IJDC | *Conference paper*

Developing Specialized Data Curation Curricula to Meet Growing Demands: A Community-based and Evolving Approach

Sophia Lafferty-Hess
Duke University

Seth Erickson
University of California, Santa Barbara

Neggin Keshavarzian
Princeton University

Wanda Marsolek
University of Minnesota

Jennifer Moore
Washington University

Mikala Narlock
Indiana University

Abstract

Data curation is "the encompassing work and actions taken by curators of a data repository in order to provide meaningful and enduring access to data" (Johnston et al., 2018a). It can be multifaceted and complex based on the types of data, the expertise of the curator, disciplinary expectations, and repository policies. With evolving data sharing practices and standards, ensuring data curators and stewards have access to high-quality, extensible instruction on specific data types is essential for supporting the goals of open research and accessible data sharing, particularly in the landscape where funders (National Institutes of Health, 2020) and journals (Naughton & Kernohan, 2016) are mandating data publication for the purpose of reproducibility, reuse, and external validation. In brief, data need to be curated for effective re-use and in alignment with the FAIR principles (Wilkinson et al., 2016). The Data Curation Network (DCN) (Johnston et al., 2018b) has been actively developing education and training programs to expand capacity in data curation along multiple axes. This paper will explore the progression of the DCN's education program based primarily within the United States, highlighting a recent effort to develop specialized data curation education for four specialized data types. We will conclude with lessons learned, reflections on growth of education efforts in the DCN more broadly, and potential next steps.

Submitted 10 February 2025 ~ *Accepted* 20 February 2025

Correspondence should be addressed to Sophia Lafferty-Hess, Email: sophia.lafferty.hess@duke.edu

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Data curation education can trace a lineage from archival theory to digital curation to data management to an emerging field of study in its own right (Noonan & Chute, 2014). Within library and information schools (LIS) the field of data curation has been steadily growing (Harris-Pierce & Quan Liu, 2012; Yoon, Murillo, & Jettpace, 2023); however, there are still many opportunities for continued growth to serve both new and seasoned information professionals being situated to support data sharing requirements. While librarians and data professionals graduating from LIS programs with focuses in data curation may be better equipped with the service-oriented aspects of data curation there are still gaps in gaining technical skills (Yoon, Murillo, & Jettpace, 2023). With these gaps in formal LIS training, continuing professional development and training programs for information professionals in data management and sharing best practices have been developed (Tibbo & Jones, n.d., EDINA and Data Library, 2017, Bakker & Bryant, n.d); the Data Carpentries provide opportunities for skilling up both researchers and librarians in technical skills including programming languages, data cleaning techniques, and database management (Pugachev, 2019); and other programs focus on curation specifically for enabling verification of reproducibility of data, code, and software (Christian, Arguillas, & Peer, 2018).

However, training focused specifically on data curation as performed by information professionals working to prepare and optimize data for sharing via established repositories to meet the FAIR principles has been an area for growth. This type of training necessitates inclusion of both conceptual skills to communicate effectively with researchers and functional skills to review data files coming in a myriad of forms, formats, and disciplines.

The Data Curation Network: A Community of Practice

Given the complexity of these tasks and the need to build community around data curation, the Data Curation Network (DCN) (Johnston et al., 2018b) was built to support the needs of professional data curators and stewards in the United States. Through generous funding from the Alfred P. Sloan Foundation, which supported the planning and implementation of the DCN, the DCN transitioned to be a member-funded organization in July 2021. While the DCN hosts numerous offerings, including educational opportunities described below, sustaining institutional members have access to a cross-institutional data curation workflow (Coburn & Johnston, 2020). Through this model, members can request expert curation of a dataset from a data curator at a different institution. This is facilitated by DCN staff, who monitor submissions, ensure deadlines are met, and track overall metrics on the types of datasets curated by the DCN against the expertise of data curators.

One of the guiding principles of the DCN is the power of community and one of the strengths of the DCN is the ability to practice "radical interdependence." This mechanism of collaborating moves beyond cooperation and recognizes the interconnectedness of the work we are all doing: "[Radical interdependence] is about collaboration, trust, vulnerability, and accountability; not from a punitive perspective, but from an intrinsic desire to show up for and serve one another based on clear expectations, commitment, and the shared goal of advancing the field" (Carlson et al., 2023).

The DCN Education Committee, which was formalized in the DCN Governance structure in 2021, helps guide the strategic development and implementation of the program (DCN, 2021). The Committee provides feedback on associated grant funded projects, teaches CURATED workshops, and advises on other initiatives related to the DCN education program. The DCN has been developing curation education materials for more than seven years at this point and continues to grow and develop a structured education program. Throughout the periods of development, we have relied on the guiding values and central principle of radical interdependence to launch and refine educational offerings.

Building the CURATED Training

In 2017, members of the DCN piloted a workshop applying the CURATED steps to conceptualize the curation of data (Moore, 2018). Funded by IASSIST, this event led to an implementation grant from the Institute of Museum and Library Services (IMLS) grant [#RE-85-18-0040-18] to build a full two-day workshop offering and deliver it across the United States. The CURATED steps guide learners through what might be involved in the curation process, including:

- Check files and read documentation
- Understand data (or trying to)
- Request missing information
- Augment metadata for findability
- Transform file formats for reuse
- Evaluate for FAIRness
- Document all curation activities[1]

The CURATED steps were based upon research performed by Johnston et al. (2018b) where researchers identified important curation activities and their satisfaction with those activities, which led to several high priority areas for data curation in practice (Hudson-Vitale et al., 2020). While these steps are presented linearly, they can often be iterative and serve as the foundation for data curation education as well as the curation processes performed by member institutions within the DCN.

The CURATED workshops were initially designed as a two-day in-person workshop that intermingled lecture, peer-to-peer discussion, and hands-on activity with datasets in different formats including checklists at each step that reinforce curation actions. Key learning objectives for the workshops included increasing the understanding of data curation practices and tools, sharing expertise, and meeting like-minded colleagues interested in building and extending curation practices. Workshops also included a capstone project where participants formed groups to develop a "primer" over the next 6 months with a DCN mentor and a peer review process built into the development process. Primers are a community-based resource aimed at supporting the curation of a particular data type, format, or concept in data curation.

Outcomes from CURATED workshops

From 2017-2022, the DCN held 6 in-person workshops and engaged with more than 250 data professionals through in-person and online training using the CURATED framework. In all the workshops, the DCN has emphasized collaboration and conversations amongst peers, which has led to further educational developments to benefit the broader data curation community. Some of the results of the workshops and community engagement include:

- 47 published primers on a range of topics[2]
- The creation of a set of self-guided online modules based on the CURATED steps[3] (Blake et al., 2022)

[1] Data Curation Network CURATED Steps: https://datacurationnetwork.org/outputs/workflows/
[2] Data Curation Network Primers: https://datacurationnetwork.org/outputs/data-curation-primers/
[3] CURATED online modules: https://datacurationnetwork.github.io/CURATED/

- The development of an abbreviated 3-hour long training that has been provided to a virtual audience expanding the reach and impact

While DCN education is primarily based within the United States, there have been several opportunities to bring this training to an international audience. In 2024, members of the DCN collaborated with the Digital Research Alliance of Canada to offer an abbreviated CURATED workshop at IASSIST. A key benefit of this experience was attempting to capture some of the complexity of working with and curating data from an international context as well as providing some written materials in French (Clary, 2024). This workshop built on the previous translation of the CURATED steps into French (CURATION) developed by the Portage Network for the Canadian Dataverse community (Cooper et al., 2021).

These experiences have not only been useful to further develop and formalize the DCN's role in providing education for data curators, but also reaffirmed the importance of relying on the community of experts to pool our knowledge and resources.

Building Specialized Data Curation Trainings

Building on these foundational concepts, members of the DCN, including Duke University (lead), University of California - Santa Barbara, Washington University in St. Louis, University of Minnesota, Princeton University, and the Association of Research Libraries, have been working on an IMLS-funded pilot project [#RE-252343-ols-22] to expand the foundational curricula for four specific data types: geospatial data, scientific images, code, and simulations-based research. The project drew together data curators and information professionals from across the United States to create in-depth training resources for these four specialized data types, offer a pilot workshop, and refine outputs for distribution.

This project selected the four data types based on analysis of curation requests submitted to the DCN (from 2018-2022) indicating some of the most frequently submitted datasets by data type include code, simulations, scientific images and geospatial data (see Table 1), a trend which has continued in submissions since 2023-2024 (Data Curation Network, 2024). The higher frequency of these datasets requiring DCN curation support and the lower percentage of curators (2021-2022) with this expertise reflect areas for additional support for curating research data.

Table 1. 2018-2022 percentage of DCN curation requests per data type compared to curator expertise as of January 2022 and a total of 265 datasets.

Data Type	DCN Curation Requests	DCN Curator Expertise
Code	31.6%	ranging from 52% for R to 7% for C++
Simulations	10.9%	16%
Scientific Images	6.4%	19%
Geospatial	6%	33%

Instructional Cohorts

In order to develop these specialized data curation curricula and recognizing that the DCN had limited expertise in curating these data types, the project employed a community-driven approach. This approach aimed to develop four cohorts, one for each data type, that would include a

mentor who was part of a project design team. These cohorts would work collaboratively to develop the curricula for their data type over a 6–8-month period. A project design team guided the project and included a Project Coordinator, two liaisons to the DCN Education Committee (the current Director of the DCN and the committee chair), and all four mentors. The project was supported by an advisory board including individuals that developed the original CURATED training as well as representatives from new DCN member institutions. The approach centered on the importance of communication both between the mentors, the members of the cohorts, and the overall project design team. This communication was enabled by an in-person kick-off planning meeting, virtual meetings across cohorts to share information, engagement on the DCN Slack channel, and a pilot workshop.

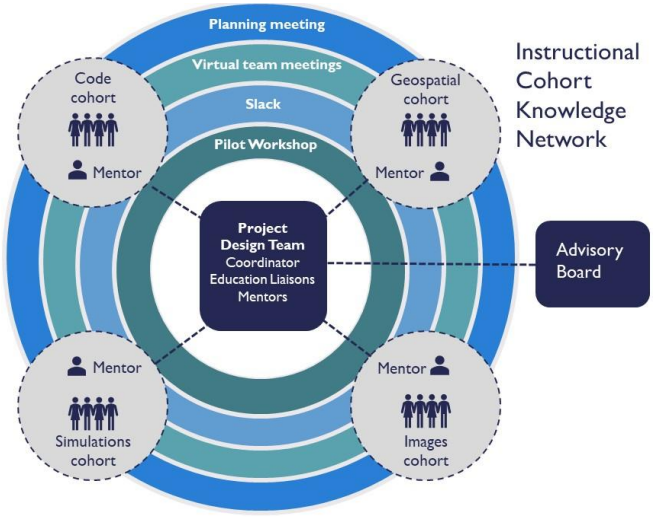

Figure 1. Structure and communication channels for the instructional cohort approach.

Recruitment

Recruitment for the cohorts began with an open call for volunteers for instructional cohort members in October 2022 that was sent to various listservs used by information and data professionals. This included RDAP, IASSIST, DataCure, DLF, MLA, and RDA-US, as well as posts on the Data Curation Network blog. The call aimed to center the benefits to the applicants including travel costs for attending the in-person events; a modest stipend; and citable products for demonstrating contributions to the field. A short statement of interest was requested including information regarding their experience with the data type they selected for consideration and their experience with instruction and training. The project design team then individually ranked each volunteer through anonymous scoring and made final determinations for offers to participate as a group. The final cohorts included individuals with a range of experience levels but primarily coming from the data management, librarian, and archive fields of study.

Kick-off Meeting

The project design team and the cohorts first came together at a 1.5 day in-person kick-off meeting in January 2023 in Washington DC. The meeting included a facilitator to encourage full participation across attendees. Two noteworthy activities that participants commented upon in feedback included an initial team-building exercise where each cohort developed a "team mascot"—this exercise helped setting a congenial tone that continued throughout the meeting. Attendees also participated in an exercise that encouraged individuals to reflect on both project level and individual ways we might measure our success. A key benefit of hosting a kick-off meeting was the ability to get cohort members on the same page, establish shared goals and expectations, and generally form relationships between teams. We recognized the value of

community and collaboration in previous DCN education efforts and wanted to be sure that these cohorts had the same chance to connect as colleagues.

Curricula Development

In order to facilitate, streamline, and ensure consistency across data types as teams developed their curriculum, each cohort was provided a semi-structured Lesson Plan template. Outside of this, though, the cohorts largely had freedom to determine the development of learning content and materials. While encouraged to use the CURATED framework where appropriate, without simply repeating the framework, this was not a requirement, and different cohorts approached the integration of CURATED in different ways. A short overview of each cohort's approach and some unique challenges for those data types are discussed below.

Code

The code cohort consisted of librarians and research data specialists with backgrounds in data curation and research software development. In shaping the curriculum, the cohort chose to focus on foundational concepts needed to work effectively with code-intensive datasets. The goal was less to teach programming than to equip curators with the vocabulary and "interactional expertise" needed to document and improve code written by scientists (Collins, Evans, & Gorman, 2007). The learning outcomes developed by the cohort emphasize activities like identifying, explaining, and documenting aspects of research software most relevant to reproducibility and reuse. Existing frameworks, including the FAIR Principles for Research Software (Hong et al., 2022) significantly informed the curriculum.

The cohort identified foundational concepts and used these to organize the lesson materials; the resulting curriculum includes modules on computing platforms, programming languages, dependencies, and documentation (Cooper, Janée, Maye, Ruhs, Erickson, 2024). Additional modules on licensing and project structure are planned for the future. The material for each lesson was largely developed by individual cohort members, with input and feedback from other members of the cohort and other cohorts. The workshop concludes with a hands-on exercise, reviewing a code-intensive dataset previously submitted to the DCN (Bump et al., 2019).

Simulation-based research

The simulation cohort included a spectrum of expertise. Two of the cohort members identified as Research Data Librarian or Specialist, two as Engineering liaisons, and one as Mapping & Geospatial Data Librarian. All brought expertise in data management and consulting to some degree. Of the five-person cohort, three had experience curating data in general. Of the three with previous experience, two people had experience with simulation data either creating, analyzing or curating. Those who had more experience helped the others catch up and those with less experience provided space to slow the others down (in a good way) to help make connections and provide context—the cohort had a built-in peer review system in this way.

Bringing together a cohort of members from across the U.S. can be difficult when working across time zones and existing responsibilities. Many meetings did not have full attendance but those present still worked to move the project forward. Towards the end of curricula development time, more and longer meetings were added with the idea that members would attend when it worked for them and could attend some part of the meeting. Different cohort members led the meetings. This was helpful to distribute the weight of leading a meeting and encouraged participation (knowing how hard it is to run a meeting sometimes can help aid in participation versus when one is 'just' a member of the meeting).

Developing a workshop on good practices for curating simulation research, that can be held within 3 hours, can be overwhelming. The first major task was to come to an understanding together of what would be in scope. Once the cohort identified what was out of scope, learning objectives were developed defining expectations and what learners would walk away from this workshop with. Much of the curriculum focused first around how to identify simulation research data and what simulation data consists of. To do this the cohort relied heavily on a seminal article on the topic (Mullendore, Mayernik, & Schuster, 2021). The next priority was how to make said

data understandable and reusable focusing primarily on documentation and metadata. The cohort developed a glossary of common terms associated with simulation data research as well as a helpful metadata checklist to help guide the learners (Wynholds, Shimon, Rios, Misgna, Marsolek, 2024). During the workshop, attendees worked through a couple of datasets in different states of curation - this helped learners see areas of the dataset that could be improved.

Geospatial data

Geospatial cohort members brought a wealth of knowledge and experience in GIS, geography, and data curation, respectively. Given that, parsing out what is critical to know about using geospatial data versus curating geospatial data for curriculum development was nontrivial. Therefore, the curriculum development plan evolved over months of drafting and paring down content. Because reviewing geospatial data often requires complicated, specialized software, the cohort knew the curriculum had to include orientation on those tools. Therefore, there are three areas of concern: a basic understanding and identification of geospatial elements; what a curator should look for when presented geospatial data; how to use the tools to effectively review it. The curriculum is broken into: i) environmental setup, which is focused on QGIS, an open source tool; ii) introduction to GIS, which has no shortage of content; iii) ethics and GIS data, which describes how GIS data might need to be adjusted or restricted to protect people or sites; iv) common GIS types, raster, vector, and containers are addressed; v) GIS metadata, which focuses on reproducibility and usability; vi) transformations, when to and when not to, advantages and disadvantages (Christiansen, Grove, Kernik, Norris, & Moore, 2024). Through this shared understanding, the cohort developed the stated learning objectives including recognizing geospatial data terms, recognizing essential geospatial metadata elements, differentiating between geospatial file types, using basic transformation tools, and understanding ethical, equitable, and accessible considerations.

Pacing the curriculum was tricky; GIS tools and data for complete beginners often present a steep learning curve and the missteps can take a curator into a loop of failure. The experience of getting comfortable in this space can be misleading for experts who have been using geospatial data and tools for a long time, which was the case for all but one of the cohort members. Plenty of time must be built into the curriculum to support learning by failing, a roaming instructor (or more depending on the group size) needs to be on deck to help folks out when they hit barriers. To orient participants on the software, a DRUM dataset was supplied, "Data supporting the comparison of golden-winged warbler and American woodcock productivity in northern Minnesota, USA," in shapefile format, which is a ubiquitous, commonly used file format that a curator might encounter (Kramer et al., 2019). Shapefiles are notoriously difficult since the .shp file is accompanied by several other file dependencies. This file is also used in a readme file exercise and a transformation exercise and provides ample opportunity for demonstrating challenges and approaches in geospatial curation. Although this format was the central tool leveraged for the curriculum, the team discussed and built in content to orient participants to other file types they might encounter. Building exercises like those in the curriculum using other geospatial file types is certainly attainable.

This high-functioning cohort worked well together, met regularly, and shared the burden of developing the curriculum as evenly as possible. They balanced a holistic presentation of key information with a modular approach full of examples, exercises, and engagement. Each module references the CURATED model, to anchor participants to when and why they might make use of the information.

Scientific images

A challenge from the beginning was figuring out how to scope and focus the data type more. Considerations for curating scientific images can vary depending on the field of research and instrument producing the image. The first few months the group tackled defining what a scientific image is and scoping the focus of the workshop. The group used the CURATED steps to explore some of the complexity around curating scientific images as well as training on common image curation tools, particularly ImageJ (Fiji) (Gignac, Wright, Kenney, Schuler, & Keshavarzian, 2024).

The cohort used an example dataset that had been submitted to the DCN to focus the workshop lectures and activities (Moniri et al., 2020). Attendees were able to see the steps the original curator took in curating that dataset to the published version. The cohort approach was to have each cohort member focus on one of the CURATED steps and apply that step to the example dataset. From this, the team developed the lecture slides, small group discussion prompts and activity worksheets. One activity that was developed was a participant-focused role-playing exercise on the "Request" step of CURATED that used an AI chatbot. The learner just pastes the instructions to a chatbot of their choice and then asks "Dr. Roe Bott" questions about the dataset that they are curating.

Peer Review

After approximately five months of development, each cohort submitted their materials for peer review. This is a standard process in the development of DCN primers: by reviewing these community-based resources from different perspectives, we can enhance the usability and utility of our resources. Unlike other peer review processes, the DCN prefers materials to be in a mostly, but not fully, developed state. This provides opportunities for intervention prior to investing a significant amount of time.

In the Summer of 2023, every member participated in this review process to receive and provide feedback. Each data type was reviewed by one person from each of the other cohorts (e.g., the code content was reviewed by a geospatial, simulations, and scientific images cohort member). Prior to submitting the materials for review, team members could leave specific comments or questions that they would like input on. Reviewers were able to provide both in-line feedback as well as general feedback by an anonymous form; a focus on the importance of constructive feedback was stressed by the design team and reflected by the peer reviewers.

This review process was useful for not only improving the curricula, but also allowing for cross-learning between the cohorts. By having a better understanding of what the other curriculum development teams were teaching, the activities for attendees, and suggested resources in each curriculum, it helped to inspire additional changes for each cohort.

Piloting Materials

The curriculum materials were piloted at a 2-day in-person workshop in October 2023 at Duke University. The cohorts, core project team, and 15 other archive and library professionals came together to test out the new curriculum materials. Participants from the archive and library community were recruited via an open call to the listservs mentioned above. Applications were scored by the project design team using a standard DCN Workshop Application Rubric (Wright, 2023). Pilot workshop participants were provided a small stipend for their contributions as well as four travel scholarships awarded to diversify participation.

Each cohort was given 2.5-3 hours to teach the materials of their choosing. Feedback was collected from all the participants both using a paper formative feedback form focused on each individual module, completed after each module to ensure the content was fresh on participants minds, and a final aggregate summative survey sent after the completion of the workshop (see Appendix). The feedback on the individual modules were collated into a digital format and shared with respective cohort members in order to further refine their materials before project end in Spring 2024. This method of having peer instructors observe each other's training and provide feedback has been used as pedagogical practice to enhance instruction and teaching (Alabi et al., 2012; Yiend, Weller, & Kinchin, 2012).

Dissemination and Outcomes

A goal of the project has been to make any materials generated available to the broader community for reuse. The core materials generated and made available to the community include:

- A lesson plan including learning objectives, key terminology, dataset citations, and content outlines
- Slide decks used in instruction (with speaker notes)
- Associated learning materials including templates, exercises, and training datasets

The dissemination of the materials includes a GitHub repository containing the materials generated for the Specialized pilot workshops[4] in addition to the CURATED workshop slides.[5] In addition, the materials are archived with the University of Minnesota Digital Conservancy Repository DCN Collection providing a DOI for the resources for stable citation and attribution to instructional cohorts.[6] A core audience for these materials include those within the community who will continue to enhance and provide specialized training in collaboration with the DCN.

Throughout this pilot project, the design team prioritized data and feedback to improve not only the modules, but also the curriculum development process. Regarding the pilot workshop (see Appendix for aggregate results), 29 participants out of 31 completed the final feedback survey. When assessing the success of the workshop 16 agreed and 12 strongly agreed that the workshop better prepared them to approach the curation of these unique data types (see Figure 2). When asked about their general satisfaction with the pilot workshop, 17 were extremely satisfied and 12 were satisfied (see Figure 3).

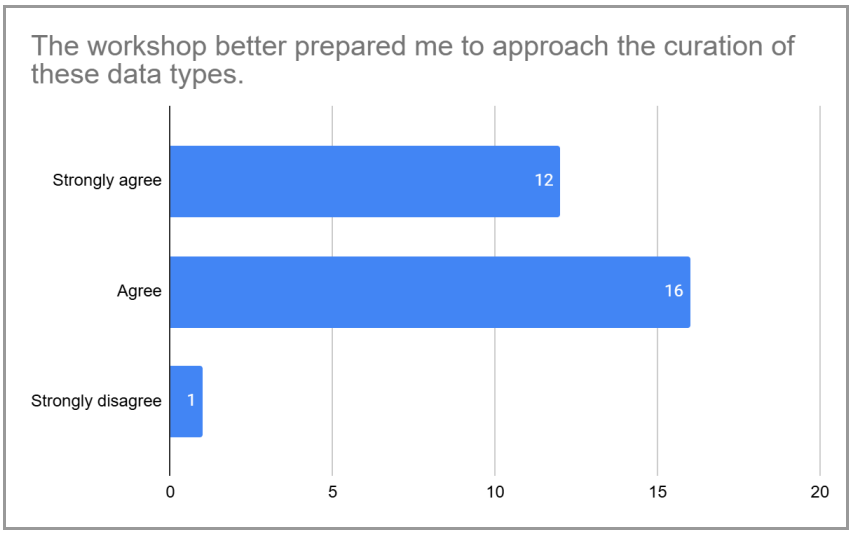

Figure 2. Results of pilot workshop summative survey regarding how well the workshop prepared attendees to approach the curation of these data types.

[4] DCN Specialized Data Types GitHub repo: https://github.com/DataCurationNetwork/curation-curriculum/tree/main/Specialized%20Data%20Types

[5] CURATED Fundamentals GitHub repo: https://github.com/DataCurationNetwork/curation-curriculum/tree/main/Curation%20Fundamentals%20Workshop

[6] DCN Education Archived Collection: https://hdl.handle.net/11299/265761

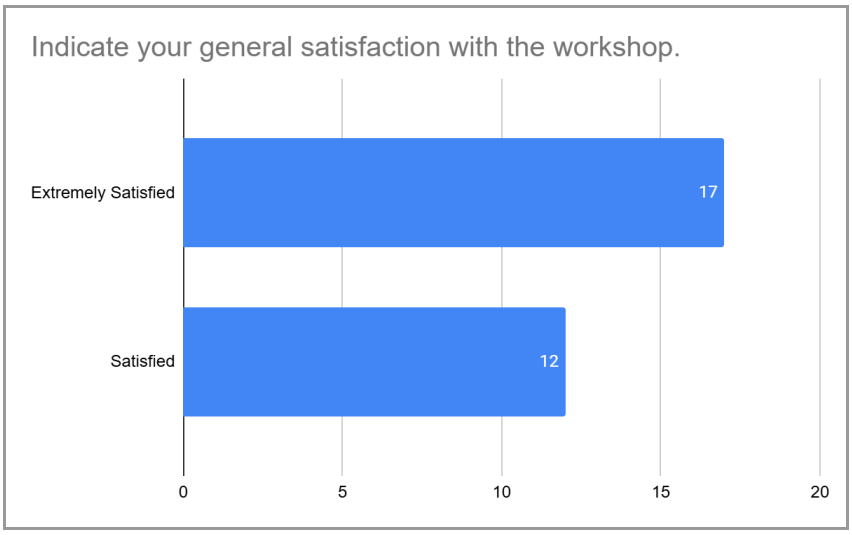

Figure 3. Results of pilot workshop summative survey regarding general satisfaction with the workshop.

Additionally, cohort members were surveyed to learn more about their experience to support the DCN iteratively enhancing and evolving our education approach in the future (see instrument in Appendix). Since we wanted cohort members to feel confident that their answers would remain anonymous, and therefore their feedback can be as positive or negative as possible, we are sharing only a portion of the feedback below in the form of quotes and broad lessons learned. While there was room for improvement in the cohort structure and processes, discussed below, numerous cohort members highlighted in their feedback that the peer review process aided them in curriculum development. Cohort members also valued their mentor individually and as a liaison to the broader project and added that they were grateful to have been part of the project.

Lessons Learned: Reflecting and Moving Forward

In order to ensure we are meeting the needs of the DCN, as well as those of the larger data management community, we are engaging in "reflection as a practice." As Donald Schön defined in 1983, "reflective practice is the ability to reflect on one's actions so as to engage in a process of continuous learning" (Schön, 1983). Our approach to continuous learning includes structured peer review processes, feedback from the workshop participants, a cohort member survey, and self-reflection. Below we reflect on some initial lessons learned as we consider the progression of the DCN's education program, highlighting qualitative data from cohort members and pilot workshop participants.

Above all, one of the guiding principles of the DCN is **the power of community** and the ability to expand capacity via harnessing the expertise broadly found across the library and archive profession. This unique position and perspective of the DCN was also reflected in a pilot workshop participant's comment:

> 'For me, this workshop showcased the DCN's unique position and ability to bring experienced practitioners and novice curators together in a joint learning experience, and to advance the tools, techniques, and principles of data curation.'

However, one of the things we have also learned is that bringing people together is only one part of developing shared resources. **Structures and scaffolding of deliverables** are also essential. This is an area where we see room for improvement. As one cohort member noted:

> 'Overall this was a wonderful experience. There were times that the communication could have been better, just some confusion about documents and how to structure some things.'

The pilot project provided an opportunity to begin to develop more structures that can then be iterated upon and improved as we move toward further implementation of specialized data curation workshops. Moving forward, we will expand not only our communication strategy, but also the templates and expectations, which some cohort members could have benefitted from.

When considering a pedagogical approach to teaching more in-depth content on specific data types, we also learned that **time is an asset** and more time for learning is valued in synchronous learning environments. As one pilot workshop participant wrote:

> 'There simply wasn't enough time. Each topic was given 2.5 hours, but I'm thinking some of the topics would have benefited from having 4 hours, or even 5 hours. 5 hours (with suitable breaks) would have given the chance for participants to engage more deeply and more hands-on with a topic, especially one as complex as geospatial.'

Given this feedback, future iterations of these workshops will not include all four data types, instead pairing two data types together in a 1.5 to 2-day workshop such as code with simulations and geospatial data with scientific images. This new iteration of the workshop materials is being integrated into a new workshop series in Spring 2024-2025 that is in partnership with the National Institutes of Health.[7] This next workshop series will provide an opportunity to refine the materials with an eye on creating a cohesive learning experience with two complimentary data types.

Considering the workshop wholistically, one other area where more concentrated work could have been focused was the **harmonization of materials across data types**. While certain templates and expectations were set, including the use of a standard lesson plan, branded DCN slide decks, and a requirement to use datasets in hands-on learning, instructional cohorts were provided agency to develop the content how they saw fit. During an organized share-out across cohorts, potential overlaps in concepts and how that might be handled was discussed. For instance, during the simulation session the instructors did not go deep into the general concepts for code curation (such as dependencies) and instead referenced the upcoming code session. The peer review process also aided groups in having a better shared understanding of what would be covered by other teams, lessening the potential for duplicate information being presented. Generally, the materials for each data type were designed for a cohesive 2-5 hour long in-person training session; however, in the future there is an opportunity to "modularize" the content, which could aid in mixing and matching concepts that may span data types and support the development of asynchronous learning modules.

As the DCN looks towards the future of our education program, there are a number of key priorities and goals that will guide our development. First, and foremost, is **valuing the people** that commit their time and resources into contributing to data curation initiatives. The cohort model with a guiding mentor has been successful for generating shared resources both with our primers and now with instructional materials. The DCN has also attempted to provide incentives to contributors including stable citations to materials, sponsoring presentations to the community, attending in-person events to connect with colleagues, authoring blog posts, and providing stipends to content developers. Providing incentives including monetary and professional development opportunities is one mechanism to demonstrate the value of the people needed to develop shared resources. However, it would be useful to formalize and document this model with more explicit descriptions of roles and responsibilities, benefits to contributors, and opportunities for future engagement.

Second, is to focus more explicitly on **sustainability and maintenance of materials**. Both our primers and instructional materials were created by community volunteers; therefore, we have not

[7] Spring-Summer 2025 NIH and DCN Data Curation Training Opportunities: https://datacurationnetwork.org/spring-summer-2025-nih-and-data-curation-training-opportunities/

required ongoing contributions past the publication date. However, as technology changes and curricula grow and adapt, a more formal approach to the maintenance of instructional materials is needed. One model worth examining, is the "maintainers" approach used by The Carpentries. Additionally, as materials are updated or adapted, ensuring proper versioning and credit for future contributors will be important.

Finally, while the audience for our current education offerings has primarily been information professionals, there is an opportunity to **engage more directly with researchers**. While CURATED is generally focused on information professionals curating data coming into a repository, there are potentials for broader applications. For instance, Duke University has used CURATED in their Responsible Conduct of Research program for graduate students (Lafferty-Hess & Darragh, 2019). There is a huge potential impact to collaborate with disciplinary societies and individual researchers to curate data as it is produced.

Conclusion

As the need for data curation education continues to grow in response to open science policies and mandates, the DCN remains committed to refining educational offerings, prioritizing sustainability, and fostering a collaborative environment that values the expertise and contributions of community members. The DCN's educational efforts have evolved over the years and will continue to do so through community-driven approaches. By continuing to reflect on DCN practices, adapt to feedback, and embrace the principle of radical interdependence, the DCN aims to further enhance the field of data curation and support the goals of open research and accessible data sharing. The lessons learned from this pilot project will inform future iterations of workshops and curriculum development to continue to meet the evolving needs of data curators and stewards in an increasingly data-driven research landscape.

Acknowledgements

The development of Data Curation Network education materials has been supported in part by the Institute of Museum of Library Services [#RE-85-18-0040-18 and #RE-252343-OLS-22]. The authors would also like to thank Lisa Johnston, Cynthia Hudson-Vitale, Briana Wham, members of the Data Curation Network Education Committee, attendees of the pilot workshop, and the Data Curation Network generally for their feedback and engagement with the DCN education program. Finally, the authors would like to acknowledge the instructional cohorts (listed alphabetically):

- Leighton Christiansen (Geospatial)
- Talya Cooper (Code)
- Paul Gignac (Scientific Images)
- Kelly Grove (Geospatial)
- Mariah Kenney (Scientific Images)
- Melinda Kernik (Geospatial)
- Greg Janée (Code)
- Kay P Maye (Code)
- Girmaye Misgna (Simulations)

- Timothy Norris (Geospatial)
- Fernando Rios (Simulations)
- Nick Ruhs (Code)
- Amy Schuler (Scientific Images)
- Heather Shimon (Simulations)
- Sarah Wright (Scientific Images)
- L. Wynholds (Simulations)

Without the essential contributions of the instructional cohorts the development of the specialized data curation materials would not have been possible.

References

Alabi, J., Huisman, R., Lacy, M., Miller, W., Snajdr, E., Trinoskey, J., & Weare, Jr., W. (2012). By and for us: The development of a program for peer review of teaching by and for pre-tenure librarians. *Collaborative Librarianship, 4*(4), 165–174. doi:10.29087/2012.4.4.06

Bakker, C. & Bryant, P. (n.d.). *Introduction to research data management.* Medical Library Association. Retrieved from https://www.mlanet.org/courses/introduction-to-data-management/

Blake, M., Borda, S., Carlson, J., Darragh, J., Fearon, D., Hadley, H., Herndon, J., Johnston, L., Kalt, M., Kozlowski, W., Lafferty-Hess, S., Moore, J., Narlock, M., Scott, D., Hudson-Vitale, C. Wham, B., & Wright, S. (2022). *CURATED training*. Data Curation Network. Retrieved from https://datacurationnetwork.github.io/CURATED/

Bump, J. K., Beyer, D., & O'Neil, S. (2019). *Code, data, and metadata document for the manuscript: Territorial landscapes: incorporating density-dependence into wolf resource selection study designs* [Data set]. Data Repository for the University of Minnesota. doi:10.13020/s40h-fv72

Carlson, J., Narlock, M., Blake, M., Herndon, J., Imker, H., Johnston, L., ..., Wright, S. (2023). *The art, science, and magic of the Data Curation Network: A retrospective on cross-institutional collaboration.* Ann Arbor, MI: Michigan Publishing Services Maize Books, p. 6. doi:10.3998/mpub.12782791

Christian, T., Arguillas, F., & Peer, L. (2018). *Curating for reproducibility: Producing high quality data and code for transparent and reproducible research*. Workshop at International Digital Curation Conference IDCC18, Barcelona, Spain. Retrieved from https://osf.io/j3rpv/

Christiansen, L., Grove, K., Kernik, M., Norris, T., & Moore, J. (2024). *Geospatial curation curriculum.* University of Minnesota Digital Conservancy. doi:10.24926/2024.265767

Clary, E. & Narlock, M. (2024, August). *Information exchange and cross-pollination of idea: Reflections on the IASSIST & CARTO 2024 Workshop*. Data Curation Network Blog. Retrieved from https://datacurationnetwork.org/2024/08/06/knowledge-exchange-and-cross-pollination-of-ideas-reflections-on-the-iassist-carto-2024-workshop/

Collins, H., Evans, R., & Gorman, M. (2007). Trading zones and interactional expertise. *Studies in History and Philosophy of Science Part A*, *38*(4), 657–66. doi:10.1016/j.shpsa.2007.09.003

Coburn, E. & Johnston, L., (2020). Testing our assumptions: Preliminary results from the Data Curation Network. *Journal of eScience Librarianship, 9*(1), 4. doi:10.7191/jeslib.2020.1186

Cooper, A., Steeleworthy, M., Paquette-Bigras, È., Clary, E., MacPherson, E., Gillis, L., & Brodeur, J. (2021). Creating guidance for Canadian Dataverse curators: Portage Network's Dataverse curation guide. *Journal of eScience Librarianshi*p, 10(3), 1201. doi:10.7191/jeslib.2021.1201

Cooper, T., Janée, G., Maye, K. P., Ruhs, N., & Erickson, S. (2024). *Code curation curriculum*. University of Minnesota Digital Conservancy. doi:10.24926/2024.265764

Data Curation Network. (2021). *Data Curation Network governance model*. University of Minnesota Digital Conservancy. Retrieved from https://hdl.handle.net/11299/225742

Data Curation Network. (2024). *Annual report 2023: End of year highlights of the Data Curation Network.* University of Minnesota Digital Conservancy. Retrieved from https://hdl.handle.net/11299/260285

EDINA and Data Library. (2017). *Research Data Mantra (Online Training Units & Data Handling Tutorials)*. University of Edinburgh. Zenodo. doi:10.5281/ZENODO.1035217

Gignac, P., Wright, S., Kenney, M., Schuler, A., & Keshavarzian, N. (2024). *Scientific images curation curriculum.* University of Minnesota Digital Conservancy. doi:10.24926/2024.265766

Harris-Pierce, R. L., & Quan Liu, Y. (2012). Is data curation education at library and information science schools in North America adequate? *New Library World, 113(11/12),* 598–613. doi:10.1108/03074801211282957

Hong, C., Katz, N., Baker, D., Barker, M., Lamprecht, A., Martinez, C., Psomopoulos, F., Harrow, J., Castro, L., Gruenpeter, M., Martinez, P., Honeyman, T., Struck, A., Lee, A., Loewe, A., van Werkhoven, B., Jones, C., Garijo, D., Plomp, E., & Genova, F. ...Yehudi, Y. (2022). FAIR Principles for Research Software (FAIR4RS Principles) (1.0). RDA FAIR4RS Working Group. Zenodo. doi:10.15497/RDA00068

Hudson-Vitale, C., Hadley, H., Moore, J., Johnston, L., Kozlowski, W., Carlson, J., Blake, M., & Herndon, J. (2020). Extending the research data toolkit: Data curation primers. *International Journal of Digital Curation, 15*(1), 14. doi:10.2218/ijdc.v15i1.713

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2018a) How important is data curation? Gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly Communication, 6*(1). doi:10.7710/2162-3309.2198

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2018b). Data Curation Network: A cross-institutional staffing model for curating research data. *International Journal of Digital Curation, 13*(1), 125–140. doi:10.2218/ijdc.v13i1.616

Kramer, G., Peterson, S., Daly, K., Streby, H., & Andersen, D. (2019). *Data supporting the comparison of golden-winged warbler and American woodcock productivity in northern Minnesota, USA* [Data set]. Data Repository for the University of Minnesota. doi:10.13020/znag-tn48

Lafferty-Hess, S. & Darragh, J. (2019). Data Management 201: Preparing for Data Publishing [slides]. Retrieved from https://duke.box.com/s/0h6g5464h1gjdhoe11lzi7ltcocapx3d

Moore, J. (2018). *IASSIST/DCN Data Curation workshop*. IASSIST Blog. Retrieved from https://iassistdata.org/blog/2018/02/22/iassistdcn-data-curation-workshop/

Moniri, S., Bale, H., Volkenandt, T., Wang, Y., Gao, J., Lu, T., Sun, K., & Shahani, A. J. (2020). *Dataset for 'Multi-Step crystallization of self-organized spiral eutectics'* [Data set]. University of Michigan - Deep Blue Data. doi:10.7302/day1-6d63

Mullendore, G. L., Mayernik, M. S., & Schuster, D. C. (2021). Open science expectations for simulation-based research. *Frontiers in Climate, 3*, 763420. doi:10.3389/fclim.2021.763420

National Institutes of Health (NIH). (2020). *Final NIH policy for data management and sharing.* Office of The Director, National Institutes of Health. Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

Naughton, L., & Kernohan, D. (2016). Making sense of journal research data policies. *Insights, 29*(1), 84–89. doi:10.1629/uksg.284

Noonan, D., & Chute, T. (2014). Data curation and the university archives. *The American Archivist, 77*(1), 201–240. doi:10.17723/aarc.77.1.m49r46526847g587

Pugachev, S. (2019). What are "The Carpentries" and what are they doing in the library? *Portal: Libraries and the Academy, 19*(2), 209–214. doi:10.1353/pla.2019.0011

Schön, D. (1983). *The Reflective Practitioner: how professionals think in action.* London: Temple Smith: 102-104.

Tibbo, H. & Jones, S. (n.d.). *Research data management and sharing*. Coursera. Retrieved from https://www.coursera.org/learn/data-management

Wright, S. (2023). *DCN workshop application rubric.* University of Minnesota Digital Conservancy. Retrieved from https://hdl.handle.net/11299/252652

Wilkinson, M. D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., ..., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. doi:10.1038/sdata.2016.18

Wynholds, L., Shimon, H., Rios, F., Misgna, G., & Marsolek, W. (2024). *Simulations curation curriculum*. Retrieved from the University of Minnesota Digital Conservancy, doi:10.24926/2024.265765

Yiend, J., Weller, S., & Kinchin, I. (2014). Peer observation of teaching: The interaction between peer review and developmental models of practice. *Journal of Further and Higher Education, 38*(4), 465–484. doi:10.1080/0309877X.2012.726967

Yoon, A., Murillo, A. P., & Jettpace, T. (2023). Data curation education: Cross-disciplinary analysis of master's programs. *Journal of Education for Library and Information Science, 64*(4), 476–511. https://doi.org/10.3138/jelis-2022-0025