

ESTS Open Data Policy Analysis

Analysis of Open Data policies collected [here](#) based on the analytic [Reading OA-OD Technical Infrastructures](#) by Angela Okune

Points to raise:

- “Opening” Data vs “Sharing” Data
 - Doesn’t mean that we require everything to be open. Open Data is modular.
- What community is imagined by the policies? (One that has space for acts of refusal vs one that takes open ended sharing as the fundamental good).
- Keywording is so powerful in what it indicates. Small practice that could be easily infrastructured.

Ideas for our own ESTS policy:

- Data supplement a la <https://qdr.syr.edu/ati/ati-instructions> to accompany submissions
- Ethnographic data glossary (rewriting of meta values statement)
- Develop keyword taxonomy

Prompts for Oct 2021 ESTS Open Data Roundtable

- Describe the systems/initiative(s) you are working on and with what social, political and economic commitments?
- How have you been able to make those commitments explicit/visible?
- What do you think is at stake in open data? How have you and your colleagues been thinking about publishing open data?
- Do any of you have experience and initiatives related to peer-review of data?
- Particular challenges you have faced in credentialing and infrastructuring your initiative?

Develop nested data values statement:

Sharing data is better.

One step towards recognizing the importance of the relationality of data and people. Not just data as an object of valuation.

Whatever data can be shared, should be shared.

Leave traces with your data about your ethical decisions (where did you get this data, how did you decide inclusion and exclusion of data, why are you sharing it, what do you hope with it).

Design/Visions/Goals

This section asks the following questions:

1. Who was the system built to serve and why?
2. Who built the system and with what social, political and economic commitments?
3. What assumptions about language and knowledge are built in?
4. What functions are provided by the system?
5. What is the business model?

User Engagement

This section asks the following questions:

1. What user data is captured and displayed?
2. Who uses the system and how often?
3. How is the system actually used? What counts for divergence between intended and actual use?
4. What pathways are users directed to go through the system?
5. What other pathways seem to have emerged?
6. What opportunities are there for user feedback?
7. How do users reference their use of the system?

[Inclusions & Exclusions of Data Types](#)

This section asks the following questions:

1. What kinds of data do users enter?
2. What best practices are recommended for different data types?
3. What data types are not mentioned in the policy?

[Data Organization, Preservation & Movement](#)

This section asks the following questions:

1. Where and how is data stored (backend technologies, formats) and preserved?
2. How does data move through the system? What functions allow for data discovery?

Data Sharing & Management

This section asks the following questions:

1. When and how is data publicly shared? At what cost?
2. Who can data be shared with?
3. Are there different levels of access to data? What are the terms of use and licensing?

Metadata and Documentation

This section asks the following questions:

1. What metadata is required or recommended for different data types?
2. What documentation is required or recommendation for different levels of access?

Citation and Attribution

This section asks the following questions:

1. How is data and metadata to be cited?
2. How is data and metadata to be attributed?
3. How are collaborations cited and attributed?

Collaborative Practice

This section asks the following questions:

1. How does collaboration seem to be conceived and how is it implemented?
2. Does the system have explicit partners? Who?
3. How are partnerships and collaborations decided on?
4. What governance mechanism structures collaborations?
5. How is collaboration to be sustained?

Technical Infrastructure

This section asks the following questions:

1. When did the system go live?

2. When was the system updated most recently? Who makes changes to the system?
3. Where are the servers hosted (geographically and institutionally)?
4. Who responds to platform issues/breakdowns/bugs?
5. Is the work done on the platform (revisions, updates) visible or hidden?

Data Relationalities

Questions:

How are people working with indigenous and/or community data demonstrating benefit of data collection, archiving and purposing for communities? (data sovereignty being synonymous with nation-state building and how tribal nation-building and citizenship unsettles that, concern about data suzeiranty)

What is the scale of aggregation of data? Why is it aggregated at that scale? How might it be repurposed or aggregated differently? (concern about disaggregation of indigenous data under federal policy of deductive disclosure)

Where is indigenous or community data stored? What is the data infrastructure scaffolding it? Who governs the data infrastructure? (concern about data hoovering)

What narratives surround data, or give data meaning?

Recommendations:

-Instead of using data availability statement, using [indigenous data agreements](#) that in addition to this template, asks people to reflect on their relationship with data they collect, purpose, and store

-marking people and marking their contributions as “good metadata” (In Pollution is Colonialism, Liborion marks people as how they identify, for e.g., settler/diasporic settler of color/Black/Indigenous and associated tribal citizenship, and for those who don’t, they use the category “unmarked”: how does this look as a metadata practice?)

-What different labors are involved in the article? [CLEAR Lab manual](#): “We distinguish an author as someone who is accountable to the paper. If it bombs and is deemed unethical, who is accountable? If someone walked up to someone on the author list and said

“congrats on the new pub!” would that person know what they were talking about? All authors have to have read the paper and consented to be on it (this is the last step).” As metadata, clustering around different labors?

Equity in Author Order: A Feminist Laboratory’s Approach: <https://catalystjournal.org/index.php/catalyst/article/view/28850>

-Mukurtu recommends [media content warnings](#) as metadata

-the debate on TK labels?

-the role of [OCAP training](#)?

-QDR further has “[data project types](#)” that aid authors in depositing their data. The type “[ATI Data Supplement](#)” can be useful for PECE-style annotation/analytic process. Requires a Data overview Statement of 1,000 words that asks authors to think about why they chose particular data/datasets/evidence, instrumentation, and omissions/possibilities; why they made analytic choices; and “logic of annotation” i.e. why a particular passage was highlighted for annotation, etc. [Sample ATI project](#)

<https://blog.thegovlab.org/post/selected-readings-on-indigenous-data-sovereignty> (annotated bibliography)

[CARE Principles for Indigenous Data Governance](#)

“existing FAIR principles—that open data should be findable, accessible, interoperable and reusable—focus on data characteristics that facilitate increased sharing while ignoring historical context and power differentials.”

“the addition of CARE: that open data should be for collective benefit, recognize Indigenous peoples’ authority to control their own data, carry a responsibility to demonstrate how they benefit self-determination, and have embedded ethics prioritizing the rights and wellbeing of Indigenous people.”

[Good Data](#)

“Open Data Charter’s commitment to free use, reuse, and redistribution by anyone, at any time, and anywhere, for example, is in direct conflict with the rights of Indigenous Peoples to govern their own data and control how and by whom it is accessed”

[Indigenous Data Sovereignty: An Agenda](#)

Context of the book: “The purpose of the workshop, sponsored by the Academy of the Social Sciences in Australia (ASSA) and the Centre for Aboriginal Economic Policy Research (CAEPR) at The Australian National University, was to identify and develop an indigenous data sovereignty agenda, leveraging international instruments such as the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP)” (3)

-disaggregation of indigenous data

-non-anonymization for targeted intervention

- “indigenous peoples face the potential of neo-data colonialism from the globalisation of information systems and ‘big data’.” (7)

-‘deficit data–problematic people’ correlation when you search for “indigenous data” on google

-what forms of sociality do data indicators point to for indigenous people?

-case studies advocate for collective conceptions of planning that inform indigenous aspirations and agendas

-need to develop data capacity within indigenous people for self-determination and control over data

- “data sovereignty means managing information in a way that is consistent with the laws, practices and customs of the nation-state in which it is located.” (39)

- “That indigenous people are typically poorer than the surrounding settler state has important implications for data sovereignty. This is because collecting data that can be turned into information and later organised into meaningful knowledge is a costly process.” (40)

-the problem of deductive disclosure for indigenous communities: federal statistics offices use this kind of data to protect small indigenous communities from being identified based on a combination of specific indicators and neighborhoods. However, this data is not available to indigenous communities even with their consent

-indigenous IRBs and the capacity of indigenous people to collect their own data about issues relevant to them

-data suzerainty: replacement of indigenous data systems (art, oral histories, genealogies, for example) with imperial information systems

-where is indigenous data stored? Ian Pool: “Data hoovering” For e.g. credit card companies might store data offshore. Or historical data converted into digital formats. Something that has already happened with tattoo designs in NZ

“some hapū ‘books’ contain information that has scientific or other properties, sovereignty and pecuniary values, all of which are disputed in cross-national trade negotiations under way at present” (71)

-what kinds of data do indigenous communities need? Disaggregated data, relational data, population projections, ecological and environmental data,

Desi Rodriguez-Lonebear: “Tribal data building continues as a critical catalyst for tribal nation rebuilding”

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3754667

“the University of Wyoming Art Museum (USA) holds a large collection of objects from Rapa Nui (colonially known as Easter Island) and is working collaboratively with the Father Sebastian Englert Anthropological Museum and other island stakeholders on a digital repatriation project. Questions rising from the project include: What are the legal and political issues that complicate sharing objects across borders? What are the obstacles and advantages of involving individuals who are non-indigenous by birth but are welcomed by the community? Who owns the data of digitized cultural objects?” (78)

Digitized data crosses b/w territorial boundaries of Wyoming, US, Chile, Rapa Nui

“Who has the right to make the choice to digitize cultural objects?”

<https://meridian.allenpress.com/american-archivist/article/74/1/185/24254/Opening-Archives-Respectful-Repatriation>

- “Reciprocal curation, then, is a set of practices through which both tribes and scholars annotate item-level records within specific collections with the aim of producing a rich, layered, dynamic set of knowledge.”