


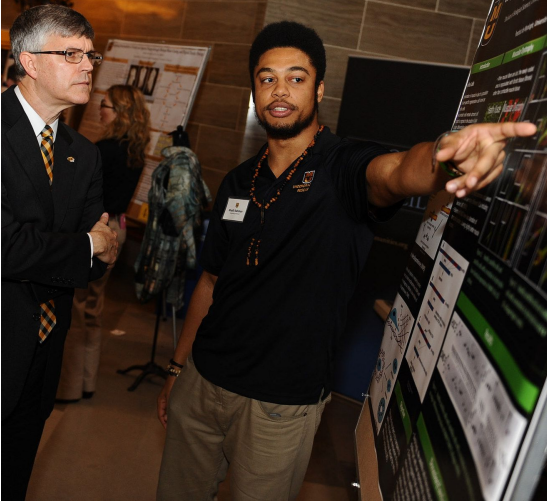
Collections as Data Personas

persona /ˌpəˈsɒnə/: an archetypal user

Faculty	Motivations	Specific Goals
	<ul style="list-style-type: none"> • Desire to uncover previously unknown information and create new insight • Need to earn the respect and trust of colleagues by producing high quality, interpretive scholarship based on solid research • Ambition to push the boundaries of traditional methods and take advantage of new research opportunities • Eagerness to design new assignments that will harness their students' interest in technology and teach them how to be effective and critical digital citizens • Interest in helping to develop better collections 	<ul style="list-style-type: none"> • Locate and download data that is trustworthy and appropriately formatted to work with their preferred software • Develop assignments for a new course with the Library's recently appointed Digital Humanities Specialist • Submit request to the IT department for 1 Terabyte of online storage for data projects so it is easy to review student work and so that it is easy to share with peers during peer review or after publication
<p>Dr. Kylie Yu is an associate professor of US history at a small liberal arts college in the midwest. Their research is focused on representations of immigrants in the media in the 19th century and they teach the survey of US history, 1865 - Present, as well as upper level courses for history majors on immigration, identity, and urban life. They are not technophobic, but they have very little formal technical training.</p>	<p>"I am very excited about the potential for digital methods to allow me to ask different questions than I am used to, but I don't think I have the tech skills to create useful datasets so I can begin experimenting."</p> <p>"I was initially very excited about the digitized newspaper collections here, but the software I wanted to use only works on plain text, not PDFs, and I simply don't have time to download each page individually. I was also disappointed in the poor quality of the OCR text. Then I discovered the enhanced availability of computationally-accessible newspaper content provided by the Library of Congress and it blew my mind to consider all the research possibilities. This is going to be fun!"</p>	


Collections as Data Personas

persona /ˌpəˈrɪsənə/: an archetypal user

Undergraduate Student	Motivations	Specific Goals
	<ul style="list-style-type: none">● Gaining valuable skills & experience for future employment● Graduate with a basic understanding of practical applications for the range of data processing techniques● The need to balance numerous curricular and extracurricular demands● Very limited time for research due to the length of the semester	<ul style="list-style-type: none">● Find data for a class assignment● Create and document a derivative dataset● Learn methods that support assessment of data representativeness and relevance● Complete a beginner's workshop on R and Jupyter Notebooks● Analyze the data with topic modeling● Credit data source in the final project
<p>Julian Santino is a junior majoring in political science with a minor in music. He plays intramural tennis and is involved in several student organizations. He has already secured a summer internship with an NGO working on poverty in Central America and hopes to work for them full time after graduation.</p>	<p>"I know that being able to work with data will be an important part of whatever job I have in the future."</p> <p>"People keep saying that I'm a 'digital native' and they assume I know more than I do. Just because I'm comfortable with social media doesn't mean I'm ready to be a programmer."</p> <p>"Until I took this class, I didn't know that librarians could help me find data and teach me software tricks. I thought the library was just books, online articles and study space."</p>	


Collections as Data Personas

persona /ˌpɜːrˈsɒnə/: an archetypal user

Data Curator	Motivations	Specific Goals
	<ul style="list-style-type: none">• A sense of responsibility for ensuring the quality of data collections and their descriptive metadata• A commitment to data preservation throughout its lifecycle• A desire to make access as frictionless as possible for users, requiring less mediation• Being able to provide helpful answers to requests	<ul style="list-style-type: none">• Collect and make available data that are high quality (reasonably free of errors and omissions), well described and easy to access• Maintain and preserve digital collections as well as archiving versions and derived datasets• Manage data acquisitions more efficiently• Create documentation that informs patrons how repository data can be acquired and used
<p>Ava Wright is a Data Curator at a public research library in New York. While the majority of her work is devoted to working with colleagues to acquire new collections and making sure they are processed correctly, she also regularly engages with visiting researchers. Her intimate knowledge of the collection helps her answer user questions and her interactions with users inform her thinking about description and access. She is seen as the local expert on text and data mining (TDM), and is recognized for collaborating with others on several successful pilot projects.</p>	<p>“Our workflow is constantly evolving as the nature of collections and user requests change and we struggle to balance our desire to release high quality data with the fact that it will never be perfect. Data cleaning often takes more time than we have. We try to make up for this with great documentation.”</p> <p>“We use GitHub to keep track of the dozens of scripts we’ve developed to process new collections and slice and dice those collections based on user requests.”</p> <p>“We generally trust our repository system, but we know that our data are getting out into the world. We are committed to open access and love that our data are being used and reused and remixed, but we worry that they will lose their context once separated from our system.”</p> <p>“I work with subject specialists to negotiate data acquisitions, to get access to data via an array of delivery mechanisms (email, mail, FTP). We have many different types of license arrangements that we need to track, and sometimes renegotiate.”</p>	


Collections as Data Personas

persona /ˌpəˈsɒnə/: an archetypal user

Repository Services Manager	Motivations	Specific Goals
	<ul style="list-style-type: none"> • Keep up with growing demand of curators, faculty & students who express interest in computational analysis of repository datasets • Share optimal practices on file formats, metadata standards & other key curation tips with researchers <i>early in their projects</i> to facilitate and prep for later ingest • Develop the general profile of the repository to be a core library service, not an invisible backend system 	<ul style="list-style-type: none"> • Offer clear requirements for system's new external API design, and create user-friendly documentation • Work with system administrator on connecting to central high performance computer and serving GIS data in the repository • Tell his supervisor about key takeaways from a conference panel on digital library service management • Build partnerships with data librarians, the digitization unit, and the digital information systems center
<p>Dustin Bishop works at a large research institution in the U.S. He obtained his MLIS in archival studies and digital preservation after working as a system administrator for a content management software firm. In his free time, Dustin is taking Python tutorials to expand his scripting skills and is reading up on the implications of cloud storage and services for repository systems.</p>	<p>“Funder-mandated data management plans propel researchers to figure out early on what to do with the output of their work and to seek someone who can help. It turns out, that’s me! I use these opportunities to advise on the many ways to make data open.”</p> <p>“Each quarter I send my boss a report of service support requests and usage analytics for repository content discovery and file downloads. I wish I had a better way to report on known interest in collections that are difficult to actually use as is and would benefit greatly from enhanced computational access services. We can offer more!”</p> <p>“Our unit’s proven ability to collaborate with and support the growing digital scholarship program is seen as a big win by the University Librarian and two deans in the humanities and social sciences. We hope that the Dean of Engineering will get on board soon.”</p>	


Collections as Data Personas

persona /ˌpɜːrˈsɒnə/: an archetypal user

Metadata Analyst	Motivations	Specific Goals
	<ul style="list-style-type: none">● Increase broad content discovery and use by fully exposing metadata for our digital objects whenever possible● Foster development of interoperable metadata by establishing and proliferating efficient and consistent descriptive practices● Produce accurate metadata that supports (re)use	<ul style="list-style-type: none">● Develop reliable connections to authority services and reconciliation tools● Create interoperable metadata that standardizes and harmonizes local & global description● Assign rights statements and licenses to repository metadata to encourage reuse● Initiate automated methods for doing qualitative/computational analysis on legacy data to fix errors, especially as standards change
<p>Becky Roberts, a metadata librarian at a large research institution, enjoys tackling difficult problems and finding solutions to them. She is familiar with XSLT, Python, and SPARQL, and enjoys browsing GitHub and professional lists to learn new tools and approaches, though she finds her time is limited for exploration. Becky spends most of her time analyzing and transforming metadata, ensuring it maintains integrity as it is processed for different endpoints. She is keenly aware of how her decisions affect end user access and discovery.</p>	<p>“As I work with metadata and come across issues, I make an effort to give feedback to the standards communities so we can make things work better.”</p> <p>“I work with metadata in all forms, from all sources. I need to know the standard followed in creating it (but I find that rarely is the standard followed completely). So you deal with what you are given.”</p> <p>“I work with others at my institution to make as much metadata freely available for bulk download with a clear market indicating that the metadata is free of rights restrictions.”</p> <p>“I’m excited about exposing our metadata at scale, but we are still figuring out how best to estimate the resources required to get it done.”</p>	


Collections as Data Personas

persona /ˌpəˈrɒnə/: an archetypal user

University Archivist	Motivations	Specific Goals
	<ul style="list-style-type: none"> • Promote the college's collections, locally and internationally • Ensure digitized and born digital material from the archives are presented and understood in their appropriate context • Develop new, diverse ways in which archival collections can be used • Encourage faculty, student, and broader community use of the collections 	<ul style="list-style-type: none"> • Create metadata for digitized and born-digital objects, including provenance, reuse rights, technological context, and relationship to other materials • Produce data visualizations for archival collections that support archival processing and provide context to researchers • Enable use of computational methodologies with archival collections while still protecting property and access rights of the content
<p>Torsten Melhus, a college archivist. With 15 years in this job, he possesses extensive knowledge of the college and is deeply committed to preserving its history. He manages the work of 2 full-time staff and 5 part-time student workers. Torsten is involved in all aspects of the department, from acquisition of new collections, to processing, to working with researchers. He is interested in exploring new ways to promote use of the college's collections, but has limited time and resources.</p>	<p>"I am thrilled that we have implemented standardized rights statements from rightsstatements.org. It was a fair amount of work to get it in place, but scholars really appreciate knowing the rights status straightaway."</p> <p>"Last semester the archives collaborated with Dr. Marshall on a digital humanities mapping project. It was a smashing success and we all learned a lot. I want more faculty to incorporate the college's collections in their classes."</p> <p>"I feel archivists and scholars are really hampered by the lack of an easy way to produce useable data from scanned manuscripts and other texts. There must be a techy solution!"</p>	

Collections as Data Persona

persona /ˌpɜːrˈsɒnə/: an archetypal user

Data Reporter	Motivations	Specific Goals
	<ul style="list-style-type: none">● Produce attractive, informative visualizations with cutting-edge techniques to accompany data-driven stories● Improve quality of data wherever possible, developing solid workflows, documentation, and practices● Share work products via a public repository to promote, extend, and persist the impact of journalistic work	<ul style="list-style-type: none">● Find well-documented data from trustworthy sources● Move data to and from our cloud service, where it is processed, with minimal overhead and technical headaches● After a piece is published, transfer the processed data, documentation and scripts used to develop visualizations to GitHub or another repository● Keep track of data downloads and use
<p>Eva Delgado works at a North American newspaper with a wide circulation in a major metropolitan area. Part analyst, part developer, part data scientist, she also keenly understands state and regional politics and education issues, frequently networking with researchers and policy makers in these areas. She is starting her 3rd year within a growing data journalism department that is increasingly central to the production of the big stories getting lots of attention from readers.</p>	<p>“I am excited when I publish a citable data visualization, with links to my source data, alongside an article, because I know the viz will catch a lot of eyes and interest in my piece.”</p> <p>“Up to now, I’ve found a lot of my data on government agency web sites, though recent political developments make future access less certain. The rest I find at research centers and policy institutes. It’s getting easier to find things online on my own, which saves me a lot of time.”</p> <p>“I try to share most of my data on GitHub. Sometimes the data is too large. It might be nice to find research institutions that can host the data for reuse by others. GitHub is great, but libraries are likely better for persistent, long-term access to data.”</p>	