

LG-74-18-0173-18

Lewis & Clark College and The Aubrey R. Watzek Library

Data in the Disciplines

Project Abstract

Data intensive research is altering scholarly practices across a broad range of disciplines. Libraries are experimenting with ways that they can support research data management and data information literacy but have yet to find successful, tested models even at large research institutions. To meet the needs of their faculty and student populations, there is an opportunity for smaller institutions to take part in the experimentation around providing these services. The overall goals of this project are a) to advance the ways that small college libraries can support research data management and data information literacy, and b) to expand the kinds of research-based learning that are available to undergraduate students at small colleges.

In this project, Lewis & Clark College and four regional partner colleges (Reed College, University of Puget Sound, Whitman College, and Willamette University) will develop two 1.5 day workshops to provide faculty and students with data management and data information literacy support in two disciplinary domains, Chemistry and Ethnography. In the winter and spring of 2019, librarians from the partner colleges will design the workshops in collaboration with external consultants with specific domain expertise in Chemistry and Ethnography. Faculty, student research assistants, and librarians from the partner schools will participate in the workshops in the summer and fall of 2019. Following the workshops, the project team will assess the effectiveness of the workshops through surveys of workshop participants and widely disseminate both the curriculum produced at the workshops and all materials as well as lessons learned from putting on the workshops themselves.

The project will provide 1) A tested model for establishing regional collaborations among small colleges/university libraries to support data management and literacy in specific academic disciplines, 2) data management curriculum modules tailored to the particular curation needs and challenges of two disciplinary data communities, and 3) an outline of a one-hour data literacy curriculum in each discipline co-developed by librarians and faculty.

The project will advance the research data management skills and data information literacy knowledge of faculty and student populations at small colleges and universities by offering libraries at those institutions a viable model for supporting such advancement. The actual curricular modules in data management and data information literacy for Chemistry and Ethnography should be applicable in many undergraduate settings and will consequently touch students beyond the universe of small colleges and universities. This project and its expected outcomes also have the potential to serve as a model for the development of undergraduate data information literacy curriculum in other disciplines and at other institutions.

Lewis & Clark College and The Aubrey R. Watzek Library

Data in the Disciplines: Developing a Network to provide Data Management and Data Information Literacy Services at Small College and University Libraries

June 2018

Statement of National Need

As data intensive research alters the methodology of scholarship, college students need an understanding of how to find, apply, generate, share and manage data in their field of study. While libraries at large research universities are at the forefront of research data management, small college libraries have an opportunity to build on a culture of student/faculty research to introduce research data management and data information literacy into the undergraduate curriculum. In this project, Lewis & Clark and four partner colleges will develop a series of workshops that provide students and faculty with data management and data information literacy support in specific disciplinary domains and offer a model of regional collaboration that can be replicated by other small colleges. As part of the process, we will develop and share curricular modules dealing with chemistry and ethnographic data for undergraduates.

While it is clear that data-intensive research is becoming more prevalent across disciplines, the mechanisms that colleges and universities use to support that research are evolving. Researchers need help collecting, organizing, and sharing data. Both scholars and students need to understand how to manage, publish, find and reuse data sets to move research agendas forward. Academic libraries have been experimenting with ways that they can support data in scholarship through education, consultation, and technical support. This is an evolving area of practice, one that still lacks successful, tested models even at large research institutions.¹ This project will advance understanding of effective methods.

There is a particular need for smaller institutions to take part in the experimentation around providing these services. Data management practices and data information literacy bring undergraduate students in touch with the research process and can be a meaningful part of undergraduate research experience. Undergraduate research is recognized as a high impact educational practice by the American Association of Colleges and Universities. Liberal arts colleges are increasingly built around a scholar/teacher model in which faculty pursue original research while creating opportunities for students to engage in the process. According to the AAC&U, in the case of undergraduate research with faculty, “The goal is to involve students with actively contested questions, empirical observation, cutting-edge technologies, and the sense of excitement that comes from working to answer important questions.” High impact educational practices, including undergraduate research, have a disproportionately positive impact on students from disadvantaged backgrounds.²

Lewis & Clark College has a longstanding and robust tradition of faculty-mentored student research, a vibrant and growing library data services program, and a track record of successful collaboration with regional peers. In the summers of 2015 and 2016, Lewis & Clark, Reed College, University of Puget

¹ B Latham. (2017). Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges. *The Journal of Academic Librarianship*, 43(3), 263–265. <https://doi.org/10.1016/j.acalib.2017.04.004>

² GD Kuh. (2008). *High-impact educational practices: what they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.

Sound, Whitman College, and Willamette University, co-organized four workshops on data management in the sciences attended by a total of 101 faculty, students research assistants, and librarians.³ These events pooled together library expertise in data management, fostered collaboration between researchers and librarians, and improved practices in research laboratories.⁴ The final reports from each are attached to this application as Supporting Documents.

The workshops proposed a model of collaboration that responded to the situational and resource challenges frequently encountered by research data management librarians working in small college environments.⁵ We worked with a consultant from a local Research 1 university to develop our foundational curriculum in an area where our team lacked prior training; we brought together teams of faculty and student research assistants, acknowledging students as critical partners in research data management in the small college environment; and we recruited faculty partners who in turn became advocates and evangelists for the role of the library in providing research data services at our respective campuses. Moreover, the workshops responded to the recognition of the undergraduate research experience as an opportunity to integrate instruction in data management skills and data management literacy into the undergraduate experience.⁶

Following the workshops, we found that faculty participants had increased confidence and investment in the importance of teaching data management concepts. Within our five institutions, we have successfully leveraged the workshops to help faculty integrate data management lessons into their courses. Data fluency is increasingly recognized as a core competency for undergraduates, and there is a need to develop more data management literacy lessons that can be readily integrated into the undergraduate curriculum.

Feedback from the first set of workshops, as well as another documented case of data management training,⁷ has indicated that a more focused, disciplinary approach would prove valuable to scholars working in areas where data practices are very distinct. Data curation challenges are often specific to data types and experimental methodology, and opportunities for data-sharing may be particular to cultures within research communities. Furthermore, the data curation landscape is evolving rapidly as research communities adopt and/or refine data sharing practices in response to mandates by federal funders. Consensuses are forming in communities around metadata schemas and discipline specific repositories, and, in many cases, data curation best practices have evolved and/or solidified over the

³ The Mellon Foundation supports collaboration between these five liberal arts colleges in the Pacific Northwest as the Northwest Five Consortium and funded this workshop series.

⁴ The team that developed the workshop reported on it in the March 1 2017 issue of the *IFLA Journal* <https://doi.org/10.1177/0340035216678239>. The final reports for each workshop are attached to this application as Supporting Documents.

⁵ M Toups and M Huges. (2013). What data curation isn't: A redefinition for liberal arts universities. *Journal of Library Administration*, 53(4), 223-233.; K Partlo, D Symons and JD Carlson. (2015) Revolutionary or evolutionary? Making research data management manageable. In *Creating Research Infrastructures in the 21st Century Academic Library: Conceiving, Funding, and Building New Facilities and Staff*. Lanham (MD): Rowman & Littlefield, 171-201.

⁶ R Clement, A Blau, P Abbaspour and E Gandour-Rood. (2017) Team-based data management instruction at small liberal arts colleges. *IFLA Journal*, 43(1), 105-118.

⁷ J Qin and J D'Ignazio. (2010). "Lessons learned from a two-year experience in science data literacy education." *International Association of Scientific and Technological University Libraries, 31st Annual Conference*, Paper 5, Lesson Learned 3. <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>

past couple years . To best serve researchers and teachers, there is a need to supplement the generalized, one-size-fits-all data management curriculums⁸ with guidance tailored to particular research communities.

Building on a collaboration with Dr. Anne Bentley,⁹ a previous workshop attendee with whom we developed a data management lesson for integration into an upper-level Inorganic Chemistry course, one of the two proposed workshops will focus on chemical data. Chemistry's notably regimented undergraduate curriculum creates ready opportunities for integration of the data literacy lesson we will develop across our five schools. Chemistry is also a discipline where data management has been recognized as a core competency; the most recent iteration of the American Chemical Society's guidelines for Bachelor's Degree Programs acknowledges that, "instruction should be provided in data management and archiving" as well as the more traditional information literacy skills of "managing citations and related information."¹⁰ To provide some degree of breadth across the liberal arts and sciences, we plan to offer a second workshop in ethnographic research data. Ethnographic research has the potential to touch the curriculum of many fields within the social sciences (Political Science, Economics, Sociology, Anthropology, etc.) and has been recognized as an irreplaceable data type.¹¹

Project Design

The proposed project will be a collaboration with the institutions named above and will result in two 1.5-day workshops for faculty, students and staff to increase understanding of research data management practices in particular disciplinary domains and develop undergraduate curricula in data information literacy specific to those domains.

The overall goals of the project are a) to advance the ways that small college libraries can support research data management and data information literacy, and b) expand the kinds of research-based learning that are available to undergraduate students at small colleges. The project will provide 1) A tested model for establishing regional collaborations among small colleges/university libraries to support data management and literacy in specific academic disciplines, 2) data management curriculum modules tailored to the particular curation needs and challenges of two disciplinary data communities, and 3) an outline of a one-hour data information literacy curriculum in each discipline co-developed by librarians and faculty.

The workshop design will follow the model we piloted in the 2015 and 2016 workshops. We will bring together teams composed of a faculty researcher, their student research assistants, a librarian, and, as appropriate, an information technology staff member or supporting technician from institution. This

⁸ There are various generalized data management curricular modules available online including the New England Collaborative Data Management Curriculum (NECDMC); the Data Observation Network for Earth (DataONE) Data Management Modules; and the MIT Open Courseware Data Management Course.

⁹ Dr. Anne Bentley, Associate Professor of Chemistry, Lewis & Clark College attended our workshops with her research team in the summers of 2015 and 2016.

¹⁰ American Chemical Society Committee on Professional Training. (2015) *ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*, pg 18.

<https://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acg-guidelines-for-bachelors-degree-programs.pdf>

¹¹ A Asher and LM Jahnke. (2013). Curating the Ethnographic Moment. *Archive Journal*.

<http://www.archivejournal.net/essays/curating-the-ethnographic-moment/>

workshop design models the type of librarian-faculty-student collaboration that can be the hallmark for data management services in these small settings, while taking advantage of a unique opportunity for student learning and training. Student participants will gain skills to apply in their research teams and be valuable stakeholders in conversations about both research data management and curriculum development.

The first day of each workshop will be devoted to developing participant awareness and understanding data management issues in their disciplinary fields. We will adapt our previous generalized workshop curriculum (itself an adaptation of the New England Collaborative Data Management Curriculum as well as the DataONE modules), which includes lecture and discussion sessions interspersed with group activities, to maximize its benefit for teams from each disciplinary domain.¹² As we did in previous years, the curriculum will be calibrated based on participating teams' interests or stated data curation challenges.

The following half-day will be devoted to developing a data management module suitable for integration into an undergraduate curriculum by attending faculty members. Where librarians have employed data management literacy in classroom settings they have found that lessons have more impact if tied closely to the "disciplinary context, terminology, and workflow,"¹³ and when engaging directly with data otherwise encountered in the curriculum.¹⁴

As shown in the reports (Supporting Documents), our earlier workshops were highly subscribed, and we believe that faculty will be eager to attend the proposed workshops because of the opportunity to improve their research processes, develop curricula for teaching, and develop ongoing connections with regional colleagues in their area of research/teaching. Librarians from each of our five schools will participate in recruiting from our respective faculty.

Chemistry Workshop

On the first day of the Chemistry Research Data Curation Workshop, we will tailor and deliver our generalized data management curriculum to more deeply address the needs of chemists. For example, we will introduce file-naming and organizational strategies for commonly encountered chemical data types (spectra and other instrument output, physical samples, assays), explore best practices for capturing and cross-referencing key details of the experimental narrative in lab notebooks, and consider current practices for data-publishing and data-sharing within the discipline including emerging repositories in the field. Breakout sessions will provide an opportunity to explore strategies and tools specific to various analysis outputs such as SEM images or protein conformations. On Day Two, we will distill some general concepts around chemical data literacy into a one-hour curriculum module for chemistry majors.

¹² The curriculum for our 2015 and 2016 workshops was adapted from the NECDMC modules. Our workshop slides are available here: <http://ir.library.oregonstate.edu/concern/defaults/sx61dq99g>

¹³ J Qin and J D'Ignazio. (2010). "Lessons learned from a two-year experience in science data literacy education." *International Association of Scientific and Technological University Libraries, 31st Annual Conference*, Paper 5, Conclusion. <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>

¹⁴ BA Reisner, KTL Vaughan, and YL Shorish. (2014). Making Data Management Accessible in the Undergraduate Chemistry Curriculum. *Journal of Chemical Education*, 91, 1943-1946.

Ethnography Workshop

The Ethnographic Data Curation workshop will address how to handle the unique data produced when working with human subjects. Specifically, we will share strategies for dealing with common data outputs including audio files from interviews and field notebooks, evaluate methods for safeguarding sensitive records to protect privacy, discuss coding systems at length, note requirements for obtaining consent for archiving data and anonymization strategies for preservation, review indexing resources such as subject specific controlled vocabularies and finding aids, orient participants to key data registries in the field, and reflect on the ethical considerations at play in working with this irreproducible data type. Two Lewis & Clark College faculty, Dr. Sarah Warren, Associate Professor of Sociology and Dr. Keith Dede, Professor of Chinese, have expressed interest in attending the workshop. Additionally, Dr. Warren has expressed interest in working with the library to develop a data management lesson to introduce to Lewis & Clark's Sociology and Anthropology theses students thereafter.¹⁵ (See letter of support)

Activities/Project Team

Our steering team will begin meeting in December 2018 to plan the workshops according to the schedule presented below. The team will consist of science/data librarians representing the schools in the consortium. The team will review and refine our timeline for the project and divide into subgroups to plan each workshop.

The Steering Team will be led by Mark Dahl, Director of Lewis & Clark's Watzek Library. Team members will include Parvaneh Abbaspour, the Science & Data Services Librarian at Lewis & Clark College; David Isaak, Data Services Librarian at Reed College; Eli Gandour-Rood, Science Liaison Librarian at the University of Puget Sound; John Replinger, Science Librarian at Willamette University. A librarian from Whitman College will participate as a liaison to the team in a somewhat more time-limited capacity. The team will meet weekly via videoconferencing as was our established pattern in our 2015 and 2016 workshop endeavors.

The Chemistry and Ethnography planning groups will begin work in earnest in winter and early spring 2019. In partnership with the consultants (identified below), they will design the workshop curriculum. They will also be in touch with potential faculty attendees about the workshop curriculum and the best timing for the respective workshops. The workshops will take place in summer and/or early fall semester.

The administrative coordinator at Lewis & Clark's Watzek Library will coordinate the workshop logistics including reserving space, accommodations, and establishing a system for registration. In early April, registration for the workshops will open. Librarians at the five schools will reach out to faculty whose teaching and research align with the workshop topics and encourage them to take part.

Consultants

Dr. Ye Li, Scholarly Communications and Instruction Librarian at the Colorado School of Mines, will work with us to develop the Chemistry Research Data Workshop curriculum. Dr. Li earned a Ph.D. in

¹⁵ At Lewis & Clark a thesis is required of all Sociology and Anthropology majors. Lewis & Clark awarded 46 Sociology and Anthropology (SOAN) degrees in 2015-16, the most recent year for which data are available.

Chemistry as well as an M.A. in Library and Information Science from the University of Iowa. She spent six years working as the Chemistry Librarian at the University of Michigan before moving to the Colorado School of Mines. She is a member of the Research Data Alliance Chemistry Research Data Interest Group and has published on chemistry research data management, research data sharing, and improving chemistry information literacy.¹⁶ Dr. Li's letter of commitment and CV are attached.

Celia Emmelhainz, Anthropology and Qualitative Research Librarian at the University of California, Berkeley will work with us to plan the Ethnographic Research Data Management Workshop. In addition to holding a M.L.S., Emmelhainz earned a M.A. in Cultural Anthropology from Texas A&M University and has done ethnographic research in Mongolia and Kazakhstan. At Berkeley, she leads workshops on qualitative methods and qualitative data management and has published on developing a controlled vocabulary for indexing anthropology datasets and issues surrounding ethnographic data preservation. She is a member of the Research Data Alliance Digital Practices in History and Ethnography Interest Group and serves on the board of advisors for the national Council on Preservation of Anthropological Records. Prior to moving to Berkeley, Ms. Emmelhainz worked as the Social Science Data Librarian at Colby College a small liberal arts college in Waterville, Maine.¹⁷ Ms. Emmelhainz's letter of commitment and CV are attached.

Assessment and Dissemination

Following the workshops we will survey attendees on the workshop outcomes and address the IMLS agency-level goal of learning/communities of practice. We will ask questions to determine how well the workshops informed data management practices in research and how effective they were in developing curricular modules. We will also develop a follow-up survey to assess the effectiveness of the curricular modules to be given once they are implemented in the classroom, possibly as early as fall 2019.

This proposed project has the potential to aid many audiences, among them data librarians at small colleges and universities seeking a model for extending services, data curators at all institutions seeking data management curriculum tailored to specific disciplines, and disciplinary faculty interested in adding data management lessons into their existing undergraduate curriculum. We will share our experiences and publicize the products through a multi-pronged approach to maximize impact.

To begin, we will publish the workshop curriculum and data curation literacy curricular modules on the project website, and one or two librarians will present on the workshops at a national conference, possibly the Digital Library Federation Forum, which usually occurs in October. Additionally, all the schools in this project are members of the Orbis Cascade Alliance Consortium, a group of 38 academic libraries in the Pacific Northwest, many of whom have initiatives underway in data management and data information literacy. Consortium members will be invited to attend a webinar in which we discuss this project and prospects for additional efforts like it in the region.

The project webpage will reside on the website of Lewis & Clark's Watzek Library, for which there is already an established infrastructure of support and which we can ensure will be maintained for many

¹⁶ Profile and contact information for Dr. Ye Li: <http://inside.mines.edu/~yeli/>

¹⁷ Profile and contact information for Celia Emmelhainz: <http://dlab.berkeley.edu/people/celia-emmelhainz>

years to come. We will additionally deposit our modules in one of our partner college's institutional repository, the University of Puget Sound's *Sound Ideas*. To promote findability, we will deposit the modules and lessons into the Data Management Training Clearinghouse, a registry for online learning resources about research data management.¹⁸ They will be published as open educational resources under a Creative Commons Attribution-Noncommercial license allowing for others to make further modifications to maximize their usefulness in diverse contexts.

We will work with our faculty partners and consultants to reach audiences in the disciplinary fields. Dr. Bentley will post our chemical data management curricular modules on *VIPEr*, the Virtual Inorganic Pedagogical Electronic Resource, a National Science Foundation funded website for open-education curricular resources relevant to inorganic chemists.¹⁹ The resources will be announced in a blog post, and the modules posted under "teaching resources." *VIPEr* provides a venue for directly reaching other chemistry faculty and educators.

To reach ethnographic practitioners beyond our alliance, we will also share news of the ethnographic data management curriculum through a blog post on the website *anthro{dendum}.org*. Our consultant Celia Emmelhainz has previously published about ethnographic data management on the site's precursor, *Savage Minds*, which in 2010 the *American Anthropologist* noted to be "the central online site of the North American anthropological community."²⁰ Blogging continues to be recognized as an important way to communicate research to stakeholders and strengthen communities of practice.²¹

Many members of our steering team maintain twitter accounts and we plan to supplement our more formal blogging efforts with a distributed microblogging stratagem. We've identified the hashtag #datainthedisciplines to unite posts across platforms and amongst our steering team members' accounts with the aim of creating "chatter" around the project.

Finally, we will share news of the program on assorted data librarian channels including the national "datacure" listserv - a forum for communication between data curators at institutions of all sizes. Additionally, we will post to the "datamanage" listserv which connects data curators within the Pacific Northwest. We will post to "laredas", the Liberal Arts Research Data Support listerv, to specifically reach out to our liberal arts colleges. For good measure, we will also post to the "iassist" listserv to reach internationally.

¹⁸ Data Management Training Clearinghouse: <http://dmclearinghouse.esipfed.org/>

¹⁹ VIPEr is the Virtual Inorganic Pedagogical Electronic Resource: A Community for Teachers and Students of Inorganic Chemistry; <https://www.ionicviper.org/>

²⁰ Posts by Celia Emmelhainz on *Savage Minds: Notes and Queries in Anthropology* <https://savageminds.org/author/cmommelh/>

²¹ E Brown and C Woolston. (2018). Why Science Blogging Still Matters. *Nature*, 554, 135-137. <https://doi.org/10.1038/d41586-018-01414-6>

Budget

The total request to IMLS is \$24,983, which includes funding to hire an undergraduate student assistant, for participants travel to/from workshops, conference travel for dissemination, workshop expenses, and to engage outside consultants. Lewis & Clark will also cost share the amount of \$1,580 for conference travel. Further details are included in the budget justification. In addition, Lewis & Clark and partner libraries will contribute library and staff time, and will provide a venue to host the workshops.

Project Timeline

Date	Activity
December 2018	Funding period begins; Steering team established; Subgroups for each workshop appointed.
January-February 2019	Subgroups contact potential participants, establish dates for workshops; project web site created.
March-May 2019	Workshop curriculum development with consultants; Recruitment of participants; logistical planning for workshops; survey design for workshop participants.
June-July 2019	First workshop held.
August-September 2019	Second workshop held.
August-November 2019	Dissemination of project at national conference(s).
August-November 2019	Surveys distributed to workshop attendees; results of surveys posted on project web site.
November 2019	Workshop curriculum and data literacy curricular modules published on project website and announced through the described channels.

Diversity Plan

Lewis & Clark College and the partner colleges participating in this project are liberal arts colleges that recruit students nationally and internationally. All five colleges seek to create welcoming multi-cultural, multi-denominational, multi-identity communities. Lewis & Clark has a dedicated Dean and Office of Diversity of Inclusion that works with other units in the college to recruit a diverse pool of students, support diversity in faculty and staff hiring practices, and to create a welcoming community for individuals from all backgrounds. The College has multiple programs to support minority and first generation students as they enter college. For example, the Xplore program is a seven-day orientation for first generation students that gives students the tools to succeed in science and math courses.

This initiative will reach a diverse group of students in a number of ways. The student research assistants attending the workshop will acquire valuable skills in data management that will be an advantage to them as they pursue careers past college. The model developed will particularly benefit

students in the liberal arts college sector of higher education, whose students may not typically have access to cutting-edge research practices. Data management practices can be an important aspect of undergraduate research, a high impact educational practice which, as mentioned above, has a disproportionately positive impact on students from disadvantaged backgrounds. The curricular modules we develop will enrich research-based courses in Chemistry and ethnography among the participating schools and a wide range of colleges and universities beyond the Pacific Northwest. Students of all backgrounds taking these modules will gain valuable competencies in data management and data information literacy in their discipline that should help them advance to graduate training or onto the many other career paths in which data competencies are valuable.

National Impact

In service to the National Digital Platform category, this project brings together scholars with librarians to advance data intensive scholarship and develop curricula that can have a broad impact among students and faculty at small colleges and universities.

The Data Curation Network, a consortium of seven university libraries and a digital repository, is a consortium devoted to sharing specialized expertise around data curation. This project attempts to use the same strategy of working together to accomplish shared goals in the small college/university context. With this project as a model, small colleges should be able to “scale up” their data services to faculty and students by providing professional development opportunities such as these workshops.

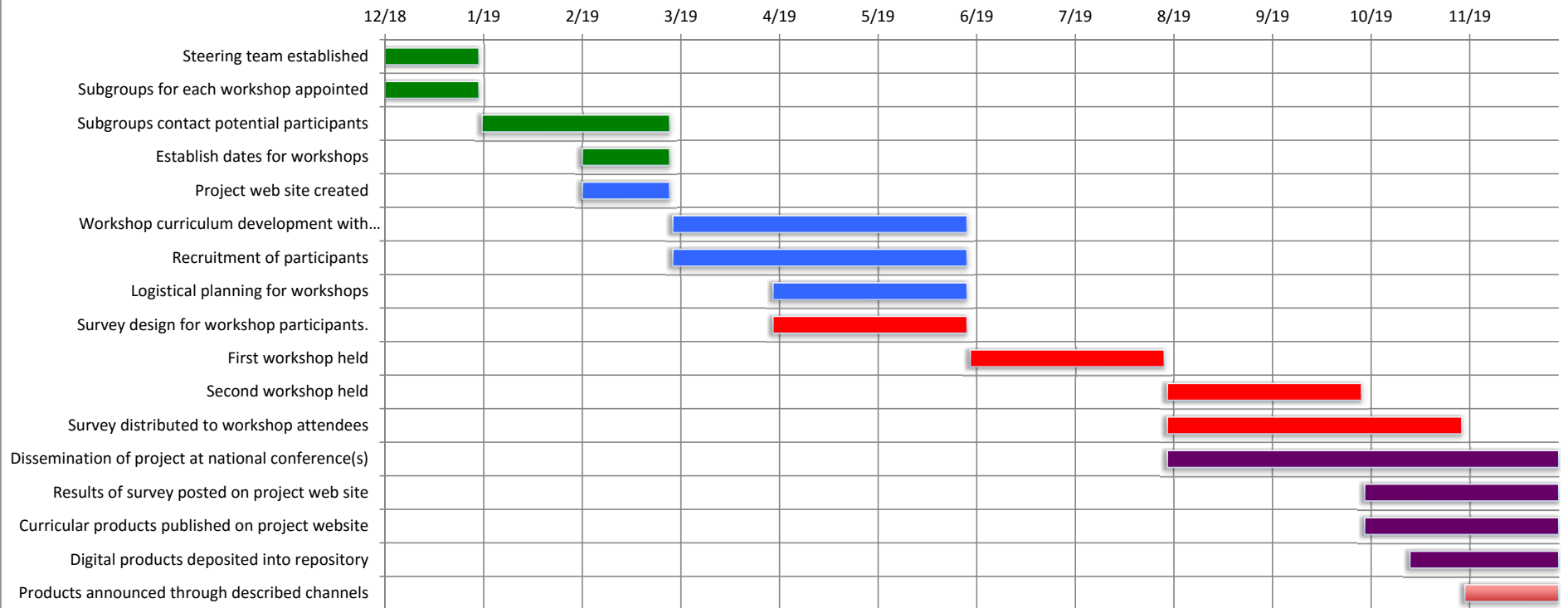
By using consultants from larger institutions our team will tap into more specialized expertise. But our approach is not simply an attempt to replicate the activities of larger institutions. The collaborative workshop model that brings faculty, staff, and students together will produce longstanding collaborative relationships between participants and lead to the creation of high-quality, student-centered curricular modules. It will be a model that plays to the strengths of small colleges and universities but also offers lessons that are applicable more broadly.

The actual curricular modules that we produce in data management and data information literacy for Chemistry and Ethnography should be applicable in many undergraduate settings, from colleges to large universities. Likewise, they will serve as models for the development of undergraduate data information literacy curriculum in other disciplines.

As the role of the academic library evolves from provider of collections towards services and expertise, this project will provide a valuable model of library services in support of scholarly communication and information literacy. Our model positions the library in a leadership role in supporting and educating faculty about emerging practices around data sharing and reuse. It also positions the library as an educator that can facilitate curriculum development for undergraduates around emerging disciplinary research practices.

Lewis & Clark College and The Aubrey R. Watzek Library
 Data in the Disciplines
 Schedule of Completion

Schedule of Completion



DIGITAL PRODUCT FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

- Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

Part I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

The digital products created for this project will hold a Creative Commons Attribution-Noncommercial 4.0 International license (CC-BY-NC): <https://creativecommons.org/licenses/by-nc/4.0/>

This license will allow others to enhance and reuse the material for their own purposes in an attempt to maximize the utility of the resources to a broad audience. Out of respect for intellectual investment required to create these resources as well as the expertise held by practitioners in our profession, we will request attribution. Additionally, we have decided to not allow for commercial reuse of our materials. We are aware that there are some for-profit data curation service initiatives out there, and to keep with the ethos of the community and the tremendous spirit of collaboration that goes into all of this, we respectfully request that our materials stay in the public domain. We have decided against a share-alike license to limit any additional restrictions on the content.

The copyright will be held jointly by the co-authors of each digital product.

To explain property permissions, we will apply the Creative Commons "CC-BY-NC" branding to all digital products and explain the licensing in text on the website. We will also link to the licensing explanation on the creative commons site, with a full description of the licensing terms.

To aid users in attribution we will provide a suggested citation for each product.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

At Lewis & Clark College (LC), the College owns all intellectual property created by its employees within the scope of their employment, by persons acting under contract with the College and by students and other persons using College facilities for research and study. The Dean of the College of Arts and Sciences, as a representative of the College administration, has approved our use of a Creative Commons license, to be held by the co-authors of the products, and has signed off on this proposal. If funded, we will formalize this agreement with a signed letter from the College president. As copyright holders, the authors will place no restrictions on use of content beyond those outlined in the Creative Commons Attribution-Noncommercial 4.0 International license.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

We will collect survey data from workshop participants. Data collection will be approved by LC's Institutional Review Board and will comply with human subjects' research protection requirements. No content created for this project will expose private information of project participants.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

1. We will create one webpage describing the project on which we will publish all of the curricular products which we will produce. The webpage will be hosted on the Watzek Library website which utilizes the Library's Springshare LibGuides Content Management System.

2. We will create a two data management workshop curriculums, one focused on chemistry research data and one focused on ethnographic research data. The curriculums will be "slide decks" or Microsoft PowerPoint file - described further below.

3. We will create two data management literacy curricular lessons, one focused on chemistry research data and one focused on ethnographic research data. The curriculums will be "slide decks" or Microsoft PowerPoint file - described further below.

4. We will collect survey data from workshop participants for each of the two workshops. We will also collect voluntary feedback from faculty whom implement our data management literacy lesson during the Fall 2019 academic term. Feedback will be collected in a Google form as described below.

5. A presentation about the project will be presented at national conference. The presentation will be a "slide deck" - described further below.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

1. The web-presence will be created and hosted on the Watzek Library's Springshare LibGuides Content Management System.

2 & 3. We will create our curricular modules in Google Slides to facilitate collaborative authoring between our team members. We will then export and publish them in Microsoft PowerPoint.

4. The survey will be created in Google Forms, which will create an associated dataset in a Google sheets file.

5. We will create our presentation in Google Slides to facilitate collaborative authoring. We will then export them to Microsoft PowerPoint.

A note on the use of Google products: Lewis & Clark College holds an agreement with Google by which it retains full copyright and administrative privileges over all files attached to LC accounts. One of the two project directors will be designated as the “owner” of all files produced for this project.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

1. The website will be a dynamically generated HTML file.

2 & 3. The curricular modules will be published as both a Portable Document Format file (.pdf) and as a Microsoft PowerPoint file (.pptx) to facilitate future revisions. Microsoft PowerPoint files can be opened in the Keynote application common to the Mac operating system and be readily converted to Google Slides.

4. The survey questions will be transcribed to a Google document and downloaded as a Text (.txt) file for deposit into a repository. The anonymized survey results will be downloaded into a Comma Separated Values file (.csv) for deposit into the repository.

5. The presentation will be saved and published a Portable Document Format (.pdf) file.

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

Each planning team (Chemistry & Ethnography) will appoint a Digital Products Steward who will take the lead on creating files, managing versions, and coordinating publication of associated products. We will abide by the project’s Schedule of Completion, which allocates time for the creation and publication of the digital products, and wherein the workshops themselves create natural checkpoints for curriculum development. At the start of each meeting of each planning team, the team will touch base on their progress towards project goals including the production of digital products.

The Steering Team, and ultimately project directors, will take responsibility for making sure all products are posted to the project website and deposited in the University of Puget Sound’s Institutional Repository as described.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

1. Watzek Library in the care of Lewis and Clark College maintains the Watzek Library website. There is an organizational infrastructure in place including dedicated staff time to website maintenance, including content migration in the event that is required in the future.

2, 3, 4 & 5. Additionally, we will deposit all the curricular modules, the survey questions and the anonymized associated dataset, as well as the resulting presentation into the University of Puget Sound's Institutional Repository, *Sound Ideas*. *Sound Ideas* runs on bepress which is LOCKSS compliant.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

1. We will use the LibGuides Administrative Interface to record Dublin Core metadata concerning the webpage.

2,3,4 & 5. All other resulting digital products will be described with Dublin Core metadata upon deposit into the University of Puget Sound's Institutional Repository, *Sound Ideas*.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

1. Watzek Library is invested in maintaining the content held in the LibGuides Content Management System. In the event the library needs to migrate the content, we will make sure to migrate the metadata as well.

2, 3, 4 & 5. The University of Puget Sound's IR includes built in preservation workflows to maintain the integrity of the digital assets, including object metadata.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

1. We will add HTML5 Meta-tags containing Dublin Core metadata to the heading of the webpage which will improve search engine optimization and enhance discoverability.

2, 3, 4 & 5. *Sounds Ideas* exposes their metadata via OAI-PMH protocol which allows other systems to harvest the records.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The webpage will be openly available online and accessible via standard web browsers. We can employ publicly available tools, such as PowerMapper, to ensure the website meets accessibility standards.

Sound Ideas is hosted on the Bepress Digital Commons platform. It is also openly available online and accessible via standard web browsers. *Sound Ideas* is managed by dedicated staff at the Collins Library of the University of Puget Sound.

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

The Watzek Library Website which will host our project website is here: <http://library.lclark.edu/>

The proposed url for project webpage is: <http://library.lclark.edu/datainthedisciplines>

For both previous iterations of our collaborative workshops we created websites, which announced the workshops and served the purpose of sharing logistical information with participants.

2015 Northwest 5 College Consortium (NW5C) Data Curation Workshop Website: <https://nw5cdata2015.wordpress.com/>

The slides created for our 2015 data management workshop are available here.

<http://ir.library.oregonstate.edu/concern/defaults/sx61dq99g>

2016 NW5C Workshop website: <https://nw5cdata2016.wordpress.com/>

Workshop materials, including slides, and final report can be found here:

<https://nw5cdata2016.wordpress.com/workshop-materials/>

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

We will collect survey data from workshop participants regarding workshop outcomes. We will also collect data from faculty who have implemented our data management literacy lesson during the Fall Term of 2019. The survey results will be used to help the grant team evaluate the success of the workshop, the effectiveness of the lesson, and inform interested communities through public presentation and publications. The results also have the potential to inform the development of new programs and products beyond the time frame of this grant.

We will distribute the survey to workshop participants coincident to the timing of each workshop. We will distribute a follow-up survey evaluating the effectiveness of the data management literacy lessons to any workshop participants who implemented the lessons during the Fall of 2019. The follow-up survey will be distributed in September and responses requested by early November 2019.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

Data collection will be approved by LC's Institutional Review Board and will comply with human subjects' research protection requirements. Once the survey instruments have been prepared (March-April 2019), we will submit for approval (April-May 2019).

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

We will collect information from respondents such as institutional affiliation, role (faculty, staff, student), disciplinary affiliation, which even in aggregate, in the context of a small college campus, could result in personal identification. In acknowledgement of this reality, we will include a question seeking consent from respondents to include their anonymized answers in public presentations or publications related to the workshop. Informing them that this may potentially include aggregations of numerical data and quotations from open-ended questions, identified by role as appropriate. We will only include answers from respondents who have given their consent in the data files prepared for public release. We have done exactly this in the past, and, based on our experience, we believe this won't overly impact the number of responses given the relatively non-sensitive nature of the questions.

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

A question seeking consent (detailed above) will be built into our survey tool such that answers become values in the resulting dataset. This will literally embed the consent agreements into the dataset. The final dataset prepared for deposit will only include values from respondents who have granted their consent.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

We will collect the data through a survey created in Google forms connected to one of the project directors Lewis & Clark Google accounts. LC maintains an agreement with Google by which it retains full copyright and administrative privileges over all files attached to LC accounts. The resulting dataset will be viewable in an associated Google Sheet that will be shared among members of the Steering Team. The final dataset for publication, with only the values from consenting respondents, will be exported as a Comma Separated Values (.csv) file, which can be read by many applications.

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

The survey questions will be transcribed to a Google document and downloaded as a Text (.txt) file for deposit into the University of Puget Sound's Institutional Repository *Sound Ideas*, alongside the final dataset. They will be associated as digital items under one metadata record, or "work". The metadata will include a description of the project to workshop event to contextualize the dataset.

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

We will deposit the final "public" dataset, alongside a document detailing the survey instrument, into the University of Puget Sound's Institutional Repository *Sound Ideas*. As a bepress Digital Commons product, *Sound Ideas* employs various preservation checks and workflows discussed above.

We will share the "doi" for the data on our project website, in presentations, and in any future publications.

A.8 Identify where you will deposit the dataset(s):

Name of repository: Sound Ideas (University of Puget Sound)

URL: <https://soundideas.pugetsound.edu/>

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

Once convened, the first meeting of the Steering Team and the Workshop planning teams will appoint Digital Products Stewards (DPS) and the DPS will lead a review of the data management plan at the following meeting. Thereafter the DPS will consider the document a guiding document and provide quarterly reports to the steering team on the status of all digital products to confirm our compliance with the plan.