Council on Library and Information Resources

Profiles in Data and Software Curation

Inna Kouper[1]

[1]Indiana University

Council on Library and Information Resources

URL: https://clir.pubpub.org/pub/v9rrmduk

> This four-part series highlights the careers and achievements of former CLIR data and software curation fellows. In 2022, Inna Kouper—a 2012-2014 CLIR data curation fellow herself—interviewed eleven former data and software curation fellows who received fellowships between the years 2015 and 2018. The work of these and 40 more fellows was made possible with the generous support of the Alfred P. Sloan Foundation. Dr. Kouper talked to each about their fellowship experiences, their careers, and the challenges of data and software curation, preparing profiles that capture the breadth and diversity of their work while summarizing their perspectives on current practice.

This research was made possible with the support of the Alfred P. Sloan Foundation.

Introduction

> This first part provides some definitions and contextualizes the parts that will follow for those who are less familiar with digital curation.

Broadly, *digital curation* refers to organizing, preserving, and improving digital information over time. It involves working with a wide range of digital formats and media, including text, audio, video, images, and data. It also involves the use of specialized tools and technologies to manage and preserve digital information. As more and more information is created and shared digitally, it becomes increasingly important to have systems and processes in place to ensure that this information maintains its value and is preserved and made accessible for future generations. Curation also helps to ensure the integrity and reliability of digital information, which is essential for many activities, including research and education.

Data curation encompasses curation activities that focus on data, including research, government, and enterprise data. It is aimed at making datasets fit-for-use and accessible over the long-term. In addition to cleaning and transformations that are applied to datasets, this work emphasizes the use of standards and metadata so that the data becomes easier to share and re-use.

Software curation is similar to data curation in the sense that it makes digital artifacts such as software and code available for future re-use. Software curation also involves processes and tools that help to describe, document, and archive the objects under curation. Similar to other types of curation, it includes caretaking practices to support the meaningful creation, use, and reuse of software as a research object. Software curation also has unique challenges, such as the need to preserve the specific hardware and software environments in which the software was developed and used.

There are many interesting initiatives that focus on digital, data, and software curation and provide resources for further reading. Some examples include:

- The [Data Curation Network](#), a collaborative project that aims to promote best practices in data curation and to support the development of data curation expertise.
- The [Software Preservation Network](#), a consortium of organizations working to preserve and make available software and other digital artifacts related to computing history.
- The [Internet Archive](#), a nonprofit organization that aims to provide universal access to all knowledge by collecting and preserving a wide range of digital content, including web pages, software, and video.

Constant Values, Changing Practices: The State of Digital Curation

> This second part focuses on data curation and is based primarily on conversations with four fellows: [Mara Sedlins](#), [Rachel Starry](#), [Zack Lischer-Katz](#), and [Fernando Rios](#).

The larger context of our conversations about digital resources and their curation with former CLIR fellows was set by the question of "why": Why engage in preserving research data and other digital content and making it available? The fellows connected their answers to the values of public education, to openness and transparency in knowledge-making, and to ethical practices in research and librarianship. Curation is not about simply making something available; it's about making sure that resources are available to as wide an audience as possible and that curation itself addresses the needs and interests of multiple communities and voices. Trust, transparency, collaboration, and engagement were among the common themes that ran through our discussions.

While the values remain stable and guide digital curation, its practices are changing. Libraries have begun to rethink the roles of institutional repositories and how they fit into curation services. One change the fellows noted is a move toward more sustainable models of building repositories, including a shift from customized in-house software development to community-supported or enterprise models (e.g., [Dryad](#) or [Figshare](#)). As Rachel Starry observed, when institutions invest in establishing centralized consortia-based services and agreements, faculty and students have stable access to data curation and publishing resources. This shift affords more scalability and sustainability, but also takes some of the burden of hands-on data curation from librarians and allows them to focus on other things, such as training, evaluation, and advocacy.

Another change is the expansion of types of data and digital content that libraries work with and blurring of the boundaries between content. For example, research data is now often curated together with the code that was used to generate or process it. It is also important, said Rachel, to think about data in a broader sense and go beyond quantitative datasets produced by scientific disciplines. Outputs from audio, video, virtual reality, 3D models, and other emerging technologies are now taking a more central part in curation and preservation discussions. All these changes require policies and workflows so that they become more institutionalized and less dependent on specific individuals to address them in an *ad hoc* fashion. As Zack Lischer-Katz put it,

> We're kind of rethinking the archival idea – you save something, you put it on a shelf, or you document it, you do the assessment of what it is. [Now it's more about] how it's going to decay, how we do

> conservation treatments, migrate the format, think about dynamics, networks, media that are potentially different every time you open them up. You have Facebook or Twitter and you're getting dynamic information from multiple databases. The algorithms are being updated. … [Data, media, software] they all kind of bleed together.

In addition to this forward-looking view, curation of digital content also sometimes involves going "back in time" as data curators deal with legacy data. Such data could be left to the library a long time ago, without proper curation or documentation. The data still has value, but it may be in the analog (paper) form or a scanned form that didn't go through character recognition. It may be in a digital format that is difficult or impossible to open. In addition to these technical difficulties, the original owner may not be available to address questions about why or how a digital artifact was created. Curation projects centered on this "legacy data" can become data curation failures despite all the efforts that a contemporary curator can reasonably invest. Through the outreach efforts of librarians who can tell the stories of these seeming failures, universities slowly come to realize the value of data curation. As Mara Sedlins stated, if curation and documentation does not happen right away--if it is not in time--then the data becomes unusable.

Libraries are also changing their services more and more to be data-oriented and collaborative. Data science techniques are becoming increasingly important, and there is a need for curators to learn to program and do other software-related tasks. The challenges of cross-campus collaboration are fairly new and unique to the digital world, and also require some training. More collaboration and communication would include creating machine actionable data management plans and interoperability across grant management, computing, institutional review boards, and other university systems. Historically, this has been difficult to achieve, and, according to Mara, "the best that we can work toward right now is just to have the right people at least talking to each other, [and] have these informal groups work together, even if it's not formalized from the top down."

In this changing context of services, types of data, and skills, the training of data curators becomes vitally important. As Fernando Rios pointed out, there is still a wide variety of approaches to it and an uneven emphasis on various types of data and skills. Thus, commonly used data representations such as tabular data so far have received much more attention in the library-based data curation community than other less common (but equally important) types of data. Or, the skills that are needed to manage a data repository are not as clearly delineated in data curation training as other aspects of it. As data curation becomes more and more complex and diversified, having just disciplinary knowledge of data or being willing to learn on the job is not enough. There is a lot of work to be done in standardization and professionalization of data curation.

Finally, digital curation is a practice that is still in need of more research and reflection, something that each of these four former fellows are engaged in actively and deeply. They are interested in more systematic understandings of their own practices and in more studies of "digital curation in the wild," focusing on not only producers of data, but also on users and re-users of all curation outputs. As curators create standards, guidelines, best practices, and digital artifacts, the question remains: Do researchers, scholars, communities,

and even curators themselves, actually use them? If so, how? Further examination of these issues would allow future curators to better connect the values of digital curation with the needs of all stakeholders.

Suggested further readings

- Kouper, I. (2013). CLIR/DLF digital curation postdoctoral fellowship—The hybrid role of data curator. Bulletin of the American Society for Information Science and Technology, Dec/Jan(2), 46–47. https://doi.org/10.1002/bult.2013.1720390213.
- Kouper, I. (2016). Professional participation in digital curation. *Library & Information Science Research*, *38*(3), 212–223. https://doi.org/10.1016/j.lisr.2016.08.009.
- Borgman, C. L. (2019). The lives and after lives of data. *Harvard Data Science Review*, *1*(1), 10-1162. https://doi.org/10.1162/99608f92.9a36bdb6.

Cloud, Obsolescence, and Collaboration: The Challenges of Software Curation

> This third part focuses on software curation and is based primarily on conversations with three fellows: Alexandra Chassanoff, Eric Kaltman, and Seth Erickson.

Our conversations about software curation began with drawing parallels between data and software curation. Fellows most familiar with software curation described it as even more complex and challenging than data curation.

Software is a form of data that needs to be described, stored, and then recovered for use in a digital environment that may be very different from the environment for which it was created. As former CLIR fellow Eric Kaltman put it, software is "data executed by other data," so its description needs to capture all the necessary linkages between its components and related files to get it running outside of the original context. Running a legacy program or interpreting code written in a programming language often requires first tracking down all of the code's dependencies and reconstructing the environment within which it was written to work, including specific versions of related libraries and packages. Alternatively, running legacy software requires an emulator, a system that can behave as another system (emulate it). The complexity of software systems is hard to capture in metadata, so curation involves not only metadata work, but also infrastructure work, programming, license negotiation, and many other activities.

To make software curation and preservation possible, improved documentation and coding practices are the best immediate steps that everyone creating software can take. For example, if software creators draft thorough "README" text files that capture software versions, dependencies, and licensing terms, eventually others will have the information they need to run the code on their machines and interpret the results of its operation. A description of the intended behavior and output for the software code would save a lot of time for future users

and increase the likelihood of future adaptation and re-use. If software producers incorporate established coding practices such as commenting, naming conventions, portability, and modularity (scalability), mistakes can be avoided during development and transfer rather than corrected at the end of the curation cycle.

At the same time, it is important to acknowledge that often code is written for some immediate tasks at hand, rather than for preservation and reuse in the future. As Seth Erickson noted, in such cases, it may not be realistic to ask researchers to re-write their code or make significant changes to documentation. Software curators advocate for software's enduring value while balancing competing constraints, especially those limiting researchers' time and effort. Often, the outcomes are modest improvements to the code's metadata and documentation.

As software curation professionals, Seth Erickson, Alexandra Chassanoff, and Eric Kaltman have deep expertise at the intersection of libraries, history, and computer and social sciences. Such expertise has allowed them to be successful in studying and implementing software curation. To them software is an artifact that connects past, present, and future, so crafting an appropriate curation strategy requires both theoretical and practical work. All three of them are involved in the Software Preservation Network, a community that documents existing practices, develops policies, and advocates for improved infrastructure and other resources. This community views software as a shared responsibility between producers, users, and curators. Its members recognize that as the variety and complexity of software increases over time, the variety and complexity of tasks involved in curating and preserving software also increases. As a result, the prioritization and completion of those tasks requires substantial consultation and collaboration among individuals with expertise in many different areas.

One such area is preservation and cloud computing. As more and more research relies on cloud platforms, software becomes even harder to preserve. The cloud provider rather than the researcher (the producer) has control over how changes are implemented and tracked. Software environments that are connected to the cloud get updated automatically, so no one can access historical versions or dependencies. With limited control of how software environments get set up, software developers and curators are losing historical access to their work rapidly. For now, software preservation focuses largely on preserving programs from between 1970 and 2010 that were created by individual developers. But it is not clear how to capture and rerun, say, Google cloud infrastructure and all the deprecated services that were used by someone just a few years ago.

The fellows agreed that the growing diversity of software tools and platforms and the policy and legal issues that affect their access and use calls for a lot more research and experimentation in the areas of software curation and preservation. Curators must understand what happens with different formats and environments, either the ones that are going obsolete or the ones that are just emerging. Alex has called for the establishment of software curation labs that would enable experimentation with formats and support multi-method studies of how people access and use data and software and what they need from their interactions with software. This type of work would need broad coalition building and a convergence of research and practice that would

involve libraries, cloud providers, research software developers, and scholars who study curation and preservation.

Suggested further readings

- Benner, J., Erickson, S., Hagenmaier, W., Lassere, M., Williford, C., and Work, L. (2022). Supporting Software Preservation Services in Research and Memory Organizations. CLIR and the Software Preservation Network. https://www.clir.org/pubs/reports/supporting-software-preservation-services-in-research-and-memory-organizations/.
- Chassanoff, A. and Altman, M. (2020). Curation as "Interoperability with the Future": Preserving Scholarly Research Software in Academic Libraries. Journal of the Association for Information Science and Technology, 71: 325-337. https://doi-org.proxyiub.uits.iu.edu/10.1002/asi.24244.
- Rios , F., Almas, B., Chassanoff, A., Contaxis, N., and Jabloner, P. (2017). Exploring Curation-ready Software: Improving Curation-readiness. Software Preservation Network. https://www.softwarepreservationnetwork.org/exploring-curation-ready-software-use-cases-2/.
- Hagenmaier, W., Williford, C., Lassere, M., Work, L., Benner, J. G., and Erickson, S. (2019). Software preservation services in cultural heritage organizations: Mapping the landscape. Paper presented at the 16th International Conference on Digital Preservation iPRES. Available at https://ipres2019.org/static/pdf/iPres2019_paper_110.pdf.
- Cardoso-Llach, D., Kaltman, E., Erdolu, E., and Furste, Z. (2021). An Archive of Interfaces: Exploring the Potential of Emulation for Software Research, Pedagogy, and Design. Paper presented at the ACM Human-Computer Interactions CSCW2. doi: 10.1145/3476035.

Data Modeling, Historical Maps, and Communities: The Many Facets of Data Curation Careers

> This conclusion focuses on digital curation career paths and is based on conversations with four former fellows: Bommae Kim, Zenobie Garrett, Jacqueline Quinless, and Jennifer Garçon.

This time, the conversations with CLIR fellows were tied to the context of their career paths and how the data they work with helps them to articulate what matters most. Working in interesting academic and commercial settings, they seem to be carving unique paths in which their professional and life goals align. They work with data through the questions of what is right, positive, and good. In addition to openness and transparency brought up in our earlier posts, these four fellows are thinking deeply about the meaning and impact of their work, who benefits from it, and how they can make a difference through their careers.

Data science, for example, has become a way to make a difference for Bommae Kim. A data scientist at a medical center, she works with electronic medical records (EMR) and builds a model that helps to find patients who would benefit from palliative care. EMR systems developed organically over the last 40-50 years, and

they have multiple known issues with interoperability, consistency, and overall architecture. As Bommae described it, she works with thousands and thousands of tables and sometimes it is hard to determine which ones to use. Data curation involves cleaning, organization, and establishing good practices and standards to make sure the analyst's work serves both the goals of the organization and the goals of helping people. Data and software reproducibility is part of data curation, and it helps to increase reliability of data modeling and build trust across the organization.

A job that allows fellows to use their advanced data and technical skills to help communities is highly rewarding. Zenobie Garrett works on projects that help cultural and research organizations to build archives and repositories that connect communities and locales through digital data. And the core issues go well beyond the technical challenges, according to Zenobie. They include thinking strategically about preserving information for future use and addressing the ethical, environmental, and equitable access issues that come with data and software curation. You have to be proactive about it, says Zenobie, who has been thinking about these issues throughout her fellowship and later career:

> I think that data curation has some inequality in it and I want that to be at the forefront of data decisions. … at Oklahoma we had these 3D VR assets, and we were very adamant that they would be used by groups that wouldn't have access to those technologies. … we were very conscious of it, we sat down and we said, Okay, we're not going to be able to get everything but what are the biases we know that happened with data and digital access and preservation? And how can we address those in the policy, even if it never comes up?

Being proactive also includes being culturally responsive and working within community-oriented ethical frameworks to ensure that research processes support accountability in community partnerships. Data curation provides a link that connects data practices to ethical and social justice frameworks, especially for curators who work with Black, Indigenous and People of Color (BIPOC) and gender diverse data and who work in service to these communities. Jacqueline Quinless works with BIPOC and gender diverse and marginalized communities in Canada and internationally, and data curation has helped her to work with those communities to frame some of the challenges communities are facing in addressing social, economic and health inequality:

> I focus on data sustainability [and] the data curation component fits really nicely. I wanted to learn about repositories, I wanted to learn about longevity of data and software and how that can be stored and still used by communities. So, if a community was interested in participating in a survey, how was that data then shared with the community? And how can that actually build their own knowledge capacity moving forward? So that curation piece was quite critical.

Through her commitment to community-driven research, data advocacy, and teaching Jacqueline promotes the idea that data curation needs to be embedded in communities. Data is not a collection of abstract facts or measurements that can be destroyed once knowledge has been generated from it. Data is everything, it comes

from various ways of being and knowing, including traditional medicines, sacred living histories, and oral traditions. Data is a living history that needs to be protected and curated over time. Research-community relationships have to be consciously addressing data-related capacity building and reciprocity so that communities can own and develop their own knowledge systems. Jacqueline suggests asking questions such as "What is the longevity of not just where the data is stored, but how it can be used by the community? Do they have capacity and ongoing support for using data? How do communities incorporate data into their daily activities and work?"

Work with communities gives rise to a new type of data curation – public and community data curation. This type of curation prioritizes data that is generated by communities rather than researchers or academic institutions. Curating such data often relies on modest resources or uneven resource pools. The role of curator then is in, as Jennifer Garçon put it, making "minor interventions that preserve or at least enhance or increase the lifecycle of materials that just won't end up in repositories for various reasons." It is in thinking about resources that each curator brings in order to accomplish very gradual steps toward preservation, conservation, discovery, and access in partnerships with individuals and community groups.

Jennifer in her work as a special collections librarian at Princeton University Library examines post-custodial practices in community archives and works toward developing practices that are less extractive and based on ethical partnerships with various communities. Similar to Jacqueline, Jennifer advocates for thinking carefully about the notion of open data. For example, details from a project that documents histories of urban black communities could be used to erase or destroy sites that are of historical significance to those communities. Or, in documenting political protests, data on individuals and their accurate positions within certain localities need to be protected, while the primary sources need to be preserved for posterity as governments can shut down mass media and remove certain messages:

> … part of the work is really thinking about the kinds of user stories, and this is kind of a practice for any kind of digital scholarship or project management, but user stories, thinking about the ways in which the data can be used, and then making the kinds of decisions about if the data is available, and what portions of the data is available and when open data is not actually the right decision.

Talking to these fellows about their data curation work and learning about their personal and professional choices in archival and curatorial practices was exciting, extremely educational, and rewarding. Whether they talked about hospital patients, ancient communities in Europe, indigenous Canadian peoples, or opposition movements in Haiti, it was obvious that they care about the communities from which their data comes. They see data as intimately linked to humans, their behaviors and decisions, and call for and practice a deeper awareness of how curation decisions affect specific groups or individuals. This adds a new dimension to digital curation as caretaking – taking care of information, humans, and the world around us.

Suggested further readings

- Quinless, J. (2022). *Decolonizing Data: Unsettling Conversations about Social Research Methods.* University of Toronto Press.
- Ithaka S+R. Big Data Infrastructure at the Crossroads Support Needs and Challenges for Universities. (2021). https://sr.ithaka.org/wp-content/uploads/2021/11/SR-Research-Report-Big-Data-Infrastructure-at-the-Crossroads-120121.pdf.
- How Jennifer Garçon Does History. (2020). *Contingent Magazine.* https://contingentmagazine.org/2020/05/15/how-jennifer-garcon-does-history/.
- Wong, J. & Henderson, T. (2020). Co-creating autonomy: group data protection and individual self-determination within a data commons. *International Journal of Digital Curation, 15*(1). https://doi.org/10.2218/ijdc.v15i1.714.

Fellow Profiles

Alexandra Chassanoff

Dr. Alexandra Chassanoff received her masters and PhD from the School of Information and Library Science at the University of North Carolina at Chapel Hill. In 2016-2018 Alex was a CLIR postdoctoral fellow in software curation at MIT Libraries, and helped investigate and model how research libraries can develop and support software preservation services. Dr. Chassanoff recently joined the faculty at UNC SILS as an assistant professor. Her current research focuses on born-digital cultural heritage and community-driven approaches to digital curation. In particular, she studies how communities use and value digital heritage artifacts, to help inform the design and development of digital knowledge infrastructures.

- Chassanoff, A. Soft Architectures: Technology + Media + Memory. [blog]. https://softarchitectures.wordpress.com/.
- Chassanoff, A., & Altman, M. (2020). Curation as "interoperability with the future": Preserving scholarly research software in academic libraries. *Journal of the Association for Information Science and Technology* 71(3), 325–337. doi: https://doi.org/10.1002/asi.24244.

- Chassanoff, A., AlNoamany, Y., Thornton, K. & Borghi, J. (2018). "Software Curation in Research Libraries: Practice and Promise." *Journal of Librarianship and Scholarly Communication* 6(1), eP2239. doi: https://doi.org/10.7710/2162-3309.2239.

Seth Erickson

Seth Erickson did his PhD at the University of California Los Angeles, Department of Information Studies. For his dissertation he studied how computational physicists make software open source and the reasons for doing or not doing that. He described their software development practices as *conspicuous computing*, a term that emphasizes a reflective engagement with software rather than a seamless and transparent relationship.

Seth graduated in 2018 and started looking for a work environment that would involve libraries and collaboration: "I like the field of librarianship as a kind of a field to participate in as a practitioner." His interest in open science and open source made the transition to a CLIR software curation postdoc position a natural and welcome fit. While at CLIR, his interest in research software expanded and allowed Seth to engage in software curation as a practitioner. During his postdoc at Penn State University, Seth organized a series of workshops about the basics of making software open source. He worked with the university office of intellectual property and other libraries to clarify and further develop university's policies on software licensing. The work involved both working out the details of the university policies and educating researchers and other stakeholders on the topics of software and its licensing.

In his current position as a Data Services Librarian at the University of California Santa Barbara Seth continues his work in the areas of software curation and reproducibility. He is also actively involved in the Data Curation Network, where he curates data and software and develops instructional materials.

- Get to know DCN Curator Seth Erickson https://libpubsdss.lib.umn.edu/dcnumn/2019/11/20/get-to-know-dcn-curator-seth-erickson/.
- Benner, J., Erickson, S., Hagenmaier, W., Lassere, M., Williford, C., and Work, L. (2022). Supporting Software Preservation Services in Research and Memory Organizations. Council on Library and Information

Resources and Software Preservation Network. doi: https://doi.org/10.5281/zenodo.7086618.

Jennifer Garçon

Jennifer Garçon received her PhD in History from the University of Miami in 2018. Her dissertation, "Haiti's Resistant Press in the Age of Jean-Claude Duvalier, 1971–1986," examined the role of independent oppositional Haitian print and radio media in the resistance to and overthrow of Jean-Claude Duvalier's regime after 1971 and its growing role in national politics. She looked at how Duvalier's policy shifts were debated and reported in print and radio newsrooms in Haiti and the diaspora. Having had extensive archival and museum curation experience, in her job search Jennifer was interested in positions in academia and memory institutions that would have resources to expand their curation services.

Dr. Garçon joined CLIR fellowship in 2018 as the Bollinger fellow in public and community data curation at the University of Pennsylvania. It was the first position of its kind in UPenn's Digital Scholarship Group. As a new fellow Jennifer was able to test many innovative ideas about what public data curation means and what kinds of tools and skills are necessary for that work. She explored the relationship between digital scholarship, data curation, and such pressing societal issues as climate change, gentrification, resource inequities, and others. In doing that, she was able to establish the path of thinking critically about the nature of collections and community archiving and shift some of the institutional thinking about archiving, discoverability, and engagement.

In 2021, Jennifer Garçon joined Princeton University Library (PUL) as the new Librarian for Modern and Contemporary Special Collections. At Princeton she works with the 20th and 21st century materials and continues to explore ways in which such materials can and need to be collected to enable research and critical analysis on this period. This includes thinking about what constitutes a primary source and what post-custodial practices are central to the contemporary archival spaces and our understanding of the historical record.

- Jennifer Garçon, Librarian for Modern and Contemporary Special Collections: Q&A https://library.princeton.edu/news/general/2022-04-11/jennifer-garcon-librarian-modern-and-contemporary-special-collections-qa.

Zenobie Garrett

Zenobie Garrett received her PhD in archaeology from New York University in 2016. Her dissertation, "Dynamic Communities in Early Medieval Aquitaine: A GIS Analysis of Roman and Medieval Landscapes in the Vézère Valley, Dordogne, France", examined changes in communities in the face of global events, including the dissolution of the Roman Empire and migration across Europe. Looking for career options that would enable using her advanced skills in spatial technologies, Zenobie expanded her job search to university libraries that were establishing new services in those areas.

In 2018 Dr. Garrett became a CLIR postdoctoral fellow at the University of Oklahoma Libraries. Her responsibilities focused on 3D and virtual reality software curation. She was involved in the creation of a 3D scanning lab in the library, which took about four years, from a pilot program through a development plan to a full lab with appropriate equipment. Using equipment and processes such as photogrammetry (taking multiple overlapping photographs and converting the 2D information into a 3D model via special software) Zenobie supported various cultural heritage and research projects at the University of Oklahoma.

Currently, Zenobie Garrett is a postdoctoral researcher at the University of Limerick in Ireland. As an archaeologist and a GIS specialist, she is involved in a project of digitally re-mapping of Ireland's historic maps and texts. It is a large 3-year project that focuses on creating a database that, in addition to the maps that go back to the 19th century, will include research documents, letters, memoirs, and books related to places and landscapes in Ireland. The goal is to develop a digital platform that will enable research and education on the various aspects of Irish localities as well as those who worked on documenting them.

- https://zenobiewan.com/
- OS200 Project: https://storymaps.arcgis.com/stories/7cedc565e15e4ba58444f9eaf435d1de
- Wyatt, K., Garrett, Z. S. (2022). Out of the Archives. In *"Innovation and Experiential Learning in Academic Libraries: Meeting the Needs of Today's Students"*. Rowman & Littlefield.

Eric Kaltman

Eric came into the CLIR fellowship with extensive experience of working in libraries. He graduated from an interdisciplinary computer science program at the University of California, Santa Cruz, where he studied how game developers produce, store, and archive software documentation. He worked with metadata for software records and interacted with librarians about how such records get ingested into library and archival systems.

His fellowship with the Carnegie Mellon University Libraries began with primarily data curation responsibilities and then expanded into software curation and preservation. Much of his work focused on liaison workflows and communication between departmental librarians and research data producers. He worked closely with computer science liaisons and raised awareness about Carnegie Mellon's data and scholarly repository. He also spearheaded the university's involvement with the Software Preservation Network and the establishment of an Emulation as a Service Infrastructure node at Carnegie Mellon.

The CLIR fellowship helped Eric to clarify his research agenda and broaden it from computer games preservation to a historical look on other types of software in the context of libraries and other archival systems. As an assistant professor of computer science at California State University Channel Islands, he taught various classes on computer games and software and studies emulation workflows and environments. One of his recent papers documented the use of software emulations for pedagogy and research through the integration of the Emulation as a Service Infrastructure (EaaSI) into a university-level course that focused on computer-aided design (CAD). The paper received a methods recognition at the 24th ACM Conference on Computer-Supported Cooperative Work and Social Computing. He plans to join the faculty of Media and Technology Studies at the University of Alberta in Edmonton, Canada, in January 2024.

- Cardoso-Llach, D., Kaltman, E., Erdolu, E., & Furste, Z. (2021). An Archive of Interfaces: Exploring the Potential of Emulation for Software Research, Pedagogy, and Design. *Proceedings of the ACM on Human-Computer Interaction*, *5* (CSCW2), 1-22.
- [Software History Futures and Technologies (SHFT) Group](#)
- http://www.erickaltman.com/portfolio/

Bommae Kim

Dr. Kim received her PhD in quantitative psychology from the University of Virginia in 2016. Having a bachelor's degree in library and information science and work experience with the UVA libraries, Bommae was interested in opportunities that would allow her to connect her expertise in data and methods with libraries and their work.

In 2016 Dr. Kim became a CLIR postdoctoral fellow at the Federal Reserve Bank at Kansas City. She worked with the Bank's research library team, which supported Bank's economists in their research. In many ways, the library was similar to a university library, providing support for a variety of data and information services. Bommae's responsibilities focused on promoting reproducible research. She developed educational and outreach materials and also worked on creating a data repository. Bommae was also instrumental in setting up and organizing a data repository to meet the researchers' needs. Later on, Dr. Kim moved to a senior data scientist position with the Medical Center at the University of Virginia.

Currently, Bommae Kim is Lead Data Scientist at Hackensack Meridian Health, where she leads data modeling and prediction efforts to improve patient care and outcomes.

- Rikard, S. M., Kim, B., Michel, J. D., Peirce, S. M., Barnes, L. E., & Williams, M.D. (2022). Identifying individual social risk factors using unstructured data in electronic health records and their relationship with adverse clinical outcomes. *SSM - Population Health, 19.* https://doi.org/10.1016/j.ssmph.2022.101210.
- How to get clinicians onboard with predictive analytics. (2021). https://www.healthcareitnews.com/news/how-get-clinicians-onboard-predictive-analytics.
- Currier, B. D., Kim, B., Edwards, C., & Butler, C. R. (2017). Research data preservation beyond data sharing and open science. Presentation to the 2017 DLF Forum, Pittsburgh, PA. http://doi.org/10.17605/OSF.IO/A8HM2.

Zack Lischer-Katz

Zack Lischer-Katz received his PhD in Communication, Information, & Library Studies in 2017 from the School of Communication & Information, Rutgers University. In his dissertation "The Construction of Preservation Knowledge in the Artisanal Digital Reformatting of Analog Video Recordings," Zack studied media preservationists and their practices, specifically, how they preserve analog video recordings. The findings point to the ongoing blending of the manual technical labor with interpretive acts and historical knowledge in the production of digital copies.

In 2016-2020 Dr. Lischer-Katz was a CLIR postdoctoral fellow at the University of Oklahoma Libraries, where his work focused on preserving virtual reality. The library was developing several new digital initiatives, including a makerspace and a virtual reality space to its range of services. Along with other colleagues, Zack set a new research agenda around the preservation of emerging technologies and media. He also developed training materials for faculty and students. With the encouragement to experiment and reach out to others who were entering this cutting-edge field, Zack gained research and practical experience necessary for the next steps in his academic career.

Since 2020, Dr. Lischer-Katz is an Assistant Professor of Digital Curation and Preservation at the University of Arizona iSchool. His interests include visual information and sensory media. He continues pursuing research and experimentation in virtual reality and plans to extend his research into the topics related to the Southwest borderlands and connections between technologies and geographic locations. He also teaches a digital curation and preservation class at the UofA and reflects on the pedagogical aspects of digital curation as it relates to educating librarians and archivists with varying career interests and goals.

- http://zacklischerkatz.com/.
- Lischer-Katz, Z. (2022). A methodological framework for studying visual information practices. *Library & Information Science Research, 44*(4), https://doi.org/10.1016/j.lisr.2022.101188.
- Lischer-Katz, Z. (2022), "The emergence of digital reformatting in the history of preservation knowledge: 1823–2015", Journal of Documentation, Vol. 78 No. 6, pp. 1249-1277. https://doi.org/10.1108/JD-04-2021-0080.
- Lischer-Katz, Z. (2020), "Archiving experience: an exploration of the challenges of preserving virtual reality", Records Management Journal, Vol. 30 No. 2, pp. 253-274. https://doi.org/10.1108/RMJ-09-2019-

[0054](#).

Jacqueline Quinless

Dr. Jacqueline Quinless lives on the Traditional Territory of the Lekwungen speaking peoples, Wsanec and Esquimalt Nations on Vancouver Island with her children. Jacqueline is a biracial person of Indian ethnicity (Hyderabad and Secunderabad, India) and Irish/British ancestry. She holds a PhD in Sociology with a focus on the health, anti-racism anti-colonialism, social inequality, data justice/ sovereignty, applied statistics, decolonization and gender from the University of Victoria. Jacqueline also completed a data fellowship during her postdoctoral work with the Council on Library and Information Resources (CLIR) in Washington DC and the University of Victoria, where she focused on data curation. Jacqueline spent 10 years working for the Federal government and has taught research methods courses extensively in Indigenous communities across Canada and Inuit Nunangat for two decades. She is a renowned public sociologist and award-winning Public Sociologist recognized by the Canadian Sociological Association (CSA) and the Angus Reid Foundation for her community-based research in the advancement of human welfare in Canada. She is the author of the book (2022) [Decolonizing Data: Unsettling Conversations about Social Research Methods](#) by the University of Toronto Press. She is an adjunct professor in Sociology, at the University of Victoria. She enjoys teaching undergraduate and graduate course at the University of Victoria and Camosun College on Vancouver Island.

- Quinless, J. and Khair, S. (2019). The enduring potential of data: An assessment of researcher data stewardship practices at the University of Victoria. http://hdl.handle.net/1828/10509.
- Quinless, J. (2022). *Decolonizing Data: Unsettling Conversations about Social Research Methods.* University of Toronto Press. https://utorontopress.com/9781487523336/decolonizing-data/.

Fernando Rios

Fernando Rios received his Ph.D. in Geography in 2015 from the University at Buffalo, SUNY. His dissertation "A new combined uncertainty and sensitivity assessment of spatial models with object-based features" focused on computational modeling and developed a framework and an implementation that improves uncertainty assessments in spatial modeling and simulations. In his job search, Dr. Rios was interested in positions within academia, but with a broader emphasis on data and software development. Data curation sounded as a good even if yet unfamiliar opportunity, and he accepted a CLIR fellowship at Johns Hopkins University. This position changed Fernando's career trajectory and opened the world of libraries to him.

At JHU Fernando investigated the domain of software curation and preservation and how the university could develop services in support of research reproducibility and reuse. He made several metadata recommendations for the repository system that were then adopted and implemented. He also developed training materials for data and software curation. Being involved in the emerging field of software curation allowed Fernando to establish a professional network and get involved in some interesting projects later on, for example, the Learning Games Initiative Research Archive, a video games archive at the University of Arizona.

Currently, Dr. Rios is the Research Data Management Specialist in the Research Engagement department at the University of Arizona Libraries. He provides support for data management planning and strategy, research workflows, and training around the topics of data and software curation. He is actively involved in research data policy and governance work. He also manages the University of Arizona Research Data Repository (ReDATA) that was launched several years ago.

- https://fernandorios.net/.
- Rios, F., & Ly, C. (2021). Implementing and managing a data curation workflow in the cloud. *Journal of eScience Librarianship*, *10*(3). doi: https://doi.org/10.7191/jeslib.2021.1205.
- Rios, F., Lassere, M., Ruggill, J., & McAllister, K. (2020). Sustaining software preservation efforts through use and communities of practice. *International Digital Curation Conference.*

http://www.ijdc.net/article/view/696.

Mara Sedlins

Mara graduated from University of Washington in 2012 with a degree in social psychology. Her dissertation "The Automatic Social Categorization Test: Validating a New Measure" focused on how people group and distinguish faces across various social categories, such as age, race, or gender. After graduation, Dr. Sedlins was looking for positions that would allow her to maintain a good work-life balance and to broaden the scope of her work beyond her immediate PhD area. In 2016 she landed a CLIR data curation fellowship at Duke University and held a joint position between Duke Libraries and the Social Science Research Institute.

One of the main goals of Mara's postdoc position was encouraging and facilitating collaborative interactions between the library's data visualization services and the Social Science Research Institute at Duke and get the full range of data management services off the ground. She gathered information about researchers' data management practices, helped establish workflows for a new digital repository, and contributed to many other data access and analysis projects. At the same time, she continued growing her own expertise in the fast-changing field of data curation. Her background in social science research helped to understand what researchers and graduate students are going through.

After the postdoc, she joined Colorado State University as Data Management Specialist, focusing on preparing datasets for the institutional repository and providing data management instruction. She has also organized introductory coding workshops in collaboration with the Statistics department. She led a reassessment of the institutional repository's data platform, which resulted in a decision to join Dryad as an institutional member. After a reorganization, her position became part of the new Research Support and Open Scholarship unit. In her current position, Mara continues her data curation work and organizes and promotes data and computing skills. She is also engaged with campus-wide efforts to support new funder requirements for data sharing and to promote open scholarship more generally.

- Meet Mara Sedlins, Innovating Minds Speaker. https://youtu.be/f8h7nb5fwp0.

- Coding and Cookies: Automating data cleaning and analysis using R https://libguides.colostate.edu/coding-cookies.
- Boice, J., Sedlins, M. & Sharp, J. L. (2023). R Workshops for Researchers: A Successful Partnership Between a Library and a Statistical Consulting Laboratory. *Journal of eScience Librarianship* 12(2), e647. doi: https://doi.org/10.7191/jeslib.647.

Rachel Starry

Dr. Rachel Starry completed her PhD in Classical & Near Eastern Archaeology at Bryn Mawr College in 2018. For her dissertation titled *"Finding the Local within the Global: A Comparative Study of Public Architecture and Urban Development in Roman-period Lycia"* she studied local urban landscapes and architecture, focusing on southwestern Turkey. She worked with inscriptions about public monuments and buildings, site plans, and architectural remains, and created a database and visualizations to compare urban growth and pedestrian experiences across the region. This work provided Dr. Starry with experience in interdisciplinary digital scholarship and as a next step, she decided to look for career opportunities that would go beyond a single academic department. An academic library position with its multidisciplinary orientation aligned with her goals of helping a broad range of academics conduct their research effectively, and she took a CLIR fellowship at the University at Buffalo Libraries.

During her fellowship in Social Science Data Curation at the University at Buffalo (SUNY) in 2018-2020, Rachel worked closely with another postdoctoral fellow developing integrated services for digital scholarship and research data support. Through extensive research and collaborations, she laid the groundwork for the new Digital Scholarship Studio and Network at the University of Buffalo, a center that provides ongoing assistance for faculty and students who are developing digital projects. She also organized data management workshops and worked with faculty to raise awareness of related issues.

After taking a position as the Digital Scholarship Librarian in the Research Services department of the University of California, Riverside Library, Dr. Starry became Head of Digital Scholarship Services at the

University of Pittsburgh Library System. There, she leads a team of librarians and staff in supporting the digital research, teaching, and publishing needs of the Pitt community.

- https://rachelstarry.org/.
- https://github.com/rachelstarry.
- "Geographies of Engaged Digital Scholarship: Remaking Space and Place in the Academic Library." (2021). https://futures.clir.org/space-and-place/.