

**Final Report - White Paper**

Grant Number: HAA-269062-20

Project Title: Data Repository Infrastructure for Prosopographic Data

Project Director: Sarah Stanley

Grantee Institution: Florida State University

Submission Date: December 2022

## **Table of Contents**

Project Summary	2
Project Origins and Goals	2
Project Activities, Team and Participants	3
Workshop Day 1	4
Workshop Day 2	5
Workshop Days 3 and 4	6
Project Outcomes	7
Project Evaluation and Next Steps	8
<b>Appendices</b>	<b>9</b>
Appendix A - Workshop Schedule	9
Appendix B - Influential Projects	11
Appendix C - User Stories and Personas	12
Appendix D - Minimum Metadata Requirements	29

## Project Summary

Data Repository Infrastructure for Prosopographic Data was a workshop that brought together experts from libraries, data humanities, and prosopography to discuss the minimum requirements for building a repository for prosopographic data. Prosopography (also sometimes called “collective biography”) involves the study of collective or group characteristics from a particular historic context, especially in cases where individual biographic study is not possible. A data repository allows for the sharing and preservation of data across disciplinary or institutional contexts, and ideally allows for the remixing and reuse of data. Designing minimum requirements for a prosopographic data repository lays the groundwork for better data sharing and dissemination practices in prosopography work.

## Project Origins and Goals

This project was initially conceived from discussions between faculty and librarians at Florida State University. In order to provide a shared space for learning, the University Libraries hosted a Prosopographies Working Group for faculty and students interested in creating data on collective biography. During these conversations, working group members discussed different data collection standards and best practices, and talked about how to create shared resources for those interested in learning about prosopography. During this time, FSU Libraries was also in the process of developing and maintaining DigiNole: FSU’s Research Repository (<https://diginole.lib.fsu.edu/>). Library developers and repository specialists were participating in discussions about how to best develop infrastructure for sharing and preserving data for members of the FSU community. These two activities led to discussions about how FSU could combine its expertise in repository development and its faculty’s interest in prosopographic data.

## Project Activities, Team and Participants

The project was designed as a three-day workshop to be held in Tallahassee during the Fall of 2020. The COVID-19 pandemic delayed many of the activities of this grant and necessitated a move from an in-person event to a virtual event. Additionally, the workshop was spread out over four half days rather than three full days to prevent exhaustion from excessive video-conferencing. The workshop was originally designed to include both invited participants, and participants who applied to join the workshop. However, the team decided that a small virtual meetings would promote deeper conversation, so the workshop was ultimately limited to invited participants.

**Project Team:** Sarah Stanley, Digital Humanities Librarian at Florida State University, served as the Project Director for this grant. Sarah designed the workshop schedule, managed communications with participants, scheduled the event, and synthesized the discussion and notes into the final outputs. Holly Horner, PhD served as the Workshop Coordinator during 2020. She helped with populating the website, doing some initial planning for a possible in-person meeting, and writing the initial call for proposals when the team still anticipated accepting applicants.

Matthew Hunter, Digital Scholarship Librarian at Florida State University, also assisted with coordination and facilitation of the virtual meeting. Adira Philyaw, Graduate Assistant for the Office of Digital Research and Scholarship, observed the meeting and took notes on the discussions that took place.

**Participants:** In addition to Sarah Stanley, Matt Hunter, and Adira Philyaw, the following six invited participants attended the virtual workshop.

- Bryan Brown, Repository Developer, Florida State University
- Dr. Alison Booth, Professor and Director of the Scholars' Lab, University of Virginia

- Dr. Will Hanley, Associate Professor of History, Florida State University
- Dr. Courtney Rivard, Associate Professor and Director of Digital Literacy and Communications Lab, University of North Carolina, Chapel Hill
- Dr. Daniel Schwartz, Associate Professor of History and Associate Director of the Center of Digital Humanities Research, Texas A&M University
- Dr. Stewart Varner, Managing Director of the Price Lab, University of Pennsylvania

Additionally, we would like to acknowledge the involvement of Dr. Tarez Graban, Associate Professor of English at Florida State University, who was unable to attend due to a last-minute conflict, but was still integral to the planning and execution of this grant.

## Workshop Day 1

The first day of the workshop consisted of participant introductions, case studies, and developing a shared framework for the discussion. The day began with the participants introducing themselves and their investment in prosopography or data repositories. Several participants had experience with one, but not the other, so there were several times throughout the workshop where participants had to check in about the specialized language being used and defining key terms.

After introductions, the conversation continued with case study presentations. Professor Alison Booth presented on the Collective Biographies of Women (CBW) project and the intellectual framework and motivations for that project's activities. Several other projects that had bearing on the work of prosopographies and data repositories were also discussed. A list of influential projects mentioned during the meetings are included in Appendix B.

Several of the participants had facilitated or participated in discussions about shared standards for prosopography in the past, so the group spent some time discussing what factors

hadn't been previously addressed or considered in those other venues. The group spent some time coming to a shared understanding of incentives and motivations for contributing to such a repository. The group agreed that there is a lot of labor necessary to do the translation work of creating data in an interoperable format. This labor is frequently not recognized, rewarded, or compensated. The discussion at the end of Day One included brainstorming ideas for functionality that could incentivize or otherwise encourage submission and participation.

## Workshop Day 2

The second day of the workshop was devoted to thinking about potential users of the service. Before the group could think through what functions the data repository would have, they needed to envision possible future users, their goals, and their background. To accomplish this, they created “user personas” for several different possible visitors to the repository. The participants based their personas off of Samvera’s Hydra in a Box personas document.<sup>1</sup> They brainstormed several user roles that might have different motivations for using the data repository, including both academic and non-academic users, as well as users at different stages of their education or careers. The full user personas generated during that discussion can be found in Appendix C. The roles identified were:

- Data repository administrator
- Big data researcher
- Data literacy educator
- Genealogist
- Independent Researcher

---

<sup>1</sup> “User Personas” can be found under the “Information Architecture” heading at the bottom of the Hydra in a Box documentation wiki:

<https://samvera.atlassian.net/wiki/spaces/samvera/pages/405212321/Hydra-in-a-Box+Design+Documents>

- Peer Reviewer
- Proxy (or “submitter”)
- Institutionally-affiliated researcher

We also discussed another user type that we ultimately didn’t design a persona for: the “bad actor.” The group agreed that, especially in the cases of collective biography for marginalized or invisibilized groups, there may be those who wish to sabotage or exploit the open data on the system. The participants grouped these bad actors into three categories: “doxxers” (or those who may try to publicize data of users for harassment), “monetizers” (or those who would try to scrape the data and put it behind a paywall), and “bad-data-submitters” (or those who would try to interrupt data integrity by flooding the system with incorrect data). While we did not create a concrete user persona for these potential users, we agreed it would be important in future phases of the project to plan for or otherwise preempt misuse and abuse of the open data this repository provides.

## Workshop Days 3 and 4

During Day Three of the workshop, the personas generated on Day Two were finalized. After the personas were created, the group worked together to generate minimum metadata standards in order for the repository to work as planned. The group discussed previous efforts by those working with prosopographies to agree upon one universal standard. In the past, these efforts had proved unsuccessful, as so many projects have different focuses. The discussion on Day Three addressed trade-offs between ease of submission and ease of querying and comparison. The group decided that fully constraining the data to ensure that all datasets met one standard would discourage submission and remove an element of experimentation and play that

is necessary for data creation. However, everyone still agreed that there needed to be at least a few requirements to allow for the data to be searchable and discoverable in the system.

In order to circumvent some of the challenges that designing a single standard poses, the group decided to generate a list of minimum metadata requirements that all projects must meet. These data requirements would allow datasets to be integrated and searched in the system, but wouldn't necessarily allow for the cross-comparison of all datasets. In addition to having a light set of minimum standards, the group also agreed upon a few data fields that submitters could optionally map their datasets onto to allow for more granular searching of the data. The group agreed that these data fields were useful, but could not be required by the system, since not all prosopographies generate the same data. The minimum requirements and data fields that could be mapped onto are provided in Appendix D.

During the final day, the workshop participants discussed the necessary next steps for building a future repository. The participants generated a list of topics that were discussed, but not fully addressed. These are included in the “Project Evaluation and Next Steps” section, as items that will need to be resolved before the creation of a repository can begin. The group also discussed stakeholders that should be included in future discussions about the data repository.

## Project Outcomes

This project set out to think through different considerations that need to be made before building a data repository for prosopographic data. The workshop organized as a part of this project generated a lot of discussion about the people who use digital humanities infrastructure and how we can center them as we build new data sharing and preservation apparatuses. The primary outputs of this project were the User Personas, which can eventually be transformed into a User Requirements document for a future repository developer. The “Specific Goals” column

for each user outlines in detail the necessary functionality for this repository. Additionally, the project resulted in a set of recommended data standards that could be applied to this repository, including requirements for the datasets themselves and the administrative metadata the system uses. All of the results of these discussions can be found at <https://prosopreponeh.github.io/>.

## Project Evaluation and Next Steps

The next steps for continuing this project will need to involve a larger number of stakeholders. Specifically, the group agreed that future iterations of this project would need to hold discussions with non-academic audiences, as one major gap in the list of participants was non-academic audiences. This list could include genealogists, independent researchers, and community historians. The group also recommended hosting workshops at the disciplinary conferences (such as MLA, AHA, ALA, and others) to discuss with researchers and librarians what would incentivize them to use the repository. Future iterations of this project will need to convert the user personas generated into User Requirements and Technical Requirements before the repository can be built.

This project should be continued by developers in the digital humanities space, and especially those who have experience with linked data and RDF. The benefit of the current project as designed was that it allowed the project team to discuss motivations and incentives for using a data repository with current faculty who know what promotes usage of this type of infrastructure. However, as a result of these discussions, the requirements laid out in the User Personas are granular and involve data sharing, data remediation, querying and even social elements. The next steps of this project will need to be completed by a group that can reconcile these “blue sky” elements with the realities of existing technical infrastructure.

## Appendices

### Appendix A - Workshop Schedule

#### **Monday**

##### **Opening Session - 9-10:30am**

- Introduction to the goals of the workshop
- Introduction of participants

##### **Break - 10:30-10:45am**

##### **Project Examples - 10:45am-1pm**

- Generating Examples of Prosopographic Data Projects
  - Synthesizing a List of Common Features
- 

#### **Tuesday**

##### **Recap and Introduction to Personas - 9am-10am**

- Recap yesterday's session
- Discuss the "user stories" and "personas" as a way to talk about how the site will be used

##### **Persona Generation, part 1 - 10am-10:30**

- Create a list of types of users for a prosopography data repository

##### **Break - 10:30-10:45am**

##### **Persona Generation, part 2 - 10:45am-1pm**

- Create detailed personas for each possible user type
-

## **Wednesday**

### **Submission Metadata Requirements - 9am-10:30am**

- Discuss metadata requirements for submission to a possible future repository
- Generate list of required metadata for submission of data sets

### **Break - 10:30-10:45am**

### **Administrative Metadata Requirements - 10:45am-1pm**

- What metadata needs to be kept about the creation and management of the data?
  - Rights statements, access controls, version management
- 

## **Thursday**

### **Wrap-up and next steps - 9am-12pm**

- What was not covered in this discussion that should be addressed in next steps?
- What stakeholders should be involved in future iterations of this discussion?
- How do we facilitate data sharing while we wait for the funding and development of this data repository?

## Appendix B - Influential Projects

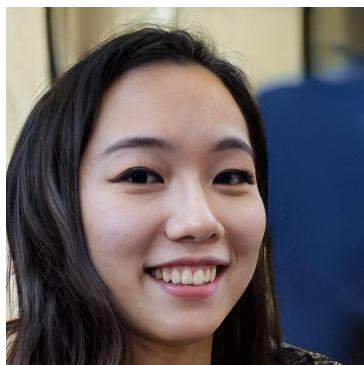
The following projects were cited as models for data modeling and data dissemination work during the workshop discussion.

- Advanced Research Consortium (ARC) at Texas A&M University -  
<https://arc.dh.tamu.edu/>
- Collective Biographies of Women - <http://cbw.iath.virginia.edu/>
- The Feminist Controversy: Text Mining the Novel -  
<http://emergengenders.dh.tamu.edu/project/>
- Layered Lives - <https://layeredlives.org/>
- Linked Infrastructure for Networked Cultural Scholarship (LINCS) -  
<https://lincsproject.ca/>
- Medieval Electronic Scholarly Alliance (MESA) - <https://mesa-medieval.org/>
- Orlando Project <https://cwrc.ca/orlando>
- Social Networks and Archival Contexts (SNAC) - <https://snaccooperative.org/>
- Syriac Biographical Dictionary - <https://syriaca.org/persons/index.html>
- Syriac Persons, Events, and Relations (SPEAR) - <https://syriaca.org/spear/index.html>

## Appendix C - User Stories and Personas

These user stories were generated on Day 2 of the workshop. The formatting for these personas were based off of Samvera's Hydra in a Box User Personas, found at <https://samvera.atlassian.net/wiki/spaces/samvera/pages/405212321/Hydra-in-a-Box+Design+Documents>. Unless otherwise indicated, the faces used for the personas were generated from <https://thispersondoesnotexist.com/>.

### **Persona #1: Administrator**



**Name:** Cindy Smith

**Role:** Administrator

**Institution:** Large Public Institution

#### **Description:**

Cindy serves as a repository administrator for several institutional repositories at her large university. Three years ago, Cindy received a PhD in History, with a focus on Digital Humanities. While she worked on her dissertation, she began taking library school courses, and specialized in Data Science and Information Technology. She has to manage a lot of different systems, so she only has the ability to check in occasionally on the Prosopographies Repository.

**Frequency of use:** As needed

**Technical Proficiency:** High

#### **Incentives:**

- Work on interesting projects in academia, without the pressure of a tenure clock
- Salary

Motivation	Scenario	Specific Goals
Technical Review of Incoming Data	A submitter wants to add a dataset, and Cindy needs to be able to review the dataset before ingest in order to ensure that the dataset is valid within the system requirements	<ul style="list-style-type: none"> <li>• See newly submitted data in a curated view</li> <li>• See outcomes of automated checks on data</li> <li>• Make necessary repairs to data and save through system-provided user interface</li> <li>• Publish data publicly</li> <li>• Inform submitter of changes made to data using system (not email)</li> </ul>
Communicate system requirements to new submitters	Cindy frequently has to consult with new users to ensure that they understand the submission requirements. Rather than reinventing the wheel for each new submitter, she relies on pre-written documentation to orient new submitters to the system requirements.	<ul style="list-style-type: none"> <li>• Link to relevant documentation about system requirements</li> <li>• Revise documentation according to questions received</li> <li>• Inform submitters about best practices for transforming metadata from original format into format required by system</li> <li>• Review submitted data for validity</li> <li>• Provide feedback to submitters about their submissions</li> </ul>
Communicate system errors and performance issues with system administrators	Cindy occasionally receives emails from users about error messages that the repository throws during queries. She needs to be able to replicate the problem and communicate the errors to the system administrators so that they can work on a solution to the problems	<ul style="list-style-type: none"> <li>• Collect reports from users regarding displayed error messages, slow system performance, crashing, etc. via internal feedback mechanism</li> <li>• Ask users for more contextual information (when necessary) - system encourages users to provide specific details on issue (such as screenshots)</li> <li>• Reproduce issue if possible to confirm</li> <li>• Forward that feedback to system administrators along with additional contextual information</li> <li>• Create mechanisms for informing users of known issues</li> <li>• Report resolution of issues to users</li> </ul>

## Persona #2 Researcher-Submitter



**Name:** Lauren Cole

**Role:** Faculty researcher

**Institution:** Midsize private institution

### Description:

Lauren is an assistant faculty member going up for tenure in the next few years, and wants to do a big data / prosopography project based on networks of authors using this repository as part of her tenure submission. She works with graduate student assistants to help with managing the data.

**Frequency of use:** Bursts of frequent use while harvesting data, then long periods of inactivity while working on the data locally.

**Technical Proficiency:** High technical proficiency for data visualization. May need help working with the API.

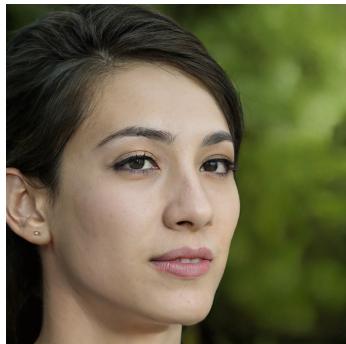
### Incentives:

- Complete her project and create a series of publications related to it
- Get tenure
- Give graduate assistants experience working on a faculty research project

Motivation	Scenario	Specific Goals
Perform a query of the repository & export the results as a data set that can be managed locally	Lauren wants to use machine learning on a subset of data in the repository. She wants to find every dataset that employs a specific RDF ontology, so she can download them, store them locally, and manipulate and analyze them on her own machine.	<ul style="list-style-type: none"><li>• Query data by data format</li><li>• Download the data that meets her parameters in multiple formats (RDF, CSV, JSON)</li></ul>

Motivation	Scenario	Specific Goals
Evaluate quality of data retrieved / understand the context in which the data was produced	Lauren needs to be able to evaluate the data at scale	<ul style="list-style-type: none"> <li>Find data dictionaries and documentation that explain the data</li> <li>Find peer review statements or contexts that confirm the validity and veracity of the data</li> <li>See track changes on the dataset so that she can see the chain of custody for the datasets</li> </ul>
Create & publish a derivative data set based on data retrieved from this repository	Lauren would like to download a data set from the repository with information relevant to her research, but then work on that data set locally to enhance it with additional information. Finally, she would like to publish this customized version as a derivative data set as a research artifact, hosted independently of the prosopographic repository (somewhere like Open Science Framework).	<ul style="list-style-type: none"> <li>Find data permissions to ensure that publishing derivative data is allowed</li> <li>Be able to cite original data in repository and link back to source</li> </ul>
Enhance and re-submit the data back to the repository with supplementary data	Lauren was able to reconcile many of the identities with VIAF numbers. She wants to enrich the core data with these supplemental data.	<ul style="list-style-type: none"> <li>Upload the data that she has transformed</li> <li>Explicitly associate her data with the source data set</li> <li>Make a record of the changes made (both track changes and prose documentation)</li> </ul>
Report data and/or system issues to administrators as they are encountered	An institution-wide change of url formatting broke the references within the database. The system stopped functioning for users, who complained to the dean.	<ul style="list-style-type: none"> <li>Write up description of issues encountered, and send to site administrators via internal feedback mechanism</li> <li>Understand the parameters necessary to fix the problem</li> <li>Provide as much detail as possible, screenshots when applicable</li> </ul>

### **Persona #3: Data Literacy Educator**



**Name:** Lian Tan

**Role:** Assistant Professor of History

**Institution:** Regional branch of a public university system

**Description:** Lian is a historian of US immigration. Census data offers important background for her research on marriage, but is not a central part of her core interests. Courses that meet the university's data literacy requirement enroll large numbers of non-majors and form a significant share of her department's enrollments. She therefore has a strong institutional incentive to offer such courses, even though they are not her first teaching preference.

**Frequency of use:** Intensive use once per semester for a several-day/week-long module within a history course.

**Technical Proficiency:** low to medium

#### **Incentives:**

- Fulfill data literacy curricular requirement for undergraduate classes
- Educate students about prosopographic data & the issues surrounding its creation/use
- Curry social and political favor within department and university by using this tool in data literacy requirement, even if not of immediate interest to Lian ("use under duress")
- Investigate the repository for possible use in her own research

Motivation	Scenario	Specific Goals
Offer course fulfilling data literacy curricular requirement for undergraduate students	Lian's university has recently implemented a data literacy requirement as a part of its core curriculum. Since history department enrollments have declined in recent years, it is important for the department to offer courses that will attract more students. As a social historian and a person known to be engaged with social media, her department head has encouraged her to design a course that can meet this new requirement. Lian has agreed to do this and sees potential for the prosopography repository as a way to engage her students with historical data useful for doing social history.	<ul style="list-style-type: none"> <li>• Browse repository for good examples to use during class that illustrate both good and bad prosopographic data</li> <li>• Teach students about metadata and how it can connect disparate datasets within a repository of this sort.</li> </ul>
Engage student curiosity in broader disciplinary questions, for which data literacy is a means rather than end	Since Lian is teaching a module on general data literacy practices, the data and tools used are secondary to the methods of critical thinking and general data manipulation. Thus, Lian and her students need to be able to make connections between prosopography data presented in the tool and major themes in social history. Her students will ultimately write a mini-biography of a person presented in the tool as the final assignment of the module, engaging both data and the ideas in their social history textbook.	<ul style="list-style-type: none"> <li>• Lead workshop session with multiple users browsing data concurrently</li> <li>• Allow student users to pull information from the tool in a generated (or recommended) citation format</li> </ul>

<p>Evaluate the repository to see if it can fit into her own research</p>	<p>Lian is intellectually committed to linking research with pedagogy. Because she will spend a considerable share of her teaching time on digital literacy, she aims to incorporate data questions into her research agenda. Many of the tools that she uses in class are rudimentary. She would like to extend them for more complex analysis, but finds herself with few opportunities and little time for further training. Nevertheless, she tries to keep up with some of the chatter about digital methods in her research circles.</p>	<ul style="list-style-type: none"> <li>• Check repository for prosopographic entities related to research topics</li> </ul>
---	--	---

## Persona #4: Genealogist



Photo by rawpixel.com from PxHere. Made available under a CC0 license.

**Name:** Lilia Johnston

**Role:** Genealogist

**Institution:** regional historical society member; affiliated with church

### Description:

Lilia is a retired school teacher who is interested in recovering silences in their community, specifically working to determine the names of those buried in a local cemetery. As a Black American woman, she has not had access to the history of her community in the same way that white Americans have. She has done a lot of work attempting to recover information about her and her community's ancestors, even when there is no census data or other official record. She works frequently with genealogical databases, such as ancestry.com, and is proficient with advanced search techniques, although she does not have experience with RDF or SPARQL queries.

**Frequency of use:** in spurts every few months

**Technical Proficiency:** familiar with some databases

### Incentives:

- Find information about her local community and family history

Motivation	Scenario	Specific Goals
Search geographically and temporally	Lilia has some family papers that she wants to use to narrow down her search of the repository	<ul style="list-style-type: none"><li>• Search by temporal range</li><li>• Search by geographical location</li><li>• When searching for large geographic area, find areas that are enclosed by that geographic region (e.g. searching for state and finding cities within)</li><li>• Search by community events and other temporal markers</li></ul>

Motivation	Scenario	Specific Goals
Find sources for data	Lilia is interested in reading the source text that the prosopographic data is drawn from. She has a lot of experience reading historical records to write narratives and reconstruct family and community networks. While the datasets can be useful for her research, she would like to see the source texts so she can read the information rather than query or visualize it.	<ul style="list-style-type: none"> <li>When online sources exist, find links</li> <li>When no online source is available, find citation</li> <li>Flag wikidata, where appropriate</li> <li>Identify pre-internet printed reference sources that can be difficult to find online</li> </ul>
Find community of people	When searching for genealogical data, Lilia finds a dead end. There are no people being returned on a query for the specific time and location she's searching in. She wants to be able to find other people who may be searching in this particular area to know if anyone is working to increase the data in this area	<ul style="list-style-type: none"> <li>Identify specific individuals working on similar research, contact those individuals</li> <li>When data is missing, create "stub" page or community page</li> </ul>
Report problems or gaps to administrator	Lilia encounters another dead end, but this time, it is for data that she has previously seen on the site. She wants to be able to contact an administrator to verify that the data is still accessible, and that no errors are occurring.	<ul style="list-style-type: none"> <li>When there are search errors, report to administrator</li> </ul>

## Persona #5: Independent Researcher



**Name:** Joaquin Rebhorn

**Role:** Independent Scholar

**Institution:** Unaffiliated

### Description:

Joaquin works in the newly-created *Public History for People* project at Vice Media's VLog division working on the *1961 Project*. He received PhD in History from UVA focusing on 20th-century counterculture movements and published a dissertation in Scalar with a GIS component mapping neo-Beat poet-activists. Another current project looks to build a prosopographical database about geolocation and demographic data of a network of contributors to three underground periodicals. He needs speedy searchability and upload of recent updates. He is known for having revived an archive of political activist art and protest songs.

**Frequency of use:** 2-3 days a week

**Technical Proficiency:** moderate

### Incentives:

- Create video essays featuring the people in the database; he is paid for each video essay, so there is a monetary incentive.
- Demonstrate commitment to open access, community-based knowledge creation and public engagement
- Expand his followers and increase his recognition in the media landscape
- Create documentary content for *1961 Project* - he has to retain copyright/IP for the documentary but wants to share underlying prosopographic data with this tool for further engagement

Motivation	Scenario	Specific Goals
To build a platform for crowd-sourcing prosopographic data	Joaquin is invested in engaging with publics and fostering community-based knowledge formation	<ul style="list-style-type: none"><li>• Easy registration for new users that encounter the repository because of his video essays</li><li>• Ensure that there are no duplicate entries related to his project and that his audience isn't adding redundant information</li></ul>

To create quick-turnaround video essays	Joaquin needs to be able to parse data and network relations in-tool so that he doesn't have to switch back and forth between a large suite of tools	<ul style="list-style-type: none"> <li>• Create network visualizations from the queries that he executes</li> <li>• Retrieve results of queries quickly</li> </ul>
To assert copyright on data and materials where appropriate	Joaquin needs to share underlying data, but he wants to retain copyright and IP of outputs generated one layer above prosop data. Needs export and reuse functionality and granular rights/sharing permissions for people data	<ul style="list-style-type: none"> <li>• To display copyright / intellectual property rights on records he's submitted</li> <li>• Needs to have clear citational information for each data set with copyright info that can be copy and pasted</li> </ul>
To share resources/data to multiple platforms	Joaquin needs to share links/data/screencaptures to multiple social media platforms to keep engaged with followers and community on Tik Tok, Twitter, Blogosphere, Instagram, and Youtube.	<ul style="list-style-type: none"> <li>• Find and use stable urls for data sets so that others can find the materials that he is citing</li> </ul>

## Persona #6: Peer Reviewer



**Name:** Emil Laurențiu

**Role:** Associate Professor

**Institution:** R2 Institution

### Description:

Emil Laurențiu is an associate professor at an R2 institution. His research focuses on medieval monastic culture, and he uses prosopography to study how different religious communities interacted. He is invested in promoting peer evaluation processes for data-driven humanities research, since his tenure review committee had trouble evaluating his digital work. He wants to ensure that future faculty going up for promotion aren't put in the same position of needing to justify data creation and curation as intellectual labor.

**Frequency of use:** semi-yearly basis when asked to review incoming projects

**Technical Proficiency:** moderate

### Incentives:

- Receiving small financial compensation
- An altruistic desire to increase the profile and validation of DH research
- Receiving credit for peer review on annual evaluations: “service to the profession”
- Demystifying the “black box” is necessary for highly technical projects

Motivation	Scenario	Specific Goals
Retrieve datasets to review	Emil is asked to serve as a peer reviewer for a project, and he needs to be able to retrieve the necessary contextual data for review such as data dictionaries, supplementary research, and methods sections.	<ul style="list-style-type: none"><li>• To be sent a package of materials including the data submitted, data documentation, change log information, and sources, as appropriate.</li></ul>
Provide corrections and suggestions	Emil notices an issue with a particular data set. There are some predicates that the researcher is using slightly incorrectly throughout the dataset, and there are a few triples in the factoid	<ul style="list-style-type: none"><li>• To provide “inline” commentary and annotation on the data using the repository interface</li><li>• To use the repository system to make general comments on the dataset and to note systemic issues that need to be fixed.</li></ul>

	prosopography that do not align with the source text that the researcher is pulling from. He needs to be able to flag both the systemic problem and the minor mistakes easily within the system, instead of communicating these errors to the researcher over email or in a separate document.	
Identify his contributions and peer reviewed suggestions	Emil has had substantive back-and-forth with a submitter about some of their data, and his feedback has had significant positive impact on the final product. He needs some personal identifier in the system that allows for the submitter to attribute changes, comments, and suggestions for the sake of acknowledging Emil and giving him credit for his feedback	<ul style="list-style-type: none"> <li>• To have his personal identifier in the system tagged and named on the dataset as a contributor</li> <li>• To have his other scholarly identifiers (such as ORCID) linked to his profile as a reviewer in the repository system</li> </ul>
Receive responses from submitter	Emil would like to receive feedback from the submitter to know whether his comments were accepted or ignored by the researcher.	<ul style="list-style-type: none"> <li>• To see corrective action taken on previous reviewer comments (such as “accepted” or “rejected” and “justification”)</li> </ul>
Export review comments	Emil has been asked to submit a portfolio of his service work, and he needs to explain the type of labor that goes into reviewing for the Prosopographies Data Repository. He needs to be able to export and share his peer-review comments outside of the repository system.	<ul style="list-style-type: none"> <li>• To export his commentary together with an abstract of the data set/project that was reviewed in an easy-to-read and shareable format</li> </ul>

## Persona #7: Proxy Submitter



**Name:** Nida Farouqi-McIntire

**Role:** Postdoc assisting mentor with research project

**Institution:** Graduate of large R1 university, currently resident at Historically Black College (HBCU)

### Description:

Nida has recently completed her PhD in literary studies focusing on life writing in the 20th century, and is in the process of transitioning to post-doc position at an HBCU where resources are more limited than at her previous R1 institution. She is finishing work on a project with her former PhD advisor in which she needs to input a number of datasets into the repository. At the same time, she is working to transition her dissertation work into a larger digital humanities project that will have both a web presence and a publication during her 2-year appointment. She wants the data that is part of the project to be stored and accessible to both scholarly and public communities.

**Frequency of use:** Daily

**Technical Proficiency:** Medium to high

### Incentives:

- Continue research from her dissertation that will eventually lead to her first monograph
- Build a CV by collaborating with a senior faculty member during her postdoc, writing articles related to her work on the repository, and present on research at major conferences

Motivation	Scenario	Specific Goals
A sense of responsibility to mentor	Nida is more technically proficient than her postdoc mentor, so her job is to submit data and shepherd it through peer review on her mentor's behalf	<ul style="list-style-type: none"><li>• Submit data on behalf of faculty member, and explicitly link data to faculty profile</li><li>• Correct inconsistencies in faculty's original data that block submission</li><li>• Create data crosswalks between bespoke faculty data and existing standards</li></ul>

Motivation	Scenario	Specific Goals
Making data available to public community who may want access to these histories and stories	Nida's work deals with the life writing of Afro-Latinas in the New York area during the 60s and 70s revolutionary movements. She wants these stories available to those continuing their work and related families and communities.	<ul style="list-style-type: none"> <li>• Submit data and associate it with her own profile (as opposed to faculty mentor's profile)</li> <li>• Receive feedback from peer reviewers and make changes accordingly</li> <li>• Track changes in dataset to reflect evolving research</li> <li>• Connect dataset to relevant community of users</li> <li>• Create separate website that pulls from her datasets</li> </ul>

## Persona #8: Researcher-Submitter



**Name:** Julius Mcmillan

**Role:** Associate Professor

**Institution:** Small Liberal Arts College

### Description:

Julius is an Associate Professor in history at a small liberal arts college who is post-tenure and interested in advancing his digital humanities project on parish records of 18th century Ireland. He has extensive archival research experience, and has attended several digital humanities workshops where he learned about metadata and basic TEI mark-up as well as how to use a few visualization tools. He has thousands of records organized with structured metadata. He wants to upload his data to the repository because he sees the importance of storage and interoperability, but he is not completely sure how to go about the process and how to use other data in the repository that may supplement his own. He also hopes that learning how to use the repository will help him teach his undergraduate research assistants about data structures.

**Frequency of use:** infrequent

**Technical Proficiency:** low/moderate

### Incentives:

- To leverage the storage and preservation resources of the repository
- To gain access to additional data for his research
- To obtain grants related to his research
- To adopt widely-used standards, allowing others to more readily engage with his work

Motivation	Scenario	Specific Goals
Sharing data he has developed and crafting a more robust access plan for future grant writing	Julius wishes to submit a grant proposal for his research, but he knows he must provide a data management plan for his grant. He needs a stable space to store his research data in the long-term.	<ul style="list-style-type: none"><li>• To find information on the repository's data preservation and access policies</li><li>• To generate a persistent link to collected data set</li></ul>

Motivation	Scenario	Specific Goals
Making his data interoperable with other datasets	Julius was at a conference recently in which he heard about a colleague who uploaded her data on tax records and legal proceedings from the same region and time period to the repository. He wants to find this data, download it, and integrate it with his own.	<ul style="list-style-type: none"> <li>• To search collection based on date, location, and key terms</li> <li>• To download useful data</li> <li>• To read documentation on any data crosswalks that the system uses</li> <li>• To use a graphical user interface to map his data onto the standards the repository uses</li> <li>• To deduplicate and merge records</li> </ul>
Creating opportunities for undergrad research assistants to learn about data structures for humanities data	Julius is working with a group of 5 undergrad research assistants who want to learn about research in the digital humanities. He wants his students to work on his data, while coming up with a project of their own that uses data in the repository.	<ul style="list-style-type: none"> <li>• To add delegate users with granular levels of read/write access to datasets</li> <li>• To constrain entries on a given project and create project-specific guidelines for data creation</li> </ul>

## Appendix D - Minimum Metadata Requirements

### Records that are required for all submissions

- Unique Identifier
  - Locally unique string to individuate different persons in your prosopography
- Source
  - The source or sources from which you are drawing evidence and data about this person

### System-provided data-mapping options

These fields are common to many prosopographic projects, but frequently are represented using different standards. The system should provide the option for users to map their data onto this internal standard. The participants recommended representing this as skos:closeMatch (<https://www.w3.org/2009/08/skos-reference/skos.html#closeMatch>) or skos:broadMatch (<https://www.w3.org/2009/08/skos-reference/skos.html#broadMatch>) rather than sameAs. This prevents problematic assertions that certain similar data fields are the same.

- Name or Label (String)
- Place (String or URI)
- Date
- Event (with the possibility of including date ranges, “before,” and “after”)
  - Birth
  - Marriage
  - Death
- Role (String or URI)
- Prose text description (String)
- External URIs (such as to existing authority records)
- Author or Creator (of Source)