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" (IMLS Objective 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), University of Minnesota (Wanda Marsolek & Mikala Narlock), Princeton University (Neggin Keshavarzian), and the Association of Research Libraries (Cynthia Hudson Vitale) 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) 15 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.

A community-driven model for curriculum building based on the data curation training offered by the DCN in our 2018-2020 IMLS-supported project will be employed to develop training programs for four areas of need 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). At the conclusion of the planning grant, a pilot workshop will be provided to the library and archives community to facilitate knowledge building, community engagement, and assessment.

Project Justification

The recent public health pandemic has underscored the necessity of broadly sharing and making accessible research outputs, especially research data. The ability to quickly and effectively share information and data about COVID-19 facilitated a rapid public health response by policy-makers, researchers, and health advocates worldwide resulting in the development of vaccines in record time.[1] Data sharing and openness in research practices are no longer "new ideas" as funders[2] and publishers/journals[3] mandate them as a key aspect of the research enterprise for the purpose of reproducibility, reuse, and external validation. Beyond merely making these data available online, research data need to be properly cared for and managed, data need to be made findable, accessible, interoperable, and reusable (FAIR).[4] In brief, data need to be curated for effective re-use.

Research data curation enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time through activities including authentication, archiving, metadata creation, digital preservation, and transformation. Assessing the value of curation is still a relatively new field of study. However, there have been studies that have found contextual information is essential to enable data reuse,[5]

---

[1] Moorthy, V., Henao Restrepo, A. M., Preziosi, M. P., & Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19). *Bulletin of the World Health Organization*, *98*(3), 150. https://doi.org/10.2471/BLT.20.251561

[2] National Institutes of Health (NIH). (2020). Final NIH Policy for Data Management and Sharing. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

[3] Naughton, Linda, and David Kernohan. 2016. "Making Sense of Journal Research Data Policies". Insights 29 (1): 84–89. DOI: http://doi.org/10.1629/uksg.284

[4] Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[5] Faniel, I.M., Frank, R.D. and Yakel, E. (2019), "Context from the data reuser's point of view", Journal of Documentation, Vol. 75 No. 6, pp. 1274-1297. https://doi.org/10.1108/JD-08-2018-0133

curated datasets have more documentation and metadata than non-curated datasets,[6] and a correlation between curated data and metrics of reuse.[7] Given the potential for enriching data description and enabling data sharing, curation is increasingly being viewed as a desirable characteristic for repositories.[8]

Information professionals, including librarians, archivists, and other data professionals, are well positioned to partner with research faculty when preparing data for sharing and archiving. Institutions are showing interest in focusing on making data publicly available[9] and there is an increase in institutionally-based repositories accepting data or supporting a standalone data repository.[10] However, while the momentum for data sharing and engaging in open science practices continues to grow, there are still numerous challenges when managing and curating data to enable a FAIR ecosystem across data types.

Multiple groups have provided data curation training over the past 15 years. Library and information science programs have developed programs to train information professionals broadly in digital curation.[11] Training programs, such as the Carpentries, provide instruction that supports the open sharing and reproducibility of data.[12] Groups such as CURE (Curating for Reproducibility) have employed social science analysis techniques to compare the outputs of a code review to a publication.[13] The Data Curation Network's approach to curation education is unique in that it stresses curation in practice where there are often heterogeneous data types being submitted to a repository or archive.

This project will fill a gap in existing data training programs by focusing on diverse types of data. An analysis of recent curation requests submitted to the DCN indicate the most frequently submitted datasets by data type are: code (31.6% or 84/265 datasets), tabular (17.7% 47/265), simulation (10.9% or 29/265 datasets), scientific images (6.4% or 17/265), and geospatial (6% or 15/265). Within the DCN, the low percentage of data curators with expertise in these data types also indicates a potential for growth (n=42): code (ranging from 52% for R to 7% for C++), simulations (16%), geospatial (33%), scientific images (19%). The high frequency of these datasets requiring DCN curation support and the lower percentage of curators with these expertise reflect areas for additional support for curating research data.

As a cooperative, the Data Curation Network focuses upon building bridges between institutions grappling with similar research data questions and building capacity through peer-to-peer instructional design models. Under IMLS grant #RE-85-18-004018, the project 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 on both the DCN website and GitHub. Each primer serves as an accessible and practical guide for information professionals to use for the curation of a specialized data format.

---

[6] Koshoffer, A., Neeser, A. E., Newman, L., & Johnston, L. R. (2018). Giving datasets context: A comparison study of institutional repositories that apply varying degrees of curation. International Journal of Digital Curation, 13(1), 15–34. https://doi.org/10.2218/ijdc.v13i1.632

[7] Hemphill, L., Pienta, A., Lafia, S., Akmon, D., & Bleckley, D. (2021). How do properties of data, their curation, and their funding relate to reuse? Deep Blue Documents. doi.org/10.7302/1639

[8] National Institutes of Health (NIH). (2020). Supplemental information to the NIH policy for data management and sharing: Selecting a repository for data resulting from NIH-supported research (NOT-OD-21-016). Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html

[9] Association of American Universities & Association of Public and Land-grant Universities. (2021). Guide to Accelerate Public Access to Research Data. https://www.aau.edu/accelerating-public-access-research-data

[10] Johnston, Lisa R., and Liza Coburn. (2020, January 15). "Data Sharing Readiness in Academic Institutions." Data Curation Network. http://hdl.handle.net/11299/211358 .

[11] Weber, N. M., Palmer, C. L., & Chao, T. C. (2012). Current Trends and Future Directions in Data Curation Research and Education. Journal of Web Librarianship, 6(4), 305–320. https://doi.org/10.1080/19322909.2012.730358

[12] The Carpentries. https://carpentries.org/workshops/

[13] Curating for Reproducibility. https://cure.web.unc.edu/

The original IMLS grant will conclude this summer with a release of the CURATED curriculum as online modules for broad adaptation and use (see the Supplemental Wireframes). CURATED (Check files/code, Understand the data, Request missing information, Augment metadata, Transform files for reuse, Evaluate for FAIRness, Document throughout) is the protocol used by the DCN to curate data submissions. This protocol was adapted into the general training curriculum for research data curators in the workshops held from 2018-2020.

One of the critical pieces of feedback the DCN received from attendees who participated in the first workshop series was how excited they were to meet new like-minded colleagues and work collaboratively on developing the data curation primers. Across the three workshops, when asked what they were most excited about after this workshop the highest percentage (36%) indicated working on the primers. In the first workshop, participants also indicated an interest in having more opportunities dive deeply into multiple data types.

The model used for both the Primers and the core CURATED curriculum will be extended and refined to advance an open community-based model for supporting information professionals to build skills and knowledge in data curation in four unique data types. The benefits from this project will extend beyond the DCN by bringing together diverse instructional cohorts from different institution types and backgrounds. Building on the open access training materials from previous DCN education initiatives, the online modules and example datasets generated in this project will provide new, rich resources that anyone can explore for self-guided learning, maximizing the impact of this project. The project team will use an iterative design model to incorporate lessons learned to future education initiatives and extend this community-based approach to develop additional resources and training for new data types.

In a recent survey of end users of DCN institutional repositories, 90% of participants indicated that they felt more confident sharing their data as a result of the curation process.[14] The DCN has also surveyed information professionals working within repositories about their perception of whether data curation adds value to the data sharing process with 82% strongly agreeing.[15] These findings suggest that the curation of data not only assists in adding value to the data themselves, but that the process also plays a central role in advancing data sharing practices. As data sharing grows as a scholarly norm, arming information professionals with curation knowledge for specific data types will be key.

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. This will contribute to a virtuous cycle where those data then have a greater value to other researchers and the public at large. Librarians and archivists are key partners in the work to share high quality data, this project will further develop a collaborative model to engage the curation community in building the resources necessary to respond to this need.

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 while guiding curators in the necessary approaches, procedures or considerations to make data more FAIR. Additionally, all four curricula will emphasize ethics, equitability and accessibility. The following project work plan will support an efficient and effective delivery of outputs.

Roles and Responsibilities

Project coordinator: Supports general project management including interfacing with IMLS staff, generating reports, arranging project team member meetings, and tracking progress to ensure the timely delivery of products and execution of the project work plan. Sophia Lafferty-Hess will serve as Project Coordinator for the project.

DCN education liaison(s): Conceptually assists in connecting the scope and direction of this project with 1) the broader DCN strategic goals and mission, 2) existing DCN materials including the CURATED training modules, checklists, and primers, 3) other initiatives or groups within the DCN, as well as refining and teaching the core curriculum. Cynthia Hudson Vitale and Mikala Narlock will serve as DCN Education Liaisons for the project.

Instructional mentors: Facilitates the creation of content within the instructional cohorts providing context regarding the scope of the project, connecting the project outputs to existing DCN curriculum and primers, and creating a welcoming environment to encourage contributions from all team members. Works with other mentors and project team members to track the overall progress of the project. Seth Erickson, Dorris Scott, Wanda Marsolek, and Neggin Keshavarzian, DCN curators with experience in these data types, will serve as Instructional Mentors for the project.

Instructional cohorts: Generates new curriculum for identified data types, actively participates in the planning meeting, monthly virtual meetings, and Slack channels. Participates in the pilot workshop (if possible) and provides feedback to other instructional cohorts. Four instructional cohort members will be recruited for each data type. Recruitment will be based upon experience or interest in working with the identified data types as well as having a broad representation of individuals from the library and archives community.

Advisory board: Provides general oversight and advice for the project. The advisory board members will consist of individuals who participated in the previous IMLS grant #RE-85-18-004018 that generated the original CURATED training utilizing a peer-to-peer instructional design model, thereby providing continuity in the visioning and strategic placement of this curriculum within the DCN portfolio. The advisory board members (see Supplemental Letters) contributing to this project include:
- Mara Blake, Head, Data & Visualization Services, NC State
- Jake Carlson, Director of Deep Blue Repositories and Research Data Services, University of Michigan
- Wind Cowles, Director, Research Data and Open Scholarship, Princeton University
- Jen Darragh, Senior Research Data Management Consultant, Duke University
- Joel Herndon, Director, Center for Data and Visualization Sciences, Duke University
- Lisa Johnston, Director, Data Repository for the University of Minnesota, University of Minnesota
- Wendy Kozlowski, Data Curation Librarian, Cornell University
- Jennifer Moore, Head of Data Services, Washington University in St. Louis

Library and archives community: Provides feedback and input on the developed curriculum during the pilot workshop and via an anonymous survey distributed after the workshop. Brings the perspectives of information professionals to allow for external validation of the curriculum. Community members with experience and interest in data curation will be recruited through national and professional associations and networks including Research Data Access and Preservation (RDAP) Association, IASSIST, DataCure, as well as local networks in North Carolina including the Triangle Research Library Network and HBCUs in the area.

Project Phases

The project will be structured into five discrete phases to enable an efficient project plan to support the timely completion of project outputs including recruitment and cohort development, curriculum development, pilot, assessment and refinement, and dissemination.

Recruitment and Cohort Development Phase: Community engagement and inclusion beyond the DCN member institutions is a priority for this project. Data curation and disciplinary expertise is spread throughout our community and through a peer-to-peer curriculum development process, we will share expertise to build the training program. Four spots for instructional team members in focused areas will be open to all academic library and archive staff with basic and advanced curation or disciplinary expertise. A call for applications will be placed on the Data Curation Network's website and advertised on relevant data librarian listservs, such as the Research Data Access and Preservation (RDAP) and DataCure, to name a few. Selection will consider both the applicant's expertise and interests as well as seeking to build teams with diversity in knowledge, expertise, and background. A cohort model (see below) for curriculum development will facilitate knowledge sharing and establish a supportive community focused on expanding curation resources and learning from each other.
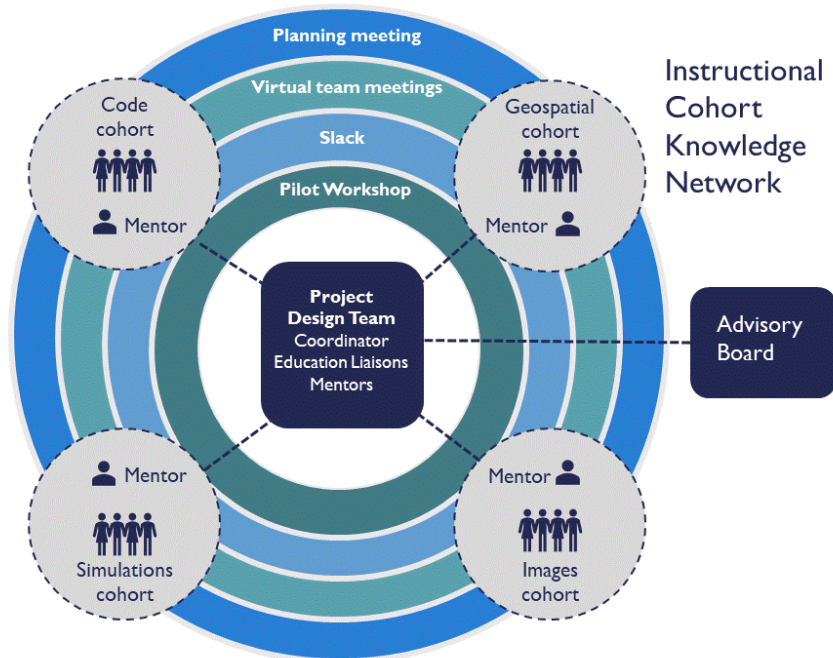
Selected individuals will be invited to participate in the project and will attend an initial 1.5-day planning meeting to establish the cohorts, network and build community, and develop an action plan for curriculum development. A skilled facilitator will lead this meeting to allow for full engagement from all participants and to maximize time spent. This meeting is currently scoped to be held in-person to allow for more meaningful community and relationship building; however, if necessary due to COVID-19 restrictions this in-person meeting will be translated to 3 half-day online meetings. During the cohort development phase, asynchronous communication channels will also be established for teams and mentors.

Curriculum Development Phase: In order to develop the curricula, we will bring together four instructional cohorts 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. The project team will build upon the existing CURATED curriculum that provided general training in methods of curation for research data and Data Curation Primers resources for curating data.

Elements of our existing training model that we intend to re-purpose include:
- Using the CURATED Protocol of the Data Curation Network to reinforce essential curation techniques and to provide a common orientation for each step of the curation process that allow learners to more easily acquire new information.
- Incorporating 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 primers), geospatial (Geodatabases, GeoJSON, GeoTIFF primers), and simulations (netCDF primer)
- Instruction interspersed with activities; such as, small group activities, demonstration of tools, and hands-on practice with datasets.
- Peer-to-peer training that places members of the DCN and subject and format experts together to learn from each other through open discussion of topics and data curation training activities.
- Capstone experience that allows participants in small groups with project mentors to deeply explore a topic while making a curator resource that will be made free and publicly available. The development of the capstone experience will be scoped during this project; however, it is outside the scope of this project to include the capstone project as part of the pilot workshop.

New curriculum design elements that will be incorporated into the training include:
- A focus on ethical, equitable and accessible considerations for research data curation
- An exploration of standardizing a pedagogical approach to instruction within the DCN education program
- A cohort approach that facilities cross-training of curators engaged in curriculum design
- An iterative design approach to assess the collaborative model for curriculum development via feedback from the advisory board, cohort members, and pilot workshop participants

Pilot Phase: After the curriculum development phase, the instructional teams will offer a pilot workshop to the community that will bring together team members and peers in the data curation community. 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.

Participants in the pilot workshop will include the instructional cohorts and mentors and 20 spots will be open to all academic library and archives staff. Participation will be heightened by partnering with national diversity initiatives as well as local HBCUs for recruitment as well as providing at least four

travel scholarships to members of historically marginalized communities, early career, non-traditional librarians, paraprofessionals, and graduate students interested in data curation,

Draft Pilot Workshop Plan:
- Day One:
  - Overview of the Core CURATED curriculum
  - Data Type 1: CURATED Applied Instruction, Hands on Exercise, Discussion
  - Data Type 2: CURATED Applied Instruction, Hand on Exercise, Discussion
- Day Two
  - Data Type 3: CURATED Applied Instruction, Hand on Exercise, Discussion
  - Data Type 4: CURATED Applied Instruction, Hand on Exercise, Discussion
  - Wrap-Up, Next Steps & Assessment

A finalized plan for the workshop agenda and pacing will be determined in collaboration with the instructional cohorts during the planning meeting.

Assessment and Refinement Phase: At the conclusion of the workshop, an survey will be sent to all participants of the pilot workshop, including instructional cohorts and information professionals who attended the workshop, to solicit feedback and input for refinement and enhancements to the curriculum. The results of the survey will then be analyzed for common themes and comments will be considered for integration into the curriculum. An iterative design approach will support applying relevant lessons learned to future workshops.

Dissemination Phase: Building on the dissemination of other digital products generated by the DCN, the new curriculum materials, which may include lecture content for specific data types, example datasets, guided discussions, or other resources, will be made open accessible under a Creative Commons license with relevant citation information provided to ensure instructional teams receive proper credit for their contributions. Links to new content will be integrated into existing DCN education and community materials including the CURATED online modules (to be published summer 2022) and relevant Primers. Portions of the pilot workshop may also be posted to the DCN YouTube channel. Dissemination and knowledge of access to these resources will be communicated via relevant listservs including RDAP, DataCure, RDA Newsletter, Carpentries Clipping, WeHere, IASSIST and others as well as through social media outlets. Likewise, we will share lessons learned regarding this peer-to-peer instructional design method and learn from others engaged in similar work by participating and presenting at relevant conferences.

Implementation Phase: 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. An online meeting during the wrap-up of the grant will be held to discuss plans to move from the planning stage to implementation. While the intention will be to provide additional rich community-based learning opportunities via in-person workshops, by making the educational resources developed during this project openly available for self-guided learning, the outputs of this planning grant will have a broad impact for the library and archives community regardless of future funding.

Timeline

Below is a general timeline that will support the project team in effectively and efficiently completing the proposed project. Note this timeline may shift based upon the COVID situation, particularly the dates of

the meetings as the project team views in-person interactions as invaluable when considering the community building aspects of this project, which are difficult to replicate in an online environment. However the project team also recognizes the need to be flexible to allow a safe working environment for participants and if required will transition the planning meeting to an online format.

- Recruitment and Team Development (August 2022-December 2022)
  - Planning Meeting (December 2022: Washington, DC)
- Curriculum Development (January 2023-September 2023)
- Pilot Workshop (October 2023: Durham, NC)
- Assessment and Refinement (November 2023-February 2024)
- Dissemination (February 2024-July 2024)
- Implementation (July 2024-Onward)

Iterative Analysis

The table below indicates key outputs generated during each phase of the project.

Phase	Outputs and Measures
Recruitment & Selection	- Four instructional team members including individuals both within and external to DCN institutions recruited - Planning and team building meeting completed - Communication channels established
Curriculum Development	Draft curriculum developed for the four data types including teaching modules and datasets for use in hands-on exercises
Pilot	- Twenty community members including four individuals from historically marginalized groups selected for pilot workshop - Pilot workshop completed
Assessment & Refinement	- Post-workshop feedback survey distributed and analyzed for trends - Feedback survey on collaborative process distributed to instructional team members - Refinements to curriculum identified and implemented
Dissemination	- Teaching modules and materials shared via the DCN established online channels - Presentations at relevant conferences
Implementation	Implementation plan generated for future DCN in-person workshops

Diversity Plan

The project team is committed to maximizing the recruitment and retention of members of historically marginalized communities. Recruitment for the instructional cohorts will be advertised broadly to encourage participation from across institutions and individuals with a variety of experience levels and focus areas. During the planning meeting, the project team will hire a skilled facilitator to help create a space where all voices will have room to contribute meaningfully to the discussions, encouraging freedom of expression and sharing of perspectives. During the curriculum development process, the instructional teams will also be encouraged to consider ethical, equitable and accessible considerations for their data type.

For participants invited to the pilot workshop, we will work with established national diversity initiatives, local HBCUs, and smaller institutions to broadly recruit participants with a diversity of experience and perspectives. To ease participation, we will also provide four scholarships for travel for at least four spots for members of historically marginalized communities, early career, non-traditional librarians, paraprofessionals, 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

Academic institutions are increasingly tasked with the stewardship of digital data to respond to a rapidly growing focus on open science, reproducibility, and FAIR data. Building the skills to curate effectively a broad array of data types is essential for supporting the diverse needs of our research community. This project will generate data type-specific educational resources and thereby expand data curation capabilities for librarians and other information professionals engaged in this work through cross-training instructional cohorts, building expertise of pilot workshop attendees, and making openly available online learning modules and training materials (i.e., datasets, etc.). The ultimate result of this project will be enhancing the quality of data curated at institutions.

This project will also 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 formation of instructional cohorts will result in both opportunities for cross-training in different data types during the pilot workshop and spaces to share experiences, lessons learned, and resources. Contributing to a shared goal and project provides a unique opportunity for community building and creating an open network for experiential learning. Asynchronous communication channels (i.e., Slack) used during the project will create a channel for continuing conversations within this knowledge network beyond the conclusion of the grant.

The following performance goals will measure the progress and success in achieving these outcomes: (1) availability of curriculum materials openly shared online as part of the existing CURATED curriculum (2) reuse of resources indicated through downloads and citation counts, and (3) assessments from participants of the pilot workshop and instructional team members, and (4) continuing engagement of the instructional cohorts.

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. The development of proficiency in data curation will help the next generation of library and archives professionals respond to the data deluge. As we move from encouraging data sharing to valuing truly

reusable, interoperable, and broadly accessible data, information professionals are natural partners with researchers and scholars. Experience with various data types and providing opportunities for community–building amongst curators is essential for developing the workforce of the future.

Phase	Aug 2022	Sep 2022	Oct 2022	Nov 2022	Dec 2022	Jan 2023	Feb 2023	Mar 2023	Apr 2023	May 2023	Jun 2023	Jul 2023
Recruitment & Team Development	■	■	■	■								
Planning Meeting					■							
Curriculum Development						■	■	■	■	■	■	■

	Aug 2023	Sep 2023	Oct 2023	Nov 2023	Dec 2023	Jan 2024	Feb 2024	Mar 2024	Apr 2024	May 2024	Jun 2024	Jul 2024
Curriculum Development	■	■										
Pilot Workshop			■									
Assessment & Refinement				■	■	■						
Dissemination							■	■	■	■	■	■

Types

The primary digital products that will be generated during this project include educational curricula for four unique data types including code, geospatial, scientific images, and simulations. The educational materials may include modules detailing suggested curation treatments for these data types as well as example datasets used for hand-on exercises.

Availability

All educational curricula developed during this project will be made openly accessible online and disseminated via alongside the CURATED curriculum (to be published Summer 2022). Links to the online modules will also be made available on the DCN website to allow for ease of access. As appropriate, materials will also be deposited into the University of Minnesota's Digital Conservancy repository within the Data Curation Network collection for long-term preservation and access including the assignment of Dublin Core metadata for discovery.

Access

All curriculum materials will be published and made accessible under a standard Creative Commons license (CC-BY or CC-BY-NC) to allow for broad reuse by the community. A standard citation of the content producers will be provided to facilitate proper acknowledgement by secondary users.

Sustainability

Sustainability of educational materials generated by the Data Curation Network is overseen by the education committee and the governance board of the organization. Curricula materials will be preserved within the University of Minnesota's Digital Conservancy repository to ensure the long-term preservation and ongoing sustainability of these materials.

Duke University traces its origins to a small school that opened in 1838 in Randolph County, North Carolina. In 1892 then Trinity College, came to Durham NC and received a land grant from Washington Duke. In 1924, Trinity College was renamed Duke University following establishment of the Duke Endowment by James Buchanan Duke. Ever evolving, Duke University strives to meet the stated aims of the university: "to foster a lively relationship between knowledge and faith; to advance learning in all lines of truth; to defend scholarship against all false notions and ideals; to develop a love of freedom and truth; to promote a respectful spirit of dialogue and understanding; to discourage all partisan and sectarian strife; and to further the advancement of knowledge in service to society." (Duke ByLaws).

In support of the overall mission of Duke, the Duke University Libraries (DUL) includes The William R. Perkins Library, Bostock Library and Rubenstein Rare Book & Manuscript Library and comprise the main West Campus library complex, which is joined by Lilly and Music libraries on East Campus, and the Pearse Memorial Library at the Duke Marine Lab. DUL provides a place for self-education and discovery, promotes scholarship and good citizenship through information literacy, acquires, organizes, preserves and delivers information resources and assists users in their effective use. As an academic library, DUL serves a wide variety of patrons including undergraduates, graduate students, faculty, and staff at Duke. DUL employees over 300 full-time staff to support the needs of our research community including over 16,000 students and 1,650 faculty.

The Data & GIS services department within DUL was formally established in 2009. Now known as the Center for Data and Visualization Sciences (CDVS), the mission of CDVS is to partner with faculty, researchers, and students through consultation, instruction, co-curricular programming, and research assistance at various stages of the research data lifecycle. Staff in CDVS have expertise in a variety of techniques to support data driven research including data visualization, GIS & mapping, data science, data management, and data curation. CDVS supports the Brandaleone Lab for Duke computing needs, holds on average over 20 open workshops a semester, and partners with a variety of groups both at Duke and beyond to advance computational approaches that support reproducibility and the extension of data in the service of society.