

ENSURING LONG-TERM REUSABILITY AND REPRODUCIBILITY:

COLLABORATIVE CURATION FOR FAIR DATA

Through the Data Curation Network (DCN), members enable findable, accessible, interoperable, and reusable (FAIR)¹ data through a shared curation model, education and outreach efforts, and research and advocacy. This work exists within a member-funded and member-driven organization, with a focus on sustainability and long-term growth. Members of the DCN help shape the future of data curation and enable FAIR data by sharing best practices, collaboratively addressing shared challenges, empowering and educating one another, and advocating for data curation broadly.

Background on the Data Curation Network

The DCN is a community-led network of curators advancing open research by making data more ethical, reusable, and understandable. Through the network, members have collectively curated datasets, shared invaluable insights, developed educational resources, and launched a robust community of practice around data curation. The work of the DCN was supported from 2016 - 2021 by the Alfred P. Sloan Foundation, and transitioned to a member-funded organization in July 2021.

Research & Advocacy

The Association of Research Libraries (ARL) led Realities of Academic Data Sharing (RADS) Initiative includes six project researchers who are also DCN members.² The DCN, as an active community of practice, engages with issues around data curation, making data and metadata FAIR, and managing and understanding the research data lifecycle. RADS leans on DCN members' pool of expertise to explore data sharing activities and infrastructure required for data sharing. Specifically, RADS team members are investigating where funded researchers across their institutions are making their research data publicly accessible, the quality of metadata of research data from their institutions, and the cost to the institution to implement federally mandated public access to research data policies.

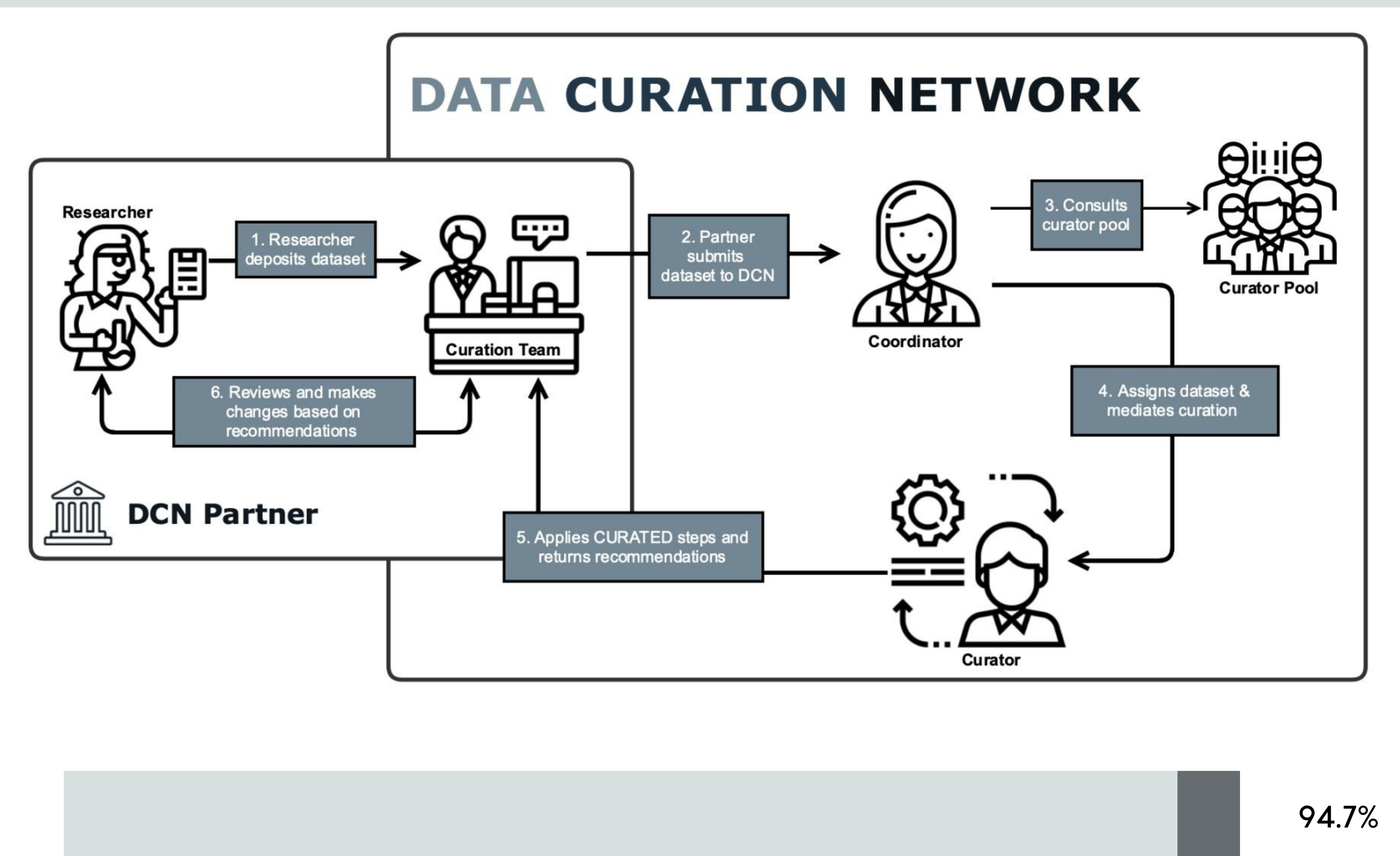
Assessment of where research data is shared and the quality of metadata will result in a set of recommended best practices for institutional repositories to improve the quality of their scholarly metadata. These recommendations will focus on improvements which specifically affect findability in the global research infrastructure, resulting in increased reusability. These best practices will be shared throughout the entire DCN and, as members of this network share their feedback, best practices will be refined. Ultimately RADS recommended best practices will have a broad impact, as they will have been tested by the DCN collective.

Shared Curation Workflow

Between July 1, 2021 and May 1, 2021 the DCN has curated 38 datasets through our shared curation workflow, representing nearly 20 formats and 18 disciplines. We've also connected our nearly 40 curators to support through consultations via collaborative software like Slack and Google Groups.

While this is a lower number of datasets than we had anticipated, in conversations with members, it is likely that there are a number of contributing factors:

- Members are up-skilling and feeling better equipped to curate datasets
- Submitting datasets for curation through the DCN requires up-front work and extends the curation timeline. Researchers often need a DOI quickly and request quick turnarounds.
- Institutional members, knowing their colleagues are similarly overwhelmed, are reluctant to pass their burdens.



During the same time period (July 1, 2021-May 1, 2021) 36/38 datasets submitted for shared curation were matched, with a success rate of 94.7%.

Education & Outreach

Over the past 12 months, members of the DCN Racial Justice Working Group, Refining the Standard Protocol Group, and the Education Committee have revised our standard CURATE(D) workflow. This tool has been designed to provide both a structured and consistent way for curators to review content, while allowing for disciplinary and format considerations. This is not only the primary teaching and education tool of the DCN, but also the workflow that we leverage to ensure consistent curation across institutions. Key updates to the workflow:

- A preamble that includes information about how to use the CURATE(D) steps, in particular the iterative nature of curation and prompting questions to consider whether data should be shared.
- Key ethical considerations at each step as well as considerations for accessible data.
- Standardized formatting and more inclusive language, to expand the applicability of the workflow beyond data curators to include data creators.

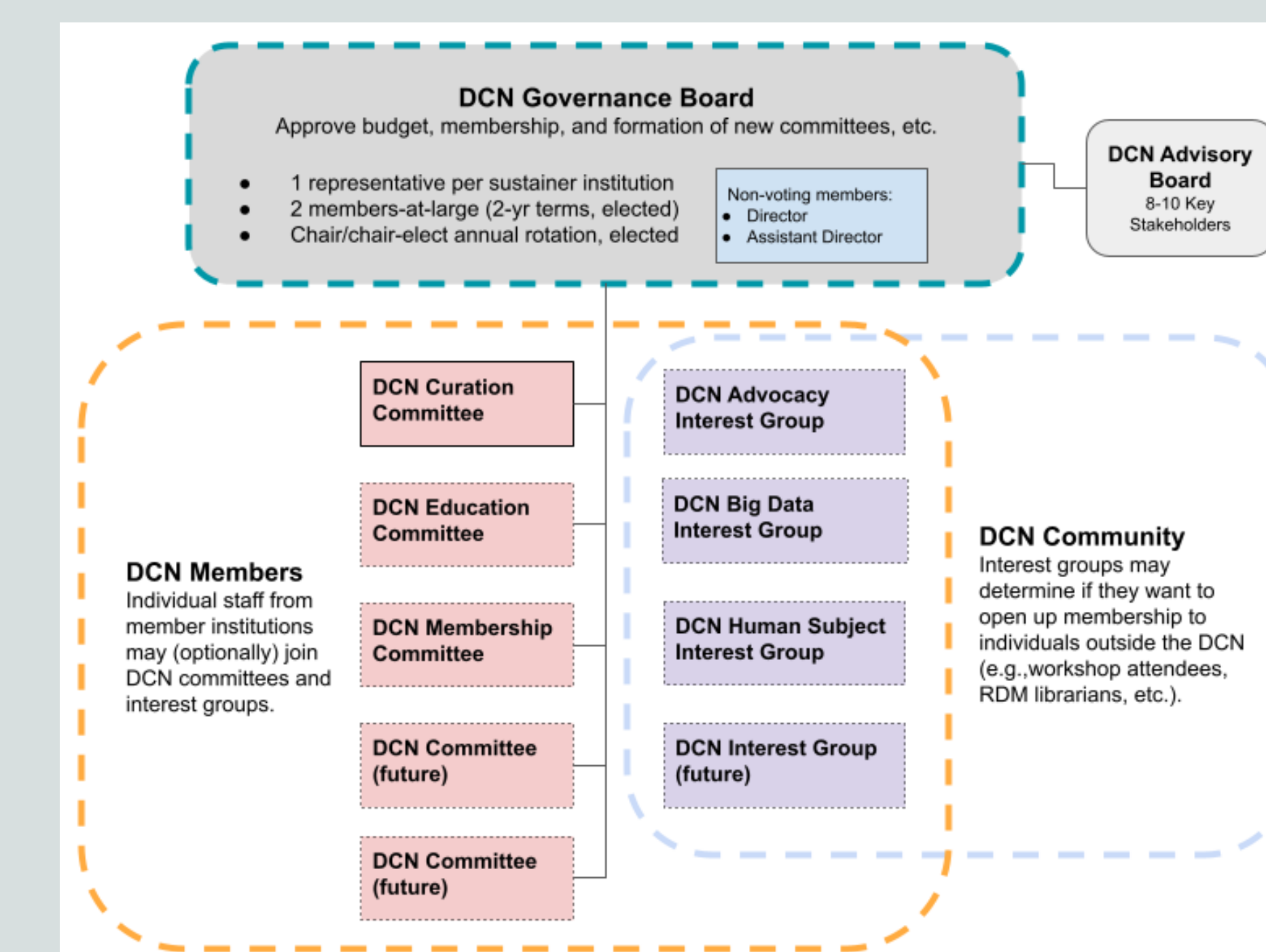
Sustainability

As DCN members engage in this work, we are acutely aware of how critical long-term sustainability is for a community. To that end, in the months leading up to and following the transition from a grant-funded to a member-funded organization,³ DCN Leadership developed a guiding, and living, Governance Model⁴ to enable member engagement and ensure that the partners guide our work.

Since then, our team has continued to iterate on the Governance Model, based on feedback from partners. We have also launched a Strategic Framework,⁵ created in consultation with DCN Representatives and Curators, to guide our work for the coming years.

References

1. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B., et al (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
2. Realities of academic data sharing (RADS) initiative. (2021, October 20). Association of Research Libraries. <https://www.arl.org/realities-of-academic-data-sharing-rads-initiative/>
3. DCN Sustainability and Transition Plan. (2020). Data Curation Network. <https://hdl.handle.net/11299/225741>
4. DCN Governance Model. (2021). Data Curation Network. <https://hdl.handle.net/11299/225742>
5. Data Curation Network. (2022). Data Curation Network Strategic Framework 2022-2025 v1.0. <https://hdl.handle.net/11299/226182>.



Authors

Mikala Narlock, MLS, Director, Data Curation Network, based at the University of Minnesota, United States of America
ORCID: 0000-0002-2730-7542



Shawna Taylor, MLIS, PMP, Project Manager, Realities of Academic Data Sharing, Association of Research Libraries, United States of America
ORCID: 0000-0002-9842-7867



Presenting Author

Mikala Narlock -- mnarlock@umn.edu -- datacurationnetwork.org/