

Metadata for Storytelling

Exploring the Collections as Data Model

Emily Baldoni, Angela Yon, and Maddi Loiselle

ABSTRACT

This article describes a case study in which a small metadata team at Illinois State University Milner Library produced a digital humanities project supporting Collections as Data (CAD) and linked data principles. Despite initial sparse descriptive content, the team recognized great potential for experimentation in a significant World War I archival collection to highlight lesser-known stories, including those of the Pioneer Infantry, women, and noncombatants. Discussion focuses on the strategic approaches in creating granular but scalable metadata for the large digital collection, and application of the data with various tools such as ArcGIS and Wikidata to construct interactive data visualizations, mapping, and digital storytelling for the Illinois State Normal University World War I Service Records collection. The article argues that even institutions without a dedicated CAD initiative can incrementally implement principles from the CAD model to add value to their digital collections. The authors first presented the project in 2024 at the Digital Library Federation Forum and the American Library Association Core Forum.¹

INTRODUCTION

For more than two decades, digitization of cultural heritage holdings has grown immensely in the galleries, libraries, archives, and museums (GLAM) community, resulting in the production and availability of large-scale digital cultural heritage data. Digital collections do not simply provide access to surrogates for unique physical materials but also increasingly include access to published associated datasets for reuse and analysis. Cultural heritage data provides rich information to gain insight and is vital for digital humanities research.² These datasets offer possibilities to open new modes of shared access, analysis, and visualizations. Since the inception of the Institute of Museum and Library Services–funded project *Always Already Computational: Collections as Data* (CAD; 2016–2018), the CAD model has been growing among the fields of digital humanities and digital libraries. The project and its subsequent effort, the Andrew W. Mellon Foundation’s *Collections as Data: Part to Whole* (2019–2021), outlined standards and best practices for turning digital collections into datasets for computational use and innovative research methods.³ Linked data is another critical component offered as collections data and expands access to and use of digital collections. The application of linked data principles on digital metadata aids in the provision of structured and interoperable data for open discovery, access, and connections in the semantic web, enriching and contextualizing the data, and creating data well-suited for digital humanities studies.⁴

Illinois State University (ISU) Milner Library first investigated principles of CAD and linked data with digital collections with the Council of Library and Information Resources (CLIR) grant-funded

About the Authors

Emily Baldoni (ejbaldo@ilstu.edu; corresponding author) is Metadata Librarian, Illinois State University. **Angela Yon** (ayon@ilstu.edu) is Cataloging and Metadata Librarian, Illinois State University. **Maddi Loiselle** (madalyn.loiselle@cpcc.edu) is Library Technical Assistant, Central Piedmont Community College. © 2026.

Submitted: 11 July 2025. Accepted for Publication: 22 October 2025. Published: 16 March 2026.



Circus Route Books collection and its accompanying digital humanities project *Agency Through Otherness: Portraits of Circus Performers, 1875–1925* in 2021.⁵ The grant supplied a unique opportunity for Milner: founded in 1857 as Illinois State Normal University (ISNU), ISU is a Carnegie R2 institution with a strong history of undergraduate education but little formal support for digital scholarship.⁶ Without this funding, the library would not have had the capacity to do such in-depth, specialized work with limited staffing resources. By leveraging the successful completion of the CLIR grant, the Technical Services Department effectively advocated for increased metadata staffing, and in 2023, the library hired a metadata librarian and metadata specialist. Together with the cataloging and metadata librarian, the pair forms the metadata team in the department. The team collaborates closely with the Digitization Center, Special Collections, and Archives Departments.

In August 2023, the digitization coordinator presented the ISNU World War I Service Records collection for the metadata team to describe. The assessment of the sizeable collection indicated metadata description could be a daunting task, and technological prospects seemed limited for the project. There was sparse descriptive content of files in the finding aid, and an extensive time commitment would be required for description. The library's digital repository, CONTENTdm, offered no accommodations for linked data and had limited visualization capabilities. Furthermore, the university's strict technology security vetting policy prevented any possibility of experimenting with new software, either open source or proprietary. Despite these factors, the project still showed promising and timely opportunities. Although creating original metadata is often seen as a barrier, the team viewed the opportunity to create original metadata to specification versus reusing and crosswalking existing metadata as a positive potential for exploration. There was rich data within the collection to capture, and the team recognized the possibility to prepare data for a linked data environment as well as to highlight lesser-known stories, including those of the Pioneer Infantry, women, and noncombatants.

This article describes local strategic approaches in creating granular but scalable metadata for the ISNU World War I Service Records collection and discusses the various methods applied to this data for a digital humanities project supporting CAD and linked data principles. Using this collection as a case study, it argues that even institutions without the resources for a dedicated, across-the-board CAD initiative can incrementally implement aspects of the CAD approach to improve collection discoverability and prepare metadata for future reuse.

LITERATURE REVIEW

Collections as Data

Since the 2017 publication of the "Santa Barbara Statement on Collections as Data" (a part of the *Always Already Computational: Collections as Data* grant project led by Thomas Padilla), the CAD model has grown internationally both individually and collectively. Small to large institutions have invested time to explore, develop, and share access to support responsible computational use of CAD. The statement proposes principles that stress interoperability, transparency, ethical commitments to communities, open and shareable data, and the elimination of barriers to access.⁷ The second grant initiative, *Collections as Data: Part to Whole*, funded two cohorts of project teams to create their own CAD projects.⁸ One of these projects, the Louisiana State University (LSU) Library's "Louisiana Digital Library as Data," focused on three principles outlined in the Santa Barbara statement: improve metadata quality, provide standardized rights statements, and ethically center vulnerable communities' needs. The grant allowed the LSU Library to examine and improve its metadata technical practices, such as bulk changes to legacy metadata and the ability

to download descriptive metadata in CSV files. The project also provided time for the library to assess collections and bring awareness to potential harmful content and oppressive metadata within digital library records and to consider the appropriateness of their digital collections for widespread dissemination.⁹

Guided by Padilla's "On a Collections as Data Imperative" for a CAD strategy pilot, the University of Utah J. Willard Marriott Library converted five library collections into datasets and openly shared the data through a public GitHub repository. The library produced datasets of geographic coordinates, genealogical information of dates and place names from newspaper obituaries, and oral history texts. They tested the data with various digital humanities tools for computational exploration, including geographic information system (GIS) mapping, text mining, and topic modeling.¹⁰

The growth of digital cultural heritage data permits increased levels of access and new research strategies of analysis and exploration such as innovative interactive applications and visualizations.¹¹ Monika Glowacka-Musial posits that these unexpected discoveries may encourage new ways to examine primary sources and historical evidence.¹² Datasets of cultural heritage collections are rich and often contain diverse metadata related to many different object types such as images, texts, audio recordings, and artworks containing historical knowledge and connections.

Linked Data as a Component of CAD

Linked data can help with this contextualization in visualizing and interpreting collections in broader cultural settings.¹³ Antonis Bikakis et al. assert, "Linked Data and Semantic Web technologies are becoming increasingly important in creating, publishing, and analyzing cultural heritage data in digital humanities."¹⁴

Ashleigh Hawkins discusses linked data as a viable means to implement the FAIR (findable, accessible, interoperable, and reusable) guiding principles for data as it is machine-readable, interoperable, and fitting for digital humanities analysis and research methods. Linked data embedded in the semantic web makes it easier to access, discover, and use. Hawkins further argues for increased engagement with archives, digital humanities, and linked data and integration of low-barrier tools like Wikidata into production workflows to expand access and use of digital archives.¹⁵

Wikidata offers a low-barrier, high-result method for creating and using linked data in libraries and cultural heritage institutions. It makes data not only open and visible but also reusable as linked data. Continued interest and engagement from libraries and other cultural heritage organizations in Wikidata and Wikimedia projects have been a recognized collaboration for quite some time. Reports published by the International Federation of Library Associations and Institutions, the Program for Cooperative Cataloguing (PCC), and the Association of Research Libraries advocate for Wikidata integration to advance discovery.¹⁶ Since 2020, PCC has launched several initiatives incorporating Wikidata for identity management, including the Wikidata Pilot Project and the Entity Management Cooperative Early Adopters Phase program.¹⁷

Wikidata also works within Wikibase, an open-source, customizable structured repository. *Enslaved: Peoples of the Historical Slave Trade* is a notable 2020 Wikibase repository funded by the Andrew W. Mellon Foundation that tells the stories of people captured by the transatlantic slave trade and is a prime example of possibilities of linked data within the CAD construct. The project contains 5 million data points, with existing and shared datasets by multiple institutions around the world, and it is made possible by linked data technologies.¹⁸

Ongoing CAD Developments and Challenges

While an essential charge in the CAD paradigm is to make collections accessible in machine-readable formats for computational research, equally vital are discussions on the processes required to turn collections into data and maintaining a methodical perspective when treating the data.¹⁹ Gustavo Candela et al. discuss the varied new methods of digital collections published as datasets amenable to computational use and propose a methodology to create machine-actionable collections following a set of tasks. To encourage institutions to adopt the publication of datasets to support computationally driven research, they applied the methodology to several use cases from GLAM institutions and observed that quality texts and metadata were essential for obtaining good results. They stress the vital importance of preparing data and its remediation for computationally driven research.²⁰

CAD certainly demonstrates momentum with these recent innovative and promising developments, yet the model is still newly emerging with challenges. At the *Collections as Data: State of the Field and Future Directions* meeting in 2023, 70 participants from 39 research institutions internationally (Australia, Canada, Hong Kong, the Netherlands, the United Kingdom, and the United States) convened. None of the participants described having a completely established CAD program at their institution. Most reported being in the early stages and having limited staffing resources to do the work. One major challenge is establishing a needs and priorities assessment when the demand for CAD is inconsistent and unpredictable.²¹ Other challenges cited include limited technical development resources and rapidly evolving user expectations.²²

COLLECTION BACKGROUND

In 1918, Milner Library namesake Ange Milner joined the ISNU War Service Committee convened by university President David Felmley shortly after America declared war on Germany. According to the report *Illinois State Normal University in the World War*, the committee's charge included five tasks:

1. Conducting the "With Our Warriors" column in the school paper, *The Vidette*, and assisting in editing special numbers.
2. Mailing copies of *The Vidette* to men and women in the service.
3. Purchasing and maintaining the Service Flag.
4. Preparing a roster of men and women in the service.
5. Arranging the War Roster and articles connected with it, in a permanent file for consultation and preservation.²³

The committee mailed 6,000 issues of the campus newspaper between January 1918 and June 1919. Chairman Douglas R. Ridgley and committee member Annette B. Cooper created the Service Flag that included a star for each service man and woman. All seven members of the committee contributed to the War Roster, but Ange Milner was tasked with maintaining it. In a December 1918 letter to Milner just days before he was sent overseas to work at the American Expeditionary Forces University, Ridgley asked Milner to create a survey to send to every individual on the roster to obtain their involvement in the war effort. "The questionnaire," Ridgley wrote, "should be printed or mimeographed on good durable paper with space sufficient for the ordinary individual and a page for extra information. . . . The record should be a serious historical survey of the military activities of the men."²⁴ Milner mailed surveys to more than 800 service members and received 685 responses. She even maintained extensive correspondence with several individuals and kept their responses for the collection. Along with her assistant, local high school student

Kenneth R. Pringle, they collected every photograph, postcard, and wedding announcement they received, as well as scouring local papers for any mention of the service men and women. In total, they collected 412 newspaper clippings, 390 questionnaires, 344 letters, 198 photographs, and many other items that pertained to the individual's time in the service. Collectively, these comprise the ISNU World War I Service Records Collection.

Digitization of the archival collection began in 2016, when the previous Digitization Center coordinator scanned only the surveys. In June 2022, the current coordinator began scanning the rest of the items in the collection, and digitization ended in December of that year. The 8,200 scanned pages were then uploaded to the crowdsourced transcription website FromthePage.com. Digitization Center student workers also took part in the transcription process, but much of the work was done by volunteers on FromthePage. In early 2023, the process was completed by several volunteers, dedicating more than 500 hours of their time to finish the project in just two months.

METADATA DESIGN AND PRODUCTION

Milner Library's metadata librarian began planning the metadata approach to the World War I Service Records collection in August 2023. From the outset, in addition to supporting discovery in Milner's digital repository and the Digital Public Library of America (DPLA), which Milner contributes to via metadata harvesting, the goal was to create metadata that would lend itself to future linked data uses and facilitate exploration of the CAD paradigm through computational use, visualization, and digital humanities projects.

Milner uses CONTENTdm as its digital repository, and this had several impacts on the metadata design of the collection. Although CONTENTdm allows individual collection fields to be mapped to Simple or Qualified Dublin Core elements, the system does not, at this time, allow for the application of a single metadata element set across the entire system; as a result, a new metadata schema must be defined for each collection in the repository. While this allows for a great deal of customization to meet the needs of individual collections, a consequence of this flexibility is a lack of standardization: a variety of different schemas are in use, developed by various staff associated with the repository at different points in time. Most metadata fields in Milner's CONTENTdm instance are mapped to standard Dublin Core fields whenever possible; however, display labels for fields are not always standardized across collections, and in some cases, Dublin Core fields mapped to a single field in one collection may be mapped to multiple fields in another. For example, in one collection, dc:subject may be mapped to a single *Topics* or *Subjects* field, while in other collections it is mapped to multiple fields representing different types of subjects (topical subjects, personal name subjects, etc.).

This tension between customization and standardization was also a factor in the design of the World War I Service Records collection. The metadata team felt it was important to customize the collection's metadata schema to the extent necessary to meet the needs of the collection while retaining a degree of interoperability with materials in the rest of the repository. The team's general approach was to keep fields closely mapped to standard Dublin Core elements that follow Dublin Core Metadata Initiative (DCMI) best practices whenever possible, creating custom fields only in cases where a greater degree of customization was necessary to meet user needs specific to the collection. One area where the team made collection-specific customizations was military unit or service branch: each survey in the collection lists the units and/or organizations (military or non-military) that the individual served with, and it was determined that users of the collection would want to be able to search this information with some granularity. To this end, a *Related*

Organization/Service Branch field (mapped to dc:subject) was added to facilitate detailed and targeted indexing, using a controlled vocabulary derived from the Library of Congress Name Authority File (LCNAF). Having a dedicated field for this information also made it possible for users to facet by military or service branch in CONTENTdm's public interface.

Another consideration in the collection's metadata design was how granular the metadata should be: on the continuum between summarization and depth indexing, how much detail should be provided? Due to the size and archival nature of the collection, the digitization and metadata teams agreed that the records should be uploaded and described at the folder level (where a folder contains all the items associated with a person), rather than the item level. This decision was made with the intention of preserving archival context and organization, so that users could more easily browse all the records documenting an individual's wartime experience. However, even with description focusing on the folder level, the team determined that two metadata fields warranted unusually granular indexing: *Related Organization/Service Branch* and *Location*. The units that an individual served with and the places where they served were considered to be key areas of user interest, and therefore, more controlled vocabulary terms were assigned for these fields than would normally be the case for folder-level description.

These considerations were primarily focused on maximizing discoverability within Milner's CONTENTdm repository. In addition to this goal, however, the metadata team also wanted to facilitate the reuse of collection metadata beyond the digital repository, in line with the CAD approach described in the literature review section of this paper. While it is impossible to anticipate all potential reuses of data, there were two specific uses that the team wanted to enable: integration with linked data applications and geospatial visualization of the locations visited by service members throughout the war.

Although CONTENTdm does not currently provide support for linked data, the team decided that it would nonetheless be valuable to embed URIs associated with controlled vocabulary terms in the metadata, so that metadata exported from the repository could be used in linked data applications. Because URIs are unlikely to be of high interest to most users, they were embedded in hidden fields that do not display publicly; instead, controlled vocabulary terms (usually from the LCNAF) were used for public display. During metadata production, URIs were added for people, corporate bodies (primarily military units and service organizations), topical subjects, and places associated with the records. For most of these fields, URIs were drawn from id.loc.gov; URIs for locations were sourced from Geonames, since the team felt this source collocated more useful collateral data for geographic places (coordinates, administrative hierarchy, geographic feature codes, etc.).

Given the heavy geographic component of the collection, the team placed a high priority on enabling geospatial visualizations. For each visit to a given location documented in an individual's file, the team not only indexed the location's name (using controlled vocabulary from the LCNAF) and URI (from Geonames) but also added geographic coordinates (from Geonames) to a dedicated *Location/coordinates* field in the metadata. As with the URI fields, the coordinates field was created as a hidden field that displays only on the backend.

The full metadata schema for the collection can be found in Table 1.

Table 1. Metadata schema for ISNU World War I Service Records digital collection.

Field/Element Name	Dublin Core Mapping	Data Type	Searchable	Hidden	Authority/Vocabulary
Title	Title	Text	Y	N	
Creator	Creator	Text	Y	N	LCNAF, http://id.loc.gov/authorities/names.html
Creator URI	None	Text	N	Y	LCNAF, http://id.loc.gov/authorities/names.html
Date	Date	Date	Y	N	W3 ISO 8601 format for year or complete date, https://www.w3.org/TR/NOTE-datetime
Decade	Coverage-Temporal	Text	Y	N	
Description	Description	Text	Y	N	
Contributor	Contributor	Text	Y	N	LCNAF, http://id.loc.gov/authorities/names.html
Contributor URI	None	Text	N	Y	LCNAF, http://id.loc.gov/authorities/names.html
Related organization/service branch	Subject	Text	Y	N	LCNAF, http://id.loc.gov/authorities/names.html
Related organization/service branch URI	None	Text	N	Y	LCNAF, http://id.loc.gov/authorities/names.html
Topics	Subject	Text	Y	N	LCSH, http://id.loc.gov/authorities/subjects.html LCNAF, http://id.loc.gov/authorities/names.html

Field/Element Name	Dublin Core Mapping	Data Type	Searchable	Hidden	Authority/Vocabulary
Topics URI	None	Text	N	Y	Getty AAT, https://www.getty.edu/research/tools/vocabularies/aat/ LCSH, http://id.loc.gov/authorities/subjects.html
Genre/format	Format-medium	Text	Y	N	Getty AAT, https://www.getty.edu/research/tools/vocabularies/aat/
Location	Coverage-spatial	Text	Y	N	LCNAF, http://id.loc.gov/authorities/names.html
Location URI	None	Text	N	Y	Geonames, https://www.geonames.org
Location/Coordinates	Coverage-spatial	Text	N	Y	Geonames, https://www.geonames.org
Language	Language	Text	Y	N	Library of Congress, ISO 639-2—in text http://www.loc.gov/standards/iso639-2/
Type	Type	Text	Y	N	DCMI types, http://dublincore.org/documents/dcmi-terms/
File format	Format	Text	Y	Y	IANA media format, http://www.iana.org/assignments/media-types
Filename	None	Text	Y	N	
Digital collection	Relation isPartOf	Text	Y	N	
Physical repository collection	Source	Text	Y	N	
Physical collection location	None	Text	Y	N	
Physical Repository	Source	Text	Y	N	

Field/Element Name	Dublin Core Mapping	Data Type	Searchable	Hidden	Authority/Vocabulary
Finding aid for physical item	Relation	Text	N	N	
Catalog record for physical item	Relation	Text	N	N	
Contact	None	Text	N	N	
Rights	Rights	Text	Y	N	Rightsstatements.org, https://rightsstatements.org
Rights statement link	Rights	Text	Y	N	Rightsstatements.org, https://rightsstatements.org

Notes. LCNAF is Library of Congress Name Authority File, LCSH is Library of Congress Subject Headings, Getty AAT is Getty Art & Architecture Thesaurus, DCMI is Dublin Core Metadata Initiative, and IANA is Internet Assigned Numbers Authority.

Following the finalization of the schema, metadata production began in September 2023. The metadata librarian and the metadata specialist cataloged the collection over the next several months, completing the main phase of description in March 2024. While the intention was to primarily draw the metadata from the contents of the files themselves, in some cases, it was necessary to do research in external databases such as Fold3 and Newspapers.com to verify or enhance information in the file. This was most common in cases where the person's military unit (one of the primary focuses of indexing) could not be readily determined due to vague or incomplete information in the file. Following final collection review and metadata quality control checks, the digital collection was published in CONTENTdm in April 2024.

FROM COLLECTION METADATA TO GEOSPATIAL DATASET

After the publication of the digital collection, the team began preparing collection metadata for reuse in additional digital humanities applications, including geospatial mapping applications. The metadata librarian began by exporting the collection metadata from CONTENTdm as a tab-delimited text file, a format which enables reviewing and editing in Excel or other spreadsheet applications. As noted by Candela et al. in their study on the reuse of digital collections data, "Preprocessing and cleaning is a crucial task to prepare the data for computationally driven research."²⁵ Having just completed metadata production and quality control, the team was confident that they had a high-quality dataset describing the archival collection. However, it soon became clear that significant restructuring and data manipulation would be necessary to transform the metadata into a geospatial dataset that could be used in GIS applications. The tabular metadata exported from CONTENTdm was structured around folder-level description: each row described an archival folder in the collection, with each folder representing a person. The result was, essentially, a table listing individual service members, one person per row, along with the places they served and the organizations they served with. To be digestible by GIS applications, however, the data needed to be structured around places, with one row per location visited, rather than one row per person.

To solve this problem, the metadata librarian used OpenRefine to restructure the dataset from person-level to location-level data. The digital collection metadata contains three fields related to geographic places: Location (a human-readable label for the place), Location URI (from Geonames), and Location/Coordinates (from Geonames). Using OpenRefine's Split Multi-Valued Cells function, each value in the Locations field was separated into its own row. The Blank Down function was then used to transfer nongeographic metadata from the original person-level record into the newly created location-level rows, so that each row would have essential collection metadata fields (e.g., name of subject, related organization/service branch, digital object URL) associated with it. However, because geographic names, coordinates, and URIs were stored in three separate fields in CONTENTdm, additional manipulation was then necessary to correctly associate each place name with its affiliated URI and coordinates. Airtable was used to reconcile the location-level file against a separate table listing place names, URIs, and coordinates, using Airtable's relational database functionality to pull in missing URIs and coordinates.

The result of this data manipulation in OpenRefine and Airtable was a geospatial dataset consisting of one row per location visited by an ISNU affiliate during the war, along with additional descriptive metadata from each individual's file. Tables 2 and 3 show a before-and-after comparison of the data for two individuals.

As shown in Tables 2 and 3, data previously represented in two rows in the person-level dataset (Table 2) must be represented in nine rows in the place-level dataset (Table 3); indeed, whereas the original (person-level) metadata export consists of 679 records, the place-level dataset uses 1,783 records to represent information from the same archival files.

Table 2. Geographic data before restructuring in OpenRefine and Airtable.

Name	Location	Location URI	Location/Coordinates
Ada Adcock	Camp Cody (N.M.); Toul (France); Brest (France)	https://www.geonames.org/5459997/ ; https://www.geonames.org/2972350/ ; https://www.geonames.org/3030300/	32.275, -107.80833; 48.68075, 5.89115; 48.39029, -4.48628
Russell R. Armstrong	Dieulouard (France); Verdun (France); Sommedieue (France); Saint-Mihiel (France); Meuse (France); Marbache (France)	https://www.geonames.org/3021394/ ; https://www.geonames.org/2969958/ ; https://www.geonames.org/2974299/ ; https://www.geonames.org/2977948/ ; https://www.geonames.org/2994106/ ; https://www.geonames.org/2996081/	48.84137, 6.06782; 49.15964, 5.3829; 49.0838, 5.46356; 48.88746, 5.55099; 48.97176, 5.36371; 48.79867, 6.106

Table 3. Geographic data after restructuring in OpenRefine and Airtable.

Name	Location	Location URI	Latitude	Longitude
Ada Adcock	Camp Cody (N.M.)	https://www.geonames.org/5459997/	32.275	-107.808
Ada Adcock	Toul (France)	https://www.geonames.org/2972350/	48.68075	5.89115
Ada Adcock	Brest (France)	https://www.geonames.org/3030300/	48.39029	-4.48628
Russell R. Armstrong	Dieulouard (France)	https://www.geonames.org/3021394/	48.84137	6.06782
Russell R. Armstrong	Verdun (France)	https://www.geonames.org/2969958/	49.15964	5.3829
Russell R. Armstrong	Sommedieue (France)	https://www.geonames.org/2974299/	49.0838	5.46356
Russell R. Armstrong	Saint-Mihiel (France)	https://www.geonames.org/2977948/	48.88746	5.55099
Russell R. Armstrong	Meuse (France)	https://www.geonames.org/2994106/	48.97176	5.36371
Russell R. Armstrong	Marbache (France)	https://www.geonames.org/2996081/	48.79867	6.106

In line with the CAD principles that inspired the project, the team made the resulting dataset openly available via a GitHub repository.²⁶ In addition to the wrangled place-level geospatial dataset described in Table 3, the team also uploaded the raw digital collection metadata, including associated transcripts. Consistent with the CAD principles espoused by the Santa Barbara statement, making the data openly available for download encourages reuse and remixing of the data by other researchers and information professionals.²⁷

PROJECTS

ArcGIS

With the newly created geospatial dataset in hand, the team began working with ArcGIS Pro to create a series of map visualizations for the collection. As an initial proof of concept, the metadata librarian imported the complete place-level dataset into ArcGIS, containing all geographic data points associated with the collection. The geographic coordinates already embedded in the descriptive metadata were used to plot points in ArcGIS, and associated metadata fields were pulled in to add contextual information for each geographic point. The result was a map displaying every location visited by an ISNU affiliate during the war (as documented by the archival collection).²⁸ The map is fully interactive: users can drag the map to reorient it and zoom into individual regions; clicking on a data point brings up a popup containing additional related information derived from the digital collection metadata, such as the person's name, the unit(s) they served with, a short description of their service, and a link to their full file in the digital collection (Figures 1 and 2). In this way, the map not only repurposes the digital collection's descriptive metadata but also provides an additional avenue of discovery, allowing users to browse geographically and then enter the digital collection to further explore the archival materials associated with a person.

Figure 1. ArcGIS map showing locations in Europe visited during World War I.

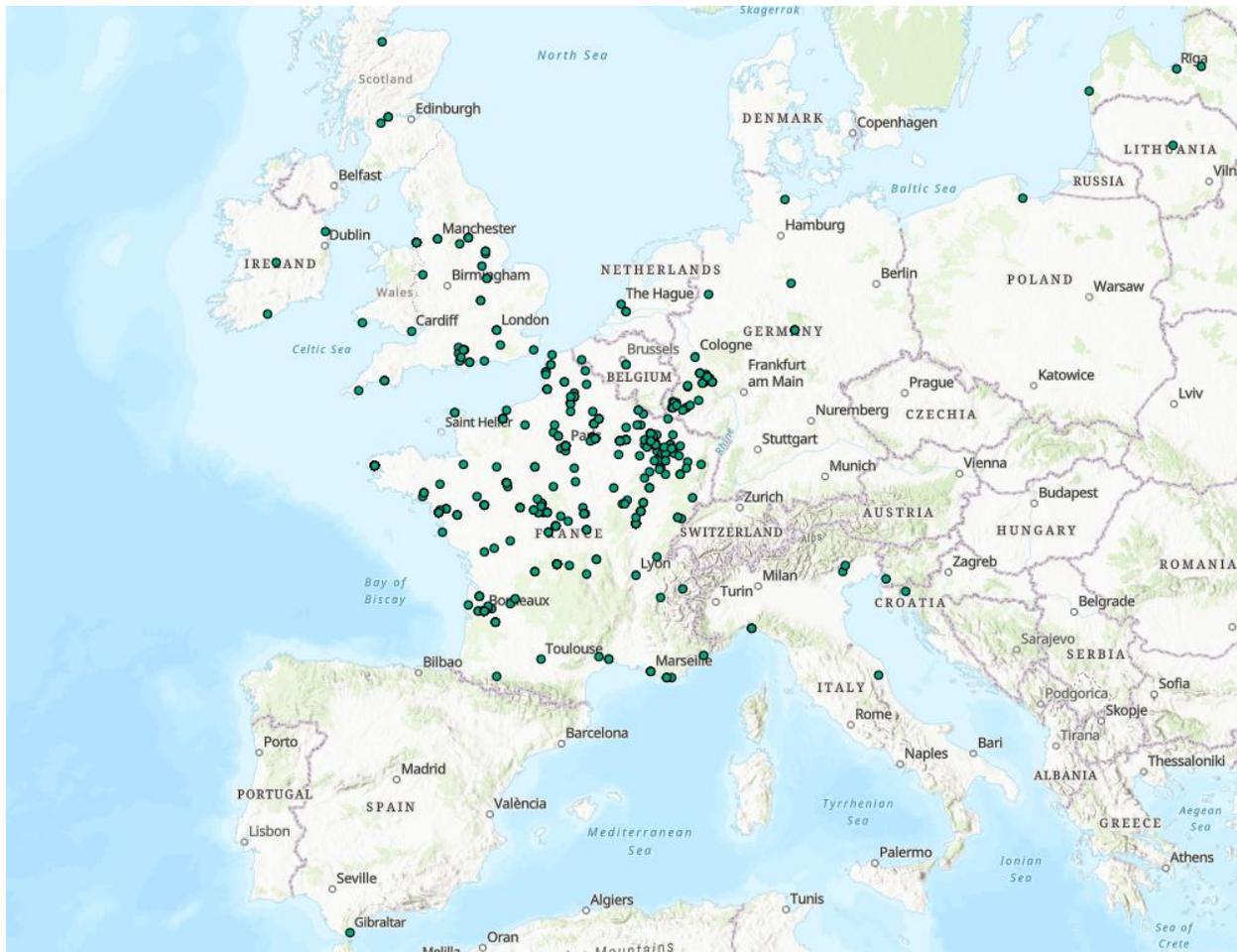
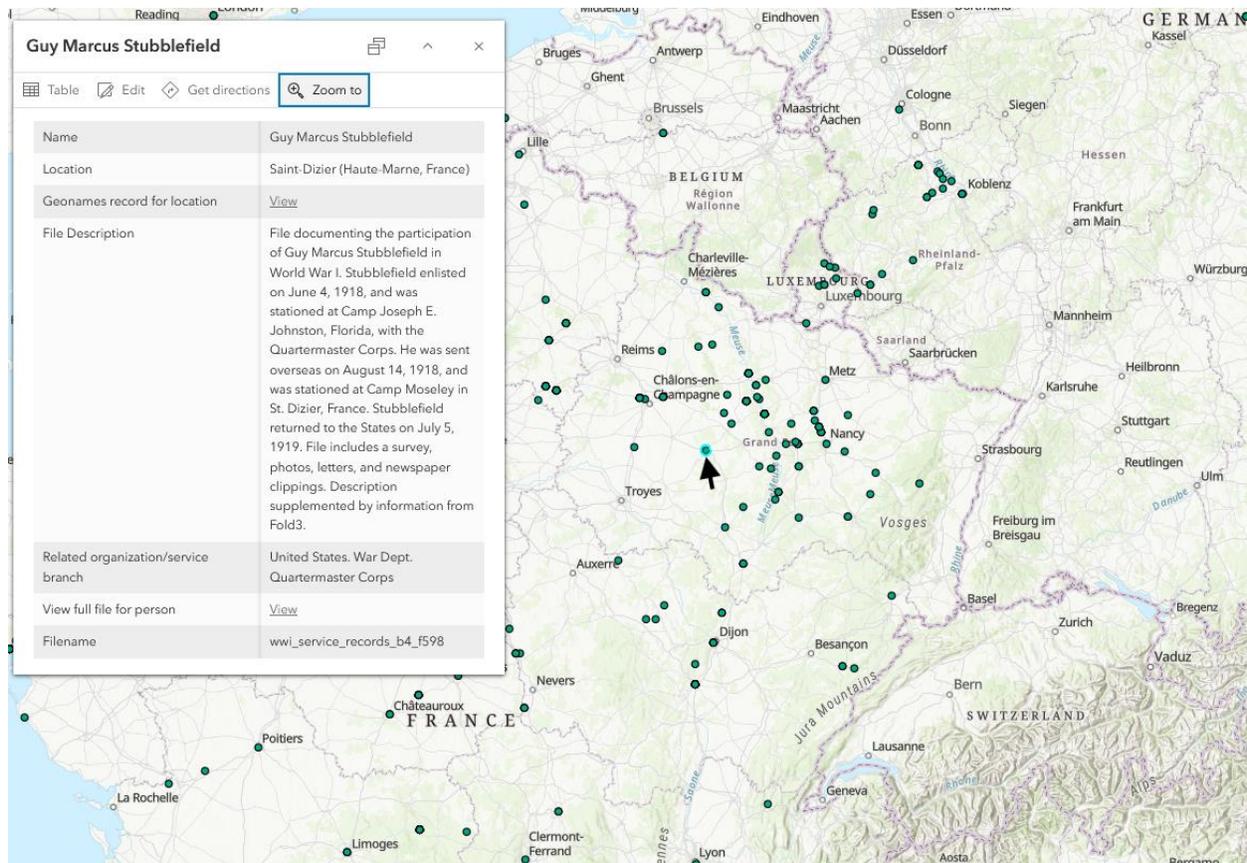


Figure 2. Example of metadata displayed for an individual data point in ArcGIS map.

After this large, collection-level map was created, the metadata librarian began creating additional map layers to highlight more specific facets of the collection. Map layers were created to visualize various subsets of the data, including the following:

- Locations visited by members of different military branches, such as the Army, Navy, Marine Corps, Air Service, and so on
- Locations visited by women in service organizations such as the Red Cross, YMCA, and Army Nurse Corps
- Locations of battles that ISNU affiliates participated in
- Locations where fatalities occurred (whether by combat, illness, or some other cause)
- Locations where ISNU affiliates were hospitalized with influenza²⁹

Like the collection-level map, each of these individual map layers is fully interactive and can be used as an additional entry point into the digital collection. Starting with well-structured, relatively clean metadata made these subset maps much easier to create: this made it possible to filter on existing metadata fields to create each new subset rather than having to rely on ad hoc keyword searches to isolate desired attributes.

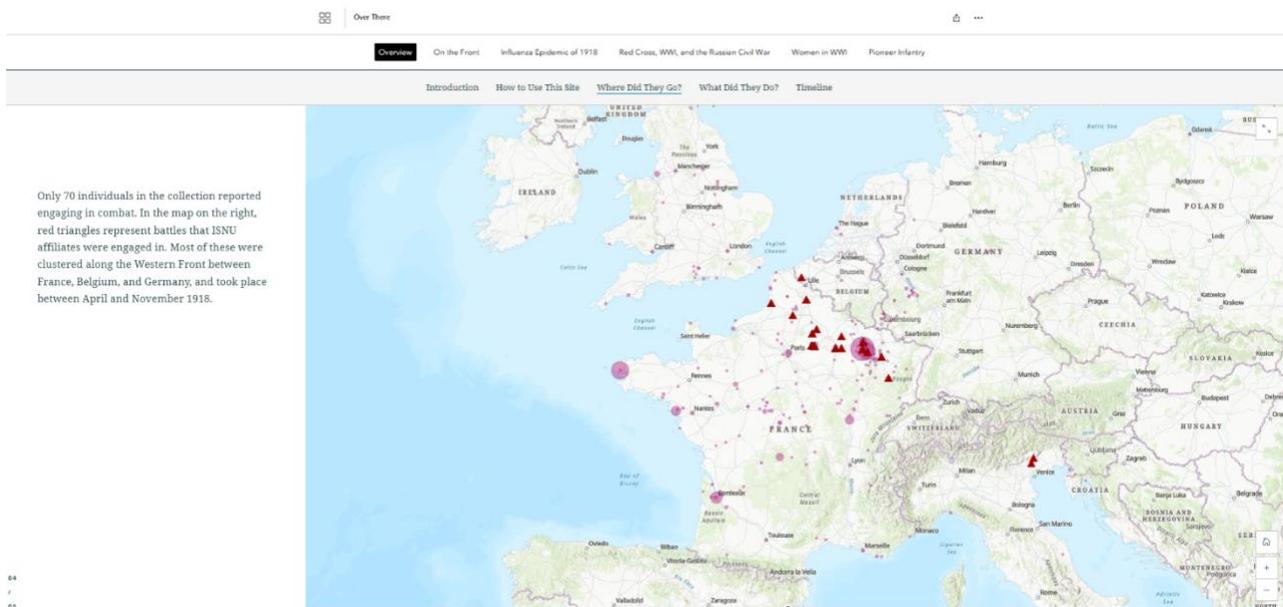
StoryMaps

After the creation of ArcGIS maps, the team needed a platform to publish the maps. Using ArcGIS's StoryMaps, from the same software family as ArcGIS Pro, afforded them the ability to seamlessly embed their work into the digital storytelling tool. The tool itself was user-friendly, allowing the team to create a dynamic digital exhibit without the need for extensive web design or coding skills.

They explored different features for presenting images, text, and geospatial data within StoryMaps, like the sidecar and map tour functions. The team also found that StoryMaps could support additional visualizations imported from other tools when telling the narratives of service members.

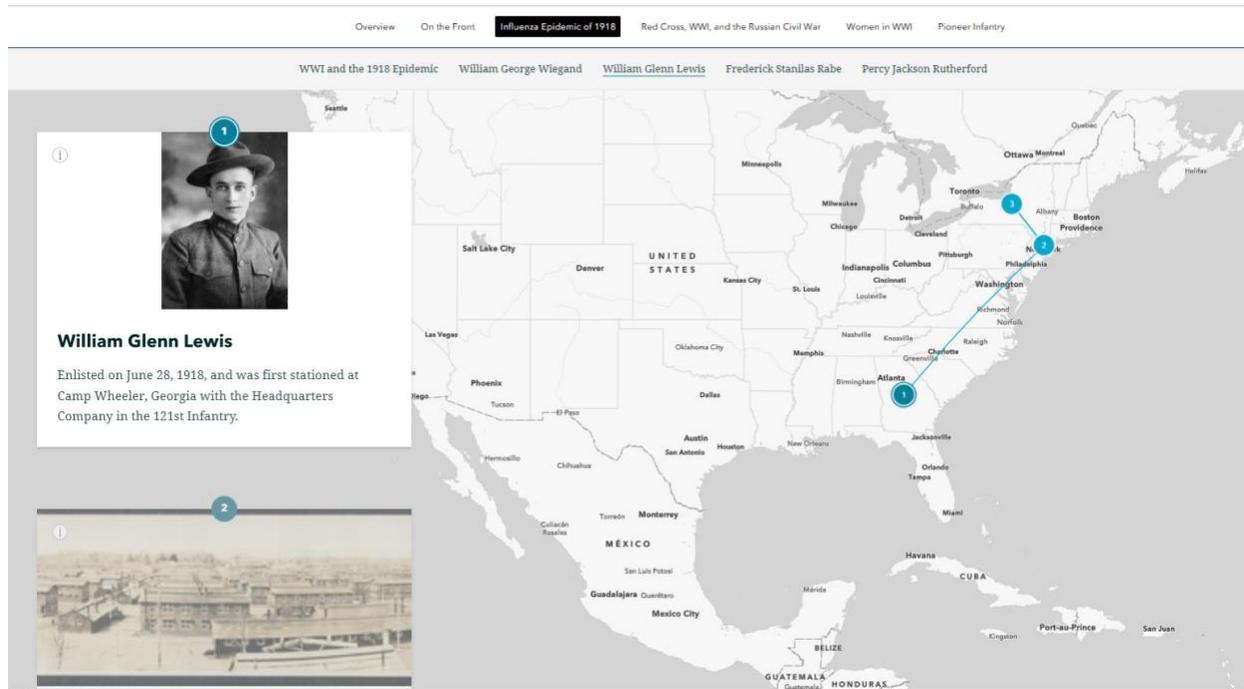
The resulting StoryMaps site, *Over There*, showcases five different accounts of the war, laid out in tab sections.³⁰ From an introduction to the site and the collection to the Influenza Epidemic of 1918 and narratives following specific individuals, the team used a variety of tools and layouts to tell the stories. The site allows for a narrative paragraph structure, shown in the “Overview” and “Influenza Epidemic of 1918” sections. It also has a presentation function known as a sidecar, shown in every section of the site. The site allows for a variety of image sizes and text spaces on each slide, which users can smoothly scroll through (Figure 3).

Figure 3. A collection-level map created in ArcGIS Pro is embedded into the sidecar tool to allow users to travel around the map and focus on different data points, showing the battles and clusters of service men and women.



The “Influenza Epidemic of 1918” section uses StoryMaps’ map tour function. The tool brings users along the journey of a soldier, from their initial training camp to their embarkation camp and overseas. For example, in the case of William Glenn Lewis from the collection (Figure 4), it shows that he was at his embarkation camp in New York when he caught the flu, was hospitalized, and was subsequently never sent overseas.

Figure 4. Screenshot from “Influenza Epidemic of 1918” showing the journey of William Glenn Lewis, who was scheduled to go overseas when he caught the flu at his embarkation camp.



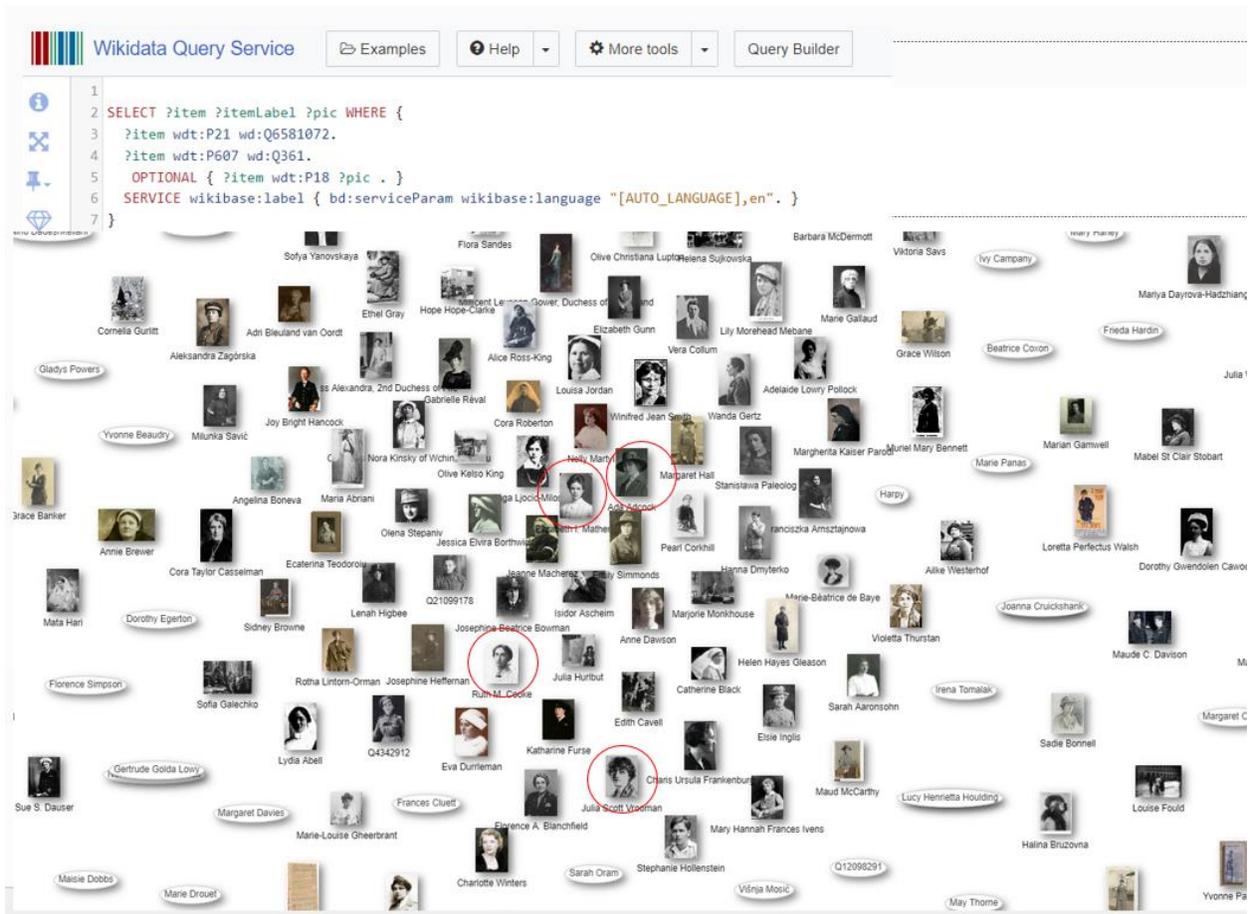
The team embedded the collection-level map created in ArcGIS Pro in the “Overview” section to show the different functionalities of the map, including the geographic data points, the concentration of individuals at a specific location, and the distribution of military units and service organizations. The section also features the embedded timeline from Histropedia (described in the Wikidata section below). StoryMaps’ ability to embed content from other sources (e.g., ArcGIS Pro and Histropedia) allowed the team to bring visualizations created in multiple tools together in one holistic, user-friendly interface.

Wikidata

Emerging standards for linked data provide new options for enhancing, contextualizing, linking, and reframing cultural heritage objects and collection data. The community-based Wikidata offers a low-barrier, high-result method for creating and using linked data in libraries and cultural heritage institutions. It makes data not only visible but reusable as linked data. For the project, the team created Wikidata entries for selected persons in the collection and also contributed their photos from the collection into Wikimedia, particularly for often-missing and underrepresented groups like women and African Americans.

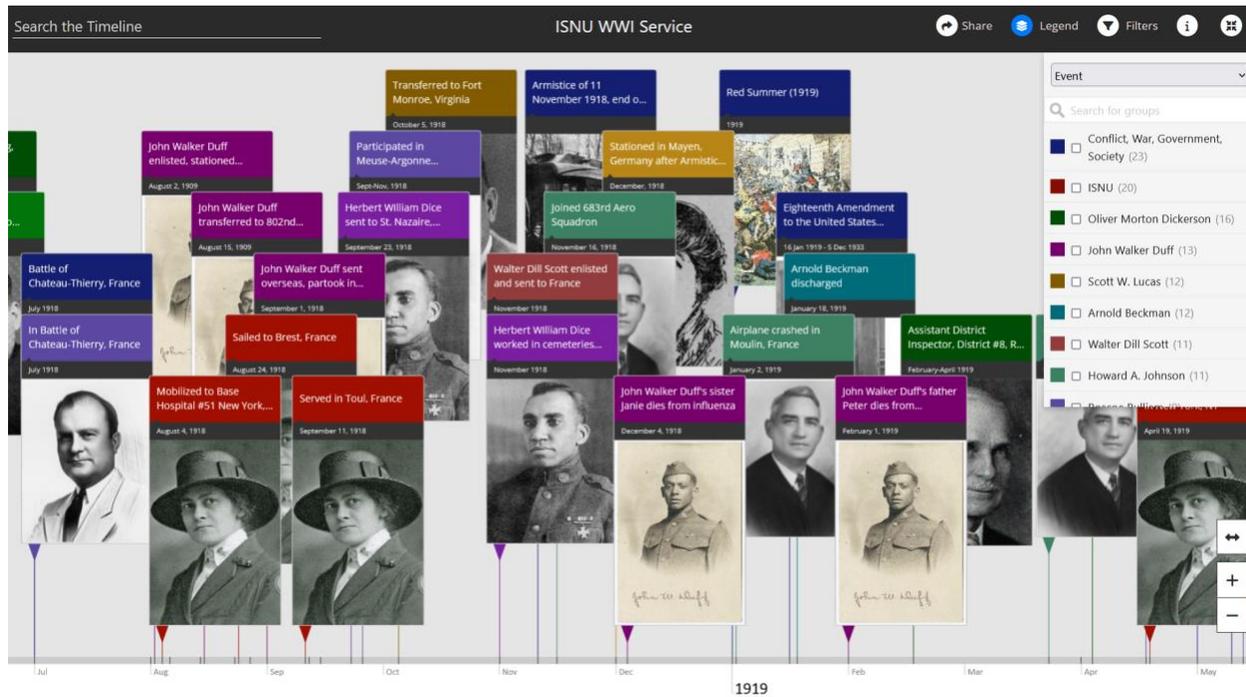
The Wikidata entries aid in higher discoverability of these persons and the collection; making these connections enriches the data, which is now shareable and integrated worldwide in the information landscape. An example of this linked data in action is through the use of SPARQL queries against the collection’s Wikidata items with all the data that exists in Wikidata. With entries pulled through Wikidata’s SPARQL endpoint, the university alumni in the collection are now featured alongside related data created by other cultural heritage institutions and exist not only within ISU collections but also in a larger world of data that is openly discoverable to the public (Figure 5).

Figure 5. A screenshot of a Wikidata SPARQL query pulling statement results of women who participated in World War I internationally displayed in a graph. Circled are women from the ISNU World War I Service Records collection.



Information from SPARQL queries with Wikidata can be pulled in numerous data visualizations, including dynamic timelines. With the newly created Wikidata entries for persons in the collection, the team utilized the open-source Histropedia tool to create a timeline that showcases ISNU individuals and embedded it into the StoryMaps exhibit, *Over There*. The timeline is generated by manual entries and a designated custom query in an Excel spreadsheet. Incorporated are the Wikidata items for underrepresented populations from the collection, such as the Pioneer Infantry and women, as well as other existing Wikidata entries for university alumni. The interactive timeline features photos of individuals from the collection added to Wikimedia and Wikidata. There is the option to display one individual with their activities during the war or to display these individuals alongside others in the collection and events described from Wikipedia pages to highlight the historical context in which these persons served and lived (Figure 6). The timeline is also a second entry into the digital collection, their Wikidata, and their Wikipedia page, if one exists.

Figure 6. Histopedia timeline displaying events of individuals from the ISNU World War I Service Records Collection.



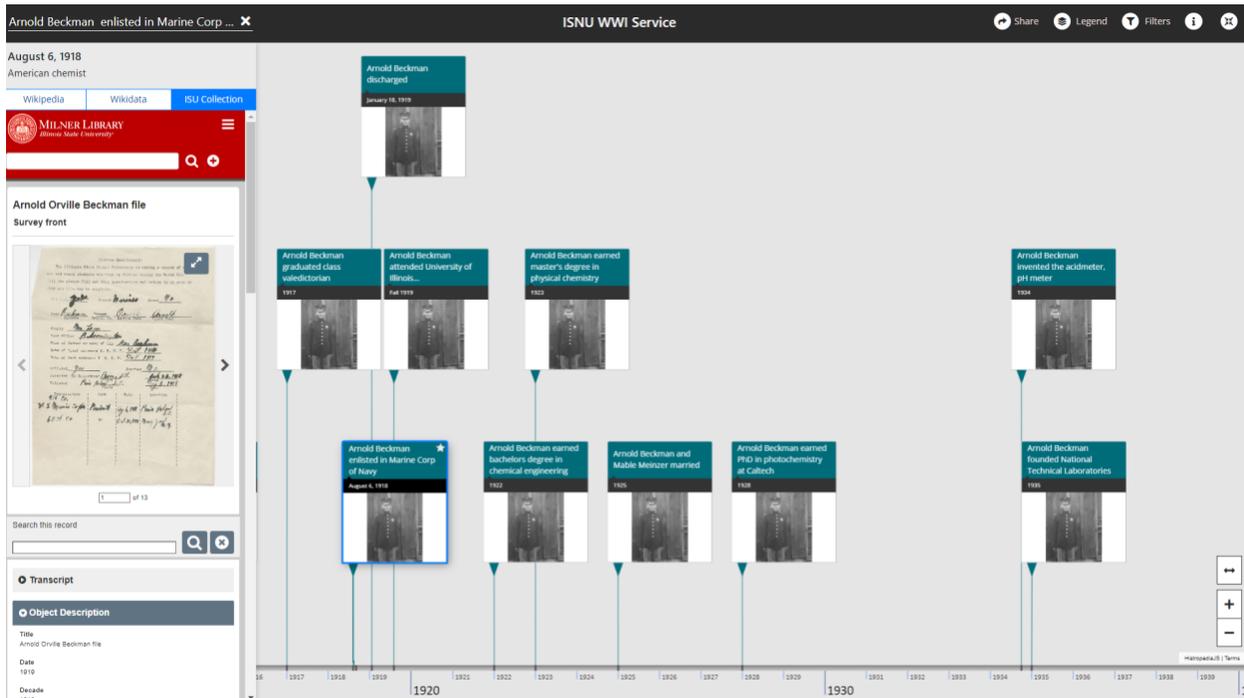
The information in the timeline varied, as data for most individuals was limited to their movements during their war service, such as Ada Adcock of the Army Nurse Corps (Figure 7) and John Walker Duff of the 802nd Pioneer Infantry.

Figure 7. Histopedia timeline for Ada Adcock.



Other individuals in the collection had well-documented and widely available information about their lives after the war. Such examples include the chemist and inventor Arnold Beckman, the inventor of the pH meter and founder of Beckman Instruments (Figure 8), and Walter Dill Scott, one of the first applied psychologists and president of Northwestern University. Their timelines show events in their lives during and after the war.

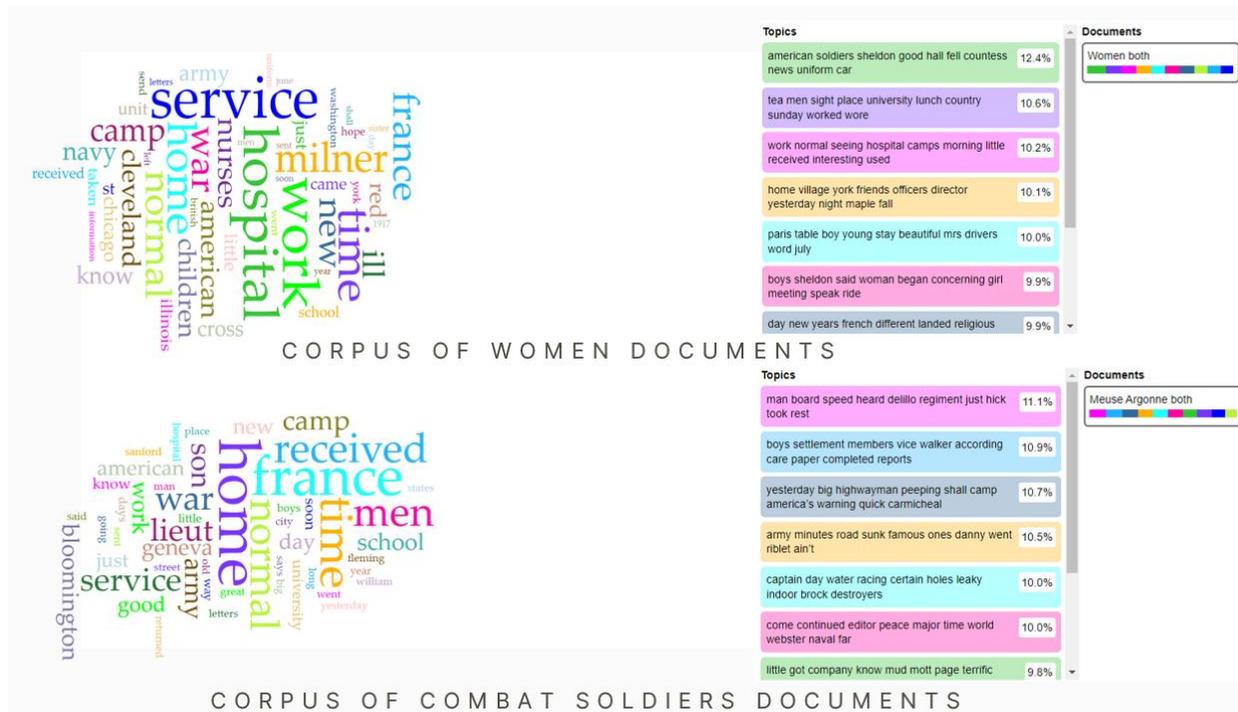
Figure 8. Histropedia timeline for Arnold Beckman and entryway into the ISNU World War I Service Records Collection, Wikidata, and Wikipedia.



Voyant

With the documents in the collection fully transcribed, it was a good opportunity to test the dataset with topic modeling. Voyant, the open-source, web-based application for performing text analysis, was used to analyze the dataset. Two corpora of newspaper clippings and letters were selected, one of women serving and the other of soldiers in the Meuse-Argonne offensive. Word frequency was also viewed in the corpora. The team approached this process as an exploration, considering whether something might be discovered that was not included in the metadata created for these files. The word clouds contain varying sizes of words based on the number of co-occurrences within the corpus; larger words indicate more occurrences and smaller ones indicate fewer.

Figure 9. Voyant text analysis results of two corpora: word frequencies displayed as word clouds on the left and topic modeling results on the right.



In Figure 9, the text findings of the women’s letters proved descriptive of the expected, traditional role of women as caregivers. The words indicated descriptors often associated with women of the period, such as *children*, *nurses*, *service*, *home*, and *hospital*. The corpus from the soldiers did not reveal much associated specifically with combat or battle conditions (Figure 9). Perhaps the letters may have been reserved and formal in this regard as they were writing to the librarian of the university. Another more likely possibility could be that the letters were self-censored and government censored, as that was the case for many correspondences during the war. The findings did not reveal anything unexpected or any themes that indicated new possibilities for access points, but it was a good exercise in exploring and analyzing the text in the letters and articles.

DISCUSSION

Institutional Support and Barriers

One major factor that shaped the team’s approach to this project was varying levels of institutional support for experimentation with digital scholarship tools, particularly GIS software. Early in the project, the team hoped to use highly customizable open-source tools such as Collection Builder to create digital exhibits and mapping visualizations for the collection. Collection Builder also allows users to easily download collection data in multiple standard data formats, further facilitating the reuse of data.³¹ However, the Library Information Technology Services department at Milner Library follows a campus policy that any downloaded software, regardless of cost, must go through a rigorous security check before the user can download the software to their computer. This clearance process, while intended to safeguard institutional data security, can present a barrier to the adoption of new tools: the process is estimated to take at least six months, and the metadata team did not want to wait that long. Moreover, library and campus IT have historically been hesitant to support open-source software, leaving considerable uncertainty as to whether a

tool such as Collection Builder would ultimately be approved even after the extensive vetting process.

In the face of this uncertainty, the team turned to other sources that the university already has access to, like the ArcGIS family of software, a powerful and user-friendly tool with existing campus IT support. It is a paid service for the Geography Department on campus, and, after contacting the appropriate staff member in the department, the team quickly received accounts and were able to start their work. The level of institutional support for software download and experimentation is likely to be an important factor for any library considering CAD and digital humanities approaches to digital collections, not only in terms of what products the institution is willing to pay and provide support for, but also in policies related to data security and software vetting.

Anticipating Potential Data Uses

One issue that recurred throughout this project was the question of to what extent the team could—or should—try to anticipate potential uses of the collection and its data. Any act of resource description by its nature entails some degree of this: information professionals try to determine the attributes of information resources that are most likely to be of interest to users, so that they can index materials in a way that facilitates retrieval. This is a central principle of user-centric metadata design. In the case of the ISNU World War I Service Records Collection, the team chose to provide more granular access to attributes such as location and service/military unit because they seemed likely to be of interest to potential researchers. Furthermore, knowing that part of the project would eventually apply digital humanities tools to the data, the team took care to add metadata in formats (e.g., geographic coordinates, URIs) that would facilitate computational reuse. This orientation, in addition to enabling our own team's experiments, has the added benefit of potentially facilitating further digital humanities adaptations of the data by other researchers.

This approach, however, has limits. It is impossible to fully anticipate all the potential uses of a collection, and the granularity of description will always ultimately be limited by time and resource constraints—catalogers cannot index everything. For example, a researcher could very conceivably be interested in details such as service members' dates of embarkation to and from the United States, names of ships, ports of entry, and so on. This data is very often included in the ISNU service record files, but the team chose not to index these attributes, deeming them too time-intensive to include. A researcher could potentially compile this information, but it is beyond what the team considered in-scope for basic resource description.

Similarly, the team's use of digital humanities tools to enhance access to the collection was intended as more of a proof-of-concept rather than a comprehensive demonstration of the collection's value for computational reuse. The hope is that these visualizations in ArcGIS, StoryMaps, Histropedia, and other tools will inspire and encourage others to do further research with the ISNU service records, but ultimately it is up to students and researchers to determine the questions they want to ask of the collection.

An Incremental Approach to a Developing Model

It has been close to a decade since *Always Already Computational* spearheaded the movement, and CAD proves to be compelling work for digital cultural heritage collections. Despite substantial developments among the GLAM community, libraries may still find the concept fairly new.

At the American Library Association Core Forum 2024 presentation, "Looking Beyond Your Digital Repository: Metadata for Storytelling," the authors began their session with a poll question to

audience members: What is your experience with CAD? Thirty-three out of the 62 audience members responded. The results of the poll are as follows:

- Never heard of it: 12%
- Heard of it, but never attempted: 55%
- Worked with it at my institution: 18%
- Have a CAD program at my institution: 15%

From this audience, results indicate that the CAD paradigm is still fairly unexplored within libraries: 67% of respondents had never attempted a CAD project.³²

The metadata team committed a substantial amount of time to the project atypical of regular workflows to design, develop, and test open datasets in multiple ways for usability. The team created and shared linked data in the broader data landscape, adapted methods to share the collection data and documentation openly with the community by creating an open GitHub repository, and produced a digital humanities exhibit, all within the confines of limited technological support. Similar to what the *Collections as Data: State of the Field and Future Directions* participants found, the library's staffing resources are limited for this work, even with the new metadata team in place. These processes are not sustainable to fold into workflows for every digital collection, nor does the library currently possess the infrastructure to maintain such an endeavor. However, the investigation produced a proof-of-concept design to show the potential uses and value of digital collections metadata in digital humanities research and serves as an example to promote digital scholarship within the library and university community. Shortly after the completion of the project, the metadata librarian and digitization coordinator were invited to discuss the project as part of a digital scholarship speaker pilot series at the library—an indication of the project's potential for raising the profile of digital humanities and CAD approaches on campus.

Digital humanities and digital scholarship are still novel concepts at the university, and the time investment to build a dedicated CAD program cannot currently be justified for the library. While the digital collection holdings are growing, to date, there has not been a high demand for digital humanities research support with requests to access digital collections data. However, although ISU does not have the same level of institutional support for CAD services as some of the projects referenced in *Collections as Data: State of the Field*, there are nonetheless substantial benefits to adopting an incremental approach, as was done in this project, where techniques and tools are strategically employed for low-barrier, high-impact tasks. Nontraditional digital and computational research and data expertise are crucial areas to develop in the metadata team as libraries face challenges of big open data, algorithms, and artificial intelligence. The team's approach to creating metadata for the collections is one that goes beyond creating access to the collection and includes fostering activity in a wider information landscape of open data access, interoperability, and reuse possibilities. The team will continue to incrementally incorporate CAD and linked data principles into regular workflows, with a focus on high-priority, low-effort tasks such as recording metadata with URIs and recording geographic coordinates. As a concluding question at the Core Forum 2024 presentation, the authors asked the audience to share a CAD project at their institution. Out of the seven responses, 43% answered with projects involving mapping geographic location components. These results may offer anecdotal evidence of mapping geographic locations as a future in-demand topic for research analysis and visualizations of collections data in libraries.

CAD principles in the “Santa Barbara Statement on Collections as Data” and the subsequent Vancouver Statement that are particularly relevant to the metadata team’s practice are iterative efforts toward interoperability, sharing and transparency of data and its provenance, working against historic and contemporary inequities represented in collections, and appropriate considerations to data consumption by technologies.³³ As a contributing partner to the Illinois Digital Heritage Hub and the DPLA, the library shares its materials and ensures the data follow Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to be harvested by these platforms to provide wider access and discovery. The metadata for the collections is also openly retrievable through DPLA, although with barriers, as it is only available through DPLA’s application programming interface (API) in JSON-LD and only for metadata fields that DPLA uses on its platform. Documentation for each digital collection and the explanation for its metadata schema and creation is a core task with the metadata team. This work makes substantial efforts to fill in gaps or biases that exist in historical collections with appropriate description to avoid potential harm and safeguard against the further amplification of these narratives through linked data, data reuse, and possible consumption by artificial intelligence.

CONCLUSION

This paper has presented a case study of CAD at a public R2 university, arguing that, even at institutions without a history of robust digital scholarship support, an incremental CAD approach can be implemented to encourage discovery of and engagement with digitized cultural heritage materials, while also facilitating future reuse of digital collections metadata. The challenges faced by the team, including a lack of institutional support for technological experimentation, anticipation of possible data uses, and perceptions of the CAD approach as still being relatively novel, are likely to be encountered by many institutions undertaking a CAD project for the first time. However, despite the library’s lack of infrastructure to fully support linked data or a free-standing CAD program, the team’s metadata strategy was grounded on the advantageous nature of structured data for interoperability, data sharing, management, and analysis—CAD and linked data principles. Likewise, the structured data will be indispensable when appropriate technologies are available to increase access and utilization of the digital collections and their associated metadata following FAIR guiding principles.

ENDNOTES

- ¹ Angela Yon, Emily Baldoni, and Maddi Loïselle, “Capturing Hidden Narratives with Metadata and Digital Storytelling,” paper presented at the Digital Library Forum, East Lansing, Michigan, July 30, 2024, <https://ir.library.illinoisstate.edu/fpml/250>; Emily Baldoni, Maddi Loïselle, and Angela Yon, “Looking Beyond Your Digital Repository: Metadata for Storytelling,” paper presented at the American Library Association Core Forum, Minneapolis, Minnesota, November 14–16, <https://ir.library.illinoisstate.edu/fpml/249>.
- ² Toma Tasovac, Sally Chambers, and Erzsébet Tóth-Czifra, “Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper,” HAL Open Science, 2020, <https://hal.science/hal-02961317>.
- ³ Thomas Padilla, “Always Already Computational,” *Always Already Computational: Collections as Data*, 2018, <https://collectionsasdata.github.io/>; Thomas Padilla, “Part to Whole,” *Collections as Data: Part to Whole*, 2019, <https://collectionsasdata.github.io/part2whole/>.
- ⁴ Ashleigh Hawkins, “Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-digital Archives via the Semantic Web,” *Archival Science* 22 (2022): 319–44, <https://doi.org/10.1007/s10502-021-09381-0>.

- ⁵ Angela Yon, "Contextualizing Performers in Circus Route Books: Linked Data Entities and the Open Data Environment," *Trends and Issues in Library Technologies* (June 2023): 40–51, <https://repository.ifla.org/handle/20.500.14598/2624>.
- ⁶ "Illinois State Normal University," Illinois State University, accessed January 22, 2026, <https://traditions.illinoisstate.edu/history/isnu>.
- ⁷ Thomas Padilla et al., "Santa Barbara Statement on Collections as Data—Always Already Computational: Collections as Data," Zenodo, May 20, 2019, <https://doi.org/10.5281/zenodo.3066209>.
- ⁸ Padilla, "Part to Whole."
- ⁹ Elisa Naquin and Leah Duncan, "Metadata for Collections as Data in a Multi-Institutional Digital Library," *Journal of Library Metadata* 23, nos. 1–2 (2023): 43–58, <https://doi.org/10.1080/19386389.2023.2229229>.
- ¹⁰ Rachel Wittmann, Anna Neatrou, Rebekah Cummings, and Jeremy Myntti, "From Digital Library to Open Datasets: Embracing a 'Collections as Data' Framework," *Information Technology and Libraries* 38, no.4 (2019): 49–61, <https://doi.org/10.6017/ITAL.V38I4.11101>.
- ¹¹ Florian Windhager et al., "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges," *IEEE Transactions on Visualization and Computer Graphics* 25, no. 6 (2019): 2311–30, <https://doi.org/10.1109/TVCG.2018.2830759>.
- ¹² Monika Glowacka-Musial, "Visualization and Digital Collections," *Library Technology Reports*, 57, no. 1 (2021): 5–10, <https://doi.org/10.5860/ltr.57n1>.
- ¹³ Windhager et al., "Visualization of Cultural Heritage," 2312.
- ¹⁴ Antonis Bikakis et al., "Editorial: Special Issue on Semantic Web for Cultural Heritage," *Semantic Web* 12, no. 2 (2021): 163–67, <https://doi.org/10.3233/SW-210425>.
- ¹⁵ Hawkins, "Archives, Linked Data and the Digital Humanities," 333–35.
- ¹⁶ Stephan Bartholmei et al., *Opportunities for Academic and Research Libraries and Wikipedia: A Discussion Paper* (IFLA 2016), <https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/info-society/iflawikipediaopportunitiesforacademicandresearchlibraries.pdf>; Program for Cooperative Cataloging, "PCC (Program for Cooperative Cataloging) Strategic Directions: January 2018–December 2021," revised January 24, 2019, <https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf>; Stacy Allison-Cassin et al., *ARL White Paper on Wikidata: Opportunities and Recommendations* (American Library Association, 2019), <https://www.arl.org/wpcontent/uploads/2019/04/2019.04.18-ARL-whitepaper-on-Wikidata.pdf>.
- ¹⁷ Program for Cooperative Cataloging, "Wikidata: WikiProject PCC Wikidata Pilot," Wikidata, accessed June 13, 2025, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot; "Early Adopters Phase," Program for Cooperative Cataloging, accessed June 13, 2025, <https://wiki.lyrasis.org/display/PFCCP/Early+Adopters+Phase>; Program for Cooperative Cataloging, "Wikidata: WikiProject PCC EMCO Wikidata CoP," Wikidata, accessed June 13, 2025, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_EMCO_Wikidata_CoP.
- ¹⁸ Enslaved: Peoples of the Historical Slave Trade, accessed June 6, 2025, <https://enslaved.org/>; Cogan Shimizu, et al., "The Wikibase Approach to the Enslaved.Org Hub Knowledge Graph," in *The Semantic Web – ISWC 2023*, edited by Terry R. Payne et al. (Cham, Switzerland: Springer, 2023), 419–34, https://doi.org/10.1007/978-3-031-47243-5_23.
- ¹⁹ Sarah Ames, "Transparency, Provenance and Collections as Data: The National Library of Scotland's Data Foundry," *LIBER Quarterly* 31, no. 1 (2021): 1–13, <https://liberquarterly.eu/article/view/10880/11789>.
- ²⁰ Gustavo Candela et al., "Reusing Digital Collections from GLAM Institutions," *Journal of Information Science* 48, no. 2 (2022): 251–67, <https://doi.org/10.1177/0165551520950246>.

-
- ²¹ Chela Scott Weber, "Collections as Data: Nascent Progress and Common Need," *Hanging Together: The OCLC Research Blog*, April 20, 2023, <https://hangingtogether.org/collections-as-data-nascent-progress-and-common-need/>.
- ²² Sally Chambers et al., "Position Statements -> Collections as Data: State of the Field and Future Directions," Zenodo, May 2, 2023, <https://doi.org/10.5281/zenodo.7897735>.
- ²³ Angeline V. Milner, World War I Illinois State Normal University Service Records Collection, Dr. Jo Ann Rayfields Archives, Illinois State University, Normal, Illinois, https://digital.library.illinoisstate.edu/digital/collection/WWI_records/.
- ²⁴ Milner, World War I Illinois State Normal University Service Records Collection.
- ²⁵ Candela et al., "Reusing Digital Collections from GLAM Institutions."
- ²⁶ "ISNU World War I Service Records," GitHub, modified June 4, 2025, <https://github.com/milner-library-metadata/isnu-world-war-i-service-records>.
- ²⁷ Padilla et al., "The Santa Barbara Statement."
- ²⁸ "All Locations Visited by ISNU Affiliates During World War I," ArcGIS Online, last modified June 4, 2025, <https://isu-geomap.maps.arcgis.com/apps/instant/basic/index.html?appid=2843c3467d9f43eb89712ba1d8b1b83d>.
- ²⁹ "Locations by Military Branch or Service Organization," ArcGIS Online, last modified July 22, 2024, <https://isu-geomap.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=faeb223654c0490bab0f9c02a90ac080>; "Locations of ISNU WWI Service with Sites of Battles, Fatalities, and Influenza Cases," ArcGIS Online, last modified June 3, 2025, <https://isu-geomap.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=170e6b9fc01d438092ab19550eef6ab3>.
- ³⁰ Emily Baldoni, Maddi Loiselle, and Angela Yon, *Over There*, ArcGIS, last modified June 16, 2025, <https://storymaps.arcgis.com/collections/c6c1bc19ebdf41548f9fd8662c1ff1e0>.
- ³¹ "About CollectionBuilder," CollectionBuilder, accessed June 26, 2025, <https://collectionbuilder.github.io/about.html>.
- ³² Baldoni, Loiselle, and Yon, "Looking Beyond Your Digital Repository."
- ³³ Padilla et al., "Santa Barbara Statement"; Thomas Padilla et al., "Vancouver Statement on Collections as Data," Zenodo, September 13, 2023, <https://doi.org/10.5281/zenodo.8342171>.