Testing Ethical Representation of Indigenous Primary Sources in Wikidata

Diana E. Marsh
University of Maryland

Ugoma Smoke
University of Maryland

Katrina Fenlon
University of Maryland

Peter Meyer
Wikimedia DC

Jesse Johnston
University of Michigan

Lucy Havens
University of Maryland

Sarah Buchanan
University of Missouri

Nikki Wise
University of Maryland

Abstract

This project aims to improve the representation of Indigenous nations and their cultural heritage collections on Wikimedia platforms by using Wikidata as a central tool. With our partners at Wikimedia DC, the new North American Wikimedians/Native American issues team, our CoPAR Working Group, and community collaborators, we will co-design a project that builds a community-focused and research driven model of integrating Indigenous archival content into Wikimedia projects via Wikidata. We ask: ***What potentialities does Wikidata, as a distributed, networked, linked data infrastructure and set of community practices, have for discovery and reuse of Native and Indigenous data, while adhering to ethical principles?*** We will utilize a semi-structured diary protocol to test and evaluate experimental data transformation workflows for integrating collections into Wikidata, drawing on an existing dataset and community-based catalogue records; throughout we will take a community-based participatory action research (CBPAR) approach in which our research process is co-designed with core community and repository stakeholders, and we emphasize community-building as part of the work. Key outcomes include: 1) Enhancing overall Indigenous representation in Wikidata; 2) evidence-based guidance on ethical Wikidata representation and data transformation for GLAM (Galleries, Libraries, Archives, & Museums) sector and beyond; 3) Community protocol development and community building; and 4) Development of wiki practices for subject matter experts and training for Indigenous editors. Our project contributes to the Wikimedia Foundation's goal of knowledge equity, testing Wikidata as a back-end data infrastructure for breaking down cultural-technical barriers to Indigenous cultural heritage, and in turn making more achievable the 'right to know' (O'Neal 2015) for Native and Indigenous communities. f metadata transformation workflows.

Introduction

There are many historical and cultural materials in museum collections, archival institutions, and community-based repositories that are underrepresented in public digital platforms. Many of these collections include historical photographs, oral histories, and museum records that reflect Indigenous cultures, experiences, and contributions.
Such primary source materials are invaluable for research and education, and are also irreplaceable cultural resources for the communities in which they originated, particularly Native and Indigenous communities (O'Neal, 2015). These communities have been disenfranchised from their own information,

data, and knowledge through the evidentiary and collecting practices of historical anthropological researchers, and of archives, museums, and other collecting institutions (Montenegro, 2019; Smith, 2006). Knowledge extraction took place in the context of assimilation, epistemicide, and genocide (Schweizer & Henry 2019); localized information (which was often never meant to be recorded) was dispersed to archives worldwide via what scholars have called an "archival diaspora" (Punzalan, 2014a; O'Neal 2014) and out of the hands of the host communities. These historical factors underscore the ethical responsibility to provide community access to archival and unpublished information (Caswell & Cifor, 2016; Odumosu, 2020) on public digital platforms.

Greater access to online and digital collections, and the growth of community-based archives and Tribal heritage centers has made more concrete the possibility of virtually reunifying (Punzalan, 2014b) primary sources with users and descendant communities. Native and Indigenous community members are increasingly using these collections for their own research and community initiatives, particularly for language revitalization (Holton 2012; Baldwin et al. 2018; Carpenter 2019), but also for Tribal and legal histories and federal recognition (e.g. Reijerkerk & Reed 2023; Lowery, 2009), artists projects, and studies on environment, and kinship (see Marsh, 2023). And of course, access to archival sources is also crucial for physical repatriation and successful Native American Graves Protection and Repatriation Act (NAGPRA) claims as well (Fforde et al., 2020; Fforde et al., 2015). Many Native and Indigenous communities are building their own digital archives using a mix of copies obtained from colonial repositories and new contemporary collections made in the community (e.g. Bhasin, Roy & Arriaga, 2011; Christen et al., 2017). But, these scattered documents, and this core epistemic infrastructure, including a wide range of cultural data from communities, are still largely held in physical repositories and underrepresented on the web. Recent NSF-funded research has shown that these materials are often not discoverable, either to anthropologists or to descendant Native and Indigenous communities (Marsh et al., 2022; Marsh et al., 2021; Marsh, 2019). Wikidata and Wikipedia have the potential to provide such discoveries worldwide.

Research has shown that a mere 20-30% of most archival collections have been processed, or rehoused in proper archival boxes, inventoried, and described in a public-facing digital cataloguing system (Bucciferro, 2008; Greene & Meissner, 2005; Panitch, 2001). An even smaller percentage of those primary source materials have digital catalogue records that are published to the web, or otherwise made accessible as a unified body of evidence that can support research. Still fewer are digitized and available for viewing online. For example, the National Archives and Records Administration (NARA) had scanned just 2.4% of its collections (over 289,500 pages) by August 2024, despite immense scanning capacity including mass digitization workflows, conveyor belt scanning, and staffed state-of-the-art facilities. Information scholar Douglas Oard has recently estimated that if NARA digitized 121,000 pages/day (a page every 2/3 of a second), it would take 375 years to digitize all of its current 12 billion pages of paper holdings (Oard 2023).
Thus, many primary source resources can only be found by the descriptions about them that are held in institutional catalogues, requiring institution-to-institution search.

Date: July 1, 2025 - June 30, 2026

Related work

In the 1990s, the Council on the Preservation of Anthropological Records (CoPAR) conducted a survey to determine where anthropological records were located and compiled a list of archival institutions and the anthropologists' papers they held. The resulting Guide to Anthropological Fieldnotes and Manuscripts in Archival Repositories (available at copar.org as a printed list and at copar.umd.edu as a hyperlinked list) is still the main resource for archival discovery for anthropology.

A 2011 AAA-organized workshop focused on archival discovery (NSF grant BCS-1159109), resulted in a wiki prototype for a registry of anthropological data that would index many repositories (Schmid & Cliggett, 2011). This crowd-sourcing initiative (anthropologists would list where their papers were) was not funded or staffed, and was not sustained. This showed the need for a sustainable and funded inter-institutional infrastructure for the discovery of archives and data. Now there are well-established platforms (discussed above) with greater infrastructural and crowd-sourced support.
More recently, Wikidata has begun to act as a core alternative to this model by helping to surface archival descriptions, connect archival collections and content to Wikipedia articles, and explore connections among distributed archives through the power of Wikidata SPARQL queries that search connections across diverse data sources (e.g., Association of Research Libraries, 2019; Szekely et al., 2014). For archivists in particular, Wikidata has emerged as a more approachable and likely sustainable platform for doing linked data work (Roke and Tillman 2022) and a core experimental tool for improving discovery and access (Babcock et al., 2021). Archives and archival content management systems such as ArchivesSpace are increasingly incorporating Wikidata into their metadata fields. OpenRefine, a core tool for batch metadata transformation in heritage fields, also has a Wikidata extension. Wikidata also interlinks with most core standards and authority files across the world. Yet, little empirical research has been conducted to test the applicability of Wikidata for Indigenous archival data.

At the same time, momentum has been growing in the creation and adoption of new protocols and models for describing and representing Indigenous knowledge. In August 2018 the Society of American Archivists adopted Protocols for Native American Archival Materials (PNAAM, a Native-led effort first drafted in 2006, see O'Neal 2014) making ethical stewardship of Native American collections a mandate; in 2019 the International Council on Archives (ICA) established the Expert Group on Indigenous Matters, in response to the UN Declaration on the Rights of Indigenous Peoples. The Steering Committee on Canada's Archives (SCCA) established the Truth and Reconciliation Commission Taskforce (TRC-TF) and released a Reconciliation Framework for archives in 2022. Scholarship in Indigenous Data Sovereignty has advanced a parallel set of principles for collective benefit, authority to control, responsibility, and ethics (CARE) focused on ethical aspects of the accessibility and reuse of cultural data (Carroll et al. 2020; Walter et al. 2019), to be adopted in tandem with the FAIR (findability, accessibility, interoperability, and reuse) principles for data stewardship (Wilkinson et al. 2016; Taitingfong 2024).

While much Indigenous knowledge is difficult to represent in Western knowledge systems and data models (Littletree et al., 2020; Littletree et al., 2015), existing Wikipedian efforts dedicated to specific communities, such as the Cherokee Wikipedia (https://chr.wikipedia.org), as well as efforts to draft Wiki protocols for Indigenous

communities elsewhere (Thorpe et al., 2023), illustrate Wikimedia's promise for community information efforts. In theory, linked data models allow multiple truths and forms of expressions, unlike many information systems that delimit by fixed categories (Moriarty & Turner, 2023; Mohan & Rodgers, 2021). And unlike many linked data environments, Wikimedia's community cooperative model and open access training model may also allow for community empowerment and sovereignty over data description, modeling, and access.

Indeed, in a report from the Association of Tribal Archives, Libraries, and Museums (ATALM) on a 2021 survey and summit with 80 Tribal archives, participants noted that funding and access to collections were among the top four most reported management challenges, with funding also acknowledged as a barrier to outreach activities that encourage engagement with collections. The top five funding priorities were "Improve collections care and conservation," "Expand technology," "Develop a greater online presence," "Train existing staff," and "Expand collections." Given the volunteer-based and open-sourced nature of Wikimedia, leveraging Wikidata could help to address Tribal archives' access and outreach needs, or "the Right to Know" (O'Neal 2015), without stretching their already limited funding (Jorgensen & Johnston, 2022). Wikidata in particular could provide an alternative avenue to other approaches to training Tribal archivists in digital collections access and cataloguing.

In 2022, our revitalized CoPAR group began collaborating with Wikimedia DC to experiment with this registry as a dataset to be integrated into Wikipedia and Wikidata. Through Wikipedia edit-a-thons in 2023, 2024, and 2025, work from student curators in graduate Information courses at the University of Maryland College of Information, and work from CoPAR collaborators, we have built that registry into a Wiki dataset and worklist for bringing community records into linked data infrastructure. In addition, in 2024, CoPAR research assistant Ugoma Smoke made crucial additions to the worklist by integrating 75 Indigenous anthropologists. The dataset now contains over 700 entries, linking anthropologists with archival repositories.

The work is based on preliminary experimental NSF-funded research (BCS 2314762) undertaken in 2023-2025 to (1) conduct a systematized search and review of the use of linked data for representing primary source data (Marsh et al., 2024), and (2) test a semi-automated workflow for deriving data from primary sources using the free version of ChatGPT (version 4o). That work illustrated some of the core affordances of linked data, and yet many of the technical and resource challenges that remain. This proposal follows on that work by partnering with Wikimedia DC and community stakeholders to collaboratively test the integration of Indigenous archival catalogue data into linked data infrastructures in Wikidata, while building a community of practice around the unique ethical considerations for this work. We ask: ***What potentialities does Wikidata, as a distributed, networked, linked data infrastructure and set of community practices, have for discovery and reuse of Native and Indigenous data, while adhering to ethical principles?***

Methods

With our partners at Wikimedia DC, the new North American Wikimedians/Native American issues team, our CoPAR Working Group, and community collaborators, we will co-design a project that builds a community-focused and research driven model of integrating Indigenous archival content into Wikimedia projects via Wikidata. To that end, we will use community-based participatory action research

(CBPAR) as our overarching project framework (Leavy, 2017; Swantz, 2008). The community-based element of CBPAR ensures a human-centered approach, with research being motivated by problems or challenges faced by communities. The participatory element of CBPAR communicates the collaborative nature of CBPAR, with a research team establishing and developing a collaboration with specific members of a community, partnering with stakeholders rather than viewing them as subjects of the research project. The action research element of CBPAR emphasizes a research project's dedication to creating change in a community. As Leavy (2017) writes, projects that use CBPAR tend to be responsive, revising research methodology and outputs based on new or clarified understandings of stakeholder needs. The methodology we outline in this project thus represents a starting point for our research, one that is informed by existing literature on challenges facing Native and Indigenous communities' archives, and will be refined through collaboration with stakeholders during the project.

In this new phase of work, we aim to test two core methods and workflows for Wikidata integration. First, we will **test our ongoing dataset of anthropological records, built by CoPAR and now transformed into a** Wiki worklist, utilizing the OpenRefine wikidata plugin to attempt batch ingest of this data into Wikidata. Second, we will test a **set of real-world descriptive catalogue representations** (in archives usually referred to as *finding aids*) of **collections from Native, Indigenous, and anthropological archives**, at varying depths of description, and evaluate the outcomes of transformations. how well linked data can be represented in Wikidata.

We will design both of these processes and their evaluation in collaboration with Native and Indigenous colleagues (including those on the CoPAR Working Group and Advisory Board) and in partnership with Indigenous community archivists, archives, and repositories holding community-relevant collections. In both processes, we will couple our CBPAR approach with semi-structured research diary protocols, in which team members document their observations during or after sessions of work on these data workflows, to allow for the study of situational, temporal, and internal phenomena (Fan et al., 2025; Czerwinski et al., 2004; Sheble et al., 2017), rendering visible the barriers and challenges that determine the success and failure of sociotechnical processes (Elliot, 1997; Waller & Ragsdell, 2012; Reis & Wheeler, 1991), and facilitating autoethnographic reflection (Garcia & Cifor, 2019) that is crucial to our community-oriented work.

In our work over Fall 2024 - Spring 2025, we piloted this approach for our image-to-linked data workflow facilitated by ChatGPT, and found it highly successful in documenting each step of the process, reflecting on the affordances and challenges of sociotechnical infrastructures, time, ethical positionings, and data or tool alignment (or incommensurability).

In this new phase of work, major gaps for workflow testing include not only broad community representation but major gaps in available data mapping and more specific areas for enhancements to description such as adding Native American languages to Properties in Wikidata.
We aim to utilize findings from this work to both inform Wikidata practices, and to develop new Protocols for working in Wikimedia with Native and Indigenous collections in the North American context, in collaboration with the North American Wikimedians/Native American issues team. In that work, we are inspired by Australian Indigenous communities to develop community-driven data protocols, and aim to look to our First Nations colleagues and

archivists in Canada, to begin shaping similar structures for Indigenous communities in the United States.
In the final two months of the project, we will host a virtual Evaluation and Protocols workshop over two days with core constituencies, including Wikimedians, Native data curators and archivists, international Indigenous Wiki-interested colleagues, and community and repository collaborators during the project. We will host four focus group discussions (FGDs) with invited participants to the workshop to reflect on our work and findings in Wikidata representation. Each focus group will meet for two sessions, once before and once after our workshop, and be comprised of 4 participants from different core audiences (Morgan 1997). All diary entries and FGDs will be transcribed using Otter.ai and coded using an inductive, grounded theory approach (Charmaz, 2008) in the qualitative coding software NVivo. Emergent codes, categories, and themes will be discussed and honed at weekly group meetings.

Research Methodology:
In summary, our research methodology includes:

- A **community-based participatory action research (CBPAR) model** in which our research process is co-designed with core community and repository stakeholders, and we emphasize community-building as part of the work.
- **Experimental workflows** of catalogue records to Wikidata representation to include batch metadata transformation (from the CoPAR dataset).
- Further **testing of a set of data transformation workflows** via the OpenRefine Wikidata plugin, here focusing on catalogue descriptions and specific Wikidata integrations.
- Both of the above will utilize a previously tested **diary protocol among four data curators** as they undertake these workflows, documenting necessary technical expertise, gaps, challenges, as well as core areas for data alignments, as they trial various tasks.
- All **diary entries will be qualitatively coded** by PIs and our lead GA, using an inductive, grounded theory approach (Charmaz, 2008).
- Finally, we will conclude the project with a two-day virtual **Evaluation and Protocols workshop** bringing together Wikimedians, data curators, and Native archivists, community members to 1) **evaluate impacts of Wikidata integration**, including conducting **four formal focus group discussions** in tandem with the workshop to reflect on outcomes, and 2) **refine drafted protocols**.

Expected output

1. **Enhance Overall Indigenous Representation in Wikidata**: Through our work to integrate the CoPAR dataset/worklist and partner community collections with Wikidata, we will increase overall Indigenous representation for a wide range of community researchers and heritage professionals; Attendant priorities include:

A. Expanding the number of Wikidata items that include Native American language Properties. Current examples: https://www.wikidata.org/wiki/Q33388, https://www.wikidata.org/wiki/Q105405
B. Working with language specialists and our networks to make constructive edits to indigenous language content on Wikimedia, e.g. the Cherokee Wikipedia at https://chr.wikipedia.org/wiki/%E1%8E%A4%E1%8E%B5%E1%8E%AE%E1%8E%B5%E1%8F%8D%E1%8F%97.
C. Adding IPA pronunciation examples to Native American terms in Wikipedias

and perhaps on Wikidata. A likely set of examples is the earliest name of the places whose name was inherited from a Native name. E.g. [[w:Wisconsin]] has the oral pronunciation of the current name but not the Native name from which it developed.

D. Adding information about specific Native American or First Nations archives into Wikidata using [[d:P485]], or perhaps along the model of representing particular artifacts in Wikidata along the model of [[d:Q66319806]].

2. **Guidance on Ethical Wikidata Representation and Transformation Workflows**: Evidence-based workflows and recommendations for ethically representing dispersed anthropological and Indigenous primary sources via Wikidata, including findings from transformations and diary protocol and work with the North American Wikimedians/Native American issues team for Wikimedia editing communities and GLAM professionals.

3. **Protocol Development**: Research and adapt existing metadata standards to create best practices suitable for representing diverse cultural narratives within Wikimedia platforms; Develop and implement protocols for representing diverse cultural narratives in Wikidata and related platforms; Working with the new Native American issues team to create guidelines and best practices for inclusive metadata description, drawing inspiration from existing models such as the ATSILIRN Protocols, and geared at Wikimedia editing communities, GLAM professionals, and subject specialists.

4. **Community Building**: Organize training sessions to recruit and educate new editors, focusing on sustainable community engagement; Establish and train a community of editors committed to long-term engagement and roles within Wikimedia projects, focused on novice and interested Native and Indigenous community members, community-based researchers and heritage professionals; we will plan presentations of this work as it is underway to anthropologists, GLAM professionals, Indigenous repositories, Wikimedians at meetings and conferences, and to experts on Native American materials.

5. **Institutional Collaboration**: Partner with libraries, archives, and museums and related professional communities to integrate authoritative data sources into Wikidata, enhancing content accuracy and depth; Collaborate with institutions with Indigenous materials and expertise such as the Library of Congress to integrate authoritative linked data sources, enhancing the quality and reliability of information, while also reflecting on institutional barriers to doing this work (e.g. resources, policy constraints, misaligned ethics).

6. **Development of Wiki Practices for Subject Matter Experts**: Project participants will learn, share, demonstrate, and train (or develop training materials) in Wikidata and Wikimedia skills related to Indigenous languages, cultural histories, and primary source materials; We will explore training alignments with needs for Tribal and community archivists for digital curation and access tools.

Risks

Given the sensitive nature of Indigenous and Tribal representation, and the long history of knowledge extraction and misrepresentation, we envision some risks to undertaking this work, which we aim to mitigate through our overall framing and collaborative approach, as well as taking specific actions in our process. Risks we envision include:

- Staff and professional employment of participants in the project may be cut in the current fiscal environment.
- Culturally-sensitive or politically fraught information being released to open platforms.
- Indigenous and community-based archivists' time is already extremely stretched, and community knowledge bearers moreso; we will need to ensure that our project not overburden collaborators

To ensure *maximal benefits and mitigate risk*, we will:

- Utilize our institutional IRB at the University of Maryland–while institutional boards such as this do prioritize institutional considerations in some cases, we have found that the UMD review boards include a range of experts on ethical risks and key aspects of community project design; they provide thorough feedback on methods, as well as project documentation such as consent forms
- We will also submit our project design to any relevant Tribal IRB or Indigenous ethics review boards, as relevant to specific collections we aim to test.
- Our overall community-based participatory action research methodology focuses us on making all project decisions with input from key community members; we will choose collections descriptions, and relevant data in them, to integrate with Wikidata, in collaboration with relevant community members and archives, to ensure that sensitive or traumatic information is not brought into the platform.
- We will defer to GLAM and community archives or repository professionals as to bandwidth for collaboration as our project and their resource contexts emerge.
- Our CoPAR Advisory Board, Working Group, and current metadata team include Native and Indigenous members, and we have continued to prioritize diversity in our leadership and teams to ensure we have a range of emic community perspectives both within our own groups and with professional and community partners.

Community impact plan

Our project contributes to the Wikimedia Foundation's goal of knowledge equity, testing Wikidata as a back-end data infrastructure for breaking down cultural-technical barriers, and in turn making more achievable the 'right to know' (O'Neal 2015) for Native and Indigenous communities. We envision that this work will lead to:

- Enhanced representation of diverse cultural narratives within Wikimedia platforms and evidence-based, cultural-technical solutions to increasing Indigenous Wikidata representation.
- A trained and active community of Indigenous and ally editors contributing to sustained Indigenous content development.
- Improved integration of authoritative linked data sources, increasing the reliability and richness of information, as well as the development of key datasets of importance to not only Wikimedia research and heritage communities, but wider Tribal and rural communities.
- Established guidelines for inclusive description, promoting best practices across the Wikimedia community, and

informing policy decisions in collaboration with the new Native American issues team and international Wikimedia groups.
- In tandem with this research project, and in partnership with Wikimedia DC and the North American Hub of Wikimedia chapters and user groups, we are eager to plan and host events for a wide audience, including international colleagues in other state contexts for Indigenous underrepresentation.

Evaluation

Overall, our research design integrates evaluation in the form of iterative semi-structured diary protocols, and collaborative meetings with community and repository partners, as well as Wikimedia partners and CoPAR boards, throughout. We are also eager to gain iterative feedback as we work with others interested in Native issues in Wikimedia throughout the project, and reflect on emergent findings. Our Focus Group Discussions will provide specific evaluation of workflows and representational issues with key stakeholder groups.

Our primary avenue for evaluation, however, will be our two-day virtual Evaluation and Protocols workshop, to take place toward the close of our project. There, we will present and gain core stakeholder feedback on findings, and successes and failures of various workflows in the domains of Indigenous knowledge representation; ease of application for other contributors, users or repositories; ethical protocols adherence; and training approaches. We also aim to utilize the workshop to plan outreach activities and additional community events beyond the scope of the funded work, including follow-on community protocols events, planning publications or white papers geared at Wikimedians, data curators, and GLAM professionals, planning presentations at relevant community-oriented conferences (e.g. Association of Tribal Archives, Libraries, and Museums (ATALM).

Budget

We request $49,890 to support key **Personnel** to carry out this project, including: 1) research assistance during the length of the academic year, including a 9-month part-time (20 hour/week) **Lead Research Assistant** to participate in metadata work and lead CBPAR research, community collaboration, and relationship building ($16393.32); three 9 month part-time (10 hour/week) **Data Curators** to undertake data transformation experiments and diary the process ($24,589.98); and 2) honoraria/incentives for both **Focus Group Discussion participants** (4 groups of 4 individuals) to evaluate Wikidata integration for Indigenous primary sources at the completion of our data integration work ($800), and for **Community Collaborators**, to compensate Tribal and Indigenous archivists or community members for consulting on use of community-based collections or information in Wikidata ($1,600); plus 15% indirect/overhead cost ($6507). See our budget template here.

References

Association of Research Libraries (2019). *ARL White Paper on Wikidata: Opportunities and Recommendations.* ARL Task Force on Wikimedia and Linked Open Data, Washington, D.C.: https://scholarworks.iupui.edu/bitstream/handle/1805/18902/ARL-white-paper-on-Wikidata.pdf

Babcock, K., Heberlein, R., McCormick, A. B., Roke, E. R., Suiter, G. K., & Tillman, R. K. (2021). *The Power of Parallel Description: Wikidata and Archival Discovery*. Retrieved

from https://stacks.stanford.edu/file/gg453cv6438/LTWHandbook.pdf

Baldwin, D., Hinton, L., & Pérez-Báez, G. (2018). The Breath of Life Workshops and Institutes. In L. Hinton, L. Huss, & G. Roche (Eds.), *Routledge Handbook of Language Revitalization* (pp. 188–196). New York: Routledge.

Bhasin, A., Roy, L., & Arriaga, S. K. (2011). *Tribal libraries, archives, and museums: Preserving our language, memory, and lifeways*. Scarecrow press.

Bucciferro, A. (2008). Attacking the Backlog: NARA Archivists Mobilize to Make Unprocessed Records Available to the Public. *Prologue Magazine, 40*(2), 46-51.

Carpenter, B. (2019). Archival Initiatives for the Indigenous Collections at the American Philosophical Society. *Case Studies on Access Policies for Native American Archival Materials*(1), 1-13. Retrieved from https://www2.archivists.org/sites/all/files/Case_1_Archival_Initiatives_for_Indiginous_Collections.pdf

Carroll, S., Garba, I., Figueroa-Rodriguez, Holbrook, J., Lovett, R., Materechera, S., . . . Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal, 19*(XX), 1-12. doi:https://doi.org/10.5334/dsj-2020-042

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific data, 8*(1), 108. doi:10.1038/s41597-021-00892-0

Caswell, M., & Cifor, M. (2016). From Human Rights to Feminist Ethics: Radical Empathy in the Archives. *Archivaria*, 23--43. https://archivaria.ca/index.php/archivaria/article/view/13557

Charmaz, K. (2008). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage.

Christen, K., Merrill, A., & Wynne, M. (2017). A community of relations: Mukurtu hubs and spokes. D-Lib Magazine, 23(5/6).

Czerwinski, M, Horvitz, E., and Wilhite, S. (2004). "A diary study of task switching and interruptions." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria. https://doi.org/10.1145/985692.985715.

Elliott, H. (1997). The Use of Diaries in Sociological Research on Health Experience. Sociological research online, 2(2), 1-11. doi:10.5153/sro.38

Fan, J., Haqq, D., Saaty, M., Wang, W.-L., & McCrickard, S. (2025). Diary Study as an Educational Tool: An Experience Report from an HCI Course. Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1, 311–317. https://doi.org/10.1145/3641554.3701845

Fforde, C., McKeown, C. T., & Keeler, H. (Eds.). (2020). The Routledge companion to Indigenous repatriation: Return, reconcile, renew. Routledge. https://doi.org/10.4324/9780203730966

Fforde, C., Ormond-Parker, L., & Turnbull, P. (2015). Repatriation Research: Archives and the Recovery of History and Heritage. In R. E.-C. null (Ed.), Human remains and the law (Vol. 1, pp. 39–59). Institute of Art and Law.

Garcia, P., & Cifor, M. (2019). "Expanding our reflexive toolbox: Collaborative possibilities for examining socio-technical systems using duoethnography." Proceedings of the ACM on Human-Computer Interaction 3 (CSCW): 1-23.

Greene, M. A., & Meissner, D. (2005). More Product, Less Process: Revamping Traditional Archival Processing. The American Archivist, 68(2), 208-263.

Holton, G. (2012). Language archives: They're not just for linguists any more. In F. Seifart, G. Haig, N. P. Himmelmann, D. Jung, A. Margetts, & P. Trilsbeek (Eds.), *Potentials of Language Documentation: Methods, Analyses, and Utilization,* (Vol. No 3, pp. 105–110): Language Documentation & Conservation Special Publication.

Jorgensen, M., & Johnston, B. (2022). *Chapter 1. Tribal Archives* (Sustaining and Advancing Cultures: Field Surveys and Summits, 2021). Association of Tribal Archives, Libraries, and Museums.

Leavy, P. (2017). Community-Based Participatory Research Design. In P. Leavy (Ed.), Research Design : Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches (pp. 224-254). New York: Guilford Publications.

Littletree, S., Belarde-Lewis, M., & Duarte, M. (2020). Centering Relationality: A Conceptual Model to Advance Indigenous Knowledge Organization Practices. *Knowledge Organization, 47*(6), 410–426.

Littletree, S., & Metoyer, C. A. (2015). Knowledge organization from an indigenous perspective: The Mashantucket Pequot thesaurus of American Indian terminology project. *Cataloging & Classification Quarterly, 53*(5-6), 640-657.

Lowery, M. M. (2009). Telling Our Own Stories: Lumbee History and the Federal Acknowledgment Process. American Indian Quarterly, 33(4), 499-522. Retrieved from http://www.jstor.org/stable/40388484

Marsh, D. E., Fenlon, K., Sorensen, A. H., & Wise, N. M. (2024). Primary Sources as Linked Data: Exploring Motives Across the Sciences and Social Sciences. *Proceedings of the Association for Information Science and Technology, 61*(1), 232-245. doi:https://doi.org/10.1002/pra2.1023

Marsh, D. E. (2023). Digital knowledge sharing: perspectives on use, impacts, risks, and best practices according to Native American and Indigenous community-based researchers. Archival Science, 23(1), 81-115.

Marsh, D. E. (2019). Research-driven approaches to improving archival discovery. IASSIST Quarterly, 43(2), 1–9. https://doi.org/10.29173/iq955

Marsh, D. E., Bell, J. A., Greene, C., & Turner, H. (2021). Bridging anthropology and its archives: An analysis from the Smithsonian's National Anthropological Archives. *Anthropology Today*, 37(2), 19–22. https://doi.org/10.1111/1467-8322.12643

Montenegro, M. (2019). Unsettling evidence: an anticolonial archival approach/reproach to Federal Recognition. *Archival Science, 19*(2), 117-140. doi:10.1007/s10502-019-09309-9

Odumosu, T. (2020). The Crying Child: On Colonial Archives, Digitization, and Ethics of Care in the Cultural Commons. *Current Anthropology*, *61*(S22), S289–S302. https://doi.org/10.1086/710062

O'Neal, J. R. (2014). Respect, Recognition, and Reciprocity: The Protocols for Native American Archival Materials. In D. Daniel & A. Levi (Eds.), Identity Palimpsests: Archiving Ethnicity in the U.S. and Canada (pp. 125–142). Sacramento, CA: Litwin Books.

O'Neal, J. R. (2015). "The Right to Know": Decolonizing Native American Archives. Journal of Western Archives, 6(1). https://doi.org/10.26077/FC99-B022

Oard, D. W. (2023). *Known by the company it keeps: Proximity-based indexing for physical content in archival repositories.* Paper presented at the International Conference on Theory and Practice of Digital Libraries.

Panitch, J. M. (2001). Special Collections in ARL Libraries: Results of the 1998 survey sponsored by the ARL Research Collections Committee. Washington, DC: Association of Research Libraries.

Punzalan, R. L. (2014). Understanding Virtual Reunification. *The Library Quarterly, 84*(3), 294-323. doi:10.1086/676489

Punzalan, R. L. (2014). Archival Diasporas: A Framework for Understanding the Complexities and Challenges of Dispersed Photographic Collections. *The American Archivist, 77*(2), 326-349. doi:10.17723/aarc.77.2.729766v886w16007

Reis, H. T., & Wheeler, L. (1991). Why a Diary Procedure? Advances in Experimental Social Psychology, 24, 270.

Reijerkerk, D., & Reed, C. M. (2023). Archives, Decolonization, and the Politics of Tribal

Sovereignty: An Examination of Accessibility Barriers to Indigenous Federal Recognition Research in the United States. The American Archivist, 86(2), 565-594. doi:10.17723/2327-9702-86.2.565

Roke, E. R., & Tillman, R. K. (2022). Pragmatic Principles for Archival Linked Data. *The American Archivist, 85*(1), 173-201.

Schmid, O., & Cliggett, L. (2011). Workshop on a Registry of Anthropological Data, American Anthropological Association, sponsored by the National Science Foundation, Arlington, VA.

Schweitzer, I., & Henry, G. J. (2019). Afterlives of Indigenous Archives: Essays in Honor of the Occom Circle. Hanover, NH: Dartmouth College Press.

Sheble, L., Thomson, L., & Wildemuth, B. M. (2017). Research diaries. In B. M. Wildemuth (Ed.), Applications of social research methods to questions in information and library science (2 ed., pp. 228-). Santa Barbara: ABC-CLIO-LLC.

Smith, L. (2006). *Uses of Heritage*. New York: Routledge.

Swantz, M. L. (2008). Participatory Action Research as Practice. In P. Reason & H. Bradbury, *The SAGE Handbook of Action Research* (pp. 31–48). SAGE Publications Ltd. https://doi.org/10.4135/9781848607934.n12

Szekely, P., Knoblock, C., Yang, F., Fink, E., Gupta, S., Allen, R., & Goodlander, G. (2014). Technology and Scaling: Publishing the Data of the Smithsonian American Art Museum to the Linked Data Cloud. *International Journal of Humanities and Arts Computing 8* 152-166. doi:https://doi.org/10.3366/ijhac.2014.0104

Taitingfong, R., Andrew, M., Maui, H., Raymond, L., Bobby, M., Jacob, P., . . . and Carroll, S. R. (2024). Aligning policy and practice to implement CARE with FAIR through Indigenous Peoples' protocols. *Acta Borealia, 41*(2), 80-90. doi:10.1080/08003831.2024.2410112

Thorpe, K., Sentance, N., & Booker, L. (2023). *Wikimedia Australia and First Nations Metadata: ATSILIRN Protocols for Description and Access*. Retrieved from Sydney, Australia:

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*(1), 1-9.

Waller, A. D., & Ragsdell, G. (2012). The impact of e-mail on work-life balance. *Aslib Proceedings*, *64*(2), 154–177. https://doi.org/10.1108/00012531211215178

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology, 22*(3), 233-243.