

## **Developing Specialized Data Curation Training to Address Needed Expertise in Focused Areas**

In order to “enhance the training and professional development of the library and archival workforce to meet the needs of their communities” ([Objectives 3.2 & 3.3](#)), Duke University Libraries (Sophia Lafferty-Hess) in partnership with members of the Data Curation Network (DCN) - Pennsylvania State University (Seth Erickson), Washington University in St. Louis (Dorris Scott), Cornell University (Wendy Kozlowski), the University of Michigan (Jake Carlson), and the University of Minnesota (Lisa Johnston) request support for a two-year planning grant (August 2022 - July 2024) to develop an enhanced, specialized curriculum for data curation training for academic library and archive staff. The DCN is a practitioner-based collaborative of (currently) 14 academic and non-profit data repositories that share staff expertise for the critical aspects of data curation including: data quality assessment, documentation, running code, file format transformation, and metadata review. The project team is requesting \$118,738 in support from IMLS.

This project will develop training programs for four areas of interest for data curators: geospatial (e.g., GIS, geodatabases, story mapping, LIDAR), scientific images (e.g., CZI, camera trap, aerial/drone imagery, NMRI, brain scans, cif), code (e.g., compilers, scripts, databases, containers, ipython/Jupyter notebooks), and simulations (e.g., netCDF, MATLAB, engineering, atmospheric). A community-driven model for curriculum building based on the data curation training offered by the DCN in our 2018-2020 IMLS-supported workshop will be employed. This project will build on the existing DCN curriculum, prioritize ethical, equitable and accessible considerations for research data curation and make the resulting educational materials openly accessible online for future use and adaptation. At the conclusion of the planning grant, the instructional teams will provide a pilot workshop for the community that will facilitate knowledge building, community engagement, and professional development.

### **Project Justification**

Under IMLS grant #RE-85-18-004018, the research team delivered workshops that trained 77 library and archive professionals. Instructors from the training program partnered with attendees to develop a collaborative capstone project, called “Data Curation Primers,” resulting in twenty-four new resources openly accessible to the curation community. Each primer serves as an accessible and practical guide for information professionals to use for the curation of a specialized data format. This project seeks to address the demand for advanced data curation training by building expertise in needed areas and filling a gap in existing data training programs by focusing on diverse types of data in the context of information professionals working within repositories and archives.

Recent curation services requested by the DCN indicates the most frequently submitted datasets by data type are: code (~33% or 82/249 datasets), tabular (~18% 45/249), simulation (~12% or 29/249 datasets), scientific images (~6% or 15/249), and geospatial (6% or 15/249). The high frequency of these datasets being submitted to the DCN may reflect areas of expertise missing from libraries curating research data as well as the relative complexity of curating these types of data. By developing more in-depth and customized training and resources in these areas, we will meet the demands of academic librarians and archivists to better support researchers at their institutions and help make data available in a form and format that can be reused by others to verify results and foster new discoveries.

### **Project Work Plan**

Building upon the existing DCN curriculum, this project will create four unique curricula for geospatial data, scientific images, code, and simulations that address curation best practices and the necessary approaches, procedures or considerations to make data more FAIR (Findable, Accessible, Interoperable, Reusable) as well as emphasize ethics, equity and accessibility. The project team will incorporate the existing open access “Data Curation Primers” resources for curating data in these specific areas,

which includes code (R and Jupyter Notebooks primers), scientific images (confocal microscopy, neuroimaging DICOM and NifTI, netCDF primers), and geospatial (Geodatabased, GeoJSON, GeoTIFF primers).

**Curriculum Development Phase:** In order to develop the curricula, we will bring together four instructional teams consisting of members from the DCN and other information professionals with expertise and interest in these data types to collaboratively develop workshop materials and data curation training activities. Workshop activities will include instruction and information sharing, but emphasize deep hands-on experience with real datasets, which is a key area that differentiates our program from similar training. The training will focus on practical data curation skills and the contextual knowledge required by data curators. The instructional teams will also explore ideal models for delivering the content (i.e., in-person, hybrid, etc.).

**Pilot Phase:** After the curriculum development phase, the instructional teams will offer a pilot workshop to the community that will bring together team members, peers in the data curation community, and an invited researcher who has experience working with the focused data types. The pilot workshop will include an overview of the curricula developed during the project - providing a mechanism to build knowledge in these areas, assess the workshop material, and strengthen community engagement.

The ultimate goal of the project, after the completion of the planning grant, will be to provide workshops including a capstone project for the four specialized data types.

### **Diversity Plan**

The project team is committed to maximizing the recruitment and retention of women, underrepresented groups, and people with disabilities. For participants invited to the pilot workshop, we will work with established national diversity initiatives and programs. We will dedicate at least three spots in our pilot workshop for minorities, early career, non-traditional librarians, paraprofessionals, curators, and/or graduate students interested in data curation. The DCN highly values inclusiveness and is committed to creating a welcoming and accessible environment, as expressed in our [Code of Conduct](#).

### **Project Results**

The curriculum developed by this program will extend research data curation expertise in needed areas. We will do this by developing instructional materials for building community knowledge of curation considerations and best practices. The curriculum generated will be shared online under an open license for reuse. This project will provide opportunities for community engagement via a peer-to-peer instructional design model and professional development opportunities to information professionals from a variety of institutions and backgrounds. The sharing of well-formed digital data can spur innovation and build trust in research in service to society; this project has the potential to broadly impact the research landscape by building expertise in data curation.

### **Budget**

The total budget for this project is \$118,738 with \$88,522 in direct costs and \$30,216 for indirect charges. A budget of \$29,774 covers the cost of the instruction planning meeting for 22 participants (meals, hotel, airfare, room rental for 6 partner representatives and 16 instructors), \$33,575 covers the costs of one pilot workshop for 6 partner representatives, 16 instructors and 3 spots of under-represented participants as described above (meals, room rental, and participants travel) and \$16,000 to provide stipends to instructors. Supplies and materials are provided at \$70. Travel to relevant conferences such as IDCC, RDAP, IASSIST, or RDA for project team members to engage with the community and disseminate the outputs of the project are provided at \$9,103 (airfare, registration, hotel, meals).