NATIONAL LEADERSHIP GRANTS FOR LIBRARIES

Sample Application LG-259718-OLS-26
Project Category: Planning

University of Arizona College of Information Science

Amount awarded by IMLS:	$199,450
Amount of cost share:	$0

The College of Information Science of the University of Arizona will develop a plan to support the curation of data and metadata related to physical samples such as rocks, fossils, biological specimens, and water samples, etc. The project will develop guidelines, training modules, and data management software to enable this work. It will conduct multiple needs analysis activities, including surveys of research data services librarians and sample user/curators, follow-up focus groups, and hands-on curation case studies with project partners. The project seeks to build community through multiple workshops with data curation librarians, sample users, and infrastructure developers. The effort will assess how data curation librarians can advance open science and data reuse for in natural science fields that rely on physical samples.

Attached are the following components excerpted from the original application.

- Narrative
- Schedule of completion

When preparing an application for the next deadline, be sure to follow the instructions in the most recent Notice of Funding Opportunity for the grant program to which you are applying.

Hidden treasures: Leveraging research data services units in libraries to preserve our nation's academic sample collections

Introduction

The University of Arizona College of Information Science, in collaboration with the University of Arizona Libraries and project partners at University of Illinois at Urbana-Champaign and University of Minnesota propose a National Leadership Planning Grant to assess how data curation librarians can enable open science and data reuse for disciplines that rely on physical samples from the natural sciences (e.g. rock samples, fossils, biological specimens, water samples, and more). University libraries and data curation librarians are increasingly being asked to support the curation of data and metadata related to physical samples, but there are few guidelines, training modules, or data management softwares available to support their work. This leaves librarians unable to support their patrons, and the users of scientific collections without access to crucial materials for research in many scientific disciplines, including Earth sciences, material sciences, agriculture, archaeology, and biomedicine. To this end, we request $199,450 from IMLS to conduct needs analysis, solidify partnerships, pilot workflows, and identify concrete steps for further action. We will conduct multiple needs analysis activities, including surveys of research data services librarians and sample user/curators, follow-up focus groups, and four hands-on curation case studies with project partners. We will build community through multiple workshops with data curation librarians, sample users, and infrastructure developers. This work will be synthesized into a roadmap for future work at a symposium at the project's end. This work will meet NLG-L program objective 1.3: *Provide broad access to and preservation of information and collections through libraries and archives*.

Project Justification

Academic libraries and research data services units play a crucial role in supporting data-intensive scholarship and enabling open science (e.g. Andrikopoulou et al., 2022; Ho et al., 2026; Liu & Liu, 2023; Scotti et al., 2025). Data curation librarians educate researchers on data sharing policies and best practices; they work directly with researchers to format, archive, and share the results of federally-funded research; and they work with administrators to set data sharing policy across the university. Research data services units have been particularly important in supporting researchers in the development and implementation of data management plans that are now required by many federal and private funders (e.g. Diekema et al., 2014; Yoon & Schultz, 2017).

However, **a significant group of researchers - those who collect and analyze physical samples from the natural sciences and produce related data and metadata - increasingly need the support of research data services units, but research data services are not yet fully equipped to provide this support.** Physical samples include fossils, rock samples, ice cores, archaeological artifacts, water samples, taxidermized animals, tissue samples, and more. These objects are foundational to scholarship in many fields, ranging from archaeology to geology to medicine to zoology. Universities are the home of untold numbers of samples, both in hundreds of curated research collections and in the uncurated personal collections of thousands of geologists, environmental scientists, and other natural scientists. These samples are valuable not only because they are expensive to collect and often rare, but also because they can be reanalyzed with new technologies long after their initial collection (**Figure 1**); indeed, many of the natural sciences, such as geology or paleontology, are fundamentally grounded in the reuse of existing sample/specimen collections (e.g. Frieman & Janz, 2018; Wyse Jackson, 1999).

Despite their great value, though, the curation and sharing of sample-related data and metadata have lagged behind data sharing efforts in the library sciences more generally. While there are many world-class curated sample collections in universities (e.g. natural history museums, geological core repositories), their collection management infrastructure has historically not been as robust as that in libraries (Thomer & Rayburn 2023, Thomer et al., 2018); further, curation of natural science samples is rarely guided by the regulatory landscape that structures the curation of samples in the biomedical field, where health privacy protection laws necessitate highly systematic and regulated curation and data sharing protocols. As a result, the ability to search for samples across collections is inconsistent and variable, and infrastructure development efforts have been piecemeal across fields. While there have been efforts to create a "union catalog" of samples (akin to WorldCat) in which one can search across many collections at once, there are no truly universal search portals for samples; rather there are siloed portals for individual, curated, natural history collections (e.g. Jackson School of Geosciences; Smithsonian National Museum of Natural History, Department of Anthropology; Virginia Institute of Marine Science), or, at best, aggregations of records for specific types of samples (e.g.

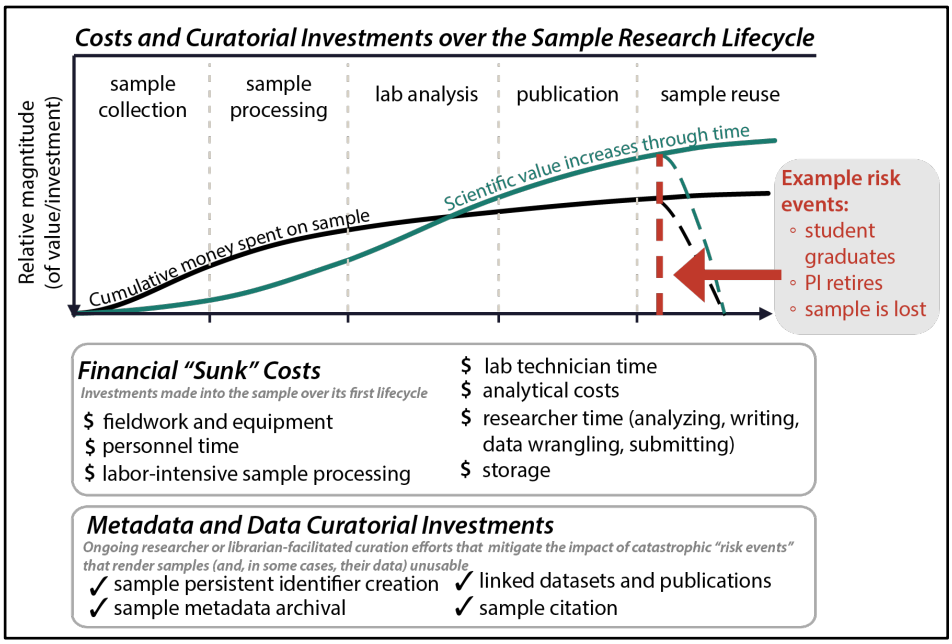


Figure 1. A graphic showing the progressive cost invested in a sample, compared with a sample's accrued value over time as more information about the sample accumulates. A typical sample collection trip in the paleosciences costs at least $3000 for travel and logistics (Thomer et al., 2025; Dowding et al,. 2026); one day of collecting samples from the ocean floor costs upwards of $10,000 in research vessel operating costs alone (UNESCO & Intergovernmental Oceanographic Commission, 2017); researcher salary and data analysis costs add tens of thousands to each of these sums.

www.iDigBio.org, an aggregation of zoological samples in the US). Additionally, university sample collections often have large collections of related digital objects, like 3D scans or photographs, but little infrastructure or expertise to manage these digital objects. Finally, in addition to curated collections, universities are also the home of thousands of uncurated "personal" collections sitting in scientists' labs, basements, and closets; these "rocks-in-closets" are scientifically valuable but are essentially invisible to potential reusers. Curated collections usually don't have capacity to accession personal collections, and scientists often don't know how to share metadata describing their personal collections.

There are now urgent reasons to move beyond these "silos of the [university] LAMs" (Zorich et al, 2008) and forge connections between academic libraries (specifically, research data services (RDS)[1] units and data curation librarians), curated natural science collections, and the individual researchers that work with, collect, and sometimes curate their own sample collections. First, RDS units are increasingly being approached by researchers with samples for support in curating data and metadata; this will only increase given new guidance from some NSF directorates that direct researchers to prepare sample management plans (in addition to data management plans) (Division on Earth and Life Studies & National Academies of Sciences, Engineering, and Medicine, 2020; Holdren, 2014; National Science and Technology Council, Committee on Science, Interagency Working Group on Scientific Collections, 2009; National Science Foundation (NSF) Office of Polar Programs, 2022). Second, due to funding and programmatic constraints limiting the new collection of samples (e.g. Announcement of the Non-Renewal of the JOIDES Resolution Operations and Maintenance Cooperative Agreement, 2023), scientists are now more reliant on reusing specimens, and therefore more in need of ways to easily search for samples online; RDS units and data curation librarians will prove critical in enabling the sharing and discovery of sample metadata. Finally, many existing data sharing initiatives and infrastructures within the sample-based sciences are facing a renewed need to build interdisciplinary interoperability and cross-institutional collaborations to improve their resilience and sustainability amidst uncertain funding landscapes and changing technologies; in short, they simply need librarians' help.

Thus, we propose a planning grant to do the following:

- We will build and solidify partnerships between data curation librarians, natural science sample users, natural science sample curators, and infrastructure developers through webinars, workshops, and a hybrid symposium.
- Through surveys and focus groups, we will analyze data curation needs in each of these stakeholder groups, and assess to what degree research data services units could practically support sample curation.
- Through case studies, we will develop and pilot training materials on sample metadata curation aimed at data curation librarians following the Data Curation Network's model of 'curation primers'.

[1] See Supporting Document 2 for a glossary of frequently referenced organizations, acronyms, and concepts.

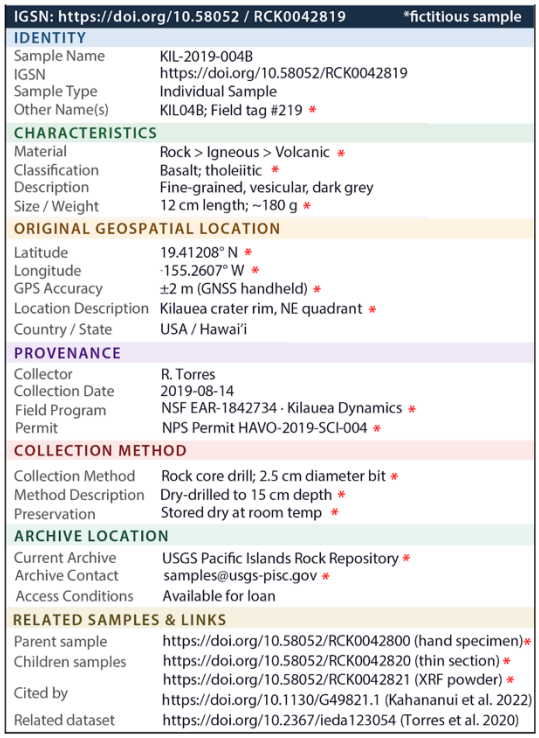
- Through a hybrid symposium, we will synthesize the results of all these activities, and plan for future implementation or research activities. We will particularly focus on identifying points where tools or connections between existing infrastructures need to be developed, and areas where larger scale training or research initiatives are needed.

In the subsections that follow, we further describe the need for this planning grant; why sample metadata curation is under supported by existing best practices, tools and resources; and how librarians, sample users and others will benefit from this project.

The need for this planning grant: prior work in sample metadata curation. The PIs of this proposed project, Andrea Thomer and Natalie Raia, are experts in both in digital curation and in physical sample metadata curation, and identified the need for this Planning Grant through years of community building, infrastructure development, and research in digital curation. Thomer is an Associate Professor at the UA College of Information Science; her teaching focuses on library and information science (LIS) topics such as information organization and digital curation, and her research focuses on digital curation and the maintenance of knowledge infrastructures (e.g. Thomer, 2017; Thomer et al., 2018; Thomer, 2022; Thomer et al., 2022). She has also worked in natural science infrastructure development for almost 20 years. Raia is a geologist-turned-information scientist who leads multiple projects at the intersection of research data management, cyberinfrastructure development, and applied research practice, with a focus on sample metadata curation. She currently serves on the Board of Directors for the American Geophysical Union. Both Thomer and Raia are on the leadership team of two major natural science data infrastructure projects: the System for Earth and Extraterrestrial Sample Registration (SESAR; www.geosamples.org) and iSamples, an experimental, interdisciplinary sample discovery portal. Thomer and Raia have also co-chaired the Physical Sample Curation Cluster at the Earth Science Information Partners (ESIP) for over two years, and as part of that work helped lead a major effort to develop sample citation guidelines. As a team, we have led dozens of workshops, conference sessions, webinars, and "data help desk" sessions on the topic of physical sample metadata curation and sharing. We are active in numerous LIS and natural science communities, including ESIP, the Research Data Alliance, the Geoscience Information Society, the Association of Information Science & Technology, the Research Data Access and Preservation (RDAP) community, the American Geophysical Union, the Society for the Preservation of Natural History Collections and the PaleoData Working Group.

Over the last two years, we have seen increased demand from librarians for training and support in sample metadata curation – as well as a growing need for collaboration between natural science data infrastructures and the library community (Borda, 2020). For instance, at our monthly meetings of the ESIP Physical Sample Cluster, there has been an uptick in attendance by librarians at a range of institutions, looking for ways to support their patrons in their sample data management. In February 2026 we presented a webinar to the RDAP community on working with SESAR (the Earth sample data management platform we both develop); over 110 data curation librarians and data managers attended – so many, in fact, that the "Zoom room" reached capacity and folks had to wait to be let in (Thomer & Raia, 2026). This clearly points to a demand for librarian-focused best practices for working with sample metadata.

Similarly, we have observed a growing need for librarians-in-the-loop in sample metadata infrastructure projects. During a standing-room-only session we convened at the Society for the Preservation of Natural History Collections in May 2025, we heard from sample collections managers who were eager to work with librarians to improve a longstanding problem: how to properly cite sample identifiers and acknowledge collections in their publications, making it difficult to demonstrate the value of sample curation. We have also observed that scientific data infrastructure communities are starting to have conversations about sustainability, resilience and digital preservation that have been central to LIS research and practice for decades. Many projects (including some of our own, e.g. SESAR) are suffering from aging codebases that threaten their continued operations, and the federal funding landscape has become uncertain – with some important natural science data resources losing funding altogether. For example, the Index of Marine and Lacustrine Geological Samples – a vital search portal for researchers in search of ocean and geologic sediment cores to reuse – was decommissioned by the National Oceanic and Atmospheric Administration in early 2025. Funding for iDigBio, an NSF-funded project that shares sample records from US zoological collections, will end this year with no clear plans for renewal. These funding cuts have inspired multiple data rescue efforts, and renewed energy to forge partnerships across disciplines and infrastructures. We believe that input and collaboration from librarians and research data services units is badly needed by these "at risk" sample infrastructure communities; sample metadata infrastructure developers should be building on prior work in LIS rather than reinventing the wheel.



IGSN: https://doi.org/10.58052 / RCK0042819	*fictitious sample
IDENTITY	
Sample Name	KIL-2019-004B
IGSN	https://doi.org/10.58052/RCK0042819
Sample Type	Individual Sample
Other Name(s)	KIL04B; Field tag #219 *
CHARACTERISTICS	
Material	Rock > Igneous > Volcanic *
Classification	Basalt; tholeiitic *
Description	Fine-grained, vesicular, dark grey
Size / Weight	12 cm length; ~180 g *
ORIGINAL GEOSPATIAL LOCATION	
Latitude	19.41208° N *
Longitude	-155.2607° W *
GPS Accuracy	±2 m (GNSS handheld) *
Location Description	Kilauea crater rim, NE quadrant *
Country / State	USA / Hawai'i
PROVENANCE	
Collector	R. Torres
Collection Date	2019-08-14
Field Program	NSF EAR-1842734 · Kilauea Dynamics *
Permit	NPS Permit HAVO-2019-SCI-004 *
COLLECTION METHOD	
Collection Method	Rock core drill; 2.5 cm diameter bit *
Method Description	Dry-drilled to 15 cm depth *
Preservation	Stored dry at room temp *
ARCHIVE LOCATION	
Current Archive	USGS Pacific Islands Rock Repository *
Archive Contact	samples@usgs-pisc.gov *
Access Conditions	Available for loan
RELATED SAMPLES & LINKS	
Parent sample	https://doi.org/10.58052/RCK0042800 (hand specimen)*
Children samples	https://doi.org/10.58052/RCK0042820 (thin section) *
Cited by	https://doi.org/10.58052/RCK0042821 (XRF powder) * https://doi.org/10.1130/G49821.1 (Kahananui et al. 2022)
Related dataset	https://doi.org/10.2367/ieda123054 (Torres et al. 2020)

Figure 2. An example scientific sample metadata record showing the types of information needed to ensure a sample is fully documented and reusable. Red asterisks denote metadata fields that do not map well to "generic" standards like Dublin Core.

We've also noted a lack of systematic research on the curation of physical sample metadata. Though there has been significant LIS research on the data curation needs of fields that *use* samples (e.g. archaeology, as studied by Faniel et al., 2021; Frank et al., 2015; or ecology, as studied by Zimmerman, 2008), there has been less research on the specific data curation needs *for* sample metadata and sample digital collections. Some notable exceptions include work by Ramdeen on the information seeking behavior of sample users (Ramdeen, 2017), Buchanan's ethnographic studies of archaeology curators (Buchanan, 2016), Darch's work on the trustworthiness of samples (Darch, 2022), Huvila and colleagues' studies of sample provenance (Huvila et al., 2022, 2024), and our own team's on-going work on sample curation and citation (Damerow et al., 2025; Raia et al., 2023; Richard et al., in review; Thomer, 2022). Additionally, there has been significant development of metadata standards by domain researchers and infrastructure developers (e.g. Hardisty et al., 2021; Klump et al., 2021; Walls et al., 2014; Wieczorek et al., 2012). Further research is needed to identify priorities for tool and best practices development. As an applied field we need research to inform our practice.

Why sample metadata and data curation is under-supported by existing data management resources and tools. We are not suggesting that all data curation librarians need to become experts in sample metadata curation. Rather, we want to build on prior approaches that have enabled data curation librarians to work with broad ranges of diverse types of data. For example the Data Curation Network is a community of data curation librarians designed to leverage the varied expertise of individual members; they draw on a "shared-curation workflow, in which datasets from one institution are matched with an expert at a different member institution" ("Better Together: The Data Curation Network," n.d.; Johnston et al., 2018). The Data Curation Network additionally facilitates the creation and peer-review of Data Curation Primers that support data curation librarians in work with domain-specific data types and practices. Approaches like these make it possible for data curation librarians to apply their expertise to a broad range of data. We are particularly interested in developing best practices or training that would help librarians take advantage of existing data infrastructures in the sample-based sciences. There are databases like our project SESAR and Open Context (an archaeology data sharing platform) that sample users and librarians could use for free, but guidelines and training targeted specifically at data curation librarians are needed.

Any new Primers, training, best practices or tools for sample metadata curation will need to account for the ways in which sample metadata curation differs from other kinds of digital curation. First, samples need to be described with specific kinds of metadata to be finable and reusable; commonly used, "generic" metadata standards like Dublin Core often don't represent these details well. Sample metadata records must include detailed information about each sample's characteristics, originally geospatial location, provenance, collection methods, and archive location (**Figure 2**); without this information both the record *and* the sample itself risk becoming unusable. Though sample metadata requirements can be highly discipline specific, there is a "core" set of metadata fields that are more-or-less common across samples; for example, the iSamples project (Thomer was a PI on this project and Raia was a postdoctoral fellow) developed a metadata standard, which proposes a common set of terms to describe samples from many fields Davies et al., 2021). These core metadata elements should ideally be paired with discipline-specific metadata (e.g. Darwin Core, a descriptive metadata standard for zoological samples) to maximize reusability.

Second, samples have different, sometimes more complicated needs regarding *identifiers,* that is, alphanumeric codes, like catalog numbers, used to refer to an object. In most, if not all, digital repositories, digital objects are assigned unique, *persistent* identifiers (PIDs), such as a digital object identifier (DOI; ISO, 2025) or an archival resource key (ARK; Kunze et al., 2025), which are universally unique and make it possible to retrieve these objects and/or their metadata quickly and persistently over time, and help mitigate the hazards of "link rot" (Ashenfelder 2011). Like articles and datasets, samples

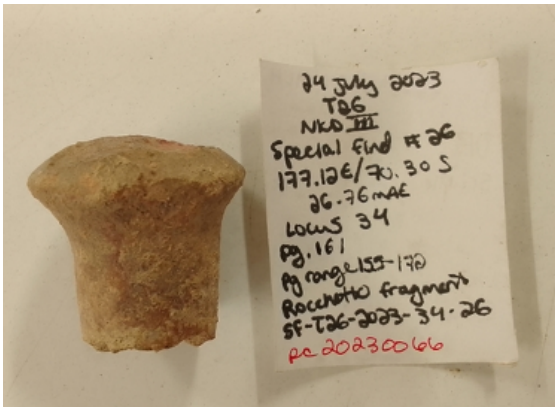


Figure 3. An example of the many identifiers that samples can accrue. This object is a spool fragment from an archaeological dig in Poggio Civitate, Italy. "Special find #26" is an informal field number used by the student that found this artifact and "SF-T26-2023-34-26" is a longer form of this number including the date; "pc 20230066" is the number the artifact was given once eventually cataloged. Image provided by Eric Kansa.

should each receive their own identifier to support unambiguous and persistent identification over time. However, best practices are still emerging regarding what kind of PID to use for samples, and how to use them. Many curated collections still prefer to use catalog numbers rather than universally unique (but longer and more opaque) PIDs, and researchers with uncurated personal collections often use *ad hoc* field or lab numbers which may or may not be recorded into a catalog. Because these identifiers are not universally unique (and because sample citation practices are highly varied (Damerow et al., 2025), there can be significant ambiguity over what sample was being referred to, making sample-based science less reproducible, and sample curation more difficult. Further, samples are known to "accumulate" identifiers over time (Damerow et al., 2021; Guralnick et al., 2015; Klump et al., 2021); researchers will often assign numbers in the field, then new numbers when processed in a lab, and new numbers still when a sample is finally cataloged (**Figure 3**). Any one of these numbers might be used in a publication, causing further ambiguity. Efforts to establish PIDs specifically for samples have seen growing success – for instance, the popular International Generic[2] Sample Number (IGSN) system, was developed for use with Earth and environmental samples (Klump et al., 2021; Klump & Huber, 2017; Lehnert et al., 2005, 2011). Other fields like archaeology and biodiversity have adopted ARKs (e.g. Koster, 2020; Kunze & Rodgers, 2008). But practices are still highly diverse across sample-using fields, making it challenging to develop best practices, let alone infrastructure that works across many sample types. Sample users and curators need guidance in choosing the best PID for their particular projects and contexts.

Samples also have a greater need to track their relationships to other samples and related research outputs. Many sample users break samples into multiple "sub-samples" or "child samples," and these relationships need to be tracked over time for the purposes of collection curation and scientific reproducibility. For instance, a researcher might collect a rock sample in the field and later chip off a portion to powder for chemical analyses, or slice off another portion to examine under a microscope. Many existing databases for sample metadata display these related identifiers as a "sample family." This makes it possible to look up one sample and see if there are subsamples available for reuse, so that expensive lab preparations can be avoided. Additionally, it's important to "link" samples to any related datasets or publications; many sample users want to find all samples associated with a particular study rather than searching for samples that match other metadata (Raia & Thomer, 2025). Some samples may have associated images like photographs or 3D scans; these too need to be "linked" to the sample's identifier so that they can be found easily in search. Thus, curation of sample metadata entails significant curation of information about relationships.

Finally, as LIS scholar Peter Darch has noted, "samples are frequently invisible to users of digital data curation infrastructure" (Darch, 2022); data that has been derived from samples often do not list the identifiers of those samples. Some disciplinary-specific data archives (e.g. the EarthChem Databases, a repository for geochemical data that is derived from samples) have tried to make links to related samples more prominent by displaying them on dataset landing pages; however, not all systems do this. Including sample PIDs in metadata of related papers and datasets is important for enabling reproducibility and discovery of samples. It's also important for enabling sample curators to "track" the use of their samples over time and thereby show the impact of collections and curation.

There are many potential ways to bridge some of these gaps and better support sample metadata curation (**Figure 4)**, but coordination and planning between libraries, curated sample collections, and sample users is an important first step. In this planning grant, we propose the coordination, needs assessment and pilot work needed between these stakeholder groups to identify what best practices and tools are most needed to bring librarians into the sample curation loop. These best practices might include guidance on selecting and using sample metadata standards, or in selecting and using appropriate

[2] Formerly the International *Geological* Sample Number

PIDs. They could also include decision trees to help select appropriate disciplinary repositories for sample metadata or data.

Who will benefit from this project. This Planning Grant will support the following groups:

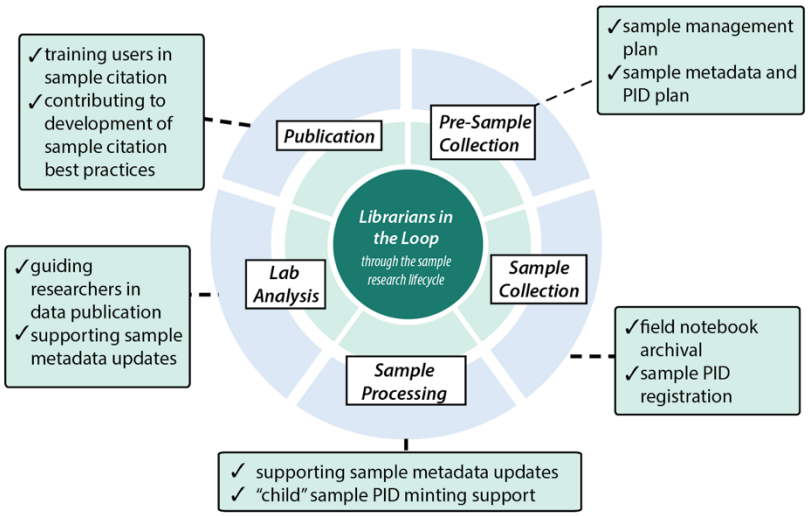


Figure 4. An illustration of some of the ways that data curation librarians might support sample users.

Data curation librarians and research data services unit in academic libraries. By conducting critical needs assessment of the data curation librarians and research data services units that are increasingly being asked to support sample metadata curation, we will better understand the requests they are receiving, their training and best practices needs, and their infrastructural constraints. This planning grant will lead to the development of future projects that will allow them to deliver more holistic and comprehensive guidance to researchers throughout the research data lifecycle.

Sample users and curators. As physical samples are increasingly held to higher standards for digital discoverability, including the usage of persistent identifiers and standardized metadata, researchers and sample curators increasingly need the expertise of librarians. This work will enable sample reuse and therefore enable further scientific discoveries, especially for the many researchers and students who do not have the funding or ability to travel needed to do field work – or are physically unable to do fieldwork – and need to rely on others' samples for their research.

The nation at large: US federal funding agencies have invested, conservatively, more than a billion dollars in the collection and study of physical material samples in the 20th and 21st century (e.g. National Academies of Science Engineering, and Medicine, 2024; The Planetary Society, n.d). Some samples represent locations or materials that can no longer be sampled either because they represent an ephemeral and discrete moment in time (e.g., water samples) or because the location or material may no longer exist (e.g., old mine sites, archaeological sites). Further, some sites are too ecologically, culturally, or politically sensitive for repeated sampling efforts (Miller, 2021); sample reuse is needed to protect cultural and natural resources. This grant will additionally have benefits to other funding agencies by laying the groundwork to help applicants meet the requirements for sample sharing, as well as better sample stewardship in general.

This grant will take the necessary and overdue first steps to bring libraries into the loop to help researchers make samples and sample-related research digitally findable. This work plans for a future in which the expertise and services of existing university research data services are better equipped to support the end-to-end sample research lifecycle of researchers across disciplines, thus ensuring our federal investments in this type of research are maximized. With researcher workflows and research data support services brought into tighter alignment, our nation's "hidden treasures" can be discovered, re-discovered, and re-used for new questions and societal challenges. Existing samples can be used to examine new national sources of critical minerals, enable new historical insights into our nation's heritage as told through the archaeological record, and more.

Project Work Plan

This Planning Grant will build partnerships between data curation librarians, natural science sample users, natural science sample curators, and infrastructure developers. We will conduct needs assessments and plan for the development of training modules, best practices, and tools for sample metadata curation in RDS settings. These plans will be informed by pilot sample curation activities with project partners at the University of Illinois at Urbana Champaign (UIUC), the University of Minnesota (UMN), and two additional universities to be determined after the launch of the grant. Below we further describe our project team, our activities in further detail, and our evaluation strategy.

Our team. The core team includes PI Thomer, co-PI Raia, and senior personnel Fernando Rios. Thomer and Raia's bios were presented on page 3. Rios is an Associate Specialist in Research Data Management at UA libraries and has familiarity with RDAP and the Data Curation Network (DCN), including prior experience with developing Primers for practitioners, and offers an important generalist perspective on data curation. Our PI team brings expertise in data curation, sample use and curation, and infrastructure development

Thomer will be the primary director of the project and will coordinate with project partners. Raia will lead stakeholder needs analysis, coordinate the Year 2 symposium, and co-oversee student interns. Rios will lead development of a DCN Primer, and co-oversee student interns along with Raia. Student interns (one per project year) will be recruited from the College of Information Science's LIS master's program or PhD program and will contribute to stakeholder analysis and DCN Primer development (detailed further below).

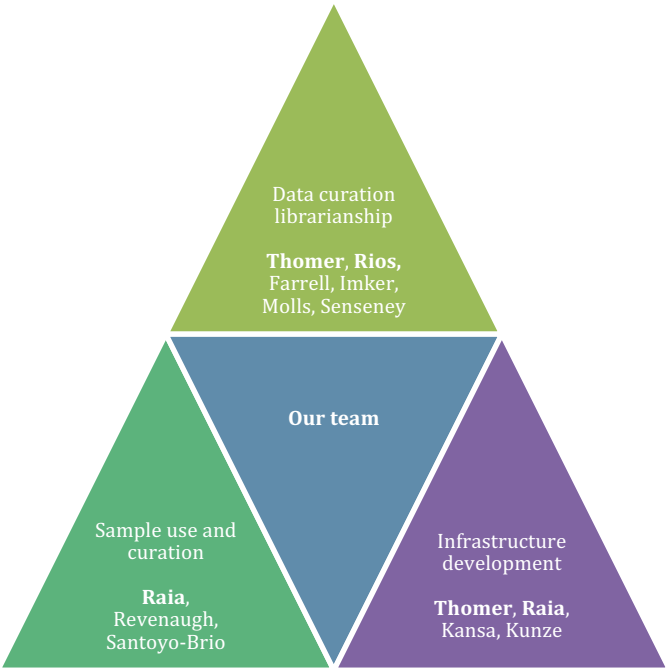


Figure 5. Areas of expertise represented by our project team. Names of PI's and senior personnel are bolded; other names represent current advisory board members (with one more infrastructure expert to be named).

Throughout the grant, we will partner with four teams of librarians and sample users from other universities to deepen needs assessment and pilot sample curation workflows. Our initial partners include teams at UIUC and UMN. Two more teams will be recruited through virtual workshops at annual RDAP and ESIP meetings. The work of these project partners, and the rest of our project activities, are described further below. Note that many of these activities will be carried out concurrently; please see the schedule of work for more detail on sequencing.

Project activities. The proposed work builds on years of Thomer and Raia's efforts in the Earth science data curation community; these activities have been developed in coordination with our project partners and advisory board members.

Activity 1: Librarian and sample user needs assessment. As outlined above, there is demonstrated need and interest from data curation librarians for training in sample metadata curation. There are also clear indications that there are numerous un- and under-curated collections in US academic institutions. However, further systematic assessment is needed of the scope of these needs. We will conduct needs analysis via surveys and focus groups to a) better understand university libraries' needs for training and readiness to support sample-based researchers in their metadata curation, and to b) understand the scope of un- and under-curated samples in US institutions. Details are as follows:

- *Survey one: data curation librarian needs assessment*. We will conduct a survey of university data curation librarians to assess how often they are asked to curate sample metadata, and to gauge their experience with sample metadata. We will ask them to what degree they work with researchers to curate sample metadata; their awareness of PIDs for samples; their awareness of best practices for curating sample metadata; and what types of sample-related data they think their researchers have. If they have an institutional data repository at their university, we will ask them what kinds of data this repository holds, and if they have any special support for people archiving sample-based research. Finally, we will ask them what kinds of training and best practices development they think is most needed to support sample metadata curation.
- *Survey two: sample user and curator needs assessment.* We will conduct a complementary survey of faculty members, research scientists, and curators within natural science departments to better understand the scope and nature of uncurated physical collections in universities. We will ask what kinds of samples they collect, where they are stored or deposited, and how sample metadata is shared. We will also ask respondents about their relationships with and awareness of their university libraries and RDS units. We will ask sample curators about their data management infrastructure and needs.

- *Deeper needs assessment via focus groups:* The surveys will give us a high-level picture of the sample metadata curation landscape, but deeper conversation and qualitative research will be needed to better understand the needs for and barriers to sample metadata data curation faced by librarians, sample users, and sample curators. Surveys will be used to recruit participants for hour-long focus groups (~3 groups of ~5 sample users/curators and ~3 groups of ~5 librarians; each participant will be offered a $50 gift card for participation); this activity will build off the surveys, provide a richer understanding of barriers to sample sharing, and aid in building consensus around ideas (Hall et al., 2019). We will use the nominal group technique to facilitate these focus groups; this technique is intended to bring together experts to brainstorm and prioritize ideas for future work (McMillan et al., 2016; Hall et al., 2019). We expect that barriers might include a need for departmental support for data curation and more training for librarians and institutional data repository staff. We hope that focus groups will surface innovative ways that researchers and librarians have already started tackling the problem of sample metadata curation within their institutions.

Results from these activities will be summarized in a white paper so we can quickly disseminate results to stakeholder groups and build on this work in other activities. We will refine the white paper into a peer reviewed publication

Activity 2: Piloting curation workflows and analyzing feasibility of different curatorial approaches. Our second group of activities focuses on piloting sample metadata curation workflows. We will first conduct four case studies of sample metadata curation and then draft a sample curation "Primer" based on the results of this work.

- *Sample curation case studies.* In Year 1, we will work with teams at partner institutions UIUC and UMN; each team will include a data curation librarian, a sample user or curator, and a student intern. The leads of each team will join the Curation Advisory Board. The UIUC team includes Enrique Santoyo-Brito, Curator at the Illinois Natural History Survey (INHS) and a TBN curator from the University of Illinois Data Bank. They will work together to assign PIDs to specimens in the INHS collections, and to link them to related data in UIUC's institutional data repository. The UMN team includes Shannon Farrell, Research Data Services Lead and Director of DRUM, University of Minnesota's institutional data repository, and Justin Revenaugh, Department Head of the UMN Department Earth and Environmental Sciences. The UMN team will curate "career" collections of individual PIs' and legacy department reference collections. Student interns will be recruited by the curation teams from their local community but paid by the University of Arizona. We will recruit an additional two curation teams from participants in focus groups (Activity 1) and webinars (Activity 3) to conduct additional case studies in Year 2. We will recruit teams that work on different types of curatorial problems and samples (e.g. archaeological samples, environmental samples, etc.).

  Thomer and Raia will meet with project teams regularly through their work and will document their successes and challenges. Each team's work will be written up (by Thomer, Raia, and team members) as a sample curation case study, which will be used to inform further planning and best practices guidelines. Case studies will be shared as part of a final white paper documenting the results of the project overall.
- *Piloting sample metadata curation Primer.* Building on the case studies, we will draft a sample curation "Primer" to be reviewed and eventually published through the Data Curation Network. Primers are, "peer-reviewed, living documents that detail a specific subject, disciplinary area or curation task and that can be used as a reference to curate research data" ("Data Curation Primers," n.d.). The DCN manages a collection of open-source providers, which anyone can contribute to, in a centralized repository. We intend this Primer as a pilot for the development of similar guidelines in the future. It is unlikely that we can account for all sample metadata curation needs in one document but drafting one following the DCN's popular and successful approach will allow us to assess the feasibility of this approach for the future. We will pick one discrete aspect of sample metadata curation to develop a Primer on, based on our needs assessment in other activities; for instance, we may develop a Primer on selecting appropriate PIDs for samples, or one on creating sufficient sample metadata – or something else entirely, depending on insights shared from our case study participants. Rios will lead this work; working with UA student interns, and consulting with Thomer, Raia, and the Curation Advisory Board, he will review case study write ups and the results of needs assessment activities to identify common curatorial needs. Focusing our efforts on creating a DCN Primer ensures that our work will be adaptable, generalizable, and usable by other institutions, as well as sustainable after the project's end. We will be able to tap into an already robust practitioner network and thereby greatly increase our chances of adoption.

Results from the sample case studies will be written up as short reports (co-authored with the curation teams) and shared via our institutional repository. We will refine these reports into a peer reviewed publication The DCN Primer draft will be submitted for review by the DCN by the end of the project.

Activity 3: Solidifying partnerships. Our final group of activities focuses on solidifying partnerships between librarians, institutional data repositories, disciplinary data repositories, and domain researchers and co-designing curation workflows informed by the outcomes from our case studies. One of the biggest obstacles to sharing sample metadata is that it requires significant interdisciplinary skill and knowledge. We will do the following to bridge communities, share knowledge, and identify pathways for future work:

- *Host workshops at virtual RDAP meetings to build community with data curation librarians.* RDAP is an important, active community for US-based data curation librarians. They run monthly webinars and annual meetings where practitioners and researchers come together to share their work on broad topics in digital curation. As described on page 3 we have previously run highly attended workshops on sample metadata curation at RDAP. We will build on our existing relationships with this community to develop solidifying partnerships between the data curation librarian community and the sample users and infrastructure developers we have pre-existing connections with. In Year 1, we will run an RDAP workshop on sample metadata curation to gauge practitioners' awareness and ability to support domain researchers in sample metadata sharing. We'll use small group activities to get a sense of what kinds of sample curation tasks participants have been asked to assist with, and use hands-on activities to begin applying existing best practices for sample curation. The results from this workshop will inform our future best practices development, and we'll use these workshops to recruit participants to other parts of the grant. In Year 2, Rios, Thomer and Raia will host a follow-on workshop at RDAP to involve the community in the development of the DCN Primer. Participants will be asked to review a draft of the Primer and provide feedback on whether it would support their curatorial needs. Again, we will use this workshop to recruit participants to other grant activities.
- *Host workshops at virtual ESIP meetings to build community with Earth science data managers and infrastructure developers.* ESIP is an interdisciplinary, international organization supporting Earth, space and environmental science data sharing, preservation and use. Members include data managers from federal agencies, data curation librarians, infrastructure developers, and academics. As previously mentioned, Thomer and Raia are co-chairs of the Physical Sample Curation Cluster, a working group focused on improving sample curation workflows and infrastructure. As such we are well positioned to build on this existing community in ESIP. Each project year, we will run sessions as part of ESIP's annual virtual meeting. The Year 1 ESIP workshop will complement the RDAP workshop; we will gauge participants' work with researchers and RDS units. The Year 2 ESIP workshop will similarly parallel the Year 2 RDAP workshop, and we will involve the ESIP community in the development of the DCN Primer. We will invite participants from RDAP workshops to also participate in the ESIP workshops, thereby building important connections between these two communities.
- *Sharing outputs and gathering feedback in other sample communities.* We will reach out to our extensive network of sample users, curators and developers to present our work and gather feedback throughout the course of this project. This includes groups like the PaleoData Working Group, the Geoscience Information Society, the International Partners of the Digital Extended Specimen, the IGSN advisory board and more.
- *Next Steps and Synthesis Symposium.* In Year 2, we will host a two-day, hybrid symposium at the University of Arizona to present results of project activities, and to create roadmaps for future work. We will invite thirty-five participants in total: 10 local participants, including Thomer, Raia, Rios, student interns, and other local domain researchers and library staff; 15 in-person external participants (including grant advisory board members and members of case study curation teams); and 10 virtual participants. The symposium will kick off with short presentations about earlier project activities, and then will focus on building roadmaps for future work. We will identify priorities for the development of sample curation best practices beyond the pilot DCN Primer, identify and prioritize needs for new tools or infrastructures, and identify opportunities for further collaboration beyond this initial Planning Grant. We particularly aim to identify projects appropriate for a future IMLS-L Applied Research or Implementation Grant, but we are also interested in brainstorming ways of formalizing ongoing collaborations between organizations like ESIP and RDAP.

Slides and tutorials from the workshops will be shared via UA's institutional repository. Thomer, Raia and Rios will synthesize notes and roadmaps from the Symposium into a white paper, also to be published in the institutional repository.

Evaluation and demonstration of impact. We will convene two advisory boards for this project; we will regularly consult with these advisory boards for feedback and evaluation. All members of the advisory boards will be invited to participate in the Year 2 Symposium.

- *The Curation Advisory Board* will include library, researcher and curator representatives from each of the Curation Case Studies at partner institutions (include University of Minnesota and University of Illinois at Urbana-Champaign to start, and two additional institutions to be determined after the grant convenes). Thomer, Raia and/or Rios will meet with the Curation Advisors biweekly (minimum) while they are actively working on their curation projects; we will also convene all Curation Advisors for quarterly meetings, where they will evaluate our work and provide feedback on priorities.
- *The Project Advisory Board* will include experts in sample infrastructure development and RDS more broadly. Our infrastructure experts include Eric Kansa (from Open Context, an archaeology data and sample publishing system) and John Kunze (founder of the ARK Alliance, a PID minting infrastructure that is used by many natural history museums). Kansa and Kunze were both senior personnel on the iSamples project (along with Thomer and Raia) and are experts on sample metadata curation and infrastructure development. We will recruit a third infrastructure expert for the board at the beginning of Year 1. Our RDS experts include Megan Senseney (Department Head of University of Arizona Libraries), Emma Molls (Director, Open Research and Publishing at UMN; Interim Director of the Data Curation Network), and Heidi Imker (Director of Research Data Service in the University Libraries, UIUC). These RDS experts have decades of experience in digital curation services. We will seek guidance from the board via quarterly meetings, where we will share project updates and seek advice.

Success of the project will additionally be measured through feedback surveys after RDAP and ESIP workshops and after the Year 2 Symposium. These surveys will ask to what extent participants felt comfortable participating in each activity; to what extent they felt they learned or contributed something valuable; to what extent they plan to stay engaged in our work going forward; and whether they have any suggestions for how we can improve our work in the future. Finally, we will also track rates of participation for each activity.

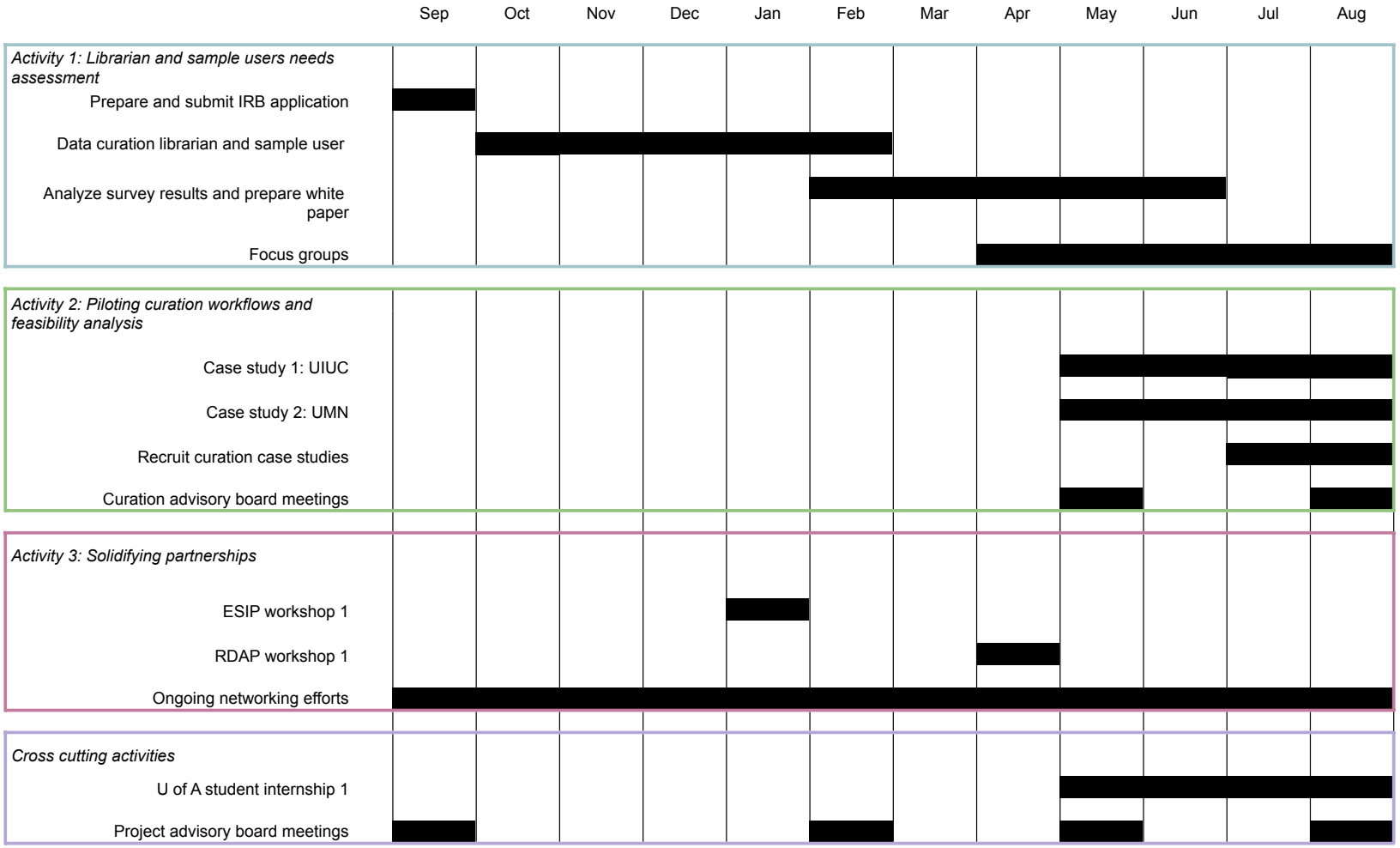
Project Results

In summary, the primary goals of this project are to conduct needs analysis, solidify partnerships, pilot workflows, and identify concrete steps for further action. We will produce several concrete deliverables (all openly published and freely available through UA's institutional repository), including:

- A white paper summarizing the results of needs assessment surveys and focus groups. This document will be used to identify priorities for best practices and tool development.
- Short reports presenting sample case studies (co-authored with the curation teams). The results of this work will further inform priorities for best practice and tool development, and constitute an important form of feasibility assessment. Through this hands-on curation work, we will gain real insight into the constraints and goals of real life RDS units and sample users and curators.
- A DCN Primer draft. The creation of this draft will demonstrate the suitability of this format for the development of future best practices; because we are contributing to an existing corpus of guidelines, we expect this work to be usable and adaptable by other institutions and communities.
- Slides and tutorials from ESIP and RDAP workshops. These will be adaptable by others running their own training workshops.
- A white paper summarizing the Year 2 Symposium, including roadmaps and plans for future work. These will build on all other project outputs to synthesize priorities for future work.
- Peer reviewed papers based on each of the white papers listed above, published in venues like the Journal of eScience Librarianship, the Data Science Journal, and the Journal of Information Science and Technology.

Perhaps most importantly, this grant will foster collaboration between data curation librarians, sample users, and sample curators. Project partners and student interns will gain valuable experience curating sample metadata, which will enable them to go back to their institutions and future jobs and improve their curatorial services. The needs assessment, feasibility analysis, and planning work this grant proposes will enable us to apply for future grants such as an NLG-L Implementation or Applied Research grant. We will leverage our existing leadership positions in the natural and information sciences to disseminate project results and lead future work.

Year One (9/1/2026 - 8/31/2027)



Year Two (9/1/2027 - 8/31/2027)