

How Data Governance Principles Influence Participation in Biodiversity Science

Authors:

Beckett Sterner
School of Life Sciences
Arizona State University
bsterner1@asu.edu

Steve Elliott
Center for Gender Equity in Science and Technology
Arizona State University

Abstract: Biodiversity science is in a pivotal period when diverse groups of actors—including researchers, businesses, national governments, and Indigenous Peoples—are negotiating wide-ranging norms for governing and managing biodiversity data in digital repositories. These repositories, often called biodiversity data portals, are a type of organization for which governance can address or perpetuate the colonial history of biodiversity science and current inequities. Researchers and Indigenous Peoples are developing and implementing new strategies to examine and change assumptions about which agents should count as salient participants in scientific projects, especially in projects that build and manage large digital data portals. Two notable efforts are the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective benefit, Authority, Responsibility, Ethics) Principles for scientific data management and governance. To characterize how these principles influence the governance of biodiversity data portals, we develop an account of fit-for-use data that makes explicit its social as well as technical conditions in relation to agents and purposes. The FAIR Principles, already widely adopted by biodiversity researchers, prioritize machine agents and efficient computation, while the CARE Principles prioritize Indigenous Peoples and their data sovereignty. Both illustrate the potency of an emerging general strategy by which groups of actors craft and implement governance principles for data fitness-of-use to change assumptions about who are salient participants in data science.

Keywords: FAIR Principles, CARE Principles, Indigenous data sovereignty, citizen science, knowledge infrastructure, colonial science, data governance

1. Introduction

The 2022 Montreal conference of the Convention on Biological Diversity (CBD) marks an important moment for negotiations over the rights of scientists, companies, nations, and Indigenous Peoples to access, contribute, and benefit from digital genetic information about biodiversity. Although biodiversity science often positions itself rhetorically as pursuing an idealistic mission of saving life and humanity, the field has flourished in part due to centuries of European and American colonialism (Agrawal, 2002; Schiebinger, 2009; Vogel, 2019). Past and present colonial activities, for example, include the extraction of specimens from Indigenous lands, the renaming of places and organisms known to Indigenous Peoples,ⁱ and the omission of Indigenous interests and contributions to knowledge. These activities contribute to broader power inequalities between the Global North and South, as non-Indigenous actors regularly overlook and override the interests and rights to self-determination of Indigenous Peoples (Rimmer, 2015; Gilbert and Lennox, 2019).

In addition to debating formal treaty or legal obligations about data and benefit sharing, multiple groups of actors are also seeking to influence the governance of biodiversity knowledge through norms called data principles that rely on endorsement and implementation at the community level. These norms are likely to become widespread and entrenched in all aspects of the biodiversity data life cycle, including planning, collection, analysis, publication, and re-use. Multiple sets of data principles have now been published and endorsed by international research societies and governments (Wilkinson et al., 2016; Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019; Lin et al., 2020), and while each set of principles has a distinct origin and purpose, they share the overarching goal of establishing new rights and priorities for researchers, businesses, national governments, and Indigenous Peoples with respect to data stored and managed in digital repositories.

We address the question of how these data principles influence who matters and who benefits in the making, sharing, and use of digital biodiversity information (Díaz-Reviriego et al., 2019). In related disciplines, one strategy to characterize the social effects of scientific norms has been to examine and critique how scientists characterize legitimate participants in their projects. Typically, participants are characterized as those who contribute research work or materials, with the most important participants being those who contribute to designing, evaluating, or funding the research. Participatory and citizen science researchers have argued this view of who matters gives insufficient recognition and power to people who make other kinds of contributions or who are affected by a project's origins, activities, and consequences (e.g., Macq et al., 2021). By assuming a limited range of participants, scientific projects employ criteria for success and evaluation that favor specific groups of people and exclude others, thus potentially continuing inequitable science. As a result, some projects now embrace expanded understandings of who matters as a participant that give power to non-professional scientists to guide the aims of research and the terms of data collection and use in collaboration with researchers (Pareja et al., 2018; Turreira-García et al., 2018; Thompson et al., 2020).

We focus on two emerging sets of data principles, the FAIR and CARE Principles, which are already influential and in close interaction with each other.ⁱⁱ The Force11 working group of scientific researchers and publishers published the FAIR Principles in 2016. The original publication (Wilkinson et al., 2016) has been cited over 6,400 times as of April 2022, and the European Union has invested substantially in developing the FAIR principles into standardized criteria and indicators that can be adopted as official policy (European Commission Expert Group on FAIR Data, 2018).

The CARE principles were initially drafted by the Indigenous Data Interest Group of the Research Data Alliance in 2018 at its plenary meeting in Botswana (Carroll et al, 2020a). The group was composed of thirteen international academic scholars and was co-led by Stephanie Russo Carroll and Maui Hudson. The drafting group included legal experts (e.g., Rodrigo Sara) with expertise in the CBD, and the CARE Principles are grounded in the United Nation’s Declaration of Rights for Indigenous Peoples (UNDRIP). CARE is gaining rapid international interest and adoption, for example by the recent United Nations Educational, Scientific and Cultural Organization Recommendation on Open Science (UNESCO, 2021).

As their principles gain community and governmental endorsements, advocates for the FAIR and CARE Principles are increasingly positioned to influence how individual projects characterize participants, for example by institutionalizing general norms for whose contributions should be explicitly recognized and who has authority over access rights and appropriate uses of data. Key leaders of the FAIR Principles, for example, explicitly advocate for machine agents—partially or wholly automated computer programs—as privileged users of scientific data for academic or commercial purposes (Mons, 2019). They state that their central aim is to enable machine agents to make authoritative judgments on the utility of the data by knowing what the data mean. Alternatively, leaders of the CARE Principles intend their principles to address the status and importance of Indigenous Peoples for making decisions about the use of data related to them and their lands, such as maps of historical hunting grounds and Indigenous Knowledge about the medicinal properties of plants (Carroll et al., 2020a).

To address our question we characterize how the growing use of the FAIR and CARE Principles by researchers, academic institutions, national governments, and Indigenous Peoples mark an important development in the governance of biodiversity data portals. In Section 2, we introduce an analytical framework of knowledge infrastructures and data governance, which we use to develop the concept of fitness-for-use to capture how governance frameworks influence the standing and rights of different stakeholders as participants. In Section 3, we show that the advocates for the FAIR and CARE Principles call for distinct but sometimes compatible classes of participants—machines and Indigenous Peoples, respectively—by prioritizing concerns about features of the data as objects versus of the peoples and purposes involved.

We then focus on the current and prospective use of FAIR and CARE in biodiversity data portals. We show that managers of internationally influential portals are adopting FAIR and its conception of machines as participants. While the CARE Principles are not yet widely adopted in biodiversity data science, partly because they are new, we indicate that their adoption could significantly alter biodiversity scientists treatment of Indigenous Peoples as salient participants (Hill et al, 2012). We discuss how the CARE Principles apply to three examples of longstanding practices of publishing data related to Indigenous cultures, lands, and economic activity. We conclude the paper by discussing why these results are significant and how they might be generalized and further explored.

2. Analytical Perspectives

We use three analytical concepts: knowledge infrastructures, data governance, and fitness-for-use. STS scholars often use knowledge infrastructure to characterize the development and application of norms about the collective production of scientific knowledge. We understand knowledge infrastructures here as ‘robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds’ (Edwards, 2010, 17). Knowledge infrastructures are increasingly recognized and studied as a class of

organizations that operate outside established government or academic institutions. Many scientific data repositories, for example, are operated by teams of scientists employed by multiple universities. While they often receive government funding, they are not directly subject in their daily operations to any particular government agency or authority. The sense of knowledge relevant to knowledge infrastructures should not be assumed to be predetermined or fixed but instead should be treated as constructed and revised over time by actors exerting influence over the development of the infrastructure.

As part of generating, sharing, and maintaining knowledge, knowledge infrastructures contend with establishing the knowledge they produce as authoritative and legitimate (Strasser et al., 2019). Christine Hine, for example, analyzed knowledge infrastructures involving citizen science participation. Hine noted two general kinds of knowledge infrastructure: those characterized by top-down relationships where the ‘ultimately authority to determine the criteria for what is to count as knowledge resides with developers working within the domain of professional science,’ (Hine, 2020, 93) and those with bottom-up relationships that ‘develop emergent standards for authenticity and accountability that differ radically from the conventional scientific model’ (Hine, 2020, 93). However, analyses like Hine’s do not address how creators of knowledge infrastructures gain authority and legitimacy for their outputs by way of external relationships with stakeholders, including funders, users, universities, and other infrastructures (Mitchell et al, 1997). There is further opportunity to study knowledge infrastructures as organizations in which a broader range of parties engage in negotiating and influencing what counts as knowledge relative to their aims.

For our purposes, biodiversity data portals are examples of knowledge infrastructures that support the construction, maintenance, and use of pooled data resources about biological species. The primary function of data portals as infrastructure is to provide online access to a pooled collection of data records. Portals are more than web interfaces for databases; they are also social organizations the activities of which constitute a spatially and socially distributed network of relationships among people and places. Some of these portals have explicit citizen science components (e.g., iNaturalist) while others do not (e.g., the Global Biodiversity Information Facility (GBIF)). Regardless, these portals establish their authoritative status as producers of knowledge not *sui generis* but instead in relation to serving the aims and values of their users and stakeholders more broadly. Put differently, the professional scientists who run the portals are rarely the sole actors who characterize the value and authority of the portals’ data resources or set portal norms (Frischmann et al, 2014, Leonelli, 2016).

Data governance norms and policies therefore are an important element of biodiversity data portals’ status as knowledge infrastructures. We define data governance as the assemblage of institutional mechanisms, norms, and policies that regulate the rights and responsibilities of people or organizations to produce and use data. For any biodiversity data portal, data governance norms can be implicit or explicit. Leonelli and Tempini (2020) have shown that knowledge infrastructures for data science are frequently precarious and face ongoing challenges delivering on their promises to stakeholders, highlighting the importance of data governance policies that support data sharing and reuse. Additional analyses have shown how data scientists refashion and renegotiate narratives of progress as they test new governance standards, for instance for metadata (e.g., Millerand et al, 2013). When followed, principles and standards for data governance like FAIR and CARE can influence the structures, functions, and practices of knowledge infrastructures to improve their utility and benefits for broad classes of actors.

We use the concept of fitness-for-use to help characterize the influence of users and stakeholders over the norms, aims, and functions of knowledge infrastructures. We borrow the term from data science, where it is widely used in evaluating scientific models and datasets (Franz and Sterner, 2018, Bokulich and Parker, 2021). Rather than treat fitness-for-use as a purely objective, mind-independent property of data, though, we understand fitness-for-use as a social as well as technical status that people attribute to a body of data in relation to some further purpose (Leonelli 2016). Biodiversity data portals are then important loci where stakeholders negotiate shared criteria for whether bodies of data are licensed for use as evidence in specific research problems, policy decisions, or commercial applications.

While scientists generally formulate fitness-for-use as a technical matter—e.g., asking whether data were sampled in a way that provides meaningful evidence for a modeling question—we use the concept to analyze how claims of usability also reflect social relationships and status. We treat being fit-for-use as a status attributed to a body of data that is analogous to a person being fit-for-work or fit-for-office. In each of these three cases, the fit-for-X label denotes that someone in a position of authority has judged the relevant person or object to have the attributes required to undertake a particular role in some collective activity. Indigenous data sovereignty, for example, asserts the authority of Indigenous Peoples to determine what counts as fit-for-use scientific knowledge in relation to the aims, use, and collection of data from their persons and lands. Many scientists and businesses are also seeking to ensure that data infrastructures provide sufficient background information, so that the scientists can designate computational proxy agents to make authoritative judgments about fitness-for-use on their behalf.

Having recognized that fitness-for-use is a socially attributed status, we can analyze data portals to determine who has the authority to make fitness-for-use judgments about data held in the repository. Additionally, we can analyze how actors use data principles (or other means) to influence the status of different groups as authorities on fitness-for-use, either in relation to a kind of data or a specific data portal. In light of common types of rights recognized for data—e.g., rights to access, change, or use (Frischmann et al., 2014)—we distinguish three sets of actors who can have formal authority on the fitness-for-use of data:

1. Actors whose input is required or legitimate in determining whether a data resource is fit-for-use by other actors, e.g. because the data are about them as subjects or because they have an organizational oversight role in the data portal
2. Actors who should be able to access a data resource and make judgements about fitness-of-use for their own purposes, e.g. because they are research experts or claim sovereign rights to the data
3. Actors whose contributions to a data resource materially affect judgements about its fitness-for-use by others, e.g. because the contribution entails legal obligations for other parties under a treaty or because of community norms about data ownership

These sets and their defining characteristics may also intersect and overlap. An example of the latter would be cases in which scientists must document permissions for collecting data in particular areas in order to meet ethical research standards for their research project (types 1 and 3 above).

We next analyze how leaders of the FAIR and CARE Principles use them to advocate for formal recognition and prioritization of different types of actors as authorities about the fitness-for-use of data.

3. Empirical Analysis

3.1. The FAIR Principles

Researchers developed the FAIR Principles for scientific data with the explicit aim of automating the work of data sharing and reuse with machine agents. We show that making data fit-for-use in this respect entails ensuring that machines can authoritatively process the meanings of data and metadata on behalf of the people or broader systems employing them. While this machine-based fitness-for-use frequently aligns with the aims and abilities of people, satisfying the FAIR Principles is not sufficient to guarantee that any datum is error-free or appropriate for use as evidence in research. Instead, the more proximate goal of FAIR is to improve the ability of computational agents to access scientific data and descriptive metadata to help determine the relevance and value of datasets for research use.

Box 1 lists the FAIR Principles (Wilkinson et al, 2016). The principle of ‘findability’ centers on having a globally unique name (identifier) for each dataset, using this identifier to locate a dataset in an online repository, and determining what kinds of information can be found in the dataset (i.e., by searching for metadata content). Identifiers are essential for addressing the ‘accessibility’ of data and metadata (see Box 1, A1). Further considerations include whether an access protocol is free, open, and universally implementable and whether metadata persist after data are deleted or lost. The principles address designating some license for the data, but they do not require open licensing in specific (Higman et al., 2019). The ‘interoperability’ principle focuses on using shared, standardized vocabularies for describing information in datasets, whether these standards are themselves FAIR and whether datasets are linked to each other using these shared vocabularies. Finally, the ‘reusability’ principle identifies several respects in which ‘rich’ description of data resources is needed, including information about how observations were made (i.e., their provenance) and the meanings of any labels or metadata categories used.

While following FAIR principles for any dataset will generally better enable humans to work with it, the overarching motivation for FAIR is to address concerns among industry, science funders, governments, and scientists that ‘the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments’ (Wilkinson et al, 2016, 1). The primary goal is to make data machine-actionable in the sense of scaffolding the capacity of machine agents to automate and improve the extraction of value from scientific data. As Berend Mons, senior author on the original paper, later explained: ‘the one-liner that captures the essence of the FAIR principles is *‘Machines know what it means’* (Mons et al, 2019, 4), where ‘it’ refers to any data and metadata and the ‘machines’ of interest are computational algorithms or agents using forms of artificial intelligence, such as symbolic reasoning or machine learning. This position ‘does not (yet) take people out of the loop. In fact the envisioned Internet of FAIR Data and Services should be an environment where our implementation choices support both machines and humans, in a tight and iterative collaboration (i.e., “Social Machines” are the end users)’ (Mons et al, 2019, 4).

Box 1: FAIR Guiding Principles

To be Findable:

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data it describes
- F4. (Meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

To be Reusable:

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

While machine agents, in principle, can serve the aims of any stakeholder seeking to use scientific data, multiple critics have noted that the FAIR principles do not address the commonsense meaning of the acronym in English, i.e., advancing goals such as fair, equitable, inclusive, or just access to data and its benefits. Datasets can be findable, accessible, interoperable, and reusable while still manifesting serious sampling biases that perpetuate existing societal injustices or inequalities (Leonelli, 2021). While compliance with the FAIR Principles will correlate positively with common measures of data openness, they do not encourage or require putting datasets in the public domain (Higdon et al., 2019).

A further feature of FAIR is that the highly abstract principles are linked to everyday concepts, such as findability, through implementation of technical solutions such as globally unique identifiers. Widespread support for FAIR by researchers has derived partly from how the principles can be locally adapted and interpreted to fit the aims and resources of particular data infrastructures. ‘No-one [sic] really argues against the idea that data, as well as the accompanying workflows and services should be findable, accessible under well-defined conditions, interoperable without data munging, and thus optimally reusable’ (Mons et al, 2019, 2). The rhetorical finesse of relying on thick concepts such as ‘richly described’ metadata helps advocates for the principles garner endorsements from researchers and stakeholders in advance of fully understanding how those concepts will be operationalized in local situations. Further

research and policy development, led especially by the European Union, has focused on how to specify domain-specific evaluative criteria for compliance with the FAIR Principles (e.g., Wilkinson et al, 2019, Le Franc et al, 2020). FAIR, therefore, establishes space for multiple stakeholders to participate in governance over knowledge infrastructures by, at least rhetorically, recognizing the need for community-level deliberation on the standards and metrics appropriate to the domain. However, the principles appeal to an abstract promise that FAIR data will be more usable and beneficial for everyone without specifying how communities should consider and address potential impacts resulting from the domain standards they develop and adopt.

3.2. The CARE Principles

As noted in Section 3.1, the FAIR Principles explicitly decline to address all data governance issues important to researchers and stakeholders, including data justice and anti-colonialism (Taylor, 2017, Chan et al, 2019, Leonelli et al, 2021). Nonetheless, the rapid, international adoption of FAIR by governments and research organizations offers an exemplary model for institutionalizing compliance to emerging data principles. Indeed, advocates of other sets of data principles have positioned themselves as articulating complementary rather than opposing priorities to FAIR. Conceiving of fitness-for-use as a social status certain actors attribute to a dataset is crucial here for characterizing a broader range of values and norms than those recognized by FAIR or the classical value-free ideal for science (Douglas, 2009).

Indigenous data sovereignty, for example, expands the possibilities of participation in knowledge infrastructures. While there are varying conceptions of Indigenous data sovereignty in the literature, an overlapping area of agreement is that Indigenous Peoples have inalienable rights and interests ‘relating to the collection, ownership and application of data about their people, lifeways and territories’ (Kukutai and Taylor, 2016a, 2), such that Indigenous Peoples can ‘control the collection, access, analysis, interpretation, management, dissemination and reuse of Indigenous data’ (Walter and Carroll, 2020, 2; see also Snipp, 2016; Kukutai and Taylor, 2016b; Tsosie, 2019). The United Nation’s Declaration on the Rights of Indigenous Peoples (UNDRIP) provides an international framework that formally acknowledges the sovereignty and cultural and intellectual property rights of Indigenous Peoples, rights that the UN’s Convention on Biological Diversity’s 2011 Nagoya Protocol specifies cover genetic data (United Nations, 2007; 2011; Davis, 2016). UNDRIP complements more specific national policies and laws that address indigenous data rights and sovereignty, for instance, in Canada and New Zealand.

Indigenous Peoples and scholars developed concepts of data sovereignty to characterize Indigenous rights and interests given historical trends (Carroll et al, 2020a). These trends include power imbalances favoring non-Indigenous over Indigenous Peoples; the collection and use of data about Indigenous Peoples without their consent; the use of that data to portray Indigenous Peoples solely as deficient in some characteristic or another; the use of that data and those deficiency narratives to further the aims and values of non-Indigenous People and governments; and the lack of control of and access to Indigenous data by Indigenous Peoples.

The CARE Principles build on the right of sovereignty as a foundation to articulate principles specifically addressing data from or about Indigenous Peoples and their lands. Between 2017 and 2019, the group collected principles and statements of interest from groups that focus on Indigenous data in different parts of the world, including Aotearoa (or New Zealand), Canada, Australia, and the United States. They then compared these statements to more widely-used principles of open data governance, such as FAIR.

As minimal norms for open data and metadata, FAIR Principles do not explicitly encourage the perpetuation of oppression and harm to Indigenous Peoples; rather, they do not prioritize redressing these issues. As demonstrated in Section 3, the FAIR Principles focus on features of data and metadata abstracted from any person or group affected by the collection and use of data. These abstractions present scientific data as objectively measurable and similar from any perspective, obscuring the relationality of data as fit-for-use for specific aims. Such abstractions have been theorized more generally as techniques by which socially dominant groups foster ignorance about how seemingly raceless contemporary institutions perpetuate privileged power and access to resources for those dominant groups and thus preserve their dominance (Mills, 2015). These abstractions obscure arguments made by scholars of Indigenous data that pooled data, and the infrastructures used to store and transmit them, are inextricably from their social and historical context. Further, the use of these data differently impact those with unequal power relations (Kukatai and Taylor, 2016b, Walter and Andersen, 2016).

Box 2: The CARE Principles for Indigenous Data Governance

Collective Benefit: Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

- C1. For inclusive development and innovation
- C2. For improved governance and citizen engagement
- C3. For equitable outcomes

Authority to Control: Indigenous Peoples' rights and interests in Indigenous data must be recognized and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data

- A1. Recognizing rights and interests
- A2. Data for governance
- A3. Governance of data

Responsibility: Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.

- R1. For positive relationships
- R2. For expanding capability and capacity
- R3. For Indigenous languages and worldviews

Ethics: Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

- E1. For minimizing harm and maximizing benefit
- E2. For justice
- E3. For future use

Proponents of CARE identify the principles' value as bringing "a people-and-purpose orientation to data governance, which complements the data-centric nature of the FAIR principles" (Carroll et al, 2020b). The CARE Principles therefore serve to identify and fill gaps

unaddressed by FAIR, and do not universally oppose making Indigenous data findable or accessible or the use of technical components like unique identifiers. In joint meetings, for example, leaders of the CARE Principles have noted the value to Indigenous Peoples' sovereignty of making data about them and their lands more findable and accessible when held by colonial governments or organizations (Carroll et al, 2021).

Nonetheless, CARE asserts the rights of Indigenous Peoples to determine which, when, and how data about them and their lands will be collected, categorized, distributed, and used. Theorizing and institutionalizing Indigenous data sovereignty, therefore, begins to move Indigenous Peoples from obscured to formally recognized participants on multiple levels in data infrastructures. For instance, the idea of access in FAIR is unmarked with respect to any particular stakeholder—i.e. rhetorically it is presented as accessible for anyone—while the 'authority to control' principle in CARE amplifies accessibility with specific expectations in relation to governance by and for Indigenous Peoples:

Indigenous Peoples must have access to data that support Indigenous governance and self-determination. Indigenous Peoples must be the ones to determine data governance protocols, while being actively involved in stewardship decisions for Indigenous data that are held by other entities (Carroll et al, 2020a, 6).

CARE, unlike FAIR, thus orients issues of access toward specific peoples, purposes, governance, and self-determination. However, the scope of CARE is intentionally restricted to Indigenous Peoples and does not aim to address non-Indigenous groups, such as African-Americans or rural communities, that have also experienced oppression or marginalization. The following section examines several implications of adopting CARE for the explicit representation and inclusion of Indigenous Peoples as participants in biodiversity knowledge infrastructures.

3.3. Use of FAIR to harness efficiencies by prioritizing computational agents

Although scientists frequently refer to a global commons of biodiversity knowledge, the idea of a single global commons oversimplifies the actual governance of information about species in several potentially misleading ways. First, there is no single repository where all information about biodiversity is pooled and subject to shared institutional arrangements. Second, a substantial proportion of data describing the locations and traits of species is held by governments or privately by companies and is not licensed for use in the public domain. Third, there is no overarching institutional arrangement regulating the collection, maintenance, and use of biodiversity data except in limited respects related to bioprospecting and wildlife trade (i.e. the Convention on Biological Diversity and Convention on International Trade in Endangered Species of Wild Fauna and Flora). Given these three facts, any group confronts challenges when it seeks to influence global governance of biodiversity data.

In this context, biodiversity scientists and data infrastructures rapidly endorsed the FAIR Principles as a framework for establishing international but domain-specific governance mechanisms (Gries et al, 2019; Lannom et al, 2019; Penev et al, 2019; Hardisty et al, 2019). These researchers acknowledge it is insufficient for their purposes for biodiversity data to be open only in the sense of accessible freely online with few or no legal restrictions on how the data may be reused. Such openness does not guarantee that users of the data will be able to identify redundant records, rely on data providers to follow standardized formats, or harmonize information about the geolocations or taxonomic identities of observed organisms across sources. The work of data collection and sharing is highly decentralized and spatially distributed. The largest international data portal, GBIF, aggregates about 2.2 billion data records as of April 2022

by combining 68,000 datasets sourced from 1,800 data publishers. These publishers are predominantly scientific organizations, academic institutions, government agencies, and citizen science projects. GBIF's coverage is nonetheless biased toward certain regions and taxonomic groups, and it does not preserve datasets deleted or lost by data publishers (Hortal et al, 2015). After decades of work, the transaction costs of finding, harmonizing, and reusing biodiversity data remain prohibitive for researchers. The FAIR Principles function to reduce these costs by enabling researchers to use computational methods in place of human labor.

Researchers and data portal leaders are using FAIR to set norms and rules that prioritize the ability of computational agents to access and evaluate biodiversity datasets. The principle of Findability, for example, addresses basic challenges to determining how many unique and relevant records exist, especially when modified or redundant copies exist in different sources. Meeting this prerequisite has been a major challenge for biodiversity data portals, which are still in the process of developing and adopting globally unique identifiers for specimens (Guralnick et al, 2015), and many online data sources continue to coin their own local identifiers. In contrast, Accessibility through free and open Application Programming Interfaces (APIs) has been easier to achieve, and international biodiversity data portals today generally support automated online queries using APIs.

To automate determinations of fitness-for-use for particular research problems, biodiversity scientists have widely adopted several metadata standards for describing the contents and collection methods of species observations (Hardisty et al, 2019). They increasingly invoke FAIR to justify the development and adoption of these more specific standards, especially for shared, standardized vocabularies. The Darwin Core format, for example, imposes minimal required metadata information such as a taxonomic name (Wieczorek et al, 2012). Darwin Core does not designate a standardized biological taxonomy, so the meanings and validity of taxonomic names in specimen records frequently change across data sources and over time, even if those data sources all follow Darwin Core Standards (Vaidya et al, 2018; Franz et al, 2020). As a result, there are ongoing debates about whether interoperability demands convergence on a single global taxonomy for species and if this truly optimizes fitness-for-use among all stakeholders (Garnett and Christidis, 2017, Franz and Sterner, 2018, Sterner et al, 2020). Another set of relevant data standards are trait ontologies, which provide regulated vocabularies for describing characteristics of single organisms or whole species, e.g. average body mass or dispersal range. These trait ontologies currently provide patchwork coverage for taxonomic groups and their phenotypic characteristics. When the Open Trait initiative launched, it cited FAIR to motivate its efforts to coordinate global interoperability for trait data (Gallagher et al, 2020).

These examples show that researchers use FAIR to prioritize the needs of computational agents to access and make determinations about the fitness-for-use of data. To the extent that the computational agents automatically update data records in a repository with labels describing fitness-for-use, they act as authorities on behalf of others (i.e. in the first way we identified). To the extent they make independent judgments that are not further shared or made available to others, they exercising the second type of authority we identified. The ultimate equitability and desirability of these practices, however, are disputed.

3.4. Use of CARE to prioritize Indigenous Peoples in biodiversity data governance

As leaders of CARE move toward implementing technical standards and criteria for compliance, they follow a similar strategy as FAIR by seeking to translate initial widespread

stakeholder support (e.g., from organizational leaders in academic libraries and museums) into bureaucratic mechanisms for driving institutional change. While most attention has focused on Indigenous genetic, cultural, and demographic data, many actors—e.g., scientists, conservationists, and businesses—continue to value the collection and study of organisms on Indigenous lands. Similarly, scientific knowledge of biodiversity has influenced the establishment of protected areas that led to forcible removal of Indigenous Peoples from their lands.

This section examines examples involving Indigenous sovereignty over biodiversity data of the kind commonly included in data portals. The examples illustrate how Indigenous Peoples assert their data sovereignty by using the three forms of authority over fitness-for-use discussed in Section 2. The CARE principles help managers and stakeholders of biodiversity data portals legitimize and justify use of these three strategies for influencing infrastructure policies and behaviors. From these examples, we anticipate how CARE will apply to biodiversity data infrastructures. Collectively the examples show how CARE can establish norms or rules about the input of Indigenous Peoples on the fitness-for-use of data, and about their recognition as participants in biodiversity data portals.

3.4.1. Indigenous cultures

The scientific value of taxonomic names rests on their ability to link information held in diverse repositories accumulated over centuries. Every biological specimen or occurrence record derives its value for broader scientific use through categorization in a taxonomic group, typically at the species rank. A substantial proportion of biological specimen collections have been collected on Indigenous lands and informed by Indigenous knowledge (Vogel, 2019), although it is hard to precisely quantify the proportion because of the required provenance information crediting Indigenous sources is often lacking. Many biodiversity data portals as organizations have not formally recognized Indigenous peoples, knowledges, and languages as authorities over and contributors to their data (Foster, 2017).

The absence of formal recognition illustrates the second type of authority we identified in Section 2, by which Indigenous Peoples advance their data sovereignty by establishing their authority to access and make judgements about the fitness-for-use of data about species on their lands. While the FAIR Principles do not provide detailed rules for how biodiversity data should be linked to cultural categories and stakeholder knowledge, the CARE Principles directly address obligations for generating ‘data grounded in the languages, worldviews, and lived experiences (including values and principles) of Indigenous Peoples’ (R3) and ensuring ‘any value created from Indigenous data should benefit Indigenous communities in an equitable manner and contribute to Indigenous aspirations for wellbeing’ (C3, Box 2). Applying the CARE Principles establishes formal recognition of Indigenous peoples as a category of actor who should be able to access and use biodiversity data portals according to Indigenous names and categories.

One context for these obligations involves strengthening rather than erasing Indigenous knowledge and cultures, especially names for organisms and places. For example, the Māori knowledge system, mātauranga Māori, has been largely unacknowledged or discussed by ecological researchers in Aotearoa New Zealand (McAllister et al, 2020). Nonetheless, Veale et al. identify

five central ways in which te reo and ta re [the Māori and Moriori Indigenous languages] have been incorporated, including the use of (1) variations of the words “Māori ” and “Moriori” to designate Aotearoa New Zealand origins, (2) Māori / Moriori vernacular

names for species, (3) Māori / Moriori place names associated with species, (4) novel descriptive names created from Māori and Moriori words, (5) novel names suggested by Māori in collaboration with taxonomists (Veale et al, 2019, 2).

The CARE Principles, therefore, indicate the importance of documenting and providing access to biodiversity data labeled with Indigenous names in metadata, publications, and data collection (Wehi et al., 2019). A limitation of FAIR is the absence of substance supporting inclusive and equitable access to scientific knowledge. For biodiversity data science, the names scientists attribute to organisms are essential metadata for realizing all four pillars of FAIR, but this scientific terminology and language perpetuates barriers for particular communities. As a result, standards of FAIR biodiversity data developed by biodiversity researchers are unlikely to recognize Indigenous Peoples as an actor category that should be prioritized for accessing and using data from Indigenous territories.

3.4.2. Indigenous lands

Open access to biodiversity data collected on or near Indigenous Peoples' lands can aid or infringe upon their rights to self-determination, depending on how access to the data affects existing power relationships. Inequitable outcomes may result, for example, if systemic inequalities and impoverished metadata forestall relevant Indigenous groups from accessing and using the information, or if the portals enable businesses, states, or conservation organizations to implement new resource exploitations or land exclusions. Simply increasing the FAIR-ness of biodiversity data does not establish norms or rules recognizing Indigenous Peoples as authorities on the appropriate use of data regarding their lands and people. Open science, moreover, is not necessarily consistent with Indigenous sovereignty. For example, a recent 'manifesto' published by biodiversity scientists predominantly from Europe and the U.S. argues that

'data should be mobilised and processed from the point of production to ensure they are available in a timely manner for research and policy needs. There should not be undue delays or hindrances for reasons other than simply the time it takes to perform the procedures. Appropriate attribution should be given and the fewest possible limitations placed on use' (Hardisty, et al 2019, 28).

Examples of data about Indigenous lands illustrate the first type of authority we identified, by which Indigenous Peoples advance their data sovereignty by establishing norms and rules about whose input is required or legitimate in making determinations on fitness-for-use for other actors. While biodiversity researchers exert their authority to restrict who can use high-resolution spatial data about protected species, the potential harms of excluding Indigenous Peoples from similar authority over Indigenous biodiversity data are typically overlooked. The CARE Principles set more robust expectations than many legal and ethical standards such as U.S. copyright law or informed consent protocols (United Nations General Council, 2007; Davis, 2016).

Leaders of CARE have co-authored and highlighted the Traditional Knowledge and Biocultural Labels as a means to articulate conditions under which Indigenous data may be viewed and used (Liggins et al., 2021). These conditions may be specific times of year or requirements for researchers to contact and establish relationships with relevant Indigenous groups. The labels illustrate how compliance with CARE is likely to regularize and document Indigenous participation in the governance of knowledge infrastructures—see especially CARE Principles C1 and E3 in Box 2.

Governing access to and use of Indigenous knowledge is critical to the future of biodiversity because a large proportion of species that exist today live on or near Indigenous-stewarded lands. Decades of empirical and theoretical studies have explored how mapping can both advance and harm Indigenous claims to land, citizenship, and sovereignty.

Activities such as participatory mapping can establish new claims to rights for Indigenous Peoples by providing factually documented information about historical land use that contests non-Indigenous-produced maps on their terms (Peluso, 1995; Harris and Hazen, 2005; Vos, 2018). When constructive relationships exist between Indigenous Peoples and local governments, mapping traditional hunting grounds, for instance, can enable collaborative management and monitoring of natural resources threatened by development or climate change (e.g., Johnson et al., 2015; Raymond-Yakoubian et al., 2020).

Conversely, producing and circulating maps can bring previously isolated peoples and natural areas into greater contact with economic and political forces that do not prioritize collective benefits with Indigenous groups or biodiversity. The formalization of land ownership itself can lead to longer-term exploitation. As Rosanne de Vos notes in a study of Indonesian counter-mapping, a ‘potential weakness of village-level spatial planning [to establish land rights] is that land can still be sold to outsiders by villagers who hold formal land titles, who in practice may convert land into plantations’ (de Vos, 2018, 627). De Vos notes that such exploitation may be driven by groups within Indigenous populations who have relatively more resources to use information from maps to pursue legal and economic interests.

Indigenous communities frequently navigate conflicts with national and regional governments and businesses that seek to control geographic information describing territories and resources. International conservation efforts and agreements increasingly shape how national governments represent their territories. In Indonesia, for example, the international REDD+ program (Reducing Emissions from Deforestation and forest Degradation, and enhancing forest carbon stocks and conservation) has centralized forest governance in the country’s federated system. This centralized program includes the One Map Initiative, which ‘aims to consolidate spatial data in order to develop one integrated geographical information system’ (Astuti and McGregor, 2015; Mulyani and Jepson, 2017). In combination with Indonesian recognition of some Indigenous land rights, this has helped advance new land claims for Indigenous communities, albeit not without concerns about green grabs of desirable land by local elites (Astuti and McGregor, 2017).

3.4.3. Commercialization

The framework of Indigenous data sovereignty asserts the critical importance of documenting the provenance of data collected on or about Indigenous lands and Peoples for achieving just and ethical relationships. Bioprospecting provides an important example. Many commercial products have been developed from studies of the properties or behaviors of biological species, products such as biofertilizers, nutritional supplements, industrial chemicals, and medicinal drugs or treatments (Efferth et al, 2016). Bioprospecting is the search for novel molecules, biochemicals, or genetic information in biological species that can be developed into commercial products for the pharmaceutical, agriculture, nanotechnology, and other industries. Since 1993, the international Convention on Biological Diversity (CBD) has governed the legal extraction of genetic resources from a country for the purpose of research and commercialization. To comply with the CBD, researchers are increasingly required to link

digitized collection permits with individual data records in biodiversity databases, a previously uncommon practice among biologists (Zimkus, 2021).

Examples of commercialization illustrate the third type of authority we identified, by which Indigenous Peoples advance their data sovereignty by establishing norms and rules about formally documenting when Indigenous Peoples or persons have contributed to biodiversity data. Although FAIR includes requirement R1.2, ‘(meta)data are associated with detailed provenance,’ current practices among biodiversity data collectors and managers are rarely sufficient to ensure that provenance information related to Indigenous knowledge and sovereignty are incorporated as metadata. Two examples of bioprospecting illustrate how the presence or absence of metadata recognizing how Indigenous Peoples contributed to biodiversity data can influence their authority over the process and results of bioprospecting. By normalizing requirements to document Indigenous permission, the CARE Principles also institutionalize recognition of Indigenous Peoples as participants in the collection, storage, publication, and use of biodiversity data collected from their lands.

Documenting the provenance of knowledge about species helps Indigenous groups establish their rights to data under the CBD. For biodiversity data, provenance most often includes descriptions of the person or persons who contributed an occurrence observation and identified an organism’s taxonomic group. Information about who gave permission for data collection, or to whom the species relate, is not generally required or available, nor is information about anyone who provided background knowledge leading to the observation, e.g., about the importance of the species or likely locations where it could be found.

Researchers rarely discover useful properties of species fortuitously without prior guidance, so they more commonly collect and investigate species based on local guidance, which can include Indigenous knowledge.ⁱⁱⁱ An example concerns commercialization of drugs from the *Hoodia gordonii* plant in South Africa (Wynberg and Chennells, 2009; Foster, 2017). As early as the 1770s, Dutch and British colonists and collectors had learned from local guides in the Kalahari Desert about the appetite suppressant properties of the species, which the Europeans documented in reports and colonial records. These reports later informed contemporary biomolecular researchers in South Africa’s Council for Science and Industrial Research (CSIR), who laid the groundwork for a patent and commercial licensing arrangement for the hoodia molecule isolated from the plant (Foster, 2017, 70-1). The Indigenous San people, through the South African San Council, challenged the validity of the patent on the hoodia molecule under the CBD, and they negotiated a benefits-sharing arrangement with the CSIR and its partners in the early 2000s. As a result, Indigenous San participation in colonial-era botany bridged contemporary intellectual property law with some of the earliest Western documentation of *Hoodia gordonii*.

Another example shows how gaps in provenance can fuel disputes over biopiracy. In 1993 the International Cooperative Biodiversity Group funded a team from Washington University in the U.S., led by ethnobotanist Walter Lewis, to collect and study medicinal plants in Peru (Greene, 2004). As Greene writes,

“It is impossible to verify exactly how and where [the team] collected plants during this conflictive period. Lewis...maintains that the collections were made in conjunction with local Ministry of Agriculture officials and largely without the use of native informants in the hills around a non-Indigenous settlement called Imazita” (Greene, 2004, 216).

This description is contested by Evaristo Nugkuag, the founder and president of Consejo Aguaruna Huambisa, an Indigenous organization that had entered into a collaborative agreement with Washington University and the research team. Nugkuag said

“that the ICBG researchers made a critical mistake in choosing to work with the Ministry of Agriculture and that ‘without having authorization to enter into communities with the community chiefs they went astray in order to collect orchids. They collected other species of medicinal plants in what could be called a discrete fashion’” (Greene, 2004, 215).

A discrete sampling strategy is targeted rather than comprehensive. This dispute contributed to the eventual withdrawal of Consejo from the partnership and a battle over permissions and contracts, ultimately undermining the potential for equitable benefit sharing.

4. Discussion and Conclusion

We have shown how advocates for two sets of international principles for data governance, FAIR and CARE, are using the principles to influence which groups are recognized by data portals as having the authority to make judgments about the fitness-for-use of the data they hold. We conceived of fitness-for-use in this respect as a social and technical status that actors attribute to a body of data, analogous to the exercise of judging someone fit-for-office. We then distinguished three types of authority actors may have over fitness-for-use judgments based on whether the actors exercise their authority (1) over other actors’ potential uses of the data, (2) with respect to their own projects, or (3) in contributing to the data resource itself. We showed that advocates use the FAIR Principles to institutionalize the first two types of authority for computational agents by ensuring these agents have access to sufficient metadata that they can “know” what the data mean. We then argued that the CARE Principles address all three forms of authority on behalf of Indigenous Peoples through examples relating to Indigenous cultures, lands, and commercialization of biological resources. Notably, the two sets of data principles we considered both advocate for particular classes of actors and do not provide a universal or comprehensive basis for data governance on behalf of all stakeholders. Our analytical approach could be used to study how data principles influence how actors attribute fitness-for-use to data resources beyond FAIR and CARE and in fields other than biodiversity.

More broadly, we have illustrated how data principles are an emerging means of governing which groups of actors are formally recognized as participants in scientific knowledge infrastructures. Historically, scientists managing data portals have determined the scope of participation on a project-by-project basis, with individual projects adopting a range of narrower to broader conceptions of who matters and how they can be involved. With the rise of data infrastructures as a general class of organization, though, diverse groups of international actors are using data principles and their emerging compliance standards as a way of influencing local norms and practices. The general strategy we identified for both the FAIR and CARE projects involves issuing a set of principles, convincing others that by implementing those principles data infrastructures will have good practices or outcomes, and helping top-down and bottom up efforts drive implementation of those principles in particular data infrastructures.

References

- Agrawal, Arun. 2002. “Indigenous Knowledge and the Politics of Classification.” *International Social Science Journal* 54: 287–97.
- Astuti, Rini, and Andrew McGregor. 2015. “Responding to the Green Economy.” *Third World*

- Quarterly* 36: 2273–93.
- Astuti, Rini, and Andrew McGregor. 2017. “Indigenous Land Claims or Green Grabs? Inclusions and Exclusions within Forest Carbon Politics in Indonesia.” *The Journal of Peasant Studies* 44: 445–466.
- Bokulich, Alisa, and Wendy Parker. 2021. “Data Models, Representation and Adequacy-for-Purpose.” *European Journal for Philosophy of Science* 11: 31.
- Carroll, Stephanie Russo, Ibrahim Garba, et al. 2020a. “The CARE Principles for Indigenous Data Governance.” *Data Science Journal* 19: 1–12.
- Carroll, Stephanie Russo, Maui Hudson, et al. 2020b. “Working with the CARE Principles.” URL: <https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenous-data-governance/>. Accessed August, 2020.
- Carroll, Stephanie Russo, Edit Herczog, et al. 2021. “Operationalizing the CARE and FAIR Principles for Indigenous Data Futures.” *Scientific Data* 8: 108.
- Chan, Leslie, Angela Okune, et al. 2019. *Contextualizing Openness*. Ottawa: University of Ottawa Press.
- Davis, Megan. 2016. “Data and the United Nations Declaration on the Rights of Indigenous Peoples.” In *Indigenous Data Sovereignty*, edited by Tahu Kukutai and John Taylor, 25–38. Canberra: ANU Press.
- De Vos, Rosanne. 2018. “Counter-Mapping against Oil Palm Plantations.” *Critical Asian Studies* 50: 615–33.
- Díaz-Reviriego, I, E. Turnhout, and S. Beck. 2019. “Participation and Inclusiveness in the Intergovernmental Science–Policy Platform on Biodiversity and Ecosystem Services,” *Nature Sustainability* 2: 457–64.
- Douglas, Heather. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press, 2009.
- Edwards, Paul N. 2010. *A Vast Machine*. Cambridge: MIT Press.
- Efferth, Thomas, Mita Banerjee, et al. 2016. “Biopiracy of Natural Products and Good Bioprospecting Practice.” *Phytomedicine* 23: 166–73.
- European Commission Expert Group on FAIR Data. 2018. *Turning FAIR into Reality*. LU: Publications Office of the European Union.
- Foster, Laura. 2017. *Reinventing Hoodia*. University of Washington Press: Seattle.
- Franz, Nico, and Beckett W Sterner. 2018. “To Increase Trust, Change the Social Design behind Aggregated Biodiversity Data.” *Database* 2018. DOI: 10.1093/database/bax100
- Frischmann, Brett M, Katherine J Strandburg, and Michael J Madison, eds. *Governing Knowledge Commons*. Oxford University Press. Oxford University Press, 2014.
- Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.
- Garnett, Stephen and Les Christidis. 2017. “Taxonomy Anarchy Hampers Conservation,” *Nature* 546: 25–27.
- Gilbert, Jérémie, and Corinne Lennox. 2019. “Towards New Development Paradigms.” *The International Journal of Human Rights* 23: 104–24.
- Greene, Shane. 2004. “Indigenous People Incorporated? Culture as Politics, Culture as Property in Pharmaceutical Bioprospecting.” *Current Anthropology* 45: 211–37.
- Gries, Corinna, Mark Servilla, et al. 2019. “Achieving FAIR Data Principles at the Environmental Data Initiative, the US-LTER Data Repository.” *Biodiversity Information Science and Standards* 3: e37047.
- Guralnick, Robert, Andrew Hill, and Meredith Lane. 2007. “Towards a Collaborative, Global

- Infrastructure for Biodiversity Assessment.” *Ecology Letters* 10: 663 – 672.
- Hardisty, Alex, William Michener, et al. 2019. “The Bari Manifesto.” *Ecological Informatics* 49: 22–31.
- Harris, Leila, and Helen Hazen. 2005. “Power of Maps: (Counter) Mapping for Conservation.” *ACME* 4: 99-130.
- Higman, Rosie, Daniel Bangert, and Sarah Jones. 2019. “Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open.” *Insights the UKSG Journal* 32 (May): 18.
- Hill, Rosemary, Chrissy Grant, et al. 2012. “A Typology of Indigenous Engagement in Australian Environmental Management.” *Ecology and Society* 17 (1).
- Hine, Christine. 2020. “Knowledge Infrastructures for Citizen Science: The Taming of Knowledge.” In *A History of Participation in Museums and Archives*. Routledge.
- Joaquín Hortal et al.. 2015. “Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity,” *Annual Review of Ecology, Evolution, and Systematics* 46: 523–49.
- Jetz, Walter, Jana McPherson, and Robert Guralnick. 2012. “Integrating Biodiversity Distribution Knowledge.” *Trends in Ecology & Evolution* 27: 151 – 159.
- Johnson, Noor, Lilian Alessa, et al. 2015. “The Contributions of Community-Based Monitoring and Traditional Knowledge to Arctic Observing Networks.” *ARCTIC* 68: 28–40.
- Kukutai, Tahu, and John Taylor. 2016a. “Data Sovereignty for Indigenous Peoples.” In *Indigenous Data Sovereignty*, edited by Tahu Kukutai and John Taylor, 1–22. Canberra: ANU Press.
- Kukutai, Tahu, and John Taylor. 2016b. *Indigenous Data Sovereignty*. Canberra: ANU Press.
- Lannom, Larry, Dimitris Koureas, and Alex Hardisty. 2019. “FAIR Data and Services in Biodiversity Science and Geoscience.” *Data Intelligence* 2: 122–30.
- Le Franc, Yann, Jessica Parland-von Essen, Luiz Bonino, Heikki Lehväslaiho, Gerard Coen, and Christine Staiger. 2020. “D2.2 FAIR Semantics.” <https://doi.org/10.5281/zenodo.3707985>.
- Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.
- Leonelli, Sabina, and Niccolò Tempini. 2020. *Data Journeys in the Sciences*. Springer.
- Leonelli, Sabina, Rebecca Lovell, et al. 2021. “From FAIR Data to Fair Data Use.” *Big Data & Society* 8: 1-14.
- Liggins, Libby, Māui Hudson, and Jane Anderson. 2021. “Creating Space for Indigenous Perspectives on Access and Benefit-Sharing.” *Molecular Ecology* 30: 2477–82.
- Macq, Hadrien, Céline Parotte, and Pierre Delvenne. 2021. “Exploring Frictions of Participatory Innovation between Sites and Scales,” *Science as Culture* 30: 161–71.
- McAllister, Tara, Jacqueline Beggs, et al. 2020. “Kua Takoto Te Mānuka: Mātauranga Māori in New Zealand Ecology.” *New Zealand Journal of Ecology* 43 (3).
- Millerand, Florence, David Ribes, et al. 2013. “Making an Issue out of a Standard.” *Science, Technology, & Human Values* 38: 7–43.
- Mills, Charles W. 2015. “Global White Ignorance.” In *Routledge International Handbook of Ignorance Studies*, edited by Matthias Gross and Linsey McGoe, 217–27. New York: Routledge.
- Mitchell, Ronald, Bradley Agle, and Donna Wood. 1997. “Toward a Theory of Stakeholder Identification and Salience.” *The Academy of Management Review* 22: 853–86.

- Mons, Barend, Erik Schultes, Fenghong Liu, et al. 2019. “The FAIR Principles: First Generation Implementation Choices and Challenges.” *Data Intelligence* 2: 1–9.
- Mulyani, Mari, and Paul Jepson. 2017. “Does the ‘One Map Initiative’ Represent a New Path for Forest Mapping in Indonesia?” *Forests* 8: 14.
- Pareja, Claudio, Jordi Honey-Rosés, et al. 2018. “What Participation? Distinguishing Water Monitoring Programs in Mining Regions Based on Community Participation.” *Water* 10: 1325.
- Peluso, Nancy Lee. 1995. “Whose Woods Are These? Counter-Mapping Forest Territories in Kalimantan, Indonesia.” *Antipode* 27: 383–406.
- Penev, Lyubomir, Mariya Dimitrova, et al. 2019. “OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science.” *Publications* 7: 38.
- Raymond-Yakoubian, Julie, Peter Pulsifer, et al. 2020. “Mapping and Indigenous Peoples in the Arctic.” In *Governing Arctic Seas: Regional Lessons from the Bering Strait and Barents Sea: Volume 1*, edited by Oran Young, Paul Arthur Berkman, and Alexander Vylegzhanin, 293–319. Cham: Springer.
- Research Data Alliance International Indigenous Data Sovereignty Interest Group. 2019. “CARE Principles for Indigenous Data Governance.” The Global Indigenous Data Alliance. URL: https://static1.squarespace.com/static/5d3799de845604000199cd24/t/5da9f4479ecab221ce848fb2/1571419335217/CARE+Principles_One+Pagers+FINAL_Oct_17_2019.pdf.
- Rimmer, Matthew. 2015. *Indigenous Intellectual Property: A Handbook of Contemporary Research*. Edward Elgar Publishing.
- Schiebinger, Londa. 2009. *Plants and Empire*. Harvard University Press.
- Snipp, Matthew. 2016. “What Does Data Sovereignty Imply.” In *Indigenous Data Sovereignty*, edited by Tahu Kukutai and John Taylor, 39–55. Canberra: ANU Press.
- Stern, Beckett, Edward Gilbert, and Nico Franz. 2020. “Decentralized but Globally Coordinated Biodiversity Data.” *Frontiers in Big Data* 3: 519133.
- Strasser, Bruno et al. 2019. “‘Citizen Science’? Rethinking Science and Public Participation.” *Science & Technology Studies*. 32: 52-76.
- Taylor, Linnet. 2017. “What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally.” *Big Data & Society* 4: 1-14.
- Thompson, Kim-Ly, Trevor Lantz, and Natalie Ban. 2020. “A Review of Indigenous Knowledge and Participation in Environmental Monitoring.” *Ecology and Society* 25 (2).
- Tsosie, Rebecca. 2019. “Tribal Data Governance and Informational Privacy.” *Montana Law Review* 80: 229–68.
- Turreira-García, Nerea, Jens Lund, et al. 2018. “What’s in a Name? Unpacking ‘Participatory’ Environmental Monitoring.” *Ecology and Society* 23 (2).
- United Nations. 1992. Convention on Biological Diversity. Vol. 31 I.L.M. 822.
- UN Convention on Biological Diversity. 2011. “The Nagoya Protocol on Access and Benefit Sharing.” Secretariat of the Convention on Biological Diversity.
- UNESCO. “Recommendation on Open Science.” Paris, France: UNESCO, 2021. <https://en.unesco.org/about-us/legal-affairs/recommendation-open-science>.
- United Nations General Council. 2007. “United Nations Declaration on the Rights of Indigenous Peoples.” <https://www.refworld.org/docid/471355a82.html>.
- United Nations Permanent Forum on Indigenous Issues. n.d. “Who Are Indigenous Peoples?” https://www.un.org/esa/socdev/unpfii/documents/5session_factsheet1.pdf.

- Veale, Andrew, Peter de Lange, et al. 2019. “Using Te Reo Māori and Ta Re Moriori in Taxonomy.” *New Zealand Journal of Ecology* 43: 1–11.
- Vogel, Gretchen. 2019. “Natural History Museums Face Their Own Past.” *Science* 363: 1371–72.
- Walter, Maggie, and Chris Andersen. 2016. *Indigenous Statistics: A Quantitative Research Methodology*. New York: Routledge.
- Walter, Maggie, and Stephanie Russo Carroll. 2020. “Indigenous Data Sovereignty, Governance and the Link to Indigenous Policy.” In *Indigenous Data Sovereignty and Policy*, edited by Maggie Walter, Tahu Kukutai, et al, 1–20. New York: Routledge.
- Wehi, Priscilla, Lyn Carter, et al. 2019. “Enhancing Awareness and Adoption of Cultural Values through Use of Māori Bird Names in Science Communication and Environmental Reporting.” *New Zealand Journal of Ecology* 43: 1–9.
- Wieczorek, John, David Bloom, et al. 2012. “Darwin Core: An Evolving Community-Developed Biodiversity Data Standard.” *PLoS ONE* 7: e29715.
- Wilkinson, Mark, Michel Dumontier, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3: 1–9.
- Wilkinson, Mark, Michel Dumontier, et al. 2019. “Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework.” *Scientific Data* 6: 174.
- Wynberg, Rachel, and Roger Chennells. 2009. “Green Diamonds of the South” In *Indigenous Peoples, Consent and Benefit Sharing*, edited by Rachel Wynberg, Doris Schroeder, and Roger Chennells, 89–124. Dordrecht: Springer.
- Zimkus, Breda, Linda Ford, and Paul Morris. 2021. “The Need for Permit Management within Biodiversity Collection Management Systems to Digitally Track Legal Compliance Documentation and Increase Transparency About Origins and Uses.” *Collection Forum* 35: 1–20.

ⁱ We follow the understanding of Indigenous peoples suggested by the United Nations Permanent Forum on Indigenous Issues: “Indigenous peoples can be understood as peoples with “Historical continuity with pre- colonial or pre-settler societies; strong links to territories and surrounding natural resources; distinct social, economic or political systems; form non-dominant groups of society; resolved to maintain and reproduce their ancestral environments and systems as distinctive peoples and communities” (United Nations Permanent Forum on Indigenous Issues, n.d.).

ⁱⁱ The more recent TRUST Principles are also relevant but outside our scope here (Lin et al., 2020).

ⁱⁱⁱ We follow (Thompson et al, 2020,1) in distinguishing between local and Indigenous knowledge “based on the histories, socio-political contexts, and self-identification of those creating and holding the knowledge” in reference to the definition of Indigenous peoples given above.