

**Linked Data and Linked Open Data Projects for Libraries, Archives and Museums: Constructing  
Pathways to Information Discovery and Cultural Heritage Sector Collaboration**

Anna P. Paquet

John Hopkins University

AS.460.674.81- Digital Curation Research Paper

Dr. Joyce Ray

December 6, 2020

## Abstract

This paper examines current Cultural Heritage-based Linked data and linked open data projects developed by Libraries, Archives and Museums (LAMs). The following research questions are explored:

R1: Are there similarities and/or differences between libraries, archives and museums in how their linked data and linked open data projects, approaches and strategies are being implemented?

R2: What specific linked data and linked open data tools and tactics are being employed, and are there key variations between libraries, archives and museums?

The linked data/linked open data landscape has advanced since Tim Berners-Lee (et al.) introduced the concept of the Semantic Web, but challenges for LAMs remain as they work with their collections' data to create new web-based projects. Fundamental to these efforts is the creation, linking, and publishing of good quality metadata that will allow LAM collections to be discovered, accessed, and disseminated through viable methods. Trends across LAM sectors for linked data and linked open data projects include: global communication and collaborative research, use of wiki-based technologies, and efforts to improve sustainability. Application concepts from the Digital Curation Centre's Curation Lifecycle Model and Adrian Brown's Digital Preservation Maturity Model may help guide LAMs toward greater sustainability of linked data and linked open data collections' projects.

*Keywords:* linked data, linked open data, libraries, archives, and museums (LAMs), digital collections, Semantic Web

**Table of Contents**

Abstract.....	2
Introduction.....	4
Background: Foundations of LOD.....	6
Research Methodologies.....	10
Literature Review.....	11
Project Websites.....	19
Discussion .....	34
Open-ended Interviews with LAM Professionals.....	35
Recommendations .....	39
Conclusion.....	41
References.....	43
Annotated Bibliography.....	50
Appendix 1: Linked data & linked open data terminology.....	55
Appendix 2: Cultural heritage linked data & linked open data resources..... (Selected Readings & Projects)	56

## **Linked Data and Linked Open Data Projects for Libraries, Archives and Museums: Constructing Pathways to Information Discovery and Cultural Heritage Sector Collaboration**

Cultural Heritage institutions are exploring new ways to engage, educate and serve their communities. In this paper I will explore the importance of both existing and ongoing linked data and linked open data projects that are designed to expand and enrich how libraries, archives and museums can better help their communities to discover new meaning in LAM collections and cultural heritage artifacts and objects. I will do this by describing the reasoning and background for this research inquiry as well as the methodologies I employ. I will highlight the current terminologies and conceptual frameworks used in discussions of these issues. Based on research of the field and interviews of LAM practitioners, with experience in linked data and linked open data projects, I will discuss my analysis and findings, make recommendations, and draw conclusions about current developments in the field.

### **Overview: How Linked Data and Linked Open Data Aid LAMs**

There is growing interest among libraries, archives and museums (LAMs), in how the use of linked data and linked open data (LOD) can aid collection-based organizations in their role of connecting communities with curated, contextualized collections. Understanding how linked data is defined helps clarify why linked data is useful for managing and contextualizing Cultural Heritage collections data. Linked data is defined as structured data which is interlinked with other data so it becomes more useful through semantic queries (W3C, 2015). Linked open data is a form of linked data that emphasizes a linked network of data where the data itself is freely available and expressed in machine readable, open-source format.<sup>1</sup> Open data means that the content is freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control, and this represents a shift in how data is traditionally managed within LAMs' collections and scholarly

---

<sup>1</sup> Definition. linked open data (LOD) is linked data which is released under an open license, which does not impede its reuse for free. <https://www.w3.org/DesignIssues/LinkedData.html>

and scientific research and publishing. Linked open data employs the same Semantic Web infrastructure and components as linked data, but uses openly available data. Linked open data is machine readable by computers, which enables its semantics to be interpreted by computers as well as accessed by human users via the Internet. Linked data and linked open data are valuable for LAM practitioners and the communities that they serve because they offer the potential for new ways of analyzing and reusing collections data that could lead to new ideas and the global exchange of knowledge.

### **LAM Linked Data and Linked Open Data Projects**

LAM organizations are holding conferences on the use of Linked Data and Linked Open Data for collections data, and highlight their growing importance for LAMs.<sup>2</sup> Across the Cultural Heritage sectors, LAM practitioners have undertaken research projects to investigate linked data and linked open implementation and technologies.<sup>3</sup> These types of projects and activities are significant because they demonstrate an investment in the vision and potential of Berners-Lee's Semantic Web, and they also exemplify real world examples of the ways in which linked data and linked open data can be used to enhance the growth, functionality, reach, relevance and usefulness of LAMs' online collections.

Exploring the value of these linked data and linked open data ventures in terms of how they are applied through digital collections projects has the potential to provide greater understanding of their current and predicted use and significance by LAMs. From this research topic, the following questions

---

<sup>2</sup> Examples of recent conferences that are actively discussing issues related to linked data and linked open data include—

2020 LD4 Conference on Linked Data in Libraries: <https://ld42020.sched.com/>

DCMI 2020: <https://www.dublincore.org/conferences/2020/>

LODLAM 2020 Summit: <https://drive.google.com/drive/u/0/folders/>

<sup>3</sup> Some examples of linked data & linked open data projects are—

Digital Public Library of America's DPLA Metadata Model <http://dp.la/info/map>

Online Computer Library Center (OCLC) Project Passage

<https://www.oclc.org/research/publications/2019/oclcresearch-creating-library-linked-data-with-wikibase-project-passage.html>

Princeton's Derrida's Margins: <https://derridas-margins.princeton.edu/>

Yale Center for British Art's Linked Open Data Service: <https://old.datahub.io/dataset/yale-center-for-british-art>

arise:

- Are there similarities and/or differences between libraries, archives and museums, where linked data and linked open data is being implemented?
- What linked data and linked open data tools and strategies are being employed and are there key variations within the sectors?

A brief explanation of linked data concepts and terminology is summarized below to place the efforts of libraries, archives, and museums within the context of the larger linked data and linked open data landscape.

### Background: Foundations of Linked Data and Linked Open Data

Berners-Lee's (2001, 2006) Semantic Web refers to: "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (W3C, 2009). To achieve this vision, linked data is expressed in the machine-readable form of an RDF (resource description framework) known as an RDF triple.<sup>4</sup>

**Figure 1**

*RDF Triple Structure*



Note: Adapted from: W3C. (2014). RDF 1.1 Concepts and Abstract Syntax.

<sup>4</sup> *RDF*: Resource Description Framework is an XML based structure for expressing relationships between entities in a machine-readable format. The semantic structure of: subject predicate object (RDF), is essentially a knowledge structure consisting of a defined entity and its relationship to another defined entity (via subject and predicate), and it can be visually represented as a graph (W3C, 2014).

<https://www.w3.org/TR/rdf11-concepts/#data-model>

To build a globally-linked data network, associations and relationships are created between data entities and datasets with the use of metadata, ontologies/vocabularies, and conceptual data models. Then a combination of applications (for data management, access, visualization etc.), database stores, SPARQL and other programming languages are used to access, query, discover and store these datasets. To Berners-Lee, people are the driving force behind the Semantic Web. People create context and meaning for the networked data, and use computers to enable the global sharing of the data via the Web (W3C, 2015).

### Importance of Berners-Lee's 5 Star Model

In 2010, Berners-Lee offered a 5 Star rating system for linked data design that encouraged a kind of "best practices" progression for publishing linked data. In Berners-Lee's 5 Star Model linked data can connect with other people's data without being open, but the use of open data enables data sharing with ability for data reuse. The levels of linked data Berners-Lee describes are:

#### Figure 2

##### *Five Star Model (Berners-Lee)*

- ★ Available on the web (whatever format) *but with an open license, to be Open Data*
- ★★ Available as machine-readable structured data (e.g., excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g., CSV instead of excel)
- ★★★★ All the above plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- ★★★★★ All the above, plus: Link your data to other people's data to provide context

Note: from the W3C. (2009). Design Issues: Linked Data.

<https://www.w3.org/DesignIssues/LinkedData.html>

For his open data version of the 5-Star Model, linked open data is the highest level: that is, that it will be open data, machine-readable and structured, in a non-proprietary format, use open standards from

W3C (RDF and SPARQL) to identify things, and that will connect your data to other people's data to provide context (W3C, 2009).

### **Semantic Reference Models for LAMs**

In order for LAMs to create linked data as described in Berners-Lee 5 Star Model, digital collections objects and their relationships are mapped using conceptual reference and ontology models. These semantic models are a means of representing and describing the relationships between linked objects and resources — they are a way of mapping data (its meaning and relationships) into machine readable frameworks so that resources and their relationships can be defined, discovered and shared on the Internet. For libraries, archives, and museums there are several important sector-specific models used for semantically structuring collections data as linked data or linked open data: CIDOC (Committee for Documentation of the International Council of Museums) - Conceptual Reference Model (CRM), the Europeana Data Model (EDM), BIBFRAME (2.0) and Records in Contexts (RiC).

#### ***CIDOC (Committee for Documentation of the International Council of Museums) - Conceptual Reference Model (CRM)***

Recognized in 2006 as an official ISO standard, the Committee for Documentation of the International Council of Museums (CIDOC) - Content Resource Model (CRM) from the International Council of Museums (ICOM) has played a significant role in the organization and transition of cultural heritage resources into RDF datasets, and the development of a semantic standardized data structure for metadata interoperability (Doerr, 2003). It includes a relational model and an ontology model to enable museum resources to be mapped into machine readable format while maintaining context and relationships between resources for human understanding (ICOM, what is the CIDOC CRM, n.d.).

#### ***Europeana Data Model (EDM)***

The Europeana Data Model was developed specifically for the Europeana project and it is designed to be interoperable with a variety of metadata standards across libraries, archives, museums



and galleries such as LIDO, EAD, METS and Dublin Core (Europeana, Europeana Data Model, n.d.).

### ***BIBFRAME (2.0)***

An initiative of the Library of Congress, this data model employs linked data concepts to structure bibliographic data standards, and it has been designated to replace MARC as the primary bibliography descriptive framework. BIBFRAME includes a conceptual reference model and a vocabulary component for standardizing the description of resources and their relationships to other entities.

BIBFRAME incorporates conceptual aspects of CIDOC-CRM (LOC, BIBFRAME, n.d.) (Branan & Futornick, 2020).

### ***Records in Contexts***

Created by the Expert Group on Archival Description (EGAD) of the International Council on Archives (ICA), Records in Contexts model has two major components: The Records in Contexts Conceptual Model (RIC-CM) and the Records in Contexts (RIC-O). Records in Context is designed to serve as a general framework for archives that are seeking to incorporate archival standards for data and datasets into the Semantic Web (ICA, 2016).

Together, these important conceptual reference and ontology models are crucial components for transforming collections data from libraries, archives, and museums into linked data, and they are a part of the existing record of research-based literature for Cultural Heritage linked data and linked open data studies.

## **Research Methodologies**

The primary research methodologies for this paper are historical analysis (literature review), content analysis of linked data and linked open data project websites, and open-ended interviews with LAM professionals. Using these methodologies, this research design will serve to determine any differences and/or similarities among the linked data and linked open data use for Cultural Heritage

institutions. And, it will identify the tools and strategies that are being employed by LAMs, and any key variations within the sectors.

The literature review will establish a historical background for development of cultural heritage-based linked data and linked open data research from which existing linked open data project websites have evolved. The discussion of publications from Berners-Lee et al. (2006, 2001), conceptual data models, metadata schemas and XML-based descriptive languages and frameworks, storage and retrieval technologies, and landmark institutional research projects (individual and collaborative), will all provide a context for understanding the current state and use of linked and linked open data within LAM-based projects.

In conjunction with the literature review, a content analysis of existing linked open data and linked data project websites will provide examples of how linked data and linked open data is valued and utilized within the Cultural Heritage sectors. This methodology will enable an exploration of LAM-based current professional practices and standards, areas of growth or stagnation in linked data and linked open data implementation, and use and practices.

Open-ended interviews with Cultural Heritage practitioners and researchers can help ensure the inclusion of professional insights, additional knowledge, and practical experience to the research topic. Engaging with professionals will help explain current trends, challenges, and related areas of inquiry that stem from prior and current websites and projects. For this paper, open ended interviews were conducted with: Kalan Knudson Davis (Special Collections Metadata Librarian, University of Minnesota Libraries), Emanuelle Delmas-Glass (Collections Manager, Yale Center for British Art), Michelle Futornick (Stanford Libraries & Project Manager for Linked Data for Production series projects), and “Scann” (Evelin Heidel, Editor Open GLAM).

In support of these methodologies, this research paper will conclude with recommendations for further linked data and linked open data research projects, and a summary conclusion of the state of the linked data and linked open data landscape for Cultural Heritage organizations.

## **Literature Review**

### **Introduction**

A review of existing linked data and linked open data research, for the Cultural Heritage domain, serves to explain how prior research has impacted the current linked data and linked open data landscape for libraries, archives and museums and what gaps exist in the present body of literature. Cultural Heritage research projects for linked data and linked open data have included all areas of Berners-Lee's 5 Star model from creating/converting datasets to RDF and other linked data format serializations, development of tools, ontologies, vocabularies, UI and visualization, storage, queries and publication, and strategies for implementation. Research surveys by the OCLC (2014, 2015, 2018), Open GLAM (2019) and the University of California's UC Linked Data Project Team Report (2018) indicate that Libraries, Archives, Museums are exploring and implementing linked data & linked open data (LOD) technologies and projects with varying levels of resources. The OCLC's 2018 survey respondents included libraries, archives and museums and the survey found that the three highest ranked barriers for cultural heritage organizations were: "steep learning curve," "inconsistency in legacy data," and "selecting appropriate ontologies" (Smith-Yoshimura, 2018).

The research topics in this section are organized into the following key areas of linked data components: semantic reference models (conceptual reference models & ontology models), vocabularies and metadata, and tools. Each LAM sector's approach, use and practices regarding each of these key areas will be described.

### **Semantic Reference Models**

#### ***Libraries***

Within the library sector, primary areas of research involve: employing BIBFRAME (2.0) to manage and express bibliographic records as linked data, reconciling MARC legacy data, and finding solutions for interoperability with MARC (as MARC-based systems are still actively used in cataloging and ILS systems). Incorporating studies from Taniguchi (2017) and Tillett (2013), Zapounidou, Sfakakis & Papatheodorou (2019) examined how bibliographic data relationships expressed as Resource Description & Access (RDA) could be successfully mapped to BIBFRAME (2.0) with exceptions of derivative relationships mapped at the RDA Works level.

BIBFRAME projects at the University of Illinois by Jin, Hahn & Croll (2016), Michael & Han (2019), have helped to identify needed areas of improvement for the BIBFRAME (2.0) model for structuring data from MARC records. Another related area of research involves developing solutions to increase interoperability between different library linked data conceptual models and cataloging standards. Ullah, Khusro, Ullah & Naeem (2018) and Samples & Bigelow (2020) have documented the work of the Program for Cooperative Cataloging (PCC) in examining areas of conceptual alignment between the Library of Congress' BIBFRAME (2.0) and the International Federation of Library Associations and Institutions IFLA-Library Reference Model (LRM). Together these projects affirm research from Park & Richards (2019) that emphasizes that BIBFRAME (2.0) is still in an early developmental stage and that additional implementation and testing reports are needed to create a more complete assessment of BIBFRAME (2.0)'s strengths and challenges as a linked data model.

Ullah, Khusro, Ullah & Naeem (2018) in their overview of the state of linked and open data within the field of cataloging, from 2014-2018, identify the need for increased conversation and collaborative efforts between catalogers and professionals outside of LAMs, as a means of advancing the development of a linked data environment for libraries with cross-domain learning and exchange and recognition of a global cataloging landscape. Projects from OCLC such as Project Passage and the CONTENTdm linked data Project Pilot, BIBFRAME (2.0), and the LD4P collaborative initiative series (phases 1-3) have implemented community components via conferences, discussion groups, advisory boards and meetings for professional dialog and exchange.

### ***Archives***

The Records in Contexts Model (RiC) has been used in a variety of archival linked data projects including: The Foundation Contemporary Jewish Documentation Center (CDEC), the Archives Nationales de France (see <https://ica-egad.github.io/RiC-O/projects-and-tools.html>), and the Social Networks and Archival Context (SNAC) cooperative. SNAC is on the forefront of U.S. national projects for the use of RiC in creating and managing linked data for archives. However, because SNAC's research is ongoing, there is currently a gap in use-cases and publications of the RiC model for U.S. institutions.

Other archives linked data projects, such as the Linked Jazz Project (<https://linkedjazz.org/>), used the Simple Knowledge Organization System (SKOS) for mapping jazz finding aids for linked data, and developed its own ontology for structuring the datasets (Adams, 2020). Zeng and Mayr (2019) have focused on the use and implications of SKOS towards linked open data and the Cultural Heritage Humanities research domain as a whole.

### ***Museums***

From CIDOC the (Committee for Documentation of the International Council of Museums) - Conceptual Reference Model (CRM), Museums have developed model extensions and compatible frameworks related to the specific requirements of their resource collections (see <http://www.cidoc->

[crm.org/collaborations](http://crm.org/collaborations)). The Smithsonian American Art Collaborative (<https://americanart.si.edu/about/american-art-collaborative>) Yale Center for British Art's online collections (<https://britishart.yale.edu/collections-data-sharing>) and the Linked Art Model (<https://linked.art/model/>) are examples of linked open data initiatives that used CIDOC-CRM as a foundational model for structuring art collections data.<sup>5</sup>

The Art Tracks project examined how art provenance data could be transformed as linked data and integrated into CIDOC-CRM. Newbury (2017) noted that while the project was successful in creating a group of provenance profiles for CIDOC-CRM, the specific nature of the provenance records themselves (due to variations with acquisition including ownership, custody, and location of a work) would require additional research to develop more effective solutions for accurately mapping and contextualizing these aspects and art provenance within CIDOC-CRM.

Managing museum-based art data through a linked data structure is the core concept of the Linked Art Model, and it has been implemented through a number of web projects and digital collection sites including the PHAROS project.<sup>6</sup> The PHAROS project is working with ResearchSpace on constructing a research platform pilot. Caraffa, Pugh, Stuber & Ruby (2020), and Delmas-Glass & Sanderson (2020) have analyzed PHAROS and the Linked Art Model through a lens of usability to demonstrate how employing a shared data model and open vocabularies serve data management goals of reuse and interoperability. The O'Keeffe Museum's Collections Online site<sup>7</sup>– and the MoMA Linked Open Data Fellowship (2018-2019)<sup>8</sup>– serve as additional examples of the Linked Art model use cases and research.

---

<sup>5</sup> For the Smithsonian Art Collaborative see Knoblock, C. A., Szekely, P., Fink, E., Degler, D., Newbury, D., Sanderson, R., ... & Krishna, R. R. (2017, October) in References.

<sup>6</sup> The Linked Art Model was developed through the Linked Art Initiative. In support of the model and initiative a Linked Art Community of participating institutions, collaborative projects and funders and an administrative editorial board (see <https://linked.art/community/index.html>).

<sup>7</sup> The O'Keeffe Museum's Collection Online site is discussed in the project websites section of this paper.

<sup>8</sup> Adams, S.A. (2019). The Project. 2018-2019 MoMA Archives Linked Open Data (LOD) Fellowship Report. <https://sites.google.com/view/momalodfellowship18-19/the-project>

## **Vocabularies & Metadata**

Models like EDM, CIDOC CRM, Record in Context, BIBFRAME (2.0), and Linked Art include either an ontology framework or the use of specific controlled vocabularies for defining and standardizing terms used to describe entities and their relationships. Alexeiv (2020), in his overview of Semantic Web museum projects, argues that in the museum sector there are multiple ontological models in use without dominance of any particular model over another (p. 21). In the archives and libraries sectors, organizations may employ different ontological models for semantically integrating their individual collections data as well.

The Getty Research Institute, whose thesauri serve as key resources for cultural heritage object terms and descriptions, converted their works into linked open data (see <https://www.getty.edu/research/tools/vocabularies/lod/>). The Getty vocabularies are an example of Linked Open Vocabularies (LOV), a reusable vocabulary for creating linked data (see <https://lov.linkeddata.es/dataset/lov/about>). Sanderson (2019) has incorporated a concept of usability into the value of using Getty open vocabularies, and he argues that embracing linked open data as Linked Open Usable Data (LOUD)<sup>9</sup>– provides a conceptual framework for valuing system interoperability, cross-discipline, collaborative research, accessibility, and consideration of users' needs (internal & external endpoints) that enhance linked open data's ability to provide increased discoverability and context.

### ***Libraries***

Research from Myntti & Cothran (2013), Crowe & Clair (2015), Neatrour & Myntti (2019), and Smith-Yoshimura (2018) demonstrates the crucial role of additional investigation in metadata towards advancing the use of linked data and linked open data for LAMs. For libraries, bibliographic metadata

---

<sup>9</sup> For a summary of Linked Open Usable Data (LOUD) principles see <https://linked.art/loud/>

records are at the center of the approach for linked data and linked open data research & publications. OCLC's Project Passage and CONTENTdm Data Pilot and LOC's BIBFRAME (2.0) project examine ways to take existing datasets from MARC and express the data as linked open data while retaining context for users. The Western Name Authority File (WNAF) project demonstrates that there is overlap with the Archives sector for working with authority records for digital libraries and special collections. Neatrou & Myntti (2019) determined that for the WNAF project the Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) was the appropriate XML standard and schema for their digital library datasets. Smith-Yoshimura (2020) makes a compelling argument in her report on the state of metadata for libraries that current concepts of how libraries engage with cataloging metadata will change with the use of linked data.

### ***Archives***

Pitti's (2015) work with the Social Networks and Archival Context (SNAC) cooperative reveals how metadata research on Archival Authority Control and social networks has the ability to shape/reshape the linked open data environment. SNAC's site aggregates linked archival Corporate Bodies, Persons, and Families (CPF) biographical data, but it is a cooperative organization with contributions and membership from libraries, archives, and museums. The SNAC project aligns with the concept of system interoperability and the value of linked open data as a shared network of discovery and context.

Crowe & Clair's (2016) research focuses on the challenge of smaller archival institutions working with unique, local archival datasets. Their conclusions support the importance of collaborative work and their tool & database design sought to provide interoperability with the larger projects of SNAC and the Virtual International Authority File (VIAF).

### ***Museums***



The Smithsonian American Art Collaborative, Europeana, and the Yale Center for British Art remain landmark and active linked open data projects for producing and disseminating museum collections' datasets as linked open data. As with projects like SNAC and BIBFRAME, these museum projects show that transforming collections' data into linked open data involves multiple areas of research and development. These museums use data models, ontologies, vocabularies, and metadata tools for creating appropriate linked open data structures and systems. Through these strategies these organizations can allow their data to be openly maintained, stored, accessed, contextualized, published, discovered, queried, and shared. Wildenhaus (2020) argues that linked data technologies may bring a new perspective to managing art exhibition histories as well potential new experiences for contextualizing and engaging with online art exhibits.

## Tools

There are an array of linked data and linked open data tools used to support linked data and linked open data projects and activities such as creation/conversion of RDF datasets and RDF serializations, metadata mapping and editing, data modeling, storage and access components, search, discovery and publishing. Common tools used within all three LAM sectors include: OpenRefine, developer Application Protocol Interface (APIs)<sup>10</sup>–, International Image Interoperability Framework (IIIF)<sup>11</sup>– and use of JSON-LD (Javascript Object Notation for Linked Data)<sup>12</sup>– for programming and publishing (see [Open Registry of LOD for GLAM Tools](#)).

---

<sup>10</sup> *Application Programming Interface (API)*: code that enables communication between software applications and provides developers with a protocol for interacting with a particular application or software component. Through an API and use of J-SON and/or other languages, developers can access, edit/use datasets (Freeman, Infoworld, 2019).

<sup>11</sup> *IIIF (International Image Interoperability Framework)*: a group of API specifications that enables its users to improve functionality, interoperability, and access of images held in digital collections/repositories ((IIIF, FAQ, 2020).

<sup>12</sup> *JSON-LD*: a framework for expressing linked data that enables the defining of elements and semantic description of relationships between objects in conjunction with the use of an URI (JSON-LD, JSON for Linking Data, 2020).

Wikidata and its underlying software Wikibase have emerged as potentially powerful research tools for linked and open data in all three LAM sectors nationally and internationally. Wikidata and Wikibase have been used to experiment with both data producer and data consumer linked data/linked open data features. Experiments such as enabling users to create & edit connections between entities, metadata and crowdsourcing. Or, using data modeling and interoperability, graph visualization, and providing interfaces for users and API endpoints (OCLC 2020) (Adams, 2020) (Hwang, 2020) (Kapsalis 2019) (Allison-Cassin & Scott 2019). Use of the wiki-based software for linked data/linked open data for cultural heritage institutions is in a state of ongoing research and development, and there is not enough existing published research to draw firm conclusions about the possibility of using wiki-based linked data/linked open data systems to connect collections' data across LAMs.

### **Summary Findings Regarding Literature Review**

Cultural Heritage institutions continue to adapt data models and ontologies such as CIDOC, EDM, BIBFRAME (2.0) and Records in Contexts (RiC) to meet the needs of their individual collections' data. Bringing together the observations of Alexiev (2020), Dobreski, Park, Leathers & Quin (2019) and Ullah, Khusro, Ullah & Naeem (2018), there does not exist a single, unifying conceptual and/or ontological model for the semantic structuring and linking of collections data for within the LAM sectors. While nationally BIBFRAME (2.0) is a primary model for semantically linking bibliographic data, the IFLA-LRM exists internationally alongside BIBFRAME (2.0) for semantically linking bibliographic data. And there are major interoperability issues between the two models that are being assessed for solutions, but the challenges remain. Archives researchers have noted the difficulties of transitioning archival data structures (collection level, item level, provenance, and other collections data) into a model for linked data structure that reflects the various archival relationships and context at work (Dobreski, Park, Leathers & Quin, 2019). Because LAMs' collections can vary in scope, size, material and context, and they are usually managed and described through different collections systems, there is still a compelling

need for continued research that covers the linked data spectrum of Berners-Lee’s 5 Star Model, from datasets to discovery to reuse.

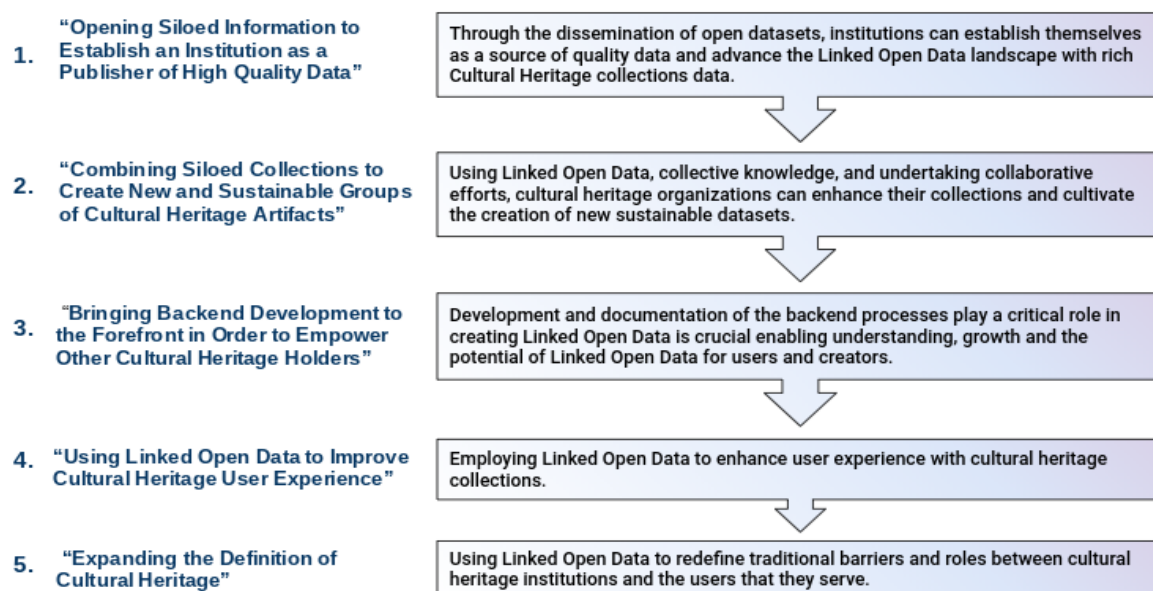
## Project Websites

### Introduction to LAM Project Website Assessment

This paper’s assessment of Cultural Heritage project websites uses five essential linked open data project stages developed by Marden, Li-Madeo, Whysel, and Edelstein (2013). Their framework, “Linked Open Data for Cultural Heritage: Evolution of an Information Technology,” serves as a conceptual guide for better understanding what elements of linked data and linked open data LAMs are actively addressing in their research projects to enhance their digital collections. The five stages (shown in Figure 3) are not intended to rank in importance one project before another, but instead they are intended to organize and document those aspects of linked data and linked open data that are being used within the cultural heritage domain (Marden, Li-Madeo, Whysel, and Edelstein, 2013).

**Figure 3**

*Five Stages — “Linked Open Data for Cultural Heritage: Evolution of an Information Technology”*



Note: Adapted from Marden, Li-Madeo, Whysel & Edelstein. (2013). linked open data for Cultural Heritage: Evolution of an Information Technology. In SIGDOC '13 Proceedings of the 31st ACM International Conference on Design of Communication. ACM. 107-112.

<http://dx.doi.org/10.1145/2507065.2507103>.

Marden, Li-Madeo, Whysel, and Edelstein's (2013) model is designed for evaluating linked open data for Cultural Heritage projects. Some of the methods and core components of linked open data that they identify are often inherent to Cultural Heritage projects that employ linked data. This includes the use of collective knowledge and collaboration, the development of backend processes, and the goal of enhancing user experience. However, the projects analyzed in this assessment do not uniformly use the principle of "open data" in creating, converting, and publishing their data as linked data.

### Barriers

The Online Computer Library Center (OCLC) international surveys (2014, 2015, 2018) identified a list of barriers for Cultural Heritage institutions that publish, ingest and use linked data. These barriers are shown below in Figure 4. As the number of Cultural Heritage institutions producing and publishing linked data has increased over time, so has the response to specific linked data barriers. The three most significant barriers for 2018 were: "steep learning curve," "inconsistency in legacy data," and "selecting appropriate ontologies" (OCLC, survey 2018) (Smith-Yoshimuri, 2018). These responses suggest that linked data use and publishing for libraries, archives, and museums has not fully advanced toward Berners-Lee's conceptual vision and infrastructure of a "web of data."

### Figure 4

*Linked Data Barriers from the OCLC International Survey for Linked Data Implementers (2014, 2015, 2018)*

Barriers to Linked Data Publishing (OCLC 2018)	Barriers to Linked Data Ingest & Use (OCLC 2018)
Total # = total number of respondents   % = percentage of responses	

	2018		2015		2014			2018		2015		2014	
Response	Total #	%	Total #	%	Total #	%	Response	Total #	%	Total #	%	Total #	%
	80		79		34			58		59		31	
Steep Learning Curve	41	51%	40	51%	4	12%	Matching, disambiguating and aligning source data and the linked data resources	28	48%	23	39%	3	10%
Inconsistency in legacy data	38	48%	33	42%	1	3%	What is published to the Internet as linked data is not always reusable or lacks URIs	18	31%	16	27%	3	10%
Selecting appropriate ontologies to represent our data	26	33%	31	39%	2	6%	Size of RDF dumps	16	28%	12	20%	2	7%
Lack of resources	23	29%	0	0%	1	3%	Unstable endpoints	16	28%	10	17%	1	3%
Little documentation or advice on how to build the systems	23	29%	21	27%	2	6%	Service reliability	15	26%	9	15%	3	10%
Establishing the links	22	28%	27	34%	7	21%	Mapping of vocabulary	15	26%	17	29%	6	19%
Lack of tools	18	23%	15	19%	5	15%	Understanding how the data is structured before using it.	14	24%	12	20%	1	3%
Immature software	17	21%	11	14%	4	12%	Lack of needed off-the-shelf tools	14	24%	10	17%	7	23%
Ascertaining who owns the data	4	5%	10	13%	1	3%	Datasets not being updated	13	22%	14	24%	1	3%
Other	19	24%	21	27%	21	62%	Lack of authority control	11	19%	15	25%	2	7%
							Volatility of data formats of dumps	10	17%	11	19%	2	7%
							Disambiguation	8	14%	6	10%	1	3%

							of terms across different languages is difficult						
							It's difficult to get other institutions to do their own harmonization between objects and concepts.	7	12%	9	15%	1	3%
							Other	13	22%	15	25%	12	39%

Note: adapted from Smith-Yoshimura, Karen. "Analysis of 2018 International Linked Data Survey for Implementers." Code4Lib Journal Issue 42, 8 November 2018. <https://journal.code4lib.org/articles/13867> and OCLC (2018). OCLC Research for Linked Data Implementers Survey. <https://www.oclc.org/content/dam/research/activities/linkedata/oclc-research-linked-data-implementers-survey-2014.xlsx>.

## Libraries

### ***OCLC CONTENTdm Linked Data Pilot***

The CONTENTdm Linked Data Pilot is organized into three phases: mapping textual metadata to entities, tools for managing metadata, and user discovery powered through use of Wikibase entities (OCLC, CONTENTdm Linked Data Pilot). Using Wikibase as a foundational platform, the pilot project connects Wikibase to the CONTENTdm management system to enable a search and discovery interface for users to access an institution's digital collection. The use of CONTENTdm includes system components for digital preservation and storage. Libraries are able to use the CONTENTdm system and accompanying technological infrastructure to map their datasets into linked data, and then use entity descriptions drawn from established authority files and existing local library vocabularies (OCLC, CONTENTdm Linked Data Pilot). The Cleveland Public Library and the Minnesota Digital Library showcased linked open data records online as sample work product for the pilot project (see <https://cplorg.contentdm.oclc.org/digital/collection/p16014coll6/id/1862>,

<https://reflections.mndigital.org/catalog/nfh:1579#/image/0>). These two partners' webpages present sample visual guides for what library users will encounter on the consumer end of the project CONTENTdm platform. In accordance with Marden, Li-Madeo, Whysel, and Edelstein's (2013) model, the following functions of linked data are identified within the OCLC CONTENTdm Data Pilot: enhancing collections, focus on backend development (such as dataset creation/editing, mapping, reconciliation etc.) and using linked open data to advance user experience. The linked data functions of the OCLC CONTENTdm Pilot underscore the value and potential value that linked data and linked open data hold for the libraries as they build digital collections for their communities. This project is important as it highlights how two well-known and accessible software platforms can be integrated and used by the libraries' sector to incorporate linked data or linked open data into their digital collections.

### **Tools & Technology.**

The OCLC CONTENTdm Linked Data Pilot program is built using several different pieces of software that are used to help create, transition and manage libraries' collections data as linked data. Identifying these tools and technological components helps reveal similarities and/or differences between projects of libraries, archives, and museums as they seek to represent their online collections as linked data. CONTENTdm provides a user interface through which information users or consumers can access and search contributors' content. The platform also provides digital preservation capabilities for its repositories (OCLC, CONTENTdm, n.d.).

Tools and technology that are employed by the pilot project and are used within other current linked data and linked open data Cultural Heritage projects include: the Wikibase platform, OpenRefine software, use of application programming interface (API), International Image Interoperability Framework (IIIF) and Javascript Object Notation for Linked Data (JSON-LD). Wikibase is used by project partners to enhance their editing and management of metadata, structured data, as well as their storage and retrieval functions (OCLC, CONTENTdm, n.d.). OpenRefine software is used to edit or clean

up datasets and a Wikidata-OpenRefine tool has been developed for reconciling metadata within Wikibase (see <https://wikidata.reconci.link/>). An IIF API is used to link digital objects with metadata (OCLC, CONTENTdm, n.d.). Participant organizations have the ability to publish the linked data for digital objects in their collections in the JSON-LD serialization, and this is useful for developers who work with editing and aggregating data from digital objects and datasets.

System improvements regarding technological infrastructure and user interface are still needed to help catalogers and other information professionals use the pilot platform to more effectively structure, link and disseminate their data to users for sharing and discovery. Use of the system by catalogers and librarians requires training. From the results of the pilot project, OCLC identified three aspects of using CONTENTdm for linked data: 1) metadata can be harvested, but it is based on participants' locally defined fields, and there can be ambiguity with object descriptions, 2) CONTENTdm discovery is enriched aggregation, but when metadata from all hosts and collections is viewed collectively there appears to be inconsistent descriptive metadata, 3) structured linked data can be developed from CONTENTdm fields and mapped to Dublin Core elements, but there is need for further work on data mapping and reconciliation (Washer, Mixburn & Einaudi, 2019).

### **Barriers.**

The challenges of the OCLC CONTENTdm Linked Data Pilot that are identified by Mixburn & Einaudi (2019) most closely align with the linked data barriers of “inconsistency in legacy data” and “steep learning curve” described by Yoshimura (OCLC, survey 2018, Yoshimura). The pilot project’s results suggest that solutions for mapping and describing existing bibliographic data into linked data are still in development but progressing. And that removing data ambiguities related to entity definitions and descriptions is an ongoing area of research. The process of transitioning data into a different framework while achieving a semantic structure that maintains meaning and context from previous ‘locally defined’ metadata requires a combination of skilled personnel and the use of computer-based



tools and technology. In response to the difficulties encountered during the pilot project, and feedback from other completed linked data projects, OCLC has implemented a new linked data initiative. The OCLC's new Entity Management program seeks to create a shared database and accompanying infrastructure that will serve as a key provider of consistent entity terms and descriptions to reduce data inconsistencies in linked data mapping (modeling) and development of user interface (UI) for catalogers and other professionals (OCLC, OCLC and Linked Data). The project's goals address the barriers of a steep learning curve and existing ambiguities associated with the creation, editing and management of linked data.

### ***Linked Data for Production 3 (LD4P3): Closing the Loop***

A consortium of Columbia, Cornell, Harvard, Library of Congress, Princeton, Stanford and other institutions are piloting an important project for the production of linked data for library resources. Linked Data for Production 3 (LD4P3) is focused on "Closing the Loop" and bridging gaps in the processes of creating datasets, building interfaces, enhancing information discovery for users, and developing a framework for sustaining the tools, research, and community that are a part of the library sectors efforts to join the linked data global landscape. LD4P3 is a part of the Linked Data for Production (LD4P) project series (LD4P1, LD4P2, LD4P Labs and LD4P3 (Futornick, personal communication, October 23, 2020). Through seven goals and five deliverables (work products), the LD4P project series aims to strengthen production tools and infrastructure for catalogers and developers. The objective is to make it easier for catalogers to use linked data for managing and describing collections data, and for empowering end users for improved search and discovery by illuminating the contexts of and connections between collection resources (Branan & Futornick, 2020). The projects' deliverables are currently under development and not available for public access.

### **Tools & Technology.**

Similar to the OCLC CONTENTdm Linked Data Pilot, a major aspect of the LD4P3 project involves improving the features and infrastructure of tools designed for creating and editing datasets so that they can be employed as linked data. A unique aspect of the project is that it brings together a group of major tools and linked data initiatives and prior collaborations in order to achieve its goals. A core technological component of the BIBFRAME initiative from the Library of Congress, the Sinopia Editor, has become a key tool for LD4P3. Sinopia is a “sandbox” or open-source collaborative environment where users can create and edit metadata and entity descriptive information for developing linked data (using the BIBFRAME model and extensions: <https://sinopia.io/>.) An online user can register an email and password to use the tool and/or use the Sinopia “stage” for practice prior to editing actual project datasets. The LD4P3 project also uses Blacklight, a software platform (<http://projectblacklight.org>), for implementing a search/discovery interface, that will be coordinated with Wikidata. It includes an international libraries initiative called Share-VDE (Virtual Discovery Environment) that is a libraries-based platform focused on developing a linked data discovery workspace to be used for goals such as converting MARC to RDF, and data enrichment: <https://www.share-vde.org/sharevde/clusters?!=en>. To help reconcile library legacy systems’ use of MARC with the shift towards BIBFRAME 2.0, the LD4P3 project involves the Library of Congress’s Program for Cooperative Cataloging venture as a stakeholder to work towards developing strategies and solutions. Lastly, the LD4P3 includes the OCLC as a partner organization and aims to incorporate the work from their new OCLC Entity Management Program.

The LD4P3 project is similar to the OCLC CONTENTdm Linked Data Pilot project, not just in terms of its core emphasis on backend development, but with its inclusion of sustainability efforts as a part of the project’s additional targeted outcomes. That aspect of the project has not been finalized and will continue through 2021 with the possible development of a new grant-based initiative LD4 (Schreur, 2020).

### Figure 3

### Screenshot of User Interface for the Sinopia linked data Editor

The screenshot shows the Sinopia linked data Editor interface. At the top, there are navigation tabs: Dashboard, Resource Templates (selected), Search, Load RDF, and Exports. Below the tabs is a search bar labeled 'Find a resource template' with the placeholder text 'Enter id, label, URI, remark, or author'. To the right of the search bar is a trash icon and a 'New template' button. The main content is a table with the following columns: Label / ID, Resource URI, Author, Date, Guiding statement, and two icons (edit and copy).

Label / ID	Resource URI	Author	Date	Guiding statement		
<a href="#">_Cartographic Instance (BIBFRAME)</a> ld4p:RT:bf2:Cartographic:Instance	http://id.loc.gov/ontologies/bibframe/Instance	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Cartographic:Instance		
<a href="#">_Cartographic Work (BIBFRAME)</a> ld4p:RT:bf2:Cartographic:Work	http://id.loc.gov/ontologies/bibframe/Work	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Cartographic:Work		
<a href="#">_Monograph Instance (BF2) Un-Nested</a> ld4p:RT:bf2:Monograph:Instance:Un-nested	http://id.loc.gov/ontologies/bibframe/Instance	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Monograph:Instance		
<a href="#">_Monograph Instance (BIBFRAME)</a> ld4p:RT:bf2:Monograph:Instance	http://id.loc.gov/ontologies/bibframe/Instance	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Monograph:Instance		
<a href="#">_Monograph Item (BF2) Un-Nested</a> ld4p:RT:bf2:Monograph:Item:Un-nested	http://id.loc.gov/ontologies/bibframe/Item	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Monograph:Item		
<a href="#">_Monograph Work (BF2) Un-Nested</a> ld4p:RT:bf2:Monograph:Work:Un-nested	http://id.loc.gov/ontologies/bibframe/Work	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Monograph:Work		
<a href="#">_Monograph Work (RIFRAME)</a>	http://id.loc.gov/ontologies/bibframe/Work	LD4P	Aug 19, 2019	based on LC template ld4p:RT:bf2:Monograph:Work		

Note: Taken from <https://sinopia.io/>

#### Barriers.

The LD4P3 project addresses some of the challenges identified in the OCLC International Linked Data Survey for Implementers (2018, 2015, 2014). The project seeks to tackle difficulties related to legacy data (particularly MARC), reconciliation of data, vocabulary mapping and authority control, lack of tools and resources. These challenges of working with linked data are important for researchers to undertake because they impact how effectively linked data consumers and end users can use and/or reuse bibliographic linked data or linked open data that is disseminated.

#### Archives

##### ***Social Networks and Archival Context (SNAC) Phase 3***

Social Networks and Archival Context project (SNAC) is an online portal that serves as a centralized source for biographical archival records that contain EAC-CPF data (Encoded Archival Context - Corporate body, Person, Family). Individual contributors develop their own organizational plans to integrate SNAC into their digital collections. Phase 3 of the SNAC project, involves incorporating

extensions for OpenRefine and ArchivesSpace software components for the portal to improve functionality for partners to contribute and edit data to the site, and work on entity reconciliation.

Phase 3 is centered on the development and deployment of APIs for linked open data producers. The APIs enable producers to easily access, edit, cleanup, and format datasets serialized in linked data form. To clarify how this impacts the SNAC portal, a summary of the portal (<https://snaccooperative.org/>) provided below.

The SNAC portal provides information to different types of linked open data users: producers (archivists, catalogers and developers) and consumers. There is tiered access to the contextualized information for each of the biographical records, known as “constellation” records. Linked open data producers can log into the site to manage and edit records and links. For general consumers (without login access), there is still access to an array of CPF (Corporate body, Person, Family) data and users can see relationships between entities, holding repositories, related authority standards, etc. Relationships between constellation records are visually represented through graphs. Users can click on the name of the bibliographic entity (in the graph) to access information about the related resources.

**Figure 4**

*Screenshot of Example SNAC API*

The screenshot shows the SNAC API documentation interface. At the top, there is a navigation bar with the SNAC logo, a search icon, and links for About, Help, and Login. Below this, the 'API Connection Information' section states that the SNAC RestAPI is available at the following URL: `http://api.snaccooperative.org/`. The 'Search Vocabulary' section provides usage information for the `vocabulary` command. It lists parameters: `term_id` (optional integer), `query_string` (string), `type` (string), and `entity_type` (optional integer). Return values include an array of objects with fields for `term`, `result`, and `timing`. An example section shows a direct term read query: `{ "command": "vocabulary", "term_id": 700 }`.

Note: from SNAC Rest API Help [https://snaccooperative.org/api\\_help](https://snaccooperative.org/api_help)

## Tools & Technology.

Like the libraries sector projects, SNAC uses well-known and accessible software tools to manage its datasets. Its OpenRefine extension serves to enable partners who contribute data to reconcile data ambiguities within their datasets, related to CPF identity description. Contributors' data is uploaded and matched and merged into central constellation records (SNAC, GitHub). Its ArchivesSpace extension is in beta testing and allows external users of ArchiveSpace (an ACMS or archival collections management system) to search and retrieve a SNAC identity record and import it into ArchiveSpace as an "agent" for archival description and record management (SNAC, GitHub). Data contributors and developers have the ability to use different access points for working with datasets and records. These are important features to ensure that metadata is accurate for aiding user discovery, and open data can be easily accessed and managed for linking and potential reuse.

**Barriers.**

SNAC Phase 3 focuses primarily on infrastructure improvements to tackle barriers regarding the need for effective tools and software within the archives sector. Such efforts support Marden, Li-Madeo, Whysel, and Edelstein's (2013) linked open data project stages that are centered on Improving the management, quality and tools used to create and edit the linked open data and corresponding metadata impacts the discoverability, usability and quality that information users will access through linked open data dissemination.

**Museums*****O'Keefe Museum Collections Online***

<https://collections.okeeffemuseum.org/>

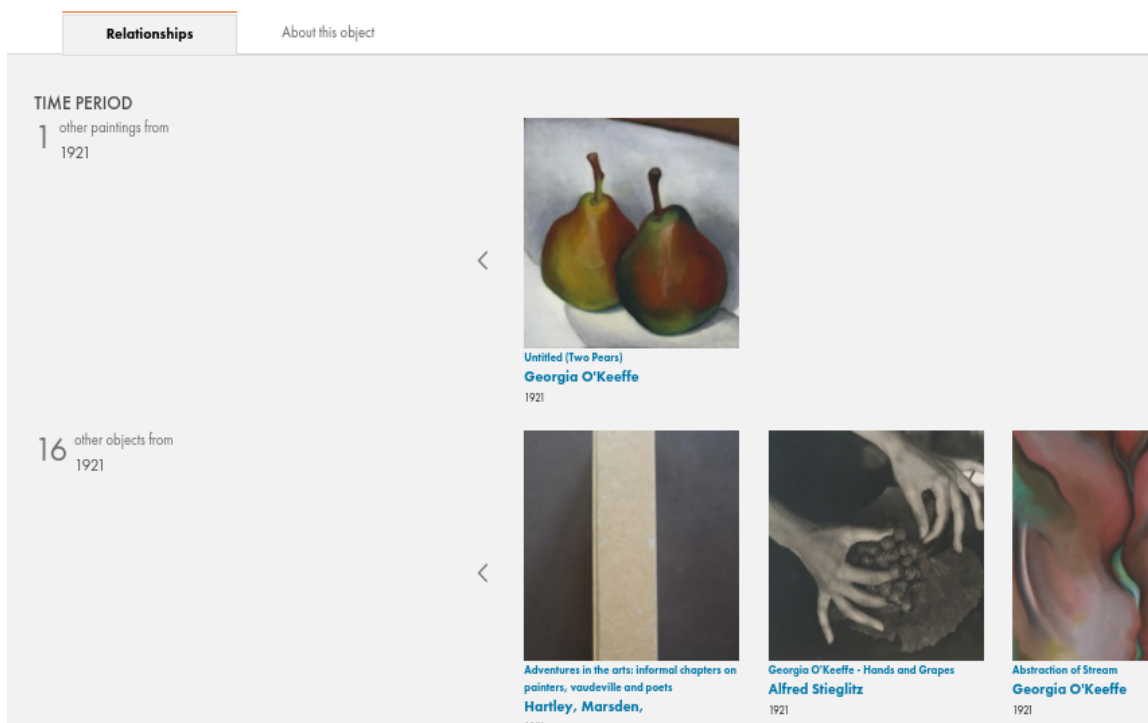
The O'Keefe Museum integrates linked data and linked open data into its "Collections Online" website. One of the distinctive aspects of the site is that a user can access different collection types, connected through linked data, using the site's interface. A user can search and access the museum's digital collections of art, artist materials, archives, books, and O'Keefe's home possessions. The O'Keefe site can serve as a model for other art museums with heterogeneous collections. Since the Linked Art model is specifically designed for museum collections, it may not fully capture or map the complexities of bibliographic and/or archival semantics that practitioners within the fields of libraries and archives use in their linked data conceptual and ontological models. However, the museum does make use of software such as ArchivesSpace, and the Liblime Koha (an integrated library system platform) to help manage archival and bibliographic data from these non-art-based digital collections.

Unlike the OCLC CONTENTdm and LD4P3 projects, the O'Keefe Museum's project is available publicly (in beta version), and the primary focus of the site through the museum's public interface is on user/audience experience and consumption of the museum's collections as linked data. The Linked Art Model serves as the primary semantic model for the O'Keefe Museum's use of linked open data. To

visualize the existing relationships and contexts between individual objects, thematic collections, object types, periods of time, and O’Keeffe’s connections with other artists, semantic relationships, the interface features a “relationships” tab for user discovery.

**Figure 5**

*Screenshot: O’Keeffe Museum Collections*



Note: from <https://collections.okeeffemuseum.org/>

By publishing its collection’s data as open through the museum’s online site and GitHub, the museum creates an opportunity for future data reuse by developers, metadata harvesters, and general users with linked data related technical skills and knowledge. Other linked data producers and publishers from within or outside the Cultural Heritage domain can specifically reuse the open data without the potential of violating digital rights terms or licensing. The O’Keeffe Museum’s site exemplifies many of the linked open data project stages identified by Marden, Li-Madeo, Whysel, and Edelstein (2013). These functions are important because they, in essence, articulate the value that

linked data and linked open data hold for museums and other cultural heritage organizations. That is, they establish the institution as a source of quality digital collections data that enriches user experience and creates space for the creation of new knowledge through the access and sharing of data.

### **Tools & Technology.**

The O’Keeffe Museum’s Collections Online site utilizes tools and schema such as JSON-LD, IIIF, and integrates multiple software platforms to bring together its different museum collections through a unified public interface as linked data and linked open data. GitHub is used for publication of open datasets and project related documentation. And other digital management systems such as Vernon Systems, ExLibris Voyager, LibLime Koha, ArchivesSpace, and Extensis are combined to function as the site’s software infrastructure (Neely, Linking O’Keefe, n.d.). For developers, linked open data serialized and published as JSON and RDF datasets can be downloaded from GitHub. JSON serialization of an object’s data is also accessible from the digital collections site from within individual object records. For general users, the IIIF Mirador Viewer allows users to zoom into images.

### **Barriers.**

Neely (2019) describes the O’Keeffe Museum’s Collections Online site through a “cultural collections as data” concept that is rooted in the approach of linked open data and the Semantic Web. Neely invites museum practitioners to think critically about how they visualize and structure their digital collections, and how their collections data is usable for discovery, consumption and connection with other museums and Cultural Heritage organizations’ collections as well. In developing its site, Neely noted that the O’Keeffe Museum faced challenges such as selecting the appropriate model and ontology, developing appropriate tools and available resources for a medium size museum, and the issue (though not referenced in Figure 3) of continued “exploration of usability and user-friendly access” (Neely, 2019).

### ***Smithsonian Open Access***



<https://www.si.edu/openaccess>

The Smithsonian Open Access site does not officially describe itself as a linked open data project. However, aspects of its use and publication of Cultural Heritage datasets complement Marden, Li-Madeo, Whysel, and Edelstein's (2013) linked open data project stages and characteristics such as advancing user experience, and enhancing/redefining the roles of Cultural Heritage institutions and consumers. The Smithsonian Open Access site features a set online tools for users (discussed below) that show an innovative way of using open data and engaging with linked open data concepts of context and reuse.

### **Tools & Technology.**

The Smithsonian Open Access site shares common tools and technology with the other linked data and linked open data projects analyzed in this paper. The Smithsonian API enables developers to access the Open Access open datasets and from their Edan repository. GitHub is used for publication of open datasets (serialized as RDF and JSON) and project related documentation. The IIIF Mirador viewer allows users to zoom into images and compare object images through their browser and the IIIF manifest view displays data structured in JSON. Additional tools for data discovery and dissemination include: the 3D Voyager, the Smithsonian Learning Lab, and the Smithsonian Flgshare.

A project such as the Smithsonian's Open Access project is a unique hybrid of publishing open datasets and tools for consumer data reuse. Other organizations can link to, use and reuse the data that the Smithsonian has made available through a CC0 license. The site does not make use of wiki-platforms for linked data, but other Smithsonian initiatives do (see

<https://confluence.si.edu/display/LODPP/Linked+Open+Data+Pilot+Project+Home>).

**Summary: Projects' Use of Linked Data & Linked Open Data Concepts, and Tools & Technology**

- OCLC's CONTENTdm Linked Data Pilot, LD4P3, and SNAC Phase 3 focus on production-end processes, while the Smithsonian's Open Access and O'Keefe Museum sites are broader in scope and include production and user-end functions.
- The Smithsonian Institution publishes open datasets in JSON-LD and via GitHub, and its Open Access site provides users with an array of tools for the discovery, use and reuse of Smithsonian-based collections data.
- Multi-institutional (national & international) collaboration for OCLC CONTENTdm Linked Data Pilot, LD4P3, and SNAC Phase 3.
- Use of GitHub, API access points, IIIF and JSON serialization
- Sustainability planning for OCLC CONTENTdm Linked Data Pilot, LD4P3, and SNAC Phase 3.
- Use of Wikibase and Wikidata infrastructure for the OCLC CONTENTdm Linked Data Pilot, LD4P3, SNAC Phase 3, and O'Keefe Museum.
- Conceptual and ontological models reflect different organizational and collection management systems employed by each Cultural Heritage sector

### **Discussion**

The projects analyzed in this paper all face issues of long-term sustainability and preservation going forward. The O'Keefe Museum, the OCLC CONTENTdm Linked Data Pilot, and Smithsonian Open Access site are integrated into their organization's repository workflows and this creates a space for critically exploring organizational engagement with linked data and linked open data concepts for their projects. Collaborative projects such as the OCLC CONTENTdm Linked Data Pilot, SNAC and LD4P3 require workflows for implementing the projects at a contributing partner level. Exploring management and digital preservation models may help project contributors or potential contributors to determine the appropriate workflows and organizational roles and resources needed to incorporate linked data or linked open data.

Two models that have been utilized for management of cultural heritage digital collections are the Digital Curation Centre's Curation Lifecycle Model (Higgins, 2008) and Brown's (2013) Digital Preservation Maturity Model.<sup>13</sup> The Curation Lifecycle Model includes stages and actions that are taken by an institution as it curates and preserves digital assets, datasets, and databases over time. It is applicable to the area of research data management and its holistic perspective of managing data could be explored with regard to managing linked data and linked open data going forward (for the Cultural Heritage domain). Brown's (2013) maturity model can be used to help cultural heritage organizations of different sizes (with access to varying resources) determine what level of capability their institution has for engaging with linked data or linked open data and strategies best meet their needs and goals.

A shared outcome of linked data and linked open data for libraries, archives and museums is the creation of a semantic infrastructure that enables a user to access LAM collections data. LAM collections data is contextualized and enriched through links/relationships to related external data (building a network of shared data). Ideally, users/consumers of linked data/linked open data platforms would be able to access LAM collections data from across cultural heritage sectors. While this ideal experience captures Berners-Lee's vision of a "web of data," it does not reflect some of the realities of managing LAM collections data and the current status of linked data tools and technologies for Cultural Heritage research. Each sector has its own established collections-based theory and application, professional standards, and data management perspectives. Therefore, as LAMs create and use linked data and linked open data, their collections data encompasses different semantics and contexts for each sector. And importantly, this would be possible only if actual and/or virtual relationships/connections existed between various LAM collections. The projects reviewed in this paper are still under production, and

---

<sup>13</sup> Brown's (2013) maturity model is organized into the following organizational assessment categories: organizational viability, stakeholder engagement, legal basis, policy framework, acquisition and ingest, bitstream preservation, logical preservation, metadata management, dissemination and infrastructure (p. 88).

they demonstrate that a linked data and linked open data landscape for cultural heritage is still being constructed and is not fully developed. LAMs are dealing with similar issues of data modeling, entity descriptions and tools to enhance editing of metadata and datasets as they seek to transform their collections data into linked data and linked open data. While they may use different conceptual and ontological models for structuring their collections, they are using common tools such as wiki technology (Wikibase and Wikidata), OpenRefine, and IIF and engaging in collaborative, multi-partner projects, and also using conferences, working groups, advisory boards and workshops for exchanging ideas and educating their members.

### **Open Ended Interviews with LAM Professionals**

I spoke with professionals from within the libraries, archives and museums sectors, who have had experience working directly on linked data and linked open data collections projects: Kalan Knudson Davis (Special Collections Metadata Librarian, University of Minnesota Libraries), Emanuelle Delmas-Glass (Collections Manager, Yale Center for British Art), Michelle Futornick (Stanford Libraries & Project Manager for Linked Data for Production series projects), and “Scann” (Evelin Heidel, Editor Open GLAM). Their collective knowledge and experiences offer insights into how linked data and open data are viewed as potentially powerful concepts and tools in shaping a new collections data experience for LAM collections’ producers and consumers.

Kalan Knudson Davis (Special Collections Metadata Librarian, University of Minnesota Libraries) worked with OCLC’s Project Passage and has worked on developing Python scripts related to linked data best practices for coding MARC subfields. Speaking from her perspective and experience of Project Passage and as professional who specializes in working with bibliographic metadata, Knudson Davis emphasized that, for the library cataloging community, linked open data is about pushing forward together, creating a group of data savvy catalogers who are engaged in making and enriching graphs of bibliographic data. It is critical that the process is collaborative. She noted the current work that is being

done with reconceptualizing bibliographic metadata with regard to linked data concepts; that it is important to think about description beyond the traditional format of MARC for bibliographic entities and to focus on community, context and entity qualities. Project Passage employed a version of Wikibase for its interface and management system, and Davis described the importance of the participants being able to edit entity descriptions and metadata through use of wiki-based tools (Personal Communication October 7, 2020).

The linked data and linked open data projects, assessed in this paper, reflect the importance of metadata, entity descriptions and data modeling on advancing the linked data and linked open data environment for Cultural Heritage collecting institutions.

The use cases that are being established through previous and current research projects and collaborations such as, the American Art Collaborative, Center for British Art, Europeana, Digital Public Library of America (DPLA), Linked Jazz, BIBFRAME, Project Passage, CONTENTdm, SNAC, and PHAROS are critical for analyzing and improving conceptual models, developing new technologies and paving the way for continued for linked data and linked open data.

Emmanuelle Delmas-Glass (Collections Manager, Yale Center for British Art) has worked with data migration, web-based knowledge representation, management of metadata throughout its lifecycle, and on the ground breaking linked data projects: the American Art Collaborative and PHAROS (Delmas-Glass, personal communication October 13, 2020). She identified key characteristics of the PHAROS project and its development including:

- All project partners agreeing to shared data model was a major achievement for the project (Linked Art Model)
- The shared data model is critical to data consistency, and enabling outside aggregators to harvest the project's data.

- The need for collaborative partners to consider what they want to accomplish together, and of equal importance, considering the needs of external users who will want to utilize the data.
- The use of linked open data can serve as a “pivotal moment” for the publishing and sharing of knowledge and data

(personal communication, October 13, 2020)

The project’s characteristics (collaboration and usability) highlighted by Delmas-Glass are at the essence of Berners-Lee’s concept of linked open data. Phase 3 of the PHAROS project is underway, and once completed, may serve as an additional linked open data project that Cultural Heritage organizations can learn from and participate in (Delmas-Glass & Sanderson, 2020).

Michelle Futornick, Project Manager (LD4P-LD4P3), identifies four core areas of focus regarding the use of linked data within the Linked Data for Production project series: models, tools, workflows and community (personal communication, October 23, 2020). For the libraries sector, Futornick describes a hybrid environment with linked data and MARC. She notes, for example, that the Library of Congress has built a tool for converting from BIBFRAME to MARC (due to the library community’s dependency on MARC-based systems) (Personal Communication, October 23, 2020).<sup>14</sup> One of the distinguishing features of the LD4P3 project is that it brings together a group of major linked data initiatives and stakeholders within the Libraries sector to advance how linked data is used to manage and enhance bibliographic collections data.

Scann (Evelin Heide, Editor Open GLAM) (<https://openglam.org/>), raises key issues that organizations should undertake when deciding on the use of open data for projects:

---

<sup>14</sup> The library sectors’ MARC 21 infrastructure is actively in use with cataloging and ILS applications, and the Library of Congress with collaboration has developed BIBFRAME to MARC 21 specifications and pilot testing to ensure community systems that are dependent on MARC maintain functionality (LOC, BIBFRAME news, n.d.). For BIBFRAME to MARC Conversion Tools see <https://www.loc.gov/bibframe/news/bibframe-to-marc-conversion.html>

- Consider how your resources are being made for people to use
- Think about why you want to use open data
- Spend time with your communities and understand who your stakeholders are (personal communication, October 25, 2020).

Scann's comments highlight the value and role of community discussion organizationally, cross-domain, and externally as Cultural Heritage organizations decide how and why to use open data. Such conversations may provide new ways of thinking about how to use and manage linked open data and its value for the Cultural Heritage domain and the communities they serve.

The SNAC, O'Keefe Museum and Smithsonian's Open Access projects specifically incorporate the publication of and the use/reuse of open datasets and are reflective of the connection between use of open data and consideration of how linked open data resources are developed for "people to use." The Smithsonian highlights examples of how its users (developers and/or general information users) have made use of some of its open data. SNAC uses linked open data in the development of its constellation records that catalogers and archivists can use to access established authority data from a centralized location. The O'Keefe Museum's use of linked open data enables viewers to see the relationships of the museum's objects/collections to various artists, periods of time, and other museum collections' data.

A major aspect of linked open data is the importance of linking to others' external data, as well as an organization making its data available for linking. The projects mentioned above have helped to create a Cultural Heritage environment where open datasets are available for use/reuse, and as datasets are published in formats such as RDF & JSON that developers can utilize. Though it is not necessarily clear the quantity of collections data that will be reused, nor how the datasets may be reused by others, these types of Cultural Heritage projects lay the groundwork for future efforts with linked open data.

### **Recommendations**

From this paper's literature review, its analysis of the OCLC CONTENTdm Linked Data Pilot, LD4P3, SNAC Phase 3, O'Keefe Museum's Collections Online, and Smithsonian Open Access projects, and insights that were gained from interviewing LAM professionals, there are three key recommendations that I suggest for LAM linked data and linked open data research:

- Increase cross-institutional and global collaboration among LAM institutions
- Continue to use shared data modeling within LAM sectors
- The use of management and preservation concepts from the Curation Lifecycle Model (Higgins, 2008) and the Digital Preservation Maturity Model from (Brown, 2013)

The first two recommendations are geared toward fostering the exchange of ideas, the development of new tools and technologies, and the creation of organizational collaborations that can help produce solutions for reducing and/or removing the barriers that currently restrict Cultural Heritage organizations that work with linked data and linked open data. Some of the primary barriers facing LAMs are: a steep learning curve for creating/using linked data, and the technical challenges associated with modeling legacy data into a linked data structure. The third recommendation is designed to help LAMs incorporate linked data/linked open data projects into the repository workflows of their digital libraries and collections. This includes the planning, assessment, and management of organizational personnel, infrastructure and additional resources that are critical to curation and preservation of digital collections that utilize linked data and/or linked open. The Digital Curation Centre's Curation Lifecycle Model suggests core actions that a digital repository should take to ensure that its datasets and digital assets are effectively managed and preserved throughout the "lifetime" of dataset or asset (from creation to access, use, reuse and/or transformation) (Higgins, 2008). This model is useful for LAMs working with linked data or linked open data because it creates a framework for critically assessing what roles, responsibilities, policies, actions and resources will be required by the organization for long-term use of linked data and linked open data. Brown's Digital Preservation Maturity Model (2013) is a useful



framework for assessing the resources, needs, and capabilities of a digital repository. If LAMs consider a maturity model type approach to their linked and/or linked open datasets and collections, they may be able to better understand how the organization is currently best situated to approach and use linked data/linked open data, and what potential capability is possible, desirable, or achievable.

### **Conclusion**

Linked data and linked open data projects from libraries, archives and museums range in focus and scale from dataset conversion and production, to open dataset publishing, to developer and user consumption via endpoints and UI. The research studied in this survey of literature reveals an existing pattern for the linked data and linked open data landscape within the Cultural Heritage domain: there are strong collaborative efforts within each of the individual sectors, as institutions employ linked data & linked open data for their collections data. And each sector has challenges with transforming their traditional systems of organization and their description for collections into systems that can manage and model linked data and linked open data structures. Wiki-based technologies are being used in all three sectors to explore interoperability and centralized management of and access to linked data and linked open data for producers and consumers nationally and internationally. Much of the large-scale research is being completed through U.S. and internationally collaborative projects. Dissemination of Cultural Heritage research data through linked data and linked open data remains an area of research growth. The use of open data for linked data to build linked open data projects has led to discussions of usability and the concept of Linked Open Usable Data with emphasis being placed on the ability of people to engage, use and reuse data. This trend highlights the need for additional research on sustainability, management and preservation of these datasets for the Cultural Heritage domain.

## References

- Adams, S. A. (2020). Linking Lost Jazz Shrines. Collections as Data, Cohort 1, Summative Forum, Jan 17, 2020, The University of Nevada, Las Vegas.  
<https://docs.google.com/presentation/d/1po6rJ4vxKCjWylKWUQPAiPGdSoRkmKKomk6H8H0bp1U/edit>
- Adams, S.A. (2019). The Project. 2018-2019 MoMA Archives Linked Open Data (LOD) Fellowship Report.  
<https://sites.google.com/view/momalodfellowship18-19/the-project>
- Alexeiv, V. (2018). Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. 8, Sofia, Bulgaria: Institute of Mathematics and Informatics – BAS, 2018. <https://www.ceeol.com/search/journal-detail?id=466>
- Allison-Cassin, S., Armstrong, A., Ayers, P., Cramer, T., Custer, M., Lemus-Rojas, M., ... & Stinson, A. (2019). ARL White Paper on Wikidata: Opportunities and Recommendations.  
<https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf>
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001, May). The Semantic Web. *Scientific American*. 29-37.
- Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. & Weitzner, D. J. (2006). A Framework for Web Science. *Foundations and Trends in Web Science*, 1(1). 1-30.  
<http://dx.doi.org/10.1561/18000000001>
- Bigelow, I. & Samples. (2020). MARC to BIBFRAME: Converting the PCC to Linked Data. *Cataloging & Classification Quarterly*. 58(3-4), 403-417. <https://doi.org/10.1080/01639374.2020.1751764>
- Branan, B. & Futornick, M. (2020). Linked Data for Production: Closing the Loop (LD4P3).  
<https://wiki.lyrasis.org/pages/viewpage.action?pageId=187176106>
- Brown, A. (2018). *Practical Digital Preservation: a how-to-guide for organizations of any size*. [ebook edition] Facet. (Original work published 2013).

<https://www.cambridge.org/core/books/practical-digital-preservation/E7E951A42FFA5BC14D412FEEA0972367>

Caraffa, C., Pugh, E., Stuber, T., & Ruby, L. W. (2020). PHAROS: A digital research space for photo archives. *Art Libraries Journal*, 45(1), 2-11. <https://doi.org/10.1017/alj.2019.34>

Clair, K. & Crowe, K. (2015). Developing a Tool for Publishing Linked Local Authority Data. *Journal of Library Metadata*. 15(3-4), 227-240. <https://doi.org/10.1080/19386389.2015.1099993>

Digital Curation Centre. (n.d.). About: What is Digital Curation?

<https://www.dcc.ac.uk/about/digital-curation>

Dobreski, B., Park, J., Leathers, A. & Qin, J. (2019). Remodeling Archival Metadata Descriptions. *Proceedings of the International Conference on Dublin Core and Metadata Applications, Seoul, South Korea*. 1-11. <https://dcpapers.dublincore.org/pubs/article/view/4223>

Delmas-Glass, E. & Sanderson, R. Fostering a Community of PHAROS Scholars through the Adoption of Open Standards. *Art Libraries Journal* 45(1), 19–23. Doi: [10.1017/alj.2019.32](https://doi.org/10.1017/alj.2019.32)

Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3), 75-92.

<https://doi.org/10.1609/aimag.v24i3.1720>

Europeana. (n.d.). Europeana Data Model Documentation.

<https://pro.europeana.eu/page/edm-documentation>

Freeman, J. (2019). What is an API? *Infoworld*. 2019, September 8.

<https://www.infoworld.com/article/3269878/what-is-an-api-application-programming-interfaces-explained.html?page=2>

Higgins, S. (2008). The DCC Life Cycle Curation Model. *International Journal of Digital Curation*, 1(3) 134-140. <https://doi.org/10.2218/ijdc.v3i1.48>

Hwang, K. (2018). Enriching Reconciled Data with OpenRefine. *Medium*, 2018, May 10.

<https://medium.com/the-bytegeist-blog/enriching-reconciled-data-with-openrefine-89b885dcadbb>

International Image Interoperability Framework (IIIF). (2020). FAQ.

<https://iiif.io/community/faq/#what-is-iiif>

International Council of Museums and International Committee for Documentation. (n.d.). What is the CIDOC CRM? <http://www.cidoc-crm.org/>

International Council on Archives. (2019). Records in Contexts: a Conceptual Model for Archival Description (v0.2).

[https://www.ica.org/sites/default/files/ric-cm-0.2\\_preview.pdf](https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf)

International Council on Archives. (2019). Records in Contexts: Ontology (RiC-O) (v0.1).

<https://www.ica.org/standards/RiC/ontology.html>

Jin, Q., Hahn, J., & Croll, G. (2016). BIBFRAME Transformation for Enhanced Discovery. *Library Resources and Technical Services*, 60(4), 223-235. <https://doi.org/10.5860/lrts.60n4.223>

Kapsalis, E. (2019). Wikidata: Recruiting the crowd to power access to digital archives. *Journal of Radio & Audio Media*, 26(1), 134-142. <https://doi.org/10.1080/19376529.2019.1559520>

Library of Congress (n.d.). BIBFRAME. <https://www.loc.gov/bibframe/docs/index.html>

Library of Congress (n.d.). BIBFRAME news.

<https://www.loc.gov/bibframe/news/bibframe-to-marc-conversion.html>

Marden, J., Li-Madeo, C., Whysel, N., & Edelstein, J. (2013, September). Linked open data for cultural heritage: evolution of an information technology. In *Proceedings of the 31st ACM international conference on Design of communication* (pp. 107-112).

<https://doi.org/10.1145/2507065.2507103>

Michael, B., & Han, M. J. K. (2019). Assessing BIBFRAME 2.0: Exploratory implementation in metadata

- maker. Proceedings of the International Conference on Dublin Core and Metadata Applications, 26-31. <https://experts.illinois.edu/en/publications/assessing-bibframe-20-exploratory-implementation-in-metadata-make>
- Myntti, J., & Cothran, N. (2013). Authority control in a digital repository: Preparing for Linked data. *Journal of Library Metadata*, 13(2–3), 95–113. <https://doi.org/10.1080/19386389.2013.826061>
- Neatrou, A. & Myntti, J. (2019). Western Name Authority File: A Pilot Regional Name Authority Project. *Journal of Library Metadata*, 19, 19-38. <https://doi.org/10.1080/19386389.2019.1589685>
- Neely, L. Linking O’Keeffe (2018, October). Generating new knowledge through linked open data from a single artist Museum. [Conference Presentation]. CIDOC 2018 Conference. [http://network.icom.museum/fileadmin/user\\_upload/minisites/cidoc/images/1.4--4--Linking\\_O\\_Keeffe\\_CIDOC.pdf](http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/images/1.4--4--Linking_O_Keeffe_CIDOC.pdf)
- Neely, L., Luther, A., & Weinard, C. Cultural Collections as Data: Aiming for Digital Data Literacy and Tool Development. (2019, February) *MW19: MW 2019*. <https://mw19.mwconf.org/paper/cultural-collections-as-data-aiming-for-digital-data-literacy-and-tool-development/>
- Newbury, D. (2017). Art Tracks: Using Linked Open Data for Object Provenance in Museums. *Museum and the Web, MW17: MW 2017, February 1, 2017*. <https://mw17.mwconf.org/paper/art-tracks-using-linked-open-data-for-object-provenance-in-museums/>
- OCLC. (2020). CONTENTdm Linked Data Pilot. <https://www.oclc.org/research/areas/data-science/linkeddatab/contentdm-linked-data-pilot.html>
- OCLC. (2018). Linked Data Survey Results: 2014, 2015, 2018. [Excel spreadsheet]. <https://www.oclc.org/content/dam/research/activities/linkeddatab/oclc-research-linked-data-implementers-survey-2014.xlsx>

Open Knowledge Foundation. (n.d.). Open Data Handbook Guide: What is Open Data?

<https://opendatahandbook.org/guide/en/what-is-open-data/#:~:text=For%20our%20purposes%2C%20open%20data,requirement%20to%20attribute%20and%20sharealike.&text=The%20data%20must%20also%20be%20available%20in%20a%20convenient%20and%20modifiable%20form>

Park, J. R., Richards, L. L., & Brenza, A. (2019). Benefits and challenges of BIBFRAME. *Library Hi Tech*.

37(3), 549-565. <https://doi.org/10.1108/LHT-08-2017-0176>

Pitti, D., Hu, R., Larson, R., Tingle, B., & Turner, A. (2015). Social networks and archival context: From project to cooperative archival program. *Journal of Archival Organization*, 12(1-2), 77-97.

<https://doi.org/10.1080/15332748.2015.999544>

Social Networks and Archival Context. SNAC Cooperative: Technical Docs. [GitHub].

<https://github.com/snac-cooperative/technical-docs>

Sanderson, R. (2018). Shout it Out LOUD. [Keynote Conference Presentation]. *EuropeanTech*, 15 & 16

May 2018, Rotterdam, the Netherlands. <https://www.slideshare.net/Europeana/shout-it-out-loud-by-rob-sanderson-europeanatech-conference-2018-98225909>

Schreur, P.E. (2020, June). LD4P3 Proposal Summary.

[https://docs.google.com/document/d/1cTXool7YE\\_Ymelo3IARhRBe1yeTRhxQIZxIJ4J5esao/edit](https://docs.google.com/document/d/1cTXool7YE_Ymelo3IARhRBe1yeTRhxQIZxIJ4J5esao/edit)

Smith-Yoshimura, K. (2018). Analysis of 2018 international linked data survey for implementers.

*Code4Lib journal*, (42). <https://journal.code4lib.org/articles/13867>

Smith-Yoshimura, K. (2018). The rise of Wikidata as a linked data source. *Hanging together. The OCLC*

*research blog*, August, 6. <https://hangingtogether.org/?p=6775>

Taniguchi, S. (2017). Examining BIBFRAME 2.0 from the viewpoint of RDA metadata schema. *Cataloging*

*& Classification Quarterly*, 55(6), 387-412. <https://doi.org/10.1080/01639374.2017.1322161>

Tillet, B. (1991). A taxonomy of bibliographic relationships. *Library resources & technical services*, 35(2),

150-158. <https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=5005983>

Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An Overview of the Current State of Linked and Open Data in Cataloging. *Information Technology and Libraries*, 37(4), 47-80.

University of California. (2018). UC Linked Data Project Team report.

[https://libraries.universityofcalifornia.edu/groups/files/doc/docs/UC%20Linked%20Data%20Final%20Report\\_01Oct2018.pdf](https://libraries.universityofcalifornia.edu/groups/files/doc/docs/UC%20Linked%20Data%20Final%20Report_01Oct2018.pdf)

Washer, B., Mixburn, J., & Einaudi, M. (2019, August). Introducing the CONTENTdm Linked Data Pilot Project. [presentation]. CONTENTdm User Group Meeting. Indianapolis, IN.

<https://www.oclc.org/research/presentations/2019/08072019-mixer-washburn-intro-cdm-linked-data-pilot.html>

Wildenhaus, K. (2019). The Possibilities of Constructing Linked Data for Art Exhibition Histories.

*Art Documentation: Journal of the Art Libraries Society of North America*. 38(1), 22-34.

<https://doi.org/10.1086/702890>

World Wide Web Consortium (W3C). (2009). Design Issues: Linked Data.

<https://www.w3.org/DesignIssues/LinkedData.html>

World Wide Web Consortium (W3C). (2013). SPARQL 1.1 Overview.

<https://www.w3.org/TR/sparql11-overview/>

World Wide Web Consortium (W3C). (2014). RDF 1.1 Concepts and Abstract Syntax.

<https://www.w3.org/TR/rdf11-concepts/>

World Wide Web Consortium (W3C). (2015). Standards: The Semantic Web.

<https://www.w3.org/standards/semanticweb/>

World Wide Web Consortium (W3C). (2015). Ontologies: Vocabulary.

<https://www.w3.org/standards/semanticweb/ontology>

World Wide Web Consortium (W3C) JSON-LD Community Group. (n.d.). JSON for Linking Data.

<https://json-ld.org/#>

Zapounidou, S., Sfakakis, M. & Papatheodorou, C. (2019). Mapping Derivative Relationships

from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*, 57(5), 278-308,

<https://doi.org/10.1080/01639374.2019.1650152>

Zeng, M.L. & Mayr, P. (2019). Knowledge Organization Systems (KOS) in the Semantic Web: a

multi-dimensional review. *International Journal of Digital Libraries* 20, 209–230.

<https://doi.org/10.1007/s00799-018-0241-2>



### Annotated Bibliography

- Anutariya, C. & Chanhom, W. (2019). TOMS: A Linked Open Data System for Collaboration and Distribution of Cultural Heritage Artifact Collections of National Museums in Thailand. *New Generation. Computing.* 37, 479–498. <https://doi.org/10.1007/s00354-019-00063-1>
- Chanhom & Anutariya discuss the research, design, development, implementation and evaluation process involved in the development of the TOMS Linked Open Data system. Expanding on previous proposal and prototype stages of the TOMS project, the peer-reviewed study evaluates the TOMS system at three national museums. The article includes an overview of related Linked Open Data projects, use of the CIDOC-CRM model, an explanation of TOMs technological infrastructure, and the impact of TOMS on user experiences (internal and external). The study provides an example of a recent international Linked Open Data project and the primary audience is an [academic international Cultural Heritage sector. However, the Linked Open Data system/technologies described in the article may have implications and insights for future research within the United States.
- Bigelow, I. & Samples. (2020). MARC to BIBFRAME: Converting the PCC to Linked Data. *Cataloging & Classification Quarterly.* 58(3-4), 403-417. <https://doi.org/10.1080/01639374.2020.1751764>
- Bigelow & Samples examine the role of the Program for Cooperative Cataloging (PCC), in collaboration with the Library of Congress (LOC), Share Virtual Discovery Environment (Share-VDE), and Swedish National Library, in the Phase 2 of the Marc to BIBFRAME project. Through a literature review and using the Phase 2 as a case study, the authors put forth a series of “key considerations” for further development of BIBFRAME as a global resource for development, use and reuse of linked data. A crucial factor in the continued development of BIBFRAME is the recognition that there is similar functionality between BIBFRAME and other data models such as the IFLA-LRM. The Bigelow & Samples indicate that additional work by the PCC, including the

Opus Model that brings together Share-VDE Opus, LC Hub and IFLA-LRM, is needed, but that the completion of the second phase of the conversion project has made significant progress toward the advancement of Linked Data bibliographic records. This article is related to my Linked Open Data research as it includes an overview of the historical development of BIBFRAME and describes the current state of the BIBFRAME project and the relationships between the collaborative partner organizations and their roles and contributions.

Clair, K. & Crowe, K. (2015). Developing a Tool for Publishing Linked Local Authority Data.

*Journal of Library Metadata*. 15(3-4), 227-240. <https://doi.org/10.1080/19386389.2015.1099993>

Crowe and Clair present an overview of Linked Data publications and projects with regard to Virtual International Authority Files and archival description standards and practices. The authors argue the need for a collaborative effort toward the design and implementation of a Linked Data tool that can be used by smaller archival and collecting institutions who may have local, non-standardized, institution specific authority records. Crowe and Clair assess the implications of previous linked data projects, and identify their core audience of archivists, librarians, and cultural heritage professionals as the primary groups that are in need of a tool for publishing linked authority records. Referencing work completed by Name Authority Cooperative (NACO) and Social Networks and Archival Context (SNAC), the authors describe strategies for advancing a linked data publishing tool. This article is valuable to my research as it documents one of the ways that linked data has been applied to archival and library work.

Dobreski, B., Park, J., Leathers, A. & Qin, J. (2019). Remodeling Archival Metadata

Descriptions. *Proceedings of the International Conference on Dublin Core and Metadata*

*Applications, Seoul, South Korea*. 1-11. <https://dcpapers.dublincore.org/pubs/article/view/4223>

The authors describe the development of the “Linked Archives Model” project at the Syracuse University’s Libraries and Special Collections and Research Center. Using existing data from

three library media collections, the research team created a pilot study and built an archival ontology model designed to represent item-level archival collections (physical & digital surrogates) that would reflect the needs of end users. To accomplish this, the Linked Archives Model project used existing library metadata to develop classes, properties and relationships, and Protege software to design an OWL-based ontology. The project pilot focused primarily on descriptive metadata and the goal of meeting users' resource discovery needs. This research is relevant to my paper as is an example of existing archival collections being transformed into a linked data format. The authors describe the process through which an institution can develop a custom linked data ontology that includes existing linked data ontology models and vocabularies.

Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3), 75. <https://doi.org/10.1609/aimag.v24i3.1720>

Doerr provides an in-depth explanation of the CIDOC Conceptual Reference Model. The publication is most accessible to an audience with familiarity of conceptual data models, ontology design and semantic structures. The publication is relevant to my research paper as it provides an overview of CIDOC-CRM which has been widely used within the Cultural Heritage Sector (particularly museums) for semantically structuring the collections data.

Mitchell, E. T. (2016). Library Linked Data: Early Activity and Development. *Library Technology Reports*. 52(1), 5-33. <https://doi.org/10.1080/02763869.2017.1294004>

Mitchell updates his previous published report (2013) on the state of Linked Open Data and libraries, archives and museums. Divided into four chapters, the publication summarizes Linked Open Data projects, research, vocabularies, standards, new developments, organizations and OCLC surveys on Linked Open Data. Written for LAMs professionals, the report is important for my research because it provides historical analysis of linked data projects and technologies that

have emerged since the Semantic Web was introduced. Mitchell concludes that the LAM sectors are generally advancing along similar Linked Data trajectories.

Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An Overview of the Current State of Linked and Open Data in Cataloging. *Information Technology and Libraries*, 37(4), 47-80.

<https://doi.org/10.6017/ital.v37i4.10432>

The authors conduct a review of linked data and open data research within the libraries sector. Compared with earlier more generalized overviews by Mitchell (2016) and Yoose & Perkins (2013), that cover the Cultural Heritage domain, this work specifically examines three research questions for library cataloging: “ How Linked Open Data (LOD) and Vocabularies (LOV) are transforming the digital landscape of library catalogs?, What are the prominent/major issues, challenges, and research opportunities in publishing and consuming bibliographic metadata as Linked and What is the possible impact of extending bibliographic metadata with the user generated content and making it visible on the LOD cloud?” (p. 49). It is helpful to read this article in conjunction with other bibliographic metadata research studies that explore the use of BIBFRAME.

Wildenhaus, K. (2019). The Possibilities of Constructing Linked Data for Art Exhibition Histories.

*Art Documentation: Journal of the Art Libraries Society of North America*. 38(1), 22-34.

<https://doi.org/10.1086/702890>

Wildenhaus analyzes the potential benefits and challenges of integrating Linked Data into art exhibition histories. Focusing primarily on an audience of museum professionals, she discusses projects such as the Smithsonian American Art Collaborative project and summarizes prior research and publications. Wildenhaus’ arguments toward the potential value that Linked Data can provide in contextualizing and expanding relationships between artists, artworks,

geographic locations for exhibitions is relevant to the question of how linked data is perceived within collecting institutions (LAMs).

Zapounidou, S., Sfakakis, M. & Papatheodorou, C. (2019). Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*. 57(5), 278-308, <https://doi.org/10.1080/01639374.2019.1650152>

As RDA was adopted as the official cataloging standard replacing AACR2, this study explores the mapping of relationships from RDA into BIBFRAME 2.0. Certain relationships expressed in RDA were more successful at being transitioned into BIBFRAME than others. Catalogers and other information and cultural heritage professionals who work with bibliographic metadata and semantic modeling are the core audience for this publication. While the study contains much technical and specialized terminology, it is useful for acquiring a better understanding of the process involved in mapping other schemas into BIBFRAME and the challenges faced by those who are transforming bibliographic metadata and relationship descriptions into BIBFRAME.

## Appendix 1

### Terminology

**Application Programming Interface (API):** code that enables communication between software applications and provides developers with a protocol for interacting with a particular application or software component. Through an API and use of J-SOON and/or other languages, developers can access, edit/use datasets (Freeman, Infoworld, 2019).

**JSON-LD:** a framework for expressing linked data that enables the defining of elements and semantic description of relationships between objects in conjunction with the use of an URI (W3C JSON-LD, JSON for Linking Data, 2020).

**IIIF (International Image Interoperability Framework):** a group of API specifications that enables its users to improve functionality, interoperability, and access of images held in digital collections/repositories ((IIIF, FAQ, 2020).

**Ontology:** a model or system for organizing specific vocabularies or terms, concepts, and relationships about a particular subject or domain of knowledge (W3C, Ontologies, 2015).

**Open Data:** “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike” (Open Knowledge Foundation, Open Data Handbook, n.d.).

**RDF:** Resource Description Framework is an XML based structure for expressing relationships between entities in a machine-readable format. The semantic structure of: subject predicate object (RDF), is essentially a knowledge structure consisting of a defined entity and its relationship to another defined entity (via subject and predicate), and it can be visually represented as a graph (W3C, RDF, 2014).

**SPARQL:** a query language used for retrieving data from stored RDF triples. SPARQL endpoints serve as an access point for conducting queries (W3C, SPARQL 2008).

## Appendix 2

### Cultural Heritage Publications, Organizations & Web Projects: Linked Data, LOD & Open Data

#### *Publications*

- Biswas, U., Marjit, U. & Sharma, K. (2018). Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet. *Information Technology and Libraries*, 37(3), 29-49.  
<https://doi.org/10.6017/ital.v37i3.10177>
- Chowdhury, S. (2019, March 11). Using Linked Data for Discovery and Preservation. *Educause Review*.  
<https://er.educause.edu/articles/2019/3/using-linked-data-for-discovery-and-preservation>
- Daquino, M., Mambelli, F., Peroni, S., Tomasi, F. & Vitali, F. (2017). Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data. *Journal on Computing and Cultural Heritage* 10(4), 1–21. <https://doi.org/10.1145/3051487>
- Davis, K. K. (2020). An insider’s look at “Project Passage” in seven linked data lessons, six constants, five changes ... and four webcomics. OCLC Next. <http://www.oclc.org/blog/main/insiders-look-at-project-passage/>
- Deng, S. (2018). A step forward: adding linked data vocabularies to digital repositories. Florida Library Association 2018 Annual Conference. May 24, 2018, Orlando, FL.  
<https://stars.library.ucf.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1652&context=ucfscholar>
- Dijkshoorn, C., Jongma, L., Aroyo, L., Van Ossenbruggen, J., Schreiber, G., Ter Weele, W., & Wielemaker, J. (2018). The Rijksmuseum collection as linked data. *Semantic Web*, 9(2), 221-230.  
<https://doi.org/10.3233/SW-170257>
- Fink, E. E. (2018). American Art Collaborative (AAC) Linked Open Data (LOD) Initiative, Overview and Recommendations for Good Practices.

<https://repository.si.edu/bitstream/handle/10088/106410/OverviewandRecommendationsAccessible.pdf?sequence=1&isAllowed=y>

Haslhofer, B. & Isaac, A. (2011). The Europeana Linked Open Data Pilot. International Conference on Dublin Core and Metadata Applications (DC 2011). The Hague.

<http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/55>

Hyvönen, E., Lindquist, T., Mäkelä, E., & Juha Törnroos. (2017). WW1LOD: an application of CIDOC-CRM to World War 1 linked data. *International Journal on Digital Libraries*. 18, 333–343.

<https://doi.org/10.1007/s00799-016-0186-2>

Godby, J., Smith-Yoshimura, K., Washburn, B., Davis, k., Detling, K., Fernsebner, C. E., Folsom, S., Li, X., McGee, M., Miller, K., Moody, H., Tomren, H., and Thomas, C. (2019). Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage. Dublin, OH: OCLC Research.

<https://doi.org/10.25333/faq3-ax08>

Greenwald, D. (2020, April 27). *What Can Data Teach Us About Museum Collections*. American Alliance of Museums Curatorial Practice. <https://www.aam-us.org/2020/04/27/what-can-data-teach-us-about-museum-collections/>

Konstantinou N., Spanos DE. (2015). Deploying Linked Open Data: Methodologies and Software Tools. In, *Materializing the Web of Linked Data*. (pp. 51-71). Springer. [https://doi.org/10.1007/978-3-319-16074-0\\_3](https://doi.org/10.1007/978-3-319-16074-0_3)

OCLC Research Archives and Special Collections Linked Data Review Group. (2020). Archives and Special Collections Linked Data: Navigating between Notes and Nodes. <https://doi.org/10.25333/4gtz-zd88>.

Okeefe, E., Wacker, M. L'Ecuyer-Coelho, M. (2019). The Outcome of the ArtFrame Project: A Domain-Specific BIBFRAME Exploration. *Art Documentation: Journal of the Art Libraries Society of North America*. 38 (1), 1-21. <https://doi.org/10.1086/703508>



Owens, T. (2016). Curating in the Open: A Case for Iteratively and Openly Publishing Curatorial Research on the Web. *Curator: The Museum Journal*. 59(4), 427-442.

<https://doi.org/10.1111/cura.12168>

Ruhland, J. & Wenige, L. (2018). Retrieval by recommendation: using LOD technologies to improve digital library search. *International Journal on Digital Libraries*. 19, 253–269.

<https://doi.org/10.1007/s00799-017-0224-8>

Zapounidou, S., Sfakakis, M. & Papatheodorou, C. (2019) Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*, 57(5), 278-308.

<https://doi.org/10.1080/01639374.2019.1650152>

### **Organizations & Conferences: Linked Data, LOD and/or Open Data**

CIDOC. (2020). CIDOC 2020 Conference: Digital Transformation in Cultural Heritage Institutions.

December 7 - 10, 2020. <http://institutions.ville-geneve.ch/fr/mah/expositions-evenements/colloques/conference-du-cidoc-2020/>

LD4. (2020). 2020 LD4 Conference on Linked Data in Libraries. <https://ld42020.sched.com/>

LODLAM. (2020). LODLAM 2020 Summit: Public Discussion Documents [Conference Notes].

<https://drive.google.com/drive/u/0/folders/1ZTapwNwlgSLMXRLVYix0cDCwgiNw7yci>

Museum Computer Network. (2020). *MCN 2020 Virtual: The Joys of Connecting Your Collections to*

*Wikidata*. November 10-12 & 17-19, 2020. <https://mcn2020virtual.sched.com/event/dfYm/the-joys-of-connecting-your-collections-to-wikidata>

Open GLAM. (2020). <https://openglam.org/>

### **Linked Data & LOD Data Projects**

Carnegie Hall Data Lab. (2020). *Carnegie Hall Data Lab*. <https://carnegiehall.github.io/datalab/>

Cornell University Library. (2020). *Cornell Hip Hop Collection*. <https://rmc.library.cornell.edu/hiphop/>

Columbia University. (2018). *GitHub: melanieWacker Art Properties*.

<https://github.com/melanieWacker/ArtProperties>

Dublin Core Metadata Initiative. (2016). *Linked Data for Professional Education (LD4PE)*.

<http://explore.dublincore.net/>

Europeana. (n.d.) Linked Open Data. Europeana Pro. <https://pro.europeana.eu/page/linked-open-data>

New York Public Libraries. (2016). *NYPL Labs*. <https://www.nypl.org/collections/labs>

Oregon State University Libraries and University of Oregon Libraries. (n.d.) OpaqueNamespace.

<https://opaquenamespace.org/>

Yale Center for British Art. (n.d.). *Collections Data Sharing*.

<https://britishart.yale.edu/collections-data-sharing>

Weeksville Heritage Center & the Semantic Lab at Pratt. (2020). Linking lost Jazz Shrines.

<https://sites.google.com/weeksvillesociety.org/linking-lost-jazz-shrines/home>

### ***Open Data Projects***

Coding da Vinci. (2020). *{CODING DA V1NC1}*. <https://codingdavinci.de/>

Open Heritage 3D. (2020). About. <https://openheritage3d.org/about>