

Vancouver Statement on Collections-as-Data

Padilla, Thomas, Scates Kettler, Hannah, Varner, Stewart, & Shorish, Yasmeeen. (2023). Vancouver Statement on Collections as Data. Zenodo. <https://doi.org/10.5281/zenodo.8341519>.

Last Update: 2023

Preceded by: Santa Barbara Statement on Collections as Data¹, 2017

Since the Santa Barbara Statement on Collections as Data (2017) was published, engagement with collections-as-data² has grown internationally. Institutions large and small, individually and collectively, have invested in developing, providing access to, and supporting responsible computational use of collections as data. An updated statement is needed in light of increased community implementation of collections as data in context of an ever more complex data landscape.

The Vancouver Statement suggests a set of principles for thinking through questions that collections-as-data work produces, as part of an expanding global, interprofessional and interdisciplinary effort to empower memory, knowledge, and data stewards³ (e.g., practitioners and scholars) who aim to support responsible development and computational use of collections as data. This stewardship role only grows in importance as artificial intelligence applications, trained on vast amounts of data, including collections as data, impact our lives ever more pervasively.

The concept of collections-as-data grows out of a particular moment in the history of galleries, libraries, archives, and museums. For decades, practitioners and scholars have been building digital collections. Simultaneously, a diverse range of user groups (e.g., scholars, artists, policy makers, entrepreneurs, journalists, and GLAM practitioners) have drawn upon computational means to derive insights from collections at scale. This work uses a variety of methods, including but not limited to: text mining, computer vision, machine learning, artificial intelligence, data visualization, mapping, image analysis, audio analysis, and network analysis. Efforts to support responsible development and computational use of collections have blossomed internationally. Yet, work remains to initiate, support, and socialize approaches that map to a range of local, regional, national, and international resources and strategic priorities.

While the specifics of how to develop, provide access to, and support the use of collections-as-data will vary, any digital material can potentially be made available as data that are amenable to computational use. Data use and reuse can be encouraged with appropriate licensing and non-proprietary formats,

¹ <https://zenodo.org/record/3066209>

² The hyphenated instance “collections-as-data” is used in this document to refer to the concept, community, and area of practice related to supporting responsible computational use of collections as data.

³ steward: a person who facilitates and advocates for the long-term access and care of collections.

particularly when data is made accessible in many different ways and access is designed to meet community needs and is mindful of ethical concerns.

Understanding and addressing ethical concerns are integral to collections-as-data. Collections-as-data efforts should make access decisions informed by a range of community priorities. Positions advanced by the Global Indigenous Data Alliance's CARE Principles⁴ and work on Maori Data Sovereignty⁵ are instructive. Care must be taken to comply with legal requirements and accessibility needs, respect diverse cultural norms and values, and consider risks to vulnerable groups and respect the rights and authority of these communities.

The scale of some collections may obfuscate the completeness - or lack of completeness - of the histories they are perceived to represent. Memory, knowledge, and data practitioners and scholars must be mindful of these absences or misrepresentations and work to prevent their repetition. Documentation should be informed by archival principles and effective reproducibility practices to ensure that users have the information they need to work with collections responsibly.

The Vancouver Statement was created to be approached both by people who are new to the idea of supporting responsible development and computational use of collections as data, as well as people who are well-versed in long established approaches to gallery, library, archive, and museum practice.

Implications of principles outlined below do not need to be addressed all at once. Indeed, some may remain aspirational for some time as people begin this work in concert with colleagues holding different experiences, values, and approaches to the work. The point is to start somewhere, assess, refine, and improve.

1. **Collections-as-data development supports responsible computational use of digitized and born digital collections.** Knowledge, data, and memory stewards expand the set of opportunities for engaging with collections by conceiving of, creating, and providing access to collections as data.
2. **Collections-as-data stewards are committed to working against historic and contemporary inequities represented in collection acquisition, scope, description, access, and use.** Commitments to working against inequities should be documented and made publicly available, except in the event that making such commitments public place communities at risk. This documentation should include, when applicable, any known silences and/or biases in the collections. Commitment details will vary across communities served by collections but will share common cause in seeking to address the needs of the vulnerable. Stewards may impose well-considered barriers to use - including refraining from making a collection available as data.

⁴ Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M., 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), p.43. DOI: <https://doi.org/10.5334/dsj-2020-043>

⁵ <https://www.temanararaunga.maori.nz/>

Stewards will respect the rights and needs of the communities who create content that constitute collections, those who are represented in collections, as well as the communities that use them.

3. **Collections as data should be widely accessible, within the bounds of ethical, legal, and community expectations.** Terms of use for collections as data must be made explicit and should be informed by community-based practices such as RightsStatements.org and the CARE principles and standard licenses such as Creative Commons⁶, Open Data Commons⁷, and Traditional Knowledge⁸ licenses. Materials should be compatible with accessibility tools and software and should be made available in multiple languages or in a format that can be machine-translated.
4. **Collections-as-data benefit from participatory design.** Rather than assuming needs or imagining user communities, stewards should collaborate with others and work to lower the barriers to use (where appropriate), consider universal design principles, and continue to assess user needs over time. Where resources permit, multiple approaches to data development and access are encouraged to facilitate interoperability (as described in Principle 6). People should be motivated and encouraged to build and share the outcome of this work.
5. **Shared documentation provides context and helps others find a path to doing the work.** Collections-as-data work involves decisions about developing, providing access to, and supporting use. Documentation should clearly explain how and why decisions were made in collections-as-data work. While no documentation can be fully comprehensive, incomplete or in-progress documentation is better than no documentation. Examples of documentation include strategic plans, job descriptions, human and machine readable metadata schemas, data sheets, workflows, application profiles, deeds of gift, and codebooks – with translations to different languages. Language choice should reflect user needs. Documentation should be publicly accessible and persistent, given appropriate ethical and legal considerations.
6. **Collections-as-data development values social and technical interoperability.** Interoperability entails alignment with emerging and/or established community models (staffing, matrixed teams, maintenance and innovation strategies), standards, and infrastructure and eases integration with centralized as well as distributed infrastructure. Consideration should be given to the biases present in standards and infrastructure. Interoperability extends beyond the technical, to social interoperability — the ability of people to interact effectively and support sustainable human infrastructure.

⁶ <https://creativecommons.org/about/clicenses/>

⁷ <https://opendatacommons.org/>

⁸ <https://localcontexts.org/licenses/traditional-knowledge-licenses/>

7. **Collections-as-data stewards work transparently in order to maintain integrity for long-term access to collections.** Collections-as-data integrity depends upon efforts to ensure and publicly document the technical integrity of the data, its provenance, and the relationships present between the collections stewards and communities and/or user groups. It requires that stewards acknowledge absences, biases, and areas of uncertainty. Transparent collections as data should include open, robust metadata (in alignment with Principle 3) and should have a preservation commitment associated with them.
8. **The progression of collections-as-data work depends on organizational commitments to sustainable infrastructure and processes.** The ongoing development of collections as data can impact staffing models, workflows, and technical infrastructure. Entities and organizations should support the development of productive collaborations and integrations between new and existing workflows, staff positions, and service models.
9. **The work of developing collections as data should balance benefits with concern for climate impact.** Digital work, by nature, relies on energy intensive data centers and uses heavy metals and other materials detrimental to the environment. Where possible, design processes that leverage computational resources that are transparent about the ways that they do or do not negatively impact the environment.
10. **The work of developing collections as data should balance benefits with concern for exploitative labor.** Creating collections as data may have been, or is, facilitated by some form of exploitative labor. Creators should acknowledge labor contexts (when known) through appropriate data documentation and the utilization of taxonomies such as CRediT⁹.
11. **Collections-as-data stewards should design access with appropriate consideration of data consumption by artificial intelligence or other technologies.** Collections as data may be used in training sets for artificial intelligence and product development. Stewards should be mindful of the ethical implications, including intellectual property and claims to data sovereignty, and should develop adequate safeguards against improper usage, detrimental loss of context, and the amplification of biases through these technologies.

⁹ <https://credit.niso.org/>