpenRefine. 30 Jan. 2021, <https://openrefine.org/download.html>.
- Parikh, N., et al. "On Segmentation of Ecommerce Queries." *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. San Francisco, California USA, 2013, pp. 493–518.
- RDF. 31 Jan. 2021, <https://www.w3.org/RDF/>.
- Resource Description Framework – RDF. 31 Jan. 2021, <https://www.w3.org/RDF/>.
- Schlötterer, J., et al. "Web-based just-in-time Retrieval for Cultural Content." *Proceedings of the 7th international ACM workshop on personalized access to cultural heritage*, 2014, Vienna, Austria, pp 805–08.
- Seifert, C., et al. "Ubiquitous Access to Digital Cultural Heritage." *Journal on Computing and Cultural Heritage*, vol. 10, no. 1, 2017, pp. 4–27. doi:10.1145/3012284.
- Sicilia, M. A. "Metadata Research: Making Digital Resources Useful Again?" *Handbook of Metadata, Semantics and Ontologies*. By Miguel-Angel Sicilia, World Scientific, 2014, pp. 1–8.
- Signore, D. O. "The Semantic Web and Cultural Heritage: Ontologies and Technologies Help in Accessing Museum Information." *Semiotik der Kultur/Semiotics of Culture*, edited by K Robering. LIT Verlag, 2008, pp. 1–31.
- Singer, R. "Linked Library Data Now!" *Journal of Electronic Resources Librarianship*, vol. 21, no. 2, 2009, pp. 114–26. doi:10.1080/19411260903035809.
- Szekely, P., et al. "Connecting the Smithsonian American Art Museum to the Linked Data Cloud." *The Proceedings of the 10th Extended Semantic Web Conference*, May 2013, Montpellier, <http://www.isi.edu/~szekely/contents/papers/2013/eswc-2013-saam.pdf>
- The J. Paul Getty Trust. "Content Program." 2014, <http://www.getty.edu/about/opencontent.html.28>.
- The Rijksmuseum. Rijksstudio. 30 Jan. 2021, <https://www.rijksmuseum.nl/en/rijksstudio>.
- Tonkin, E. L., and G. J. L. Tourte. "Using the Crowd to Update Cultural Heritage Catalogue." Presented at *Involving the Crowd in Future Museum Experience Design-CHI 2016 Workshop*, 2016, San Jose, CA, pp. 1–6.
- UKOLN. "Metadata." 2002, <http://www.ukoln.ac.uk/metadata/>
- Ullah, I., et al. "An Overview of the Current State of Linked and Open Data in Cataloging." *Information Technology and Libraries*, vol. 37, no. 4, 2018, pp. 47–80. doi:10.6017/ital.v37i4.10432.

- van der Steen, P. A. S. E. *Peer-to-peer Search over Linked Data in the Research and Education Space*. 2015. Eindhoven U of Technology, Master's Thesis, <https://pure.tue.nl/ws/portalfiles/portal/47028895/797453-1.pdf>.
- Van Kesteren, A. "Cross-origin Resource Sharing." 2013, <http://www.w3.org/TR/cors/>.
- VIAF: The Virtual International Authority File. 17 Jan. 2021, <http://viaf.org>.
- Wenz, R. "Linked Data and Libraries." *Catalogue & Index*, vol. 160, 2010, pp. 2–5.
- Wenz, R. "Linked Open Data for New Library Services: The Example of Data." *Italian Journal of Library & Information Science*, vol. 4, no. 1, 2013.
- Westrum, A., et al. "Improving the Presentation of Library Data Using FRBR and Linked Data." *Code4lib Journal*, vol. 16, 2012, pp. 1–7. <http://journal.code4lib.org/articles/6424>.
- Xu, Q., et al. "Heterogeneous Information Integration for Supply Chain." Proceedings of 2005 IEEE International Conference on Systems, Man and Cybernetics, 2005, New Jersey: IEEE P, pp. 97–105.
- Yoose, B., and J. Perkins. "The Linked Open Data Landscape in Libraries and beyond." *Journal of Library Metadata*, vol. 13, no. 2–3, 2013, pp. 197–211. doi:10.1080/19386389.2013.826075.