

Indigenous data in an institutional repository: First steps toward putting the CARE Principles into practice

Authors

Alicia Zuniga, Senior Assistant Librarian, California State University - Sacramento (CSUS)

Wanda Marsolek, Data Curation Librarian, University of Minnesota (UMN)

Shanda Hunt, Research Data Services Education & Outreach Librarian, University of Minnesota (UMN)

Background

The Data Repository for the University of Minnesota (DRUM) is an open access, curated data repository for UMN affiliates to share digital data for long-term access and reuse. DRUM launched in 2014; ten years and over one thousand submissions later, we continue to work on policy and practice improvement. One of our aims is focused on Indigenous data and implementing the CARE Principles for Indigenous Data Governance.¹ The UMN is located on the contemporary, traditional, and ancestral lands of the Dakota people. As a land grab university,^{2,3} it is our responsibility to support and advocate for Indigenous Data Sovereignty (IDSov). IDSov “upholds the rights of Indigenous Peoples, communities, and Nations to ‘govern the collection, ownership, and application’⁴ of datasets created with or about Indigenous communities, Indigenous Lands, and the community’s non-human relations.”⁵ Simultaneously, as a UMN service, DRUM is subject to University policies including the Openness in Research Policy set by the Board of Regents that states that the University will not be subject to restrictions on participation in University research or on the dissemination of the results of University research. In addition, the policy states “The University reserves the right to publish and present research results, individually and in collaboration with other researchers.”⁶ This sounds very promising and in good spirit for data loving folks. However, this policy can be in conflict with the CARE Principles.

¹ Carroll, S, Garba, I, Figueroa-Rodríguez, O, Holbrook, J, Lovett, R, Materechera, S, Parsons, M, Raseroka, K, Rodriguez-Lonebear, D, Rowe, R, Sara, R, Walker, J, Anderson, J and Hudson, M. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19: XX, pp. 1–12. DOI: <https://doi.org/10.5334/dsj2020-042>

² <https://www.hcn.org/issues/52-4/indigenous-affairs-education-land-grab-universities/>

³ <https://mn.gov/indian-affairs/truth-project/>

⁴ Stephanie Russo Carroll, Desi Rodriguez-Lonebear, and Andrew Martinez, “Indigenous Data Governance: Strategies from United States Native Nations,” *Data Science Journal* 18, no. 1 (July 2019): 31, <https://doi.org/10.5334/dsj-2019-031>; Tahu Kukutai and John Taylor, *Indigenous Data Sovereignty: Toward an Agenda* (Canberra: ANU Press, 2016).

⁵ Stephanie Russo Carroll, Marisa Duarte, and Max Liboiron. "Indigenous Data Sovereignty" in *Keywords of the Datafied State* (eds. Burrell, Singh & Davison). (April 24, 2024). *Data & Society*. <https://datasociety.net/library/keywords-of-the-datafied-state/>

⁶ <https://policy.umn.edu/research/openresearch>

Our efforts to reconsider how we handle Indigenous data began in 2021 when we accepted a dataset about the location of certain animals in the state of Minnesota. After much back and forth with the submitter, it became clear that the research was located on Indigenous lands; this was the first time we requested documentation that the Tribal Agency approved of public data sharing. Over time, we continually revisited this particular dataset wondering whether we had taken the appropriate steps. After utilizing this dataset at a Data Curation Network (DCN) workshop, the DCN participants decided to create a [data curation primer about the CARE Principles](#). They reached out to Stephanie Russo Carroll, co-creator of the CARE Principles, co-founder of the [US Indigenous Data Sovereignty Network](#), and co-founder and chair of the [Global Indigenous Data Alliance \(GIDA\)](#). Russo Carroll agreed to co-author the primer and recruited others in her network to assist. This is how we were introduced to the US Indigenous Data Sovereignty Network. Hunt and Marsolek reached out to learn more, and were, in turn, invited to speak with the group about our handling of the 2021 dataset. Graciously, 16 collaborative members met with us to reflect and advise on next steps. In the meantime, we had implemented new language into the automated acceptance email that goes out to all DRUM submitters. It states: "If you submitted data that was collected from, with, or about a Tribal agency, please provide documentation about who owns the data. We seek to apply the [CARE Principles](#) of Collective Benefit, Authority to Control, Responsibility, and Ethics." Then, in April of 2024 Hunt and Marsolek attended the US Indigenous Data Sovereignty and Governance Summit, a gathering of Indigenous leaders, community, academics, policy experts, and other data actors and rights holders from across the nation to develop guidance around appropriate data governance structures that support Indigenous Peoples' own initiatives. Attendance at this event inspired us to embark on an assessment of the Indigenous data that already exists in DRUM as a first step toward incorporating the CARE Principles into DRUM's workflow and policies.

We proposed a project in partnership with [Data Services Continuing Professional Education \(DSCPE\)](#), a ten-week experience that connected Zuniga with Marsolek and Hunt. The DSCPE project - which occurred September through November 2024 - included several vital activities performed by Zuniga which led to the following outputs:

1. list of search terms related to Indigenous data in Minnesota
2. spreadsheet of those datasets that captures the level of Indigenous consent that was given to share the data publicly
3. resource list of policies, practices, and documents related to Indigenous data
4. recommendations for DRUM based on findings and the DSCPE experience
5. map of this process so that others, including DRUM staff, can implement and build on the important groundwork established by Zuniga

Indigenous data audit process

The first step in this project was creating a list of potential search terms related to Indigenous data in Minnesota. In addition to collating a list of general Indigenous search terms (e.g., Indigenous, Tribe, etc.) and terms related to Minnesota (e.g., Grand Portage, Red Lake, etc.), Zuniga looked to Indigenous databases like [Native Health Database](#) and [Plateau Peoples' Web](#)

[Portal](#) to explore additional metadata terms and added those that were relevant to the list. It's important to note that when Zuniga compiled keywords specific to Minnesota, she looked for Tribal terms that went beyond the 11 federally recognized Tribes, finding the [Mendota Mdewakanton Dakota Tribal Community website on non-recognized Tribes](#) useful. She then used these 61 search terms to systematically search DRUM for Indigenous datasets, beginning with more broad terms and those specific to Minnesota. She compiled the datasets into a spreadsheet with pertinent information about each, including whether or not a dataset contained data about Indigenous People, land, or resources. This was determined by information included in the DRUM record, any stated data licenses or permissions, and inspecting uploaded files, including the README file. Along with the spreadsheet, Zuniga created [documentation](#) with step-by-step instructions on how to complete the audit to ensure continuity beyond the DSCPE's timeframe. The review contained 54 records that could potentially contain Indigenous data.

The scan of DRUM data uncovered quadruple the number of Indigenous datasets than was previously assumed, allowing DRUM staff to fully take stock of the breadth of its hosted data. Taking a deep dive into the DRUM datasets and examining each one often required investigative skills that went beyond a brief scan. For example, in one dataset there was no information about Indigenous data stewardship, but the two listed funding sources were a government grant and the Fond du Lac and Grand Portage Bands of Lake Superior Chippewa, which indicated that this was likely Indigenous data. Looking up the listed grant number and its publicly available description shed light on the fact that this data was collected in collaboration with students and instructors at a local Tribal college. This was not obvious upon first glance at the README file nor at the dataset landing page, and there were no additional details about the level of sharing that was allowed by the Indigenous collaborators.

Another challenge was making decisions on a case-by-case basis on whether a particular dataset contained Indigenous data. Some were obvious collaborations with Tribal Nations or funded by Tribal organizations, and some were less clear cut. For example, wild rice is an important Indigenous resource that is included in many treaties, and there are several DRUM datasets with wild rice as a focus. With this in mind, is there ever a situation where the study of wild rice would *not* be considered Indigenous data? Would solely genomic data about wild rice be considered Indigenous data? We have recently learned that the Minnesota Chippewa Tribe submitted a resolution in 2021 requesting that the university discontinue wild rice genetic research⁷. As of this writing, we are unaware if the University has complied. Lingering questions like this require both deep investigation and consultation with Tribal Nations.

While exploring research on the topic and compiling a resource list, Zuniga uncovered an article titled "Navigating University Openness in Research Policy Inconsistent with Indigenous Data Sovereignty: A Case Analysis" that documents the challenges UMN graduate students experienced when the research data permissions put forth by their Tribal Nation collaborators were at odds with UMN's Openness in Research policy.⁸ The research project delay (nearly one year) caused by UMN's initial disagreement with the Tribal Nation's data sharing agreement in

⁷ <https://hdl.handle.net/1721.1/147601>

⁸ <https://onlinelibrary.wiley.com/doi/full/10.1002/eahr.500202>

this particular case study demonstrates that this policy, despite having the best intentions for data openness, does not consider Indigenous Peoples' rights and interests.

The final deliverable for the DSCPE project was a list of high level recommendations that DRUM can implement in their policies and practices that will improve alignment with the CARE principles. The [report](#) heavily refers to the guidance of O'Brien et al.'s (2024)⁹ 47 recommended data curator activities and how they align with each category of the CARE principles. Several of their recommended activities are cited in support of each of the five recommendations: 1) update metadata to include common search terms; 2) incorporate Traditional Knowledge and Biocultural Labels and Notices; 3) retroactively investigate unclear data sharing permissions; 4) develop a policy on handling Indigenous data at submission; 5) incorporate explicit exceptions for Indigenous data in institutional policies.

Next steps

Our identities do not make us authoritative decision-makers nor experts with regards to IDSoV since we are not Indigenous. These explorations and recommendations are starting points and we recognize the importance of Indigenous Peoples' control over policies and practices to successfully incorporate the CARE principles. We aim to be mindful data stewards and allies to implement collective action. Zuniga's discoveries make clear that policies at the library and institutional levels must be reconsidered, and we acknowledge the long road ahead. Every success we've had as a data repository is due to diligent (dare we say relentless?) relationship building, and that is the path we intend to take again.

Beginning in mid-February, Hunt and Marsolek will meet weekly to continue to build on the incredible groundwork paved by Zuniga. Because our staffing and capacity are already maxed, we will send 50% of our assigned datasets to the DCN. We have begun expanding our institutional collaborators and have two important meetings on our 2025 calendar: we will meet with 1) University Library archivists who have done reparative description work in the archives to explore the addition of reparative descriptions for DRUM and/or specific datasets in DRUM and 2) Karen Diver, Senior Advisor to the President for Native American Affairs, to discuss our existing work, our aspirations, and how we can align our efforts with her office. We also hope to once again meet with the Indigenous Data Sovereignty and Governance Collaboratory and attend their spring summit in 2026. We have prioritized the following actionable next steps:

1. Complete the DRUM Indigenous dataset audit begun by Zuniga
2. Explore funding options to hire an Indigenous data curator
3. Assemble an Indigenous data community advisory board
4. Develop a DRUM policy related to Indigenous data

We hope that this initiative leads to deepened relationships, Indigenous representation on the DRUM team, and meaningful DRUM policies. We envision a future where DRUM and UMN policies align with and follow IDSoV guidelines and policies.

⁹ <https://datascience.codata.org/articles/10.5334/dsj-2024-037>

Project outputs

The protocol, recommendations, and resources resulting from this partnership can be found in the University of Minnesota Digital Conservancy.

Zuniga, Alicia; Marsolek, Wanda; Hunt, Shanda. (2025). A scan of Indigenous data in an institutional data repository. Retrieved from the University Digital Conservancy, <https://hdl.handle.net/11299/269367>.