

Bridging the Gap

Using Linked Data to Improve Discoverability and Diversity in Digital Collections

Jason Boczar, Bonita Pollock, Xiyang Mi, and Amanda Yeslibas

ABSTRACT

The year of COVID-19, 2020, brought unique experiences to everyone in their daily as well as their professional life. Facing many challenges of division in all aspects (social distancing, political and social divisions, remote work environments), University of South Florida Libraries took the lead in exploring how to overcome these various separations by providing access to its high-quality information sources to its local community and beyond. This paper shares the insights of using Linked Data technology to provide easy access to digital cultural heritage collections not only for the scholarly communities but also for those underrepresented user groups. The authors present the challenges at this special time of the history, discuss the possible solutions, and propose future work to further the effort.

INTRODUCTION

We are living in a time of division. Many of us are adjusting to a new reality of working separated from our colleagues and the institutions that formerly brought us together physically and socially due to COVID-19. Even if we can work in the same physical locale, we are careful and distant with each other. Our expressions are covered by masks, and we take pains with hygiene that might formerly have felt offensive.

But the largest divisions and challenges being faced in the United States go beyond our physical separation. The nation has been rocked and confronted by racial inequality in the form of Black Lives Matter, a divisive presidential campaign, income inequality exacerbated by COVID-19, the continued reckoning with the #metoo movement, and the wildfires burning the West coast. It feels like we are burning both literally and metaphorically as a country.

Adding fuel to this fire is the consumption of unreliable information. Ironically, even as our divisions become more extreme, we are increasingly more connected and tuned into news via the internet. Sadly, fact checking and sources are few and far between on social media platforms, where many are getting their information. The Pew Foundation report *The Future of Truth and Misinformation Online* warns that we are on the verge of a very serious threat to the democratic process due to the prevalence of false information. Lee Raine, Director of The Pew Research Center's Internet and Technology Project, warns, "A key tactic of the new anti-truthers is not so much to get people to believe in false information. It's to create enough doubt that people will give up trying to find the truth, and distrust the institutions trying to give them the truth."¹

Libraries and other cultural institutions have moved very quickly to address and educate their populations and the community at large, trying to give a voice to the oppressed and provide

Jason Boczar (jboczar@usf.edu) is Digital Scholarship and Publishing Librarian, University of South Florida. **Bonita Pollock** (pollockb1@usf.edu) is Associate Director of Collections and Discovery, University of South Florida. **Xiyang Mi** (xmi@usf.edu) is Digital Initiative Metadata Librarian, University of South Florida. **Amanda Yeslibas** (ayesilbas@usf.edu) is E-resource Librarian, University of South Florida. © 2021.



reliable sources of information. The University of South Florida (USF) Libraries reacted by expanding antiracism holdings. USF's purchases were informed by work at other institutions, such as the University of Minnesota's antiracism reading lists, which has in turn grown into a rich resource that includes other valuable resources like the Mapping Prejudice Project and a link to the Umbra Search.² The Triad Black Lives Matter Protest Collection at the University of North Carolina Greensboro is another example of a cultural institution reacting swiftly to document, preserve, and educate.³

These new pages and lists being generated by libraries and cultural institutions seem to be curated by hand using tools that require human intervention to make them and keep them up to date. This is also a challenge the USF Libraries faced when constructing its new African American Experience in Florida Portal, a resource that leverages already existing digital collections at USF to promote social justice. Another key challenge is linking new digital collections and tools to already established collections and holdings. Beyond the new content being created in reaction to current movements, there is already a wealth of information established in rich archives of material, especially regarding African American history.

Digital collections need to be discoverable by a wide audience to achieve resource sharing and educational purposes. This is a challenge many digital collections struggle with, because they are often being siloed from library and archival holdings even within their own institutions. All the good information in the world is not useful if it is not findable. An example of a powerful discovery tool that is difficult to find and use is the Umbra Search (<https://www.umbrasearch.org/>) linked to the University of Minnesota's anti-racism reading list. Umbra Search is a tool that aggregates content from more than 1,000 libraries, archives, and museums.⁴ It is also supported by high-profile grants from the Institute of Museum and Library Services, the Doris Duke Charitable Foundation, and the Council on Library and Information Resources. However, the website is difficult to find in a web search. Umbra Search was named after Society of Umbra, a collective of black poets from the 1960s. The terms Umbra and Society of Umbra do not return useful results for finding the portal, nor do broader searches of African American history. The portal is difficult to find through basic web searches. One of the few chances for a user to find the site is if they came upon the human-made link in the University of Minnesota anti-racism reading list.

Despite enthusiasm from libraries and other cultural institutions, new purchases and curated content are not going to reach the world as fully as hoped. Until libraries adopt open data formats in favor of locking away content in closed records like MARC, library and digital content will remain siloed from the internet. The library catalog and digital platforms are even siloed from each other. We make records and enter metadata that is fit for library use but not shareable to the web. As Karen Coyle asked in her LITA keynote address a decade ago, the question is how can libraries move from being "on the Web" to being "of the Web"?⁵ The suggested answer and the answer the USF Libraries are researching is with linked data.

LITERATURE REVIEW

The literature on linked data for libraries and cultural heritage resources reflects an implementation that is "gradual and uneven." As national libraries across the world and the Library of Congress develop standards and practices, academic libraries are still trying to understand their role in implementation and identify their expertise.⁶

In 2006 Tim Berners-Lee, the creator of the semantic web concept, outlined four rules of linked data:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs so that they can discover more things.⁷

It was not too long after this that large national libraries began exploring linked data and experimenting with uses. In 2010 the British Library presented its prototype of linked data. This move was made in accordance with the UK government's commitment to transparency and accountability along with the user's expectation that the library would keep up with cutting edge trends.⁸ Today the British Library has released the British National Bibliography as linked data instead of the entire catalog because it is authoritative and better maintained than the catalog.⁹

The national libraries of Europe, spurred on by government edicts and Europeana (<https://www.europeana.eu/en>), are leading the progress in implementation of linked data. National libraries are uniquely suited to the development and promotion of new technologies because of their place under the government and proximity to policy making, bridging communication between interested parties and the ability to make projects into sustainable services.¹⁰ A 2018 survey of all European national libraries found that 15 had implemented linked data, two had taken steps for implementation and three intended to implement it. Even national libraries that were unable to implement linked data were contributing to the Linked Data Open Cloud by providing their data in datasets to the world.¹¹

Part of the difficulty with earlier implementation of linked data by libraries and cultural heritage institutions was the lack of a "killer example" that libraries could emulate.¹² The relatively recent success of European national libraries might provide those examples. Many other factors have slowed the implementation of linked data. A survey of Norwegian libraries in 2009 found considerable gap in the semantic web literature between the research undertaken in the technological field and the research in the socio-technical field. Implementing linked data requires reorganization of the staff, commitment of resources, education throughout the library and buy-in from the leadership to make it strategically important.¹³ The survey of European National Libraries cited the exact same factors as limitations in 2018.¹⁴

Outside of European national libraries the implementation of linked data has been much slower. Many academic institutions have taken on projects that tend to languish in a prototype or proof of concept phase.¹⁵ The library-centric Talis Group of the United Kingdom "embraced a vision of developing an infrastructure based on semantic web technologies" in 2006, but abandoned semantic web-related business activities in 2012.¹⁶ It has been suggested that it is premature to wholly commit to linked data, but it should be used for spin-off projects in an organization for experimentation and skill development.¹⁷

Linked data is also still proving to be technologically challenging for implementation of cultural heritage aggregators. If many human resources are needed to facilitate linked data, it will remain an obstacle for cultural heritage aggregators. A study has shown automated interpretation of

ontologies is hampered by a lack of inter-ontology relations. Cross-domain applications will not be able to use these ontologies without human intervention.¹⁸

Aiding in the development and awareness of linked data practices for libraries is the creation and implementation of BIBFRAME by the Library of Congress. The Library of Congress's announcement in July 2018 that BIBFRAME would be the replacement of MARC definitively shows that the future of library records is focused on linking out and integrating into the web.¹⁹ The new RDA (Resource Description and Access) cataloging standards made it clear that MARC is no longer the best encoding language for making library resources available on the web.²⁰ While RDA has adopted the cataloging rules to meet a variety of new library environments, the MARC encoding language makes it difficult for computers to interpret and apply logic algorithms to the MARC format.

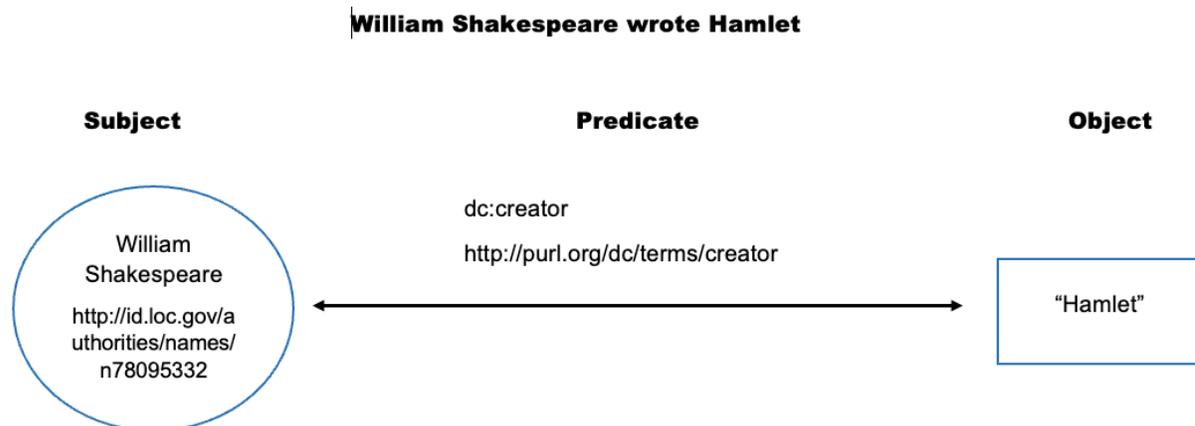
In response, the Library of Congress commissioned the consulting agency Zepheria to create a framework that would integrate with the web and be flexible enough to work with various open formats and technologies, as well as be able to adapt to change. Using the principles and technologies of the open web, the BIBFRAME vocabulary is made of "Resource Description Framework (RDF) properties, classes, and relationships between and among them."²¹ Eric Miller, the CEO of Zepheria, says BIBFRAME "works as a bridge between the description component and open web discovery. It is agnostic with regards to which web discovery tool is employed" and though we cannot predict every technology and application BIBFRAME can "rely on the ubiquity and understanding of URIs and the simple descriptive power of RDF."²²

The implementation of linked data in the cultural heritage sphere has been erratic but seems to be moving forward. It is important to pursue though because bringing local data out of the "deep web" and making them open and universally accessible, means offering minority cultures a democratic opportunity for visibility."²³

LINKED DATA

Linked data is one way to increase the access and discoverability of critical digital cultural heritage collections. Also referred to as semantic web technologies, linked data follows the W3C Resource Description Framework (RDF) Standards.²⁴ According to Tim Berners-Lee, the semantic web will bring structure and well-defined meaning to web content allowing computers to perform more automated processes.²⁵ By providing structure and meaning to digital content, information can be more readily and easily shared between institutions. This provides an opportunity for digital cultural heritage collections of underrepresented populations to get more exposure on the web.

Following is a brief overview of linked data to illustrate how semantic web technologies function. Linked data is created by forming semantic triples. Each RDF triple contains Uniform Resource Identifiers or URIs. These identifiers allow computers (machines) to "understand" and interpret the metadata. Each RDF triple consists of three parts: a subject, a predicate, and an object. The subject defines what the metadata RDF triple is about, while the object contains information about the subject which is further defined by the relationship link in the predicate.

**RDF Triple:**

`<http://id.loc.gov/authorities/names/n78095332> <http://purl.org/dc/terms/creator> <"Hamlet"> .`

Figure 1. Example of a linked data RDF triple describing William Shakespeare’s authorship of *Hamlet*.

For example, in figure 1, “William Shakespeare wrote *Hamlet*” is a triple. The subject and predicate of the triple are written as an URI containing the identifier information and the object of the triple is a literal piece of information. The subject of the triple, William Shakespeare, has an identifier which in this example links to the Library of Congress name authority file for William Shakespeare. The predicate of the RDF triple describes the relationship between the subject and object. The predicate also typically defines the metadata schema being used. In this example, Dublin Core is the metadata schema being used, so “wrote” would be identified by the Dublin Core Creator field. The object of this semantic triple, *Hamlet*, is a literal. Literals are text that are not linked because they do not have a URI. Subjects and predicates always have URIs to allow the computer to make links. The object may have a URI or be a literal. Together these URIs, along with the literal, tell the computer everything it needs to know about this piece of metadata, making it self-contained.

RDF triples with their URIs are stored in a triple-store graph style database which functions differently from a typical relational database. Relational databases rely on table headers to define the metadata stored inside. Moving data between relational databases can be complex because tables must be redefined every time data is moved. Graph databases don’t need tables since all the defining information is already stored in each triple. This allows for bidirectional flow of information between pieces of metadata and makes transferring data simpler and more efficient.²⁶ Information in a triple-store database is then retrieved using SPARQL, a query language developed for linked data. Because linked data is stored as self-contained triples, machines have all the information needed to process the data and perform advanced reasoning and logic programming. This leads to better search functionality and lends itself well to artificial intelligence (AI) technologies. Many of today’s modern websites make use of these technologies to enhance their displays and provide greater functionality for their users. The Internet is an excellent avenue for libraries to un-silo their collections and make them globally accessible. Once library collections are on the web, advanced keyword search functionalities and artificial intelligence machine learning algorithms can be developed to automate metadata creation workflows and enhance search and

retrieval of library resources. The use of linked data metadata in these machine-learning functions will add a layer of semantic understanding to the data being processed and analyzed for patron discovery. AI technology can also be used to create advanced graphical displays making connections for patrons between various resources on a research topic.

Sharing digital cultural heritage data with other institutions often involves transferring data and is considered one of the greatest difficulties in sharing digital collections. For example, if one institutional repository uses Dublin Core to store its metadata for a certain cultural heritage collection and another repository uses MODS/METS to store digital collections, there must first be a data conversion before the two repositories could share information. Dublin Core and MODS/METS are two completely different schemas with different fields and metadata standards. These two schemas are incompatible with each other and must be crosswalked into a common schema. This typically results in some data loss during the transformation process. This makes combining two collections from different institutions into one shared web portal difficult.

Linked data allows institutions to share collections more easily. Because linked data triples are self-contained, there is no need to crosswalk metadata stored in triples from one schema into another when transferring data. The URIs contained in the RDF triples allow the computer to identify the metadata schema and process the metadata. RDF triples can be harvested from one linked data system and easily placed into another repository or web portal. A variety of schemas can all be stored together in one graph database. Storing metadata in this way increases the interoperability of digital cultural heritage collections. Collections stored in triple-store databases have SPARQL endpoints that make harvesting the metadata in a collection more efficient. Libraries can easily share metadata on important collections increasing the exposure and providing greater access for a wider audience. Philip Schreur, author of "Bridging the Worlds of MARC and Linked Data," sums this concept up nicely: "The shift to the Web has become an inextricable part of our day-to-day lives. By moving our carefully curated metadata to the Web, libraries can offer a much-needed source of stable and reliable data to the rapidly growing world of Web discovery."²⁷

Linked data also makes it easier to harvest metadata and import collections into larger cultural heritage repositories like Digital Public Library of America (DPLA) which uses linked data to "empower people to learn, grow, and contribute to a diverse and better-functioning society by maximizing access to our shared history, culture, and knowledge."²⁸ Europeana, the European cultural heritage database, uses semantic web technologies to support its mission which is to "empower the cultural heritage sector in its digital transformation."²⁹ Using linked data to transfer data into these national repositories is more efficient and there is less loss of data because the triples do not have to be transformed into another schema. This increases the access of many cultural heritage collections that might not otherwise be seen.

One of the big advantages to linked data is the ability to create connections between other cultural heritage collections worldwide via the web. Incorporating triples harvested from other collections into the local datasets enables libraries to display a vast amount of information about cultural heritage collections in their web portals. Libraries thus can provide a much richer display and allows users access to a greater variety of resources. Linked data also allows web developers to use URIs to implement advanced search technologies creating a multifaceted search environment for patrons. Current research points to the fact that using semantic web technologies makes the creation of advance logic and reasoning functionalities possible. According to Liyang Yu in the book *Introduction to the Semantic Web and Semantic Web Services*, "The Semantic Web is an

extension of the current Web. It is constructed by linking current Web pages to a structured data set that indicates the semantics of this linked page. A smart agent, which is able to understand this structure data set, will then be able to conduct intelligent actions and make educated decisions on a global scale.”³⁰

Many digital cultural heritage collections in libraries live in siloed resources and are therefore only accessible to a small population of users. Linked data helps to break down traditional library silos in these collections. By using linked data, an institution can expand the interoperability of the collection and make it more easily accessible. Many institutions are starting to incorporate linked data technologies into digital collections, thereby increasing the ability for institutions to share collections. This allows for a greater audience to have access to critical cultural heritage collections for underrepresented populations. In the article “Bridging the Worlds of MARC and Linked Data,” the author states, “The shift to Linked Data within this closed world of library resources will bring tremendous advantages to discovery both within a single resource ... as well as across all the resources in your collections, and even across all of our collective collections. But there are other advantages to moving to Linked Data. Through the use of Linked Data, we can connect to other trusted sources on the Web.... We can also take advantage of a truly international Web environment and reuse metadata created by other national libraries.”³¹

UNIVERSITY OF SOUTH FLORIDA LIBRARIES PRACTICE

University of South Florida Libraries digital collections house a rich collection varying from cultural heritage objects to natural science and environment history materials to collections related to underrepresented populations. Most of the collections are unique to USF and have significant research and educational value. The library is eager to share the collections as widely as possible and hopes the collections can be used at both document and data level.

Linked data creates a “web of data” instead of a “web of documents,” which is the key to bringing structure and meaning to web content, allowing computers to better understand the data. However, collections are mostly born at the document level. Therefore, the first problem librarians need to solve is how to transform the documents to data. For example, there is a beautiful natural history collection called Audubon Florida Tavernier Research Papers in USF Libraries digital collections. The Audubon Florida Tavernier Research Papers is an image collection which includes rookeries, birds, people, bodies of water, and man-made structures. The varied images come from decades of research and are a testament to the interconnectedness of bird population health and human interaction with the environment. The images reflect the focus of Audubon’s work in research and conservation efforts both to wildlife and to the habitat that supports the wildlife.³² This was selected to be the first collection the authors experimented with to implement linked data at USF Libraries. The lessons learned from working with this collection are applied to later work.

When the collection was received to be housed in the digital platform, it was carefully analyzed to determine how to pull the data out of all the documents as much as possible. The authors designed a metadata schema of the combination of MODS and Darwin Core (Darwin Core, abbreviated to DwC, is an extension of [Dublin Core](#) for [biodiversity informatics](#)) to pull out and properly store the data.

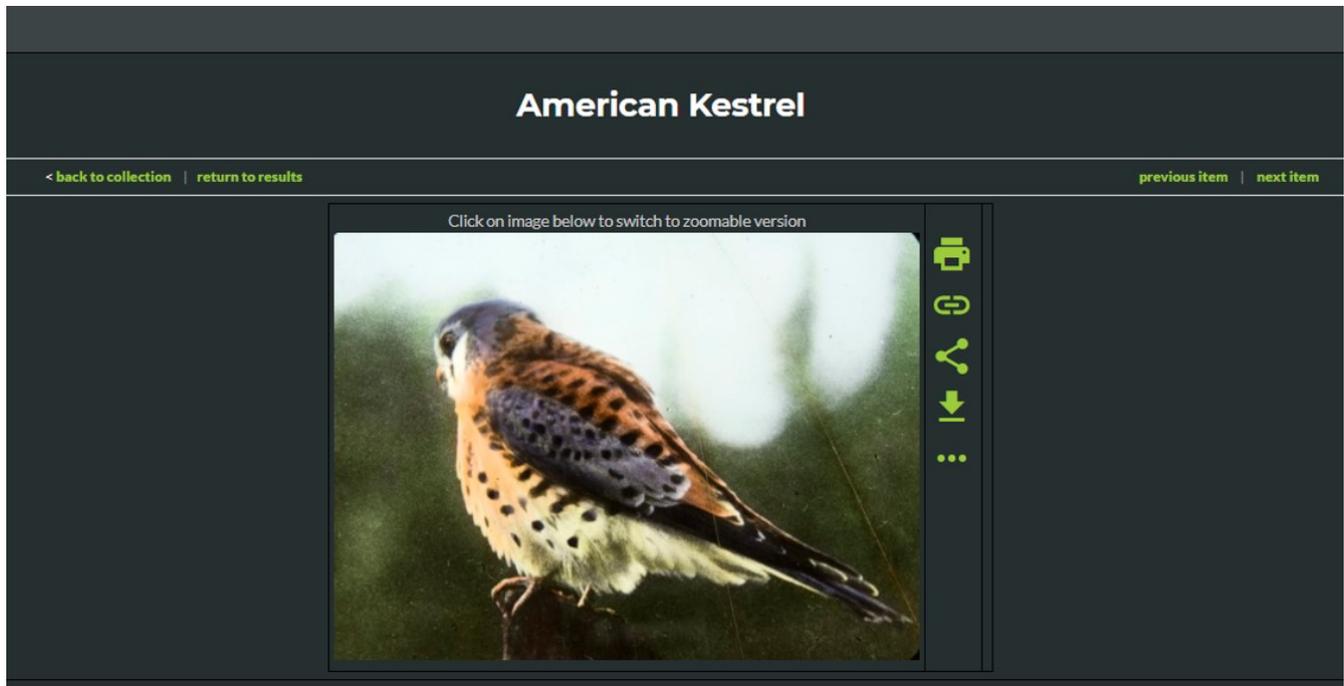


Figure 2. American kestrel.

ITEM DATA	
Title	American Kestrel
Abstract	A male American kestrel perched on a post.
Hierarchical Spatial	North and Central America United States
Holding Location	University of South Florida
Resource Identifier	A64-01016 a64.1016
Language	English
Creator	Audubon Florida
Physical Description	1 online resource (1 photo: color)
Source Institution	University of South Florida Library
Subjects / Keywords	Environmental protection Wildlife conservation American kestrel
Series Title	Audubon Florida Everglades Science Center Records
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	Falconiformes
Family	Falconidae
Genus	Falco
Species	Falco sparverius
Common Name	American Kestrel

Annotations:

- Aboutness:** Points to the Subjects / Keywords section.
- Description of the image in text:** Points to the Abstract section.
- Location information:** Points to the Hierarchical Spatial section.
- Detailed information about the bird:** Points to the taxonomic classification section (Kingdom through Species).

Figure 3. American kestrel metadata.

Figure 2 is one of the documents in the collection, which is a photo of an American kestrel. Figure 3 shows the data collected from the document and the placement of the data in the metadata schema. The authors put the description of the image in free text in the abstract field. This field is indexed and searchable through the digital collections platform. Location information is put in the hierarchical spatial field. The subject heading fields describe the “aboutness” of the image, that is, what is in the image. All the detailed information about the bird is placed in Darwin Core fields. Thus, the document is dissembled into a few pieces of data which are properly placed into metadata fields where they can be indexed and searched.

Having data alone is not sufficient to meet linked data requirements. The first of the four rules of linked data is to name things using URIs.³³ To add URIs to the data, the authors needed to standardize the data and reconcile it against widely-used authorities such as Library of Congress Subject Headings, wikidata, and the Getty Thesaurus of Geographic Names. Standardized data tremendously increases the percentage of data reconciliation, which will lead to more links with related data once published.

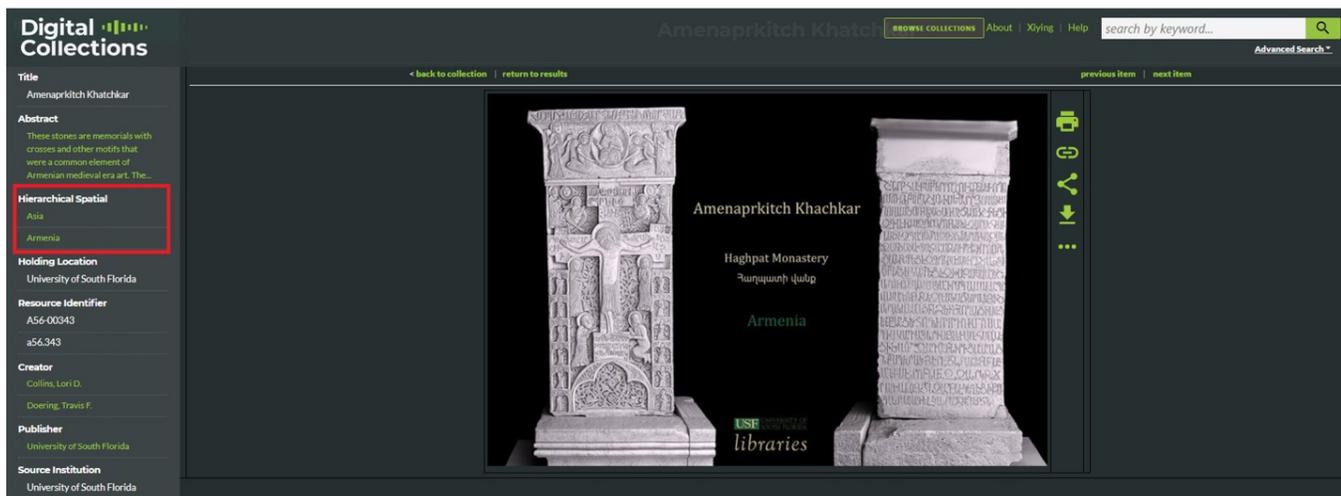


Figure 4. Amenaprkich Khachkar.

Figure 4 shows an example from the Armenia Heritage and Social Memory Program. This is a visual materials collection with photos and 3D digital models. It was created by the Digital Heritage and Humanities Collection team at the library. The collection brings together comprehensive information and interactive 3D visualization of the fundamentals of Armenian identity, such as their architectures, languages, arts, etc.³⁴ When preparing the metadata for the items in this collection, the authors devoted extra effort to adding geographic location metadata. This effort serves two purposes: one is to respectfully and honestly include the information in the collection; and the second is to provide future reference to the location of each item as the physical items are in danger and could disappear or be ruined. The authors employed the Getty Thesaurus of Geographic Names because it supports a hierarchical location structure. The location names at each level can be reconciled and have their own URIs. The authors also paid extra attention on the subject headings. Figure 5 shows how the authors used Library of Congress subject headings, local subject headings assigned by the researchers, and the Getty Art and Architecture Thesaurus for this collection. In the data reconciliation stage, the metadata can be compared against both Library of Congress subject headings authority files and the Getty AAT vocabularies so that as many URIs as possible can be fetched and added to the metadata. The focus

on geographic names and subject headings is to standardize the data and use controlled vocabularies as much as possible. Once moving to the linked data world, the data will be ready to be added with URIs. Therefore, the data can be linked easily and widely.

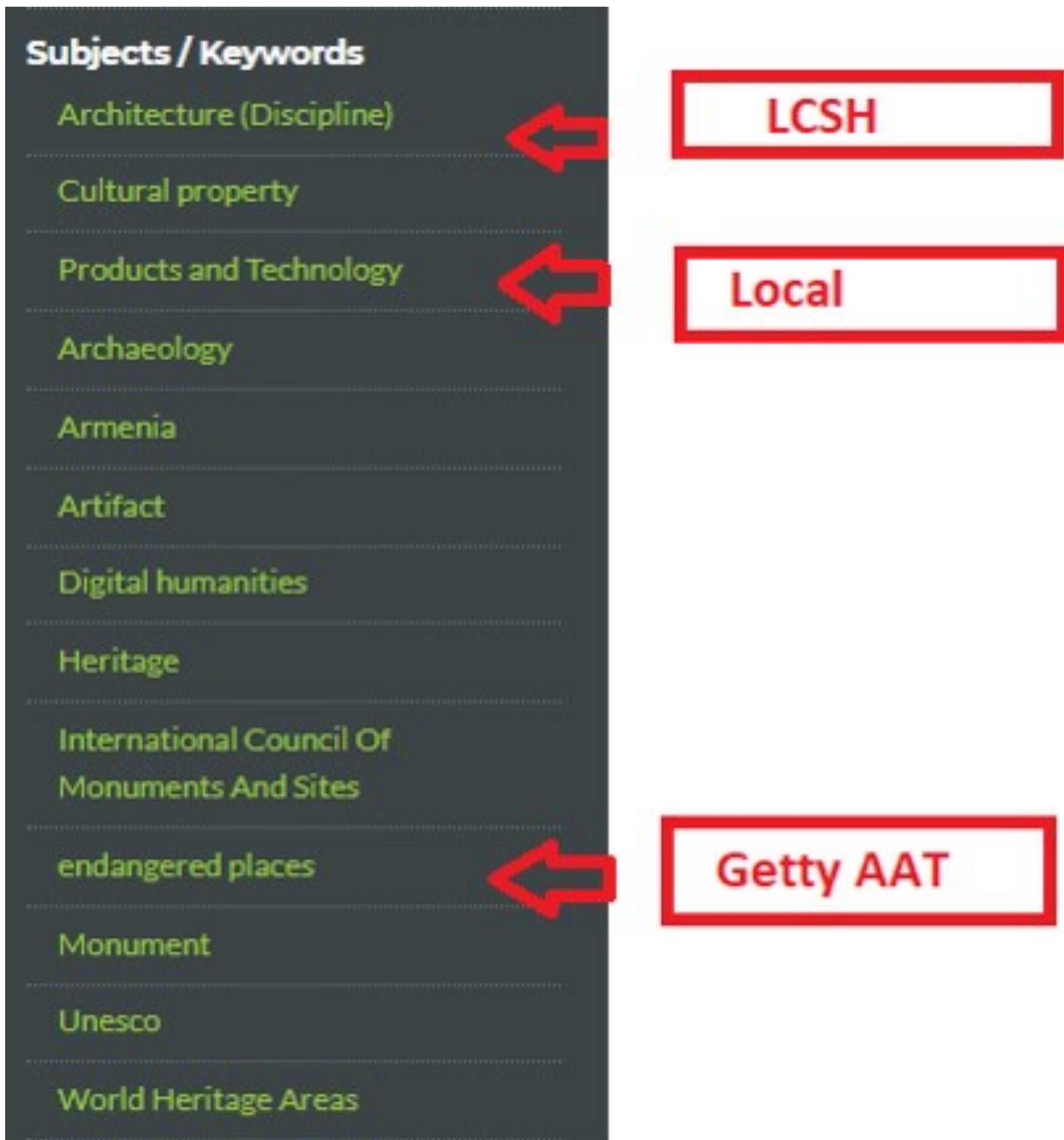


Figure 5. Amenaprkitch Khachkar metadata.

One of the goals of building linked data is to make sense out of data and to generate new knowledge. As the librarians explored how to bring together multiple USF digital collections to highlight African American history and culture, three collections seemed particularly appropriate:

- An African American sheet music collection from the early 20th century (<https://digital.lib.usf.edu/sheet-music-aa>)
- The “Narratives of Formerly Enslaved Floridians” collection from 1930s (<https://digital.lib.usf.edu/FL-SLAVENARRATIVES>)
- The “Otis R. Anthony African American oral history collection” from 1978-1979 (<https://digital.lib.usf.edu/OHP-OTISANTHONY>)

These collections are all oral expressions of African American life in the US. They span the first three-quarters of the 20th century around the time of the Civil Rights movement. Creating linked data out of these collections will help shed light on the life of African Americans through the 20th century and how it related to the Civil Rights movement. With semantic web technology support, these collections can be turned into machine actionable datasets to assist research and education activities on racism, anti-racism and to piece into the holistic knowledge base.

USF Libraries started to partner with DPLA in 2018. DPLA leverages linked data technology to increase discoverability of the collections contributed to it. DPLA employs JavaScript Object Notation for Linked Data (JSON-LD) as its serialization for their data which is in RDF/XML format. JSON-LD has a method of identifying data with IRIs. The use of this method can effectively avoid data ambiguity considering DPLA is holding a fairly large amount of data. JSON-LD also provides computational analysis in support of semantics services which enriches the metadata and in results, the search will be more effective.³⁵ In the 18 months since USF began contributing selected digital collections to DPLA, USF materials have received more than 12,000 views. It is exciting to see the increase in the usage of the collections and it is the hope that they will be accessed by more diverse user groups.

USF Libraries are exploring ways to scale up the project and eventually transition all the existing digital collections metadata to linked data. One possible way of achieving this goal would be through metadata standardization. A pilot project at USF Libraries is to process one medium-size image collection of 998 items. The original metadata is in MODS/METS XML files. We first decided to use the DPLA Metadata Application Profile as the data model. If the pilot project is successful, we will apply this model to all of our linked data transformation processes. In our pilot, we are examining the fields in our MODS/METS metadata and identify those that will be meaningful in the new metadata schema. Then we transport the metadata in those fields to Excel files. The next step is to use OpenRefine to reconcile the data in these Excel files to fetch URIs for exact match terms. During this step, we are employing reconciliation services from the Library of Congress, Getty TGN, and Wikidata. After all the metadata is reconciled, we are transforming the Excel file to triples. The column headers of the Excel file become the predicates and the metadata as well as their URIs will be the objects of the triples. Next, these triples will be stored in an Apache Jena triple-store database so that we can start designing SPARQL queries to facilitate search. The final step will be designing a user-friendly interface to further optimize the user experiences. In this process, to make the workflow as scalable as possible, we are focusing on testing two processes: first, creating a universal metadata application profile to apply to the most, if not all, of the collections; and second, only fetching URIs for exactly matching terms during the reconciliation

process. Both of these processes aim to reduce human interactions with the metadata so that the process is more affordable to the library.

CONCLUSION AND FUTURE WORK

Linked data can help collection discoverability. In the past six months, USF has seen an increase in materials going online. USF Special Collections department rapidly created digital exhibits to showcase their materials. If the trend in remote work continues, there is reason to believe that digital materials may be increasingly present and, given enough time and expertise, libraries can leverage linked data to better support current and new collections.

The societal impact of COVID-19 worldwide sheds light on the importance of technologies such as linked data that can help increase discoverability. When items are being created and shared online, either directly related to COVID-19 or a result of its impact, linked data can help connect those resources. For instance, new COVID-19 research is being developed and published daily. The Publications Office of the European Union Datathon entry “COVID-19 DATA AS LINKED DATA” states that “[t]he benefit of having COVID-19 data as Linked Data comes from the ability to link and explore independent sources. For example, COVID-19 sources often do not include other regional or mobility data. Then, even the simplest thing, having the countries not as a label but as their URI of Wikidata and DBpedia, brings rich possibilities for analysis by exploring and correlating geographic, demographic, relief, and mobility data.”³⁶ The more institutions that contribute to this, the greater the discoverability and impact of the data.

In 2020 there has been an increase in Black Lives Matter awareness across the country. This affects higher education. USF Libraries are not the only ones engaged in addressing racial disparities. Many institutions have been doing this for years. Others are beginning to focus on this area. No matter whether it’s a new digital collection or one that’s been around for decades, the question remains: How do people *find* these resources? Perhaps linked data technologies can help solve that problem. Linked data is a technology that can help accentuate the human effort put forth to create those collections.

Linked data is a way to assist humans and computers in finding interconnected materials around the internet. USF Libraries faced many obstacles implementing linked data. There is a technological barrier that takes well-trained staff to surmount, i.e., creating a linked data triple store database and having linked data interact correctly on webpages. There is a time commitment necessary to create the triples and SPARQL queries. SPARQL queries themselves vary from being relatively simple to incredibly complicated. The authors also had the stumbling block of understanding how linked data worked together on a theoretical level. Taking all of these considerations into account, we can say that creating linked data for a digital collection is not for the faint of heart. A cost/benefit analysis must be taken and assessed. The authors of this paper must continue to determine the need for linked data.

At USF, the authors have taken the first steps in converting digital collections into linked data. We’ve moved from understanding the theoretical basis of linked data and into the practical side where the elements that make up linked data start coming together. The work to create triples, SPARQL queries, and URIs has begun, and full implementation has started. Our linked data group has learned the fundamentals of linked data. The next, and current, step is to develop workflows for existing metadata conversion into appropriate linked data. The group meets regularly and has created a triple store database and converted data into linked data. While the process is slow

moving due to group members' other commitments, progress is being made by looking at the most relevant collections we would like to transform and moving forward from there. We've located the collections we want to work on, taking an iterative approach to creating linked data as we go.

With linked data, there is a lot to consider. How do you start up a linked data program at your institution? How will you get the required expertise to create appropriate and high-quality linked data? How will your institution crosswalk existing data into triples format? Is it worth the investment? It may be difficult to answer these questions but they're questions that must be addressed. The USF Libraries will continue pursuing linked data in meaningful ways and showcasing linked data's importance. Linked data can help highlight all collections but more importantly those of marginalized groups, which is a priority of the linked data group.

ENDNOTES

- ¹ Peter Perl, "What Is the Future of Truth?" *Pew Trust Magazine*, February 4, 2019, <https://www.pewtrusts.org/en/trust/archive/winter-2019/what-is-the-future-of-truth>.
- ² "Anti-Racism Reading Lists," University of Minnesota Library, accessed September 24, 2020, <https://libguides.umn.edu/antiracismreadinglists>.
- ³ "Triad Black Lives Matter Protest Collection," UNC Greensboro Digital Collections, accessed December 9, 2020, <http://libcdm1.uncg.edu/cdm/blm>.
- ⁴ "Umbra Search African American History," Umbra Search, accessed December 10, 2020, <https://www.umbrasearch.org/>.
- ⁵ Karen Coyle, "On the Web, of the Web" (keynote at LITA, October 1, 2011), <https://kcoyle.net/presentations/lita2011.html>.
- ⁶ Donna Ellen Frederick, "Disruption or revolution? The Reinvention of Cataloguing (Data Deluge Column)," *Library Hi Tech News* 34, no. 7 (2017): 6–11, <https://doi.org/10.1108/LHTN-07-2017-0051>.
- ⁷ Tim Berners-Lee, "Linked Data," W3, last updated June 18, 2009, <https://www.w3.org/DesignIssues/LinkedData.html>.
- ⁸ Neil Wilson, "Linked Data Prototyping at the British Library" (paper presentation, Talis Linked Data and Libraries event, 2010).
- ⁹ Diane Rasmussen Pennington and Laura Cagnazzo, "Connecting the Silos: Implementations and Perceptions of Linked Data across European Libraries," *Journal of Documentation* 75, no. 3 (2019): 643–66, <https://doi.org/10.1108/JD-07-2018-0117>.
- ¹⁰ Jane Hagerlid, "The Role of the National Library as a Catalyst for an Open Access Agenda: The Experience in Sweden," *Interlending and Document Supply* 39, No. 2 (2011): 115–18, <https://doi.org/10.1108/02641611111138923>.
- ¹¹ Pennington and Cagnazzo, "Connecting the Silos," 643–66.

-
- ¹² Gillian Byrne and Lisa Goddard, "The Strongest Link: Libraries and Linked Data," *D-Lib Magazine* 16, no. 11/12 (2010): 2, <https://doi.org/10.1045/november2010-byrne>.
- ¹³ Bendik Bygstad, Gheorghita Ghinea, and Geir-Tore Klæboe, "Organisational Challenges of the Semantic Web in Digital Libraries: A Norwegian Case Study," *Online Information Review* 33, no. 5 (2009): 973–85, <https://doi.org/10.1108/14684520911001945>.
- ¹⁴ Pennington and Cagnazzo, "Connecting the Silos," 643–66.
- ¹⁵ Heather Lea Moulaison and Anthony J. Million, "The Disruptive Qualities of Linked Data in the Library Environment: Analysis and Recommendations," *Cataloging & Classification Quarterly* 52, no. 4 (2014): 367–87, <https://doi.org/10.1080/01639374.2014.880981>.
- ¹⁶ Marshall Breeding, "Linked Data: The Next Big Wave or Another Tech Fad?" *Computers in Libraries* 33, no. 3 (2013): 20–22.
- ¹⁷ Moulaison and Million, "The Disruptive Qualities of Linked Data," 369.
- ¹⁸ Nuno Freire and Sjors de Valk, "Automated Interpretability of Linked Data Ontologies: An Evaluation within the Cultural Heritage Domain," (workshop, IEEE Conference on Big Data, 2019).
- ¹⁹ "BIBFRAME Update Forum at the ALA Annual Conference 2018," (Washington, DC: Library of Congress, July 2018), <https://www.loc.gov/bibframe/news/bibframe-update-an2018.html>.
- ²⁰ Jacque Samples and Ian Bigelow, "MARC to BIBFRAME: Converting the PCC to Linked Data," *Cataloging & Classification Quarterly* 58, no. 3–4 (2020): 404.
- ²¹ Oliver Pesch, "Using BIBFRAME and Library Linked Data to Solve Real Problems: An Interview with Eric Miller of Zepheira," *The Serials Librarian* 71, no. 1 (2016): 2.
- ²² Pesch, 2.
- ²³ Gianfranco Crupi, "Beyond the Pillars of Hercules: Linked Data and Cultural Heritage," *Italian Journal of Library, Archives & Information Science* 4, no. 1 (2013): 25–49, <http://dx.doi.org/10.4403/jlis.it-8587>.
- ²⁴ "Resource Description Framework (RDF)," W3C, February 25, 2014, <https://www.w3.org/RDF/>.
- ²⁵ Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284, no. 5 (2001): 34–43, <https://www.jstor.org/stable/26059207>.
- ²⁶ Dean Allemang and James Hendler, "Semantic Web Application Architecture," in *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and Owl*, (Saint Louis: Elsevier Science, 2011): 54–55.

-
- ²⁷ Philip E. Schreur and Amy J. Carlson, "Bridging the Worlds of MARC and Linked Data: Transition, Transformation, Accountability," *Serials Librarian* 78, no. 1-4 (2020), <https://doi.org/10.1080/0361526X.2020.1716584>.
- ²⁸ "About Us," DPLA: Digital Public Library of America, accessed December 11, 2020, <https://dp.la/about>.
- ²⁹ "About Us," Europeana, accessed December 11, 2020, <https://www.europeana.eu/en/about-us>.
- ³⁰ Liyang Yu, "Search Engines in Both Traditional and Semantic Web Environments," in *Introduction to Semantic Web and Semantic Web Services* (Boca Raton: Chapman & Hall/CRC, 2007): 36.
- ³¹ Schreur and Carlson, "Bridging the Worlds of MARC and Linked Data."
- ³² "Audubon Florida Tavernier Research Papers," University of South Florida Libraries Digital Collections, accessed November 30, 2020, <https://lib.usf.edu/?a64/>.
- ³³ Berners-Lee, "Linked Data," <https://www.w3.org/DesignIssues/LinkedData.html>.
- ³⁴ "The Armenian Heritage and Social Memory Program," University of South Florida Libraries Digital Collections, accessed November 30, 2020, <https://digital.lib.usf.edu/ARMENIAN-HERITAGE/>.
- ³⁵ Erik T. Mitchell, "Three Case Studies in Linked Open Data," *Library Technology Reports* 49, no. 5 (2013): 26-43.
- ³⁶ "COVID-19 Data as Linked Data," Publications Office of the European Union, accessed December 11, 2020, <https://op.europa.eu/en/web/eudatathon/covid-19-linked-data>.