# Document provenance

Copied from outline (Lourdes’s) to normal text

## Color coding:

Black: drafted

Purple: outline topics; these need text drafted

Red: missing info, refs, figs, etc.

# Candidate titles

Repository Recommendations: Implications of the CARE Principles

Recommendations for Earth and Environmental Data Repositories: Implications of the CARE Principles

Recommendations for Implementing CARE Principles by Earth and Environmental Data Repositories

Repository Recommendations: How to Implement the CARE Principles

Repository Recommendations: Implementing the CARE Principles

Link to activities list (see tab 3, “DiagramText” and “Consolidated\_CARE\_columized\_match\_Ruth\_fig“ ): <https://docs.google.com/spreadsheets/d/1kndErikgjiZMl-4tHc8RaJh5XTv5VoQpgsqNSC8F8XA/edit#gid=1565777621>

# Paper Draft

## Abstract

Datasets, from the point of collection to analysis, storage, curation, and dissemination carry cultural and political context. Historically data repositories have taken their guidance and policies as a combination of mandates from their funding agencies and the needs of their user communities, typically universities, agencies and researchers. Consequently repository practices have rarely taken into consideration the needs of other communities such as the Indigenous peoples on whose lands data are often acquired.

In recent years, a number of global efforts have worked to improve the ethical conduct of research as well as the ethical treatment of data by the repositories that hold and disseminate it. One of these efforts, initiated by the Global Indigenous Data Alliance (GIDA) [(“Global Indigenous Data Alliance” 2023)](https://www.zotero.org/google-docs/?Kz2MTe), established the CARE principles of Indigenous Data Governance [(“CARE Principles” 2023)](https://www.zotero.org/google-docs/?jfxHxN) (Collective Benefit, Authority to Control, Responsibility, and Ethics) with intent of ensuring community benefit and control over Indigenous data.

In order to support these new Principles, repositories need to update the services they provide. The question is how? Operationalizing principles into ongoing repository activities is generally a fraught process. This paper, arising from many of the repositories within the Earth Science Information Partners [(“Home - Earth Science Information Partners (ESIP)” n.d.)](https://www.zotero.org/google-docs/?6BILvC) in conjunction with members of GIDA, defines and prioritizes the set of activities repositories can take to become CARE Principles compliant in the hopes that this will help implementation in repositories globally.

## Introduction

The technological revolution of the 21st century has radically transformed our abilities to access and produce “big data” about virtually every aspect of our world. As data managers develop standards for storing and retrieving data, we need to remember that data from the point of collection to curation, storage, analysis, and dissemination carry cultural and political context. Indigenous communities are amongst those that have historically been disproportionately harmed and marginalized through research and the materials that are part of this research (Karuka 2017).

As a group of data practitioners and scholars from the Earth Science Information Partnership, the motto of which is to “make data matter” [(“ESIP Interviews Archives” n.d.)](https://www.zotero.org/google-docs/?R0RPAr), we are primarily looking to the work laid out by Indigenous activists and scholars who have formed Indigenous Data Sovereignty Networks (IDSNs) spanning across the USA, Canada, Aotearoa New Zealand, Spain, and Australia under the Global Indigenous Data Alliance (GIDA) [(“Global Indigenous Data Alliance” 2023)](https://www.zotero.org/google-docs/?Qdc0Fy) to define how repositories need to respond to the CARE principles [(Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019)](https://www.zotero.org/google-docs/?3vb2qT).

The CARE principles of Indigenous Data Governance are “grounded in community values, which extend to society at large,”[(Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019)](https://www.zotero.org/google-docs/?m4hlq6) and include collective benefit, authority to control, responsibility and ethics [(Carroll et al. 2020)](https://www.zotero.org/google-docs/?45IRj7). As we apply these CARE principles to environmental and earth data repositories, it is important to remember that CARE is rooted in the Indigenous Data Sovereignty movement and to ensure that we are not erasing the principles from their provenance as they guide us.

These guidelines are written for data repository managers, particularly managers of Earth and environmental science data in the kinds of repositories covered in [re3data.org](http://re3data.org). In general, such repositories try to take good care of data and to make it ethically available for the long-term, defined as timescales much longer than the working life of an individual researcher. These data repositories can include many different types of data; for example, biological data, records of individuals, traditional ecological knowledge, and geospatial data. As these environments are shaped by indigenous peoples and they also impact indigenous culture, access to these repositories and providing acknowledgement of the Indigenous peoples in the repository is important. However, often such repositories may be entirely unaware that they are holding Indigenous data defined as any “data, information and knowledge, in any format, that impacts Indigenous Peoples, nations, and communities at the collective and individual levels” [(Global Digital Data Alliance 2022)](https://www.zotero.org/google-docs/?MjKrEc). In addition it should be noted that these guidelines do not cover research activities that occur prior to a repository becoming involved with a project or its activities, and which are typically overseen by universities and research funding agencies.

This paper begins with a background outlining the role of repositories in the data lifecycle and why it is important to build and manage repositories according to CARE and defining several terms important to understanding the rest of the paper. Then, we provide sets of recommendations that are aligned with four categories pertaining to most data repositories: Know your community, Technology, Outreach and Repository protocols. Our aim is to provide repositories with guidelines for implementing practices that promote proper handling of data related to Indigenous Populations.

## Background

### Earth and Environmental Data Repositories

In general data repositories and archives act to curate data, be it digital or analog, for the long term making it available for use by a variety of audiences [(CCSDS 650.0-B-1 2012)](https://www.zotero.org/google-docs/?xDFHAo). The kinds of data a repository curates depends on that repository’s mission and can vary widely. For example, some repositories only acquire data from missions or projects sponsored by their funding agency, while domain specific repositories may only take data on certain topics, institutional repositories may only take data from researchers workings within that institution, and generalist repositories such as Figshare and Zenodo, may only take data that meets certain size or other constraints. Moreover, any given repository may fall under several categories, such as domain specific institutional repositories that only acquire data from within a certain domain and then only from researchers associated with that institute [(Baker and Duerr 2017)](https://www.zotero.org/google-docs/?RXjkQP).

Similarly, the services provided by repositories and the audiences they serve also vary widely and evolve over time as recognition of the utility of the data they hold changes. Often this leads to a general broadening of the audiences served and more recently increased recognition of the rights of communities and individuals from which the data came. (Baker, Duerr, and Parsons 2015?)

A repository’s funding source(s) often determines the level of curation provided for the data, though in many cases the completeness of the provenance and metadata associated with the data depends strongly on the researchers or agencies that provided it [(Baker and Duerr 2017)](https://www.zotero.org/google-docs/?AcK1bn). As a result, repositories are only rarely involved in the entire data lifecycle [(Ball 2012)](https://www.zotero.org/google-docs/?Ax1fKX), often not playing a role until well after the data has been collected.

Figure 1: The role of data repositories as mediators between data provided by researchers and agencies of all types and humanity in general

### Definitions

Provenance

Landing Page

Metadata

Education (broadly and in terms of access)

FAIR, TRUST and CARE and how they differ (use Venn diagram from Margaret’s slide deck?)

## Methods

## Recommendations for Repository Responsibilities

### Summary

We identified 51 specific activities that repositories engage in that may be related to aspects of Indigenous data governance ( Table 1). This list does not cover all possible repository activities; just those that may relate to Indigenous data. From the data repository point of view, activities were organized into four broad categories based on the areas of expertise typically represented in repository staff: Know your community, Outreach, Repository protocols and Technology.

**Know your community**. Data-centric consultations are generally carried out between depositors and a repository’s data specialists or curators who have background in a particular scientific domain. For the repositories represented in ESIP, that community historically has been environmental or Earth scientists. But when the data overlaps with the interests of indigenous populations, repositories incur a duty to expand that understanding to aspects of data that are of interest to Indigenous communities as well, making them valued voices in the data curation process.

**Outreach** activities are generally concerned with a) engaging users and new communities or sub-groups, and b) communicating to the public. Some repositories have a designated outreach officer or user services group who handles contact lists, web page construction, seminar schedules, etc; others approach communication and outreach ad hoc. Repositories are expected to practice transparency in their policies and decisions to establish and promote trust. Including Indigenous communities with appropriate communication strategies creates valued stakeholders.

All repositories have operational procedures for how decisions are made, budgets are allocated, and how day to day processes are carried out. **Repository Protocols** are generally established by repository leadership, with collaborative input from staff. New or adapted requirements (e.g., complying with FAIR or CARE principles) will require new operational protocols or checklists. An effective repository develops their protocols with their key user groups in mind. Repository policies and workflows should take into account Indigenous community knowledge and concerns. Recognizing the rights and interests of these communities and incorporating their input into the policies and protocols surrounding the repository promotes improved engagement and decision-making. Positive relationships with the Indigenous communities they serve will increase trust and reduce harm, allowing all parties to understand the implications of decisions. By placing Indigenous community members in a position to influence institutional policies and workflows, a repository can promote improved Indigenous governance, ethics and equity in their collections.

**G**enerally, **technology** is concerned with the creation and maintenance of software. Technical staff often have expertise in only their software field, without explicit scientific or data expertise. Importantly, technical features are how most people interact with the repository (e.g., through its website), and how policies are implemented (e.g, a decision to require logins). A repository must consider their potential range of stakeholders when implementing technical features, including infrastructure to establish appropriate provenance, rights, and access restrictions. Linkages between related outputs, agreements, and other significant content promote usability, benefits, and protections to the communities served. Repositories should engage with Indigenous communities to establish mechanisms to protect private information. In order to appropriately implement these technical features, repositories must understand which users require access to specific content and the appropriate circumstances to transfer these usage and access rights.

Table 1: Repository activities identified for ESIP that involve Indigenous data and people

|  |  |  |
| --- | --- | --- |
| **Category** | **ID** | **Activity** |
| Know your community | K1 | Ongoing engagement with Indigenous communities  |
| K2 | Learn enough to determine if data access restrictions are necessary |
| K3 | Learn enough to determine if data obfuscation is necessary  |
| K4 | Understand Indigenous legal rights  |
| K5 | Understand consequences of publishing the data  |
| K6 | Be aware of changing roles/relationships over time |
| K7 | Recognize when collection agreements were needed for data being deposited  |
| K8 | Ensure tool design, selection of data formats, and metadata are appropriate for communities |
| K9 | Identify types of data of interest to communities  |
| K10 | Seek to understand and respect the changing needs of communities  |
| Outreach | O1 | Be part of a broad community  |
| O2 | Identify that you hold Indigenous data |
| O3 | Advertise your governance protocols |
| O4 | Ensure practices are transparent |
| O5 | Ensure data management policies are defensible  |
| O6 | Facilitate relationships between data provider and user communities  |
| O7 | Share material with the indigenous workforce  |
| O8 | Include indigenous representatives on repository advisory board if relevant  |
| O9 | Liaise between communities and vocabulary maintainers |
| O10 | Identify communities for equitable educational engagement  |
| O11 | Identify badges that clarify types of use |
| O12 | Identify types of data that cannot be supported technically  |
| Repository protocols & procedures | P1 | Develop data policies and acquisition/curation protocols  |
| P2 | Develop procedures to work with depositors to minimize restrictions  |
| P3 | Develop procedures to request required documents during data submission |
| P4 | Develop procedures for determining if needed technologies are within technical plans |
| P5 | Develop contact protocols and follow up processes |
| P6 | Develop continuity plans |
| P7 | Allocate educational funds equitably |
| P8 | Develop processes to identify communities for engagement and collaboration |
| P9 | Include funds for specific targeted tools in technical plans as possible  |
| Technology | T1 | Enable granular embargoes  |
| T2 | Support tracking data provenance |
| T3 | Link between data, papers and other outcomes |
| T4 | Create data access control mechanisms (who/when) |
| T5 | Implement systems that allow transfer of responsibilities |
| T6 | Package required documents with their data  |
| T7 | Support communication of rights and restrictions) |
| T8 | Implement badging mechanisms |
| T9 | Use a metadata format that handles multiple languages |
| T10 | Keep a list of data “actors” and be able to link data to them |
| T11 | Use community-vetted vocabularies  |
| T12 | Support customized data views |
| T13 | Identify metadata fields for recording required documents, proper use and limitations |
| T14 | Identify appropriate metadata fields for supported badges |
| T15 | Display required documents, badges, proper use and limitations on landing pages |
| T16 | Display relevant badges on landing pages |

Figure, main graphic ? table if we need readable text (appendix?).

Overall:

Some activities map to more than one CARE principle.

Tech has the highest number of CARE-related activities (16); the other 3 groups are relatively close in totals and lower.

Of the CARE principles, there are more connections to “E” (Ethics) and “R” (Responsibility) than to C and A. This is not surprising, as repositories are focused on data preservation and access, and so already tasked with many aspects of Ethics and Responsibility. (refer to TRUST paper here)

Table 2. Number of mappings between CARE principles and each of the four categories of repository activities related to CARE.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Knowledge (10 activities) | Outreach (12 activities) | Protocols (9 activities) | Technology (16 activities) |  |
| C | 1 (10 %) | 4 (33 %) | 1 (11 %) | 6 (38 %) | 12 |
| A | 2  | 4 | 2 | 3 | 11 |
| R | 6 | 2 | 5 | 6 | 19 |
| E | 3 | 6 | 4 | 6 | 19 |

Organize details according to the CARE principles, e.g., discuss activities (in activity-groups) that go with the principle-group. Minimum: paragraph about the CARE-letter, paragraph summarizing the activities that match to it. This is pretty sparse text, though. We had talked about including some example material, as sidebars. still need to talk about that some more.

**Collective benefit (“C”)**

The “C” triad states that data ecosystems must function in ways that enable Indigenous Peoples to derive benefit from the data. A result is inclusive development and innovation that recognizes that data generate value and can enhance Indigenous self-determined development (C1). Data enriches the processes supporting Indigenous communities to enable better engagement between citizens, institutions, and governments and improve decision-making and transparency (C2). Any value created from Indigenous data should benefit Indigenous communities in an equitable manner (C3).

Twelve specific repository activities align with these principles (Figure X.a, Table 2), with 83% (10 out of 12) related to the Technical and Outreach categories. The Outreach activities are related to enhanced relationships with communities (O1, O2), such as acting as a moderator between them and data technologies (O9). The Technical activities were generally related to providing recording data sources and uses and implementation of community-specific terminology or customizations. Data curators must be able to recognize when data being deposited may have had regulations for collection, so that they can inquire about documentation (K5) and ensure that repository systems are appropriate for the data. Repository protocols would support these operations (P2).

Know your community

* Recognize when collection agreements were needed for data being deposited (C3)
	+ This is highly context dependent since the requirements for collection agreements may have legal force at the local to global level or may simply be promulgated by individual communities as expected researcher practices.
	+ Also highly context dependent are the requirements under which the data should have been collected since they can cover the entire range of not only researcher activities but also repository responsibilities. For example, such requirements could cover where and under what conditions the data should be housed or what its access restrictions are.
	+ For data without agreements that are already in the repository, the question becomes what actions should be taken to remedy the situation. This could call for working with the appropriate Indigenous community to create the needed agreements and possibly infrastructure needed to support proper curation of the data to returning the data to the community.
	+ For new data, the issue is one of either 1) ensuring that the researcher has done their due diligence and is providing any needed documentation to the repository along with the data; 2) possibly rejecting the request to deposit data if it has not been appropriately taken; or 3) working with the affected community to remedy the situation as above.

Outreach

* Be part of a broad community (C2)
	+ [When considering accepting a deposit - think about who might be interested in it and whether they are part of your current community. Accepting such data may require you to add new designated communities, including Indigenous communities (Baker, Duerr, and Parsons 2015](https://www.zotero.org/google-docs/?broken=4aKUZz)) or [(Parsons and Duerr 2005)](https://www.zotero.org/google-docs/?4rLoXV)
* Identify that you hold Indigenous data (C1, C3)
	+ Be transparent and clear in all communications about any Indigenous data holdings and what that means for such data. This includes communications in any form - your website, brochures, videos, other collateral. Such communications should indicate whether data restrictions, badging or other mechanisms are being used for these data.
	+ The question of how to deal with global datasets is fraught, since by definition they hold data relevant to every Indigenous community!
* Liaise between communities and vocabulary maintainers (C1)
	+ Vocabulary maintenance processes vary widely and only rarely are repositories directly involved with all of the vocabularies in use within that repository. The situation is even more tenuous for researchers and Indigenous communities. However, most vocabulary maintenance organizations have processes for collecting input and reviewing requested changes. Participate in these processes to ensure that vocabularies in use cover the languages and terms needed.

Protocols

* Develop contact protocols and follow up processes (C2)
	+ Develop protocols for establishing relationships with Indigenous communities both for when you determine that you already hold such data and for accepting new data deposits.
	+ Ensure contact protocols include transparent processes for following up with communities to maintain healthy relationships and to deal with changes over time.

Technology

* Enable granular embargoes (C1)
	+ Depending on the communities involved, it may be necessary to enable data restriction and access mechanisms. This could include requiring community approval before providing access via a restricted account (see for example the Gwich’in Place Names Atlas [(“Gwich’in Place Names Atlas | ELOKA” n.d.)](https://www.zotero.org/google-docs/?LWLLFU) where a request placed to NSIDC’s user services is forwarded to the community for approval); establishment of secure enclaves for data use (a’la highly sensitive data) or mechanisms to restrict access to all or part of a data set amongst others.
* Support tracking data provenance (C1)
	+ While tracking data provenance, i.e., where the data has been and what has happened to it along the way, is always important, here, depending on the communities and data involved, it may also be important to track the uses to which the data have been put and any outputs derived from the data. This can enable concerned communities to verify that agreements are being met and ensure that the data’s value is accruing to those communities.
* Link between data and papers, other outcomes (C2)
	+ Providing links between the data deposited, any derived data created, any papers written, patents developed, etc. is also important for allowing concerned communities to verify that agreements are being met and that the value of the data accrues to the community, not just those that acquired or used the data.
* Use a community-vetted vocabulary (C3, C1)
	+ Terminology matters [(Eitzel et al. 2017)](https://www.zotero.org/google-docs/?tuQGa3)! Data that comes from a community should be understandable and usable by that community! This may mean that all or part of the data, its documentation and metadata may need to be available in multiple languages or use terms relevant and understandable by those communities. For example, the website of the Clyde River Weather network [(“Clyde River Weather” n.d.)](https://www.zotero.org/google-docs/?68eseu), originally established as part of a joint project between local hunters and Elders in Nunavut and researchers from the University of Colorado Boulder and Colorado State University under the auspices of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA) is available in both English and Inuktitut.
* Support customized data views, C2
	+ Similarly, engagement and decision-making is improved when community’s access data through views that are meaningful to them. This may require the development of data derivatives and advanced visualization capabilities. If a repository does not have the capacity to develop these themselves, they may need to provide access routes (e.g., secured application programming interfaces (APIs)) to third parties trusted by communities to provide these services.

**Authority to control (“A”)**

Indigenous Peoples must have the authority and ability to control how aspects of data are represented and identified, e.g., geographical indicators, Indigenous knowledge, or material about Indigenous lands, territories and resources. Specifically, Authority includes collective and individual rights to free, prior, and informed consent in data collection and use (A1), access to data that are relevant to their world views and empower self-determination and effective self-governance (A2), and active leadership in the stewardship of, and access to, Indigenous data, especially in the context of Indigenous Knowledge (A3).

Repositories are generally not involved in early stages of the research life cycle, such as original data collection (see fig \_\_ - data life cycle), nor are they usually directly involved in the governance of their specific research communities. However, aspects of data management and archive are pertinent at these stages, and as the preservation agent for ensuing data, repositories must be aware of those aspects and have mechanisms to record pertinent features. Therefore, repository activities mapped to the recognition of Indigenous rights (A1), and the governance of data (A3), with more mappings to the Outreach category (table 2), as these are concerned with transparency, communication and engagement with new communities . Knowledge of the data must include an understanding of Indigenous legal rights and consequences (K3), and the ability to identify data of interest to specific communities (K7). Repository mechanisms are needed to determine if the repository is suitable for data that are confidential or sensitive (P9), along with appropriate protocols for their handling protocols (P1), with related Outreach activities to communicate those practices data (O3, 4, 5, 11), and Technical activities for their support (T4, 5, 6).

Knowledge of data and community

3 Understand indigenous legal rights (A1), consequences of publishing these data (E3)

7 (43) identify types for data of interest to communities (A3, E1, E3)

Outreach

3 Advertise the repository’s own governance protocols (A3)

4 Have transparent practices (A3)

5 Have defensible data management policies (A3)

11 Engage with IP review processes to assist in determining appropriateness of a repository for certain types of data, e.g, data that are confidential or sensitive. If a data type cannot be supported technically, advertise this. (A3, E1, E3))

Protocols

1 Develop data policies and collection/curation protocols (A1)

9 (43) determine if supporting the necessary tech to contain data that are confidential or sensitive is within your organization's tech plan (A3, E1, E3))

Technology

4 Create mechanism to control when/by whom data are accessed (A3)

5 Implement login systems that allow for transference of responsibilities (A3)

6 Receive (copies of) agreements between proprietors and package them permanently with data (A3)

**Responsibility (“R”)**

R1 for positive relationships

R2 for expanding capability and capacity

R3 for indigenous languages and worldviews

When working with Indigenous data, there is a responsibility to nurture respectful relationships with those from whom the data originate (Carroll, et al 2020). Relationships must be built on respect, reciprocity, trust, and mutual understanding (R1), that also include enhancement of data literacy within Indigenous workforces (R2). Further, there is a responsibility to create methods for including data in aspects of Indigenous culture (R3).

“R” principles are broad, including responsibility for many aspects of community engagement, starting with strengthening relationships, to expanding Indigenous capacity and supporting their languages and world view. A high proportion of repository activities map to “R”(19 or 40% of the total, Table 2). Only two activities fall into the Outreach area: to liaise between data providers and user communities that include Indigenous Peoples (O5), and developing shareable material for indigenous workforces (O6). The majority (17 activities) are distributed fairly evenly across the other three activity areas. Curators must have significant knowledge of the data held, in order to determine the types of restrictions that might need to be applied and how those may change over time (K1 - K4), and to determine the best data formats and metadata to target specific Indigenous communities (K6). Protocols must exist to request appropriate associated documents from submitters (P4) and to target funds for education or tool design toward Indigenous peoples (R6, R8). It should be obvious that in order to ensure continued access, that a repository also has a general continuity plan (P5). The Technology activities describe the implementation of those protocols.

Knowledge of data and community

1 Engage with Indigenous communities (R2)

2 Learn enough to determine if data access restrictions are necessary or for obfuscation of some elements (R3)

3 Work with depositors to minimize restrictions (R1)

4 Be aware of changing roles/relationships over time (R1)

6 (41) identify targeted communities with the purpose of ensuring tool design, selection of data formats, and metadata are appropriate for targeted communities and users (R3)

Protocol

4 Data submission protocols request associated docs (eg. Written agreements, permits, etc.) (R1)

5 (39) Repository has a continuity plan (R1)

6 (40) If you have educational funds, allocate them equitably, recognizing diff between “equal” and “equitable” (R2, E2)

8 (41) Repository's plan includes funds for specific targeted tools to ensure tool design, selection of data formats, and metadata are appropriate for targeted communities and users (R3)

Outreach

5 Facilitate relationships between data provider and user communities (R1)

6 Share material with an indigenous workforce (R2)

Technology

7 Ensure that rights and restrictions are clearly communicated at deposition (R1) (42) including...

have a mechanism to assign badges that clarify types of use, e.g., Biocultural and Traditional Knowledge (E3) TECH

8 Use a metadata format that handles multiple languages. (R3)

9 Keep a list of data “actors” (e.g., by role) and be able to link data to them (R3)

**Ethics (“E”)**

E1 For minimizing harm and maximizing benefit

E2 For justice

E3 For future use

Define E…

Paramount to ethical data practices is the representation and participation of Indigenous Peoples, who assess benefits, harms, and potential future uses based on their community values. Indigenous Peoples’ rights and wellbeing should be the focus in order to minimize harm and maximize benefits (E1), promote justice (E2), and allow for future use (E3) Carroll et al, 2020).

Summarize mapping to repos activities:

19 total, more in Outreach and Tech.

Several Activities that map to the Ethics group also map to other CARE principles, particularly with A (authority to control) and R (responsibility).

.

Knowledge of data and community

3 Understand indigenous legal rights (A2), consequences of publishing these data (E3)

7 (43) identify types for data of interest to communities (A3, E1, E3)

Protocols

6 If you have educational funds, allocate them equitably, recognizing diff between “equal” and “equitable” (R2, E2)

7 Establish protocol for identifying communities for engagement and collaboration (e.g, by an advisory board) (E2)

9 determine if supporting the necessary tech is within your organization's tech plan (Engage with IP review processes to assist in determining appropriateness of a repository for certain types of data, e.g, data are confidential, sensitive (A3, E1, E3))

Outreach

7 Include indigenous representatives on repository advisory board if relevant (E1, E2)

9 (40) communities for equitable educational engagement are identified (e.g, by an advisory board) COMMUNICATION AND OUTREACH, E2

10 (42) identify badges that clarify types of use, e.g., Biocultural and Traditional Knowledge (E3) COMMUNICATION

11 (43) if data type cannot be supported technically, advertise this (Engage with IP review processes to assist in determining appropriateness of a repository for certain types of data, e.g, data are confidential, sensitive (A3, E1, E3))

Technical

4 Create mechanism to control when/by whom data are accessed (A3)

\_\_ (43) outline tech to support those types (e.g., login, access rights, firewalls) (Engage with IP review processes to assist in determining appropriateness of a repository for certain types of data, e.g, data are confidential, sensitive (A3, E1, E3))TECH, A3, E3"

7 Ensure that rights and restrictions are clearly communicated at deposition (R1) (42) including... have a mechanism to assign badges that clarify types of use, e.g., Biocultural and Traditional Knowledge (E3) TECH

12 Metadata specs contain fields for recording associated material (eg. written agreements, permits, proper use, limitations, BTK, etc.) TECH

## Discussion

* 1. Overlaps and joins between our categories. activities can often be combined to enact broader goals and strategies and intersect many aspects of CARE, e.g., [pick 2 or 3 from this old list of the “mixed” activities - probably the ones that Erin and Bob already worked on! (mob deleted the others) ]
		1. Collaborative tool design - ensure selection of data formats, and metadata are appropriate for targeted communities and users (C1, A2, R3)
		2. Plan to evolve, in order to remain open and respect changing needs of Indigenous communities (C2, R1) ([CARE Paper Draft Text](https://docs.google.com/document/d/1DyWPuaHIzJaA4fsSIMl6N2aC6T9ftyh2qguFB_wJ_i0/edit#heading=h.bwq2mpqnjudm) Bob - text)

We need some sort of intro to these combined activities. They specifically came up in our discussions as ‘things to do’ - why? - check notes. Possibly related to CARE principles being somewhat general

The evolution of a data repository is critical for meeting the changing requirements of sponsors and certification instruments and for advancing technology that can be leveraged to improve the capabilities of each repository and their user communities. Such an initiative would require combining several of the activities outlined above; obviously, the repository must have a long-term continuity plan (P6), which includes technical upgrades (P4). However, there is also an opportunity to improve relationships and to better engage indigenous communities. By inviting members of indigenous communities to serve on their advisory committees (O8), data repositories gain specific feedback on enhanced features and enriching data collections to meet those needs. They also will be able to identify ways to engage more deeply, e.g.,to recruit members of indigenous communities for employment, training workshops, etc (P7, P9). Other potential opportunities include identification of new data collections, new uses for existing data, and streamlined bureaucratic procedures (e.g., contact protocols) which could build efficiency into all interactions.

* + 1. Confirm that depositing researchers have done their due diligence (C3, E1) ([CARE Paper Draft Text](https://docs.google.com/document/d/1DyWPuaHIzJaA4fsSIMl6N2aC6T9ftyh2qguFB_wJ_i0/edit#heading=h.8kwd95f71hf) ERIN)

Because repositories typically receive rather than generate data, they must ensure workflows and protocols exist to account for the authenticity and accuracy of data they accept. They must clearly identify any Indigenous data they hold (\_\_\_) and assume responsibility for putting protocols and workflows in place to facilitate appropriate understanding and data usage (\_\_\_).

Repositories must identify stakeholders and note subtle distinctions between roles like those of data depositors, subject experts, and data users (\_\_\_). Different stakeholders may more naturally fulfill some of the functions to provide due diligence, and repositories can create workflows accordingly. These roles and functions will shift over time, so repositories must pivot any assigned responsibilities for these groups as needed. This allows for repositories that have not yet designated responsible parties to establish these workflows at any point.

Specific stakeholder needs will define the information and assurances required to carry out the governance for a given repository. This may involve curation at intake, or mechanisms like flags to follow up with data entries missing certain elements after deposition. These systems should provide checks on items like identification and provenance, inclusion of keywords that Indigenous populations might use when searching a database (\_\_\_), and assurance that proper permissions or rights statements for data sharing, use, and informed individual or collective consent have been obtained (\_\_\_). Inclusion of other metadata fields may provide additional pieces of required information(\_\_\_).

Since many open repositories have fewer choices for protecting data on the front-end and have little control over how people use the data they provide, repositories can work with Indigenous populations and data depositors to adapt certain data features or add specific metadata information prior to ingestion to minimize the need for such restrictions (\_\_\_). Repositories must collaborate with stakeholders to determine how to best capture the necessary agreements, obfuscate locations of Indigenous spaces containing assets like sacred sites or seasonal movements of wildlife (\_\_\_), and provide access to data at appropriate levels. Those repositories that support embargo features and access permissions by role may choose to protect data according to agreed upon governance (\_\_\_).

Some organizations that intake or manage Indigenous data or objects form ethics boards or administrative boards to ensure their repositories follow established policies regarding tribal permissions and human subject privacy (\_\_\_). The specific instance defines who has the final say in the formation of this governance and policy, and should include Indigenous populations with an eye toward the provenance of the data.

* + 1. Ensure metadata includes responsibilities of data users (A1, R1, E3) ([CARE Paper Draft Text](https://docs.google.com/document/d/1DyWPuaHIzJaA4fsSIMl6N2aC6T9ftyh2qguFB_wJ_i0/edit#heading=h.m8xna1kemgga) ERIN)
		2. Describe appropriate use or limitations in metadata (R1, E1) ([CARE Paper Draft Text](https://docs.google.com/document/d/1DyWPuaHIzJaA4fsSIMl6N2aC6T9ftyh2qguFB_wJ_i0/edit#heading=h.utqj9alyb9ln) ERIN)
		3. Implement text or badges that clarify types of use (A1, A3, E3) ([CARE Paper Draft Text](https://docs.google.com/document/d/1DyWPuaHIzJaA4fsSIMl6N2aC6T9ftyh2qguFB_wJ_i0/edit#heading=h.qb9yi3c0b84g) ERIN)
	1. Suggestions for implementation – maybe some common threads among the sidebar examples?
	2. Potential tensions: e.g.,
		1. a policy of “Open by Default” might be incompatible with the need for authority, ownership and control
		2. s the limitations of DC for appropriately documenting Indigenous traditional knowledge (TK). (https://doi.org/10.1108/JD-08-2018-0124)

## Conclusion

## Acknowledgements

## References

[Baker, Karen S., and Ruth E. Duerr. 2017. “Research and the Changing Nature of Data Repositories.” *Curating Research Data*, 33.](https://www.zotero.org/google-docs/?A2MXSD)

[Ball, Alexander. 2012. *Review of Data Management Lifecycle Models*. Bath, UK: University of Bath.](https://www.zotero.org/google-docs/?A2MXSD)

[“CARE Principles.” 2023. Global Indigenous Data Alliance. January 23, 2023. https://www.gida-global.org/care.](https://www.zotero.org/google-docs/?A2MXSD)

[Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. “The CARE Principles for Indigenous Data Governance” 19 (1): 43. https://doi.org/10.5334/dsj-2020-043.](https://www.zotero.org/google-docs/?A2MXSD)

[CCSDS 650.0-B-1. 2012. “REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS).” CCSDS. https://public.ccsds.org/Publications/Archive/650x0m2.pdf/default.aspx.](https://www.zotero.org/google-docs/?A2MXSD)

[“Clyde River Weather.” n.d. Accessed July 10, 2023. https://clyderiverweather.org/.](https://www.zotero.org/google-docs/?A2MXSD)

[Eitzel, Melissa, Jessica Cappadonna, Chris Santos-Lang, Ruth Duerr, Sarah Elizabeth West, Arika Virapongse, Christopher Kyba, et al. 2017. “Citizen Science Terminology Matters: Exploring Key Terms.” *Citizen Science: Theory and Practice*, June, 1–20.](https://www.zotero.org/google-docs/?A2MXSD)

[“ESIP Interviews Archives.” n.d. *ESIP* (blog). Accessed May 17, 2023. https://www.esipfed.org/category/blog-archive/esip-interviews.](https://www.zotero.org/google-docs/?A2MXSD)

[Global Digital Data Alliance. 2022. “Indigenous Data Sovereignty and Governance.” Slide Deck. https://www.gida-global.org/s/Final\_Attribution\_NonCommercial\_NoDerivatives\_4\_International.pdf.](https://www.zotero.org/google-docs/?A2MXSD)

[“Global Indigenous Data Alliance.” 2023. Global Indigenous Data Alliance. January 23, 2023. https://www.gida-global.org.](https://www.zotero.org/google-docs/?A2MXSD)

[“Gwich’in Place Names Atlas | ELOKA.” n.d. Accessed July 10, 2023. https://eloka-arctic.org/product/atlas/gwichin-place-names-atlas.](https://www.zotero.org/google-docs/?A2MXSD)

[“Home - Earth Science Information Partners (ESIP).” n.d. ESIP. Accessed May 17, 2023. https://www.esipfed.org/.](https://www.zotero.org/google-docs/?A2MXSD)

[Parsons, M. A., and R. Duerr. 2005. “Designating User Communities for Scientific Data: Challenges and Solutions.” *Data Science Journal* 4: 31–38. https://doi.org/10.2481/dsj.4.31.](https://www.zotero.org/google-docs/?A2MXSD)

[Research Data Alliance International Indigenous Data Sovereignty Interest Group. 2019. “CARE Principles for Indigenous Data Governance.” The Global Indigenous Data Alliance. GIDA-global.org.](https://www.zotero.org/google-docs/?A2MXSD)

## Appendix

* 1. CARE Repository Checklist