

# The FAIR and CARE Data Principles Influence Who Counts As a Participant in Biodiversity Science by Governing the Fitness-for-Use of Data

Authors:

Beckett Sterner  
School of Life Sciences  
Arizona State University  
[bsterner1@asu.edu](mailto:bsterner1@asu.edu)

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

**Abstract:** Biodiversity scientists often describe their field as aiming to save life and humanity, but the field has yet to reckon with the history and contemporary practices of colonialism and systematic racism inherited from natural history. The online data portals scientists use to store and share biodiversity data are a growing class of organizations whose governance can address or perpetuate and further institutionalize the implicit assumptions and inequitable social impacts from this extensive history. In this context, researchers and Indigenous Peoples are developing and implementing new strategies to examine and change assumptions about which agents should count as salient participants to scientific projects, especially 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 influence the governance of biodiversity data portals, we develop an account of fitness-for-use that makes explicit its social as well as technical conditions in relation to agents and purposes. The FAIR Principles, already widely adopted by biodiversity data projects, 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 salient participants to data science.

**Keywords:** Indigenous data sovereignty, citizen science, knowledge infrastructure, colonial science, data governance

## 1. Introduction

Although biodiversity science often positions itself rhetorically as pursuing an idealistic mission of saving life and humanity, the field—and natural history more broadly—has flourished due to centuries of European and American colonialism (Agrawal, 2002; Schiebinger, 2009; Vogel, 2019). In biodiversity science colonial activities include the extraction of specimens from Indigenous lands, the renaming of places and organisms known to Indigenous Peoples,<sup>i</sup> 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). Many researchers have helped erase Indigenous culture and knowledge, and biodiversity science as a whole has yet to reckon with its history of systemic racism.

In related fields, one strategy to characterize the social effects of scientific practice 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 design, evaluate, or fund the research. Inspired by work on participatory science and citizen science, that understanding has been criticized as insufficient to describe 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 recognize stronger senses of participation that include power for 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).

There is an opportunity to use this strategy to characterize the social effects of biodiversity practices. Biodiversity science is in a pivotal period when many different groups of actors—including researchers, businesses, national governments, and Indigenous Peoples, etc.—are negotiating wide-ranging norms for governing and managing biodiversity data in digital repositories, often called biodiversity data portals. These norms are likely to become widespread and entrenched in all aspects of the data life cycle, including planning, collection, analysis, publication, and re-use. Typically, a portal participant is thought to be someone who contributes data records or metadata, but emerging governance norms have the potential to alter that typical position. A critical question, then, is how emerging governance frameworks for biodiversity data influence how portals conceptualize who matters and who benefits in the making, sharing, and use of this biodiversity information (Díaz-Reviriego et al, 2019).

Two emerging governance frameworks are the FAIR and CARE Principles. The Force11 working group of scientific researchers and publishers published the FAIR Principles in 2016. The original publication alone (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). In 2019, the International Indigenous Data Interest Group of the Research Data Alliance (RDA) published the complementary CARE Principles to advance data sovereignty and rights for Indigenous Peoples. Though more recent, the CARE principles are gaining international support and interest, and they are grounded in the United Nations Declaration of Rights for Indigenous Peoples (UNDRIP).

As they 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, i.e. partially or wholly automated computer programs, as first-class 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 usability of the data, i.e. to know what the data mean. Alternatively, the CARE Principles address the status and importance of Indigenous Peoples in 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.

We characterize how the use of the FAIR and CARE Principles mark a conceptual shift in the governance of biodiversity data portals. We introduce in Section 2 an analytical framework of knowledge infrastructures and data governance, both of which we use to develop a concept of fitness-for-use to capture the how governance frameworks are used to characterize data portal participants. In Section 3, we show that the FAIR and CARE Principles advocate 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. Next, we show that despite these distinct frameworks, the advocates and implementors of FAIR and CARE pursue a similar mechanism of institutional change for particular data infrastructures. This mechanism involves the progressive development from the principles and guidelines to evaluative criteria and maturity indicators that can be formally adopted for compliance purposes by individual organizations.

We then focus in more depth on the current and prospective use of FAIR and CARE in biodiversity data portals. We show that FAIR is increasingly adopted, especially by internationally influential portals, and that biodiversity data scientists and portals invoke motivations—such as data centralization and mass data digitization—using the machines-as-participants conceptual framework of FAIR. While the CARE Principles are not yet widely adopted in biodiversity data science, partly because of their comparatively recent development, we indicate that if they become so, they could significantly alter the background assumptions of biodiversity science about Indigenous Peoples as salient participants (Hill et al, 2012). We discuss how the CARE Principles apply to longstanding practices of publishing data related to Indigenous knowledge, lands, and cultures. 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. The concept of knowledge infrastructure is often used to characterize the development and application of norms about the collective production of knowledge in science. We define knowledge infrastructures 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 day-to-day operations to any particular government agency or authority. The sense of

knowledge relevant to knowledge infrastructures should not be assumed to be pre-given or fixed, and 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 for which 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) or those with bottom-up relationships that ‘develops emergent standards for authenticity and accountability that differ radically from the conventional scientific model’ (Hine, 2020, 93). However, analyses like Hine’s don’t address how knowledge infrastructures gain authority and legitimacy for their outputs through external relationships with a broad array of 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 can 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 are to provide online access to a pooled collection of data records. Portals are more than web interfaces for databases: they are also social organizations whose activities 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 not always the sole actors who characterize the value and authority of the portals’ data resources, nor do they alone set portal norms, which can be set through social and legal processes (Frischmann et al, 2014, Leonelli, 2016).

Data governance norms and policies are part of biodiversity data portals conceived 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 particular biodiversity data portal, data governance norms can be implicit or explicit. Leonelli and Tempini (2020) have shown that knowledge infrastructures for data science can be precarious and subject to breakdowns, a result that foregrounds questions about the effectiveness of institutional arrangements like 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.

To help characterize the influence of users and stakeholders on the norms, aims, and functions of knowledge infrastructures, we use the concept of fitness-for-use. 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). Similar to how experimenters have to

agree that an experiment is done (Galison, 1987), labelling a body of data as fit-for-use in some context represents a socially agreed-upon marker of value and legitimacy. We characterize biodiversity data portals as knowledge infrastructures in which stakeholders negotiate criteria of data as fit-for-use for particular aims, e.g. as evidence in specific research problems, as bases for policy decisions, or for other commercial and social applications generating benefits for users.

Biodiversity data portals are rarely involved in direct contests over defining knowledge generally. Instead, actors typically formulate the stakes more proximately in terms of conceptualizing and measuring progress on producing information that is usable in further research projects. Even elite, international organizations such as the GBIF do not have the sole authority to determine that their aggregated datasets are fit-for-use, nor do their datasets have statuses that generically license their use to any purpose. These portals are beholden to their users, including scientific researchers and state agencies, for judgments on fitness-for-use. In other words, no data repository that serves an audience beyond its immediate staff has a monopoly on the expertise needed to determine whether its datasets are adequate to the aims of its stakeholders, not least because the stakeholders have the final say on what their aims are.

While scientists generally formulate fitness-for-use as a technical matter—e.g. were data sampled in a way that provides meaningful evidence to address a modeling question—we extend the concept to recognize 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 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, as a relatively recent program, 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.

We apply the concept of fit-for-use data to analyze how actors negotiate over the status and priority of potential participants in knowledge infrastructures like biodiversity data portals. In doing so, we generate at least three conceptually distinct means by which stakeholders can use governance frameworks to formalize influence about who receives formal recognition as a participant of the infrastructure:

1. Establishing norms or rules about specific categories of actors whose input is required or legitimate in making determinations about the fitness-for-use of some data resource.
2. Establishing norms or rules about categories of actors who should be able to access and use some data resource
3. Establishing norms or rules about who should be formally identified in a data resource as the subject or contributor in some fashion of the included observations.

These means for codifying the status of actors as participants may also overlap, for example in cases for which scientists must include documentation of permissions for collecting data in particular areas.

We next analyze the types of actors prioritized by advocates of the FAIR and CARE Principles, and we show how those principles and associated standards influence formal recognition of who counts as a participant in biodiversity data portals.

### 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 particular 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 particular 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, being able to locate an instance of that dataset using this identifier in a repository online, and being able to determine 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, according to item A1 in the list. 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 address whether this license should be open (Higman et al, 2019). The Interoperability principle focuses on the use of 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 Reusable principle identifies several respects in which ‘rich’ description of data resources are 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 particular 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 at issue and the ‘machines’ of interest are understood as 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).

While machine agents in principle can serve the aims of any stakeholder seeking to use scientific data, multiple critics have noted that the contents of the FAIR principles do not address

### **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

the commonsense meaning of the acronym in English, i.e. in terms of advancing goals such as fair, equitable, inclusive, or just access and benefit to data. Datasets can be findable, accessible, interoperable, and reusable, for example, 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 principles are stated at a high level of abstraction that nonetheless links everyday concepts such as findability with technical concepts 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 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 the principles helps them garner endorsements from researchers and stakeholders to scientific data who use thick concepts such as ‘richly described’ metadata or ‘data munging’ that can only be fully understood in concrete practical 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 a 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 domain. However, the principles appeal to a vague sense in which FAIR data will be usable and beneficial by 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**

While the FAIR Principles have become increasingly influential internationally, they explicitly do not 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). As a result, the trajectory of FAIR offers both an exemplary model for institutionalizing compliance to emerging data governance principles. In light of the priorities, scope, and functions of the principles as described above, there has also been direct opposition and criticism. To make sense of the evolving situation, we can track how other data governance frameworks articulate their relationships to FAIR. Multiple emerging initiatives have positioned themselves as articulating complementary rather than opposed priorities for governance. The fit-for-use concept helps to characterize a broader range of values and norms than recognized by FAIR or the classical value-free ideal for science (Douglas, 2009).

Indigenous data sovereignty expands the possibilities of participation in knowledge infrastructures. While there are related but distinct concepts of Indigenous data sovereignty, 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 has interpreted to cover genetic data in the Convention on Biological Diversity’s 2011 Nagoya Protocol (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 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. The principles were initially drafted by the Indigenous Data Interest Group of the Research Data Alliance (RDA) in 2018 at the RDA Plenary in Botswana (Carroll et al, 2020a). The group was composed of international academic scholars and was co-led by Stephanie Russo Carroll from

the University of Arizona, United States, and Maui Hudson from the University of Waikato, New Zealand. Between 2017 and 2019, the group collected principles and statements of interests from groups in different parts of the world that focus on Indigenous data, including Aotearoa New Zealand, Canada, Australia, and the United States. They then compared these to more widely-used principles of open data governance, such as FAIR.

As minimal norms for open data and metadata, FAIR Principles don't explicitly encourage the perpetuation of existing oppression and harm to Indigenous Peoples so much as they do little to prioritize their remediation. As we saw in Section 3, the FAIR Principles focus on features of data and metadata abstracted away from any particular 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 toward 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 also obscure similar arguments made by scholars of Indigenous data that pooled data and the infrastructures used to store and transmit data are socially and historically situated, inextricably so, and that the use of these differently impact those with different 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 specific value of the principles in how they 'bring a people-and-purpose orientation to data governance, which complements the data-centric nature of the FAIR principles' (Carroll et al, 2020b). The critique functions to identify and fill gaps unaddressed by FAIR, and not to sweepingly oppose the making of Indigenous data findable or accessible, per se, or 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 this 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 (Carrol et al, 2020a, 6).

CARE thus orients issues of Access—which are left generic and prone to omission by FAIR—toward specific purposes, governance, and self-determination for specific peoples. 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, this 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 light of 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 that for their work, it is insufficient 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 a particular research problem, 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 in the governance of biodiversity data infrastructures. The ultimate equitability and desirability of this practice, 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 in seeking to translate initial widespread support, e.g. from many organizational leaders in academic libraries and museums, into bureaucratic mechanisms for driving institutional change. Most attention has focused on Indigenous genetic, cultural, and demographic data. Organisms on Indigenous lands have been and continue to be highly valued by scientists, conservationists, and businesses. 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 cases involving Indigenous data sovereignty for biodiversity data that are commonly included in data portals. The cases illustrate how Indigenous Peoples assert their data sovereignty by using the three strategies for fitness-of-use discussed in Section 2. When adopted by data infrastructures, the CARE principles provide institutional machinery that signal the legitimacy and justifiable use of these strategies for influencing infrastructure policies and behaviors. From these cases we reason prospectively to indicate how CARE will apply to biodiversity data infrastructures. Collectively the cases 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 of many types across diverse sources accumulated over centuries. Every biological specimen or

occurrence record derives its value for broader scientific use through identification to a taxonomic group, typically at the species rank. While currently difficult to quantify, a substantial proportion of biological specimen collections have been collected on Indigenous lands and informed by Indigenous knowledge (Vogel, 2019). Scientists have not prioritized Indigenous people, knowledge, and language as formally recognized authorities over and contributors to the data they assembled (Foster, 2017).

This case illustrates the second means by which Indigenous Peoples advance their data sovereignty by establishing fitness-for-use norms and rules that recognize their status to access and use data. While the FAIR Principles provide no detailed rules for how biodiversity data should be linked to the cultural categories and knowledge of stakeholders for the data, 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). Applying the CARE Principles in this context 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 [Māori] *te reo* and *ta re* have been incorporated, including the use of (1) variations of the words “Maori” and “Moriiori” to designate Aotearoa New Zealand origins, (2) Maori / Moriiori vernacular names for species, (3) Maori / Moriiori place names associated with species, (4) novel descriptive names created from Māori and Moriiori words, (5) novel names suggested by Maori in collaboration with taxonomists (Veale et al, 2019, 2).

The CARE Principles therefore indicate the importance of documenting and providing access to biodiversity data using Indigenous names by incorporating them into metadata, publications, and data collection (Wehi et al, 2019). A limitation of FAIR is the absence of language supporting inclusive and equitable access to scientific knowledge. For biodiversity data science, the names scientists attribute to organisms are essential metadata to realizing all four pillars of FAIR, but the terminology and language used for these names can itself perpetuate barriers for particular communities. As a result, standards for FAIR biodiversity data developed by the current community of biodiversity researchers are unlikely to recognize Indigenous Peoples as a category of actors 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 new resource exploitations or land exclusions by

businesses, states, or conservations. Simply increasing the FAIR-ness of biodiversity data, as with maps, doesn't 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).

Cases about land illustrate the second means by which Indigenous Peoples advance their data sovereignty by establishing norms and rules about whose input is required or legitimate in making determinations on the fitness-for-use of biodiversity data. 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 next to 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 don't 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 maintain control over 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, with concerns about green grabs of desirable land by local elites (Astuti and McGregor, 2017).

### *3.4.1. Commercialization*

The framework of Indigenous data sovereignty asserts that documenting the provenance of data collected on or about Indigenous lands and Peoples is essential to realizing just relationships that can and should hold between relevant agents. 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, and nanotechnology industries, among others. 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 increasingly are required to link digitized collection permits with individual data records in biodiversity databases, an uncommon practice among biologists (Zimkus, 2021).

Cases about commercialization illustrate the third means by which Indigenous Peoples advance their data sovereignty by establishing norms and rules about formally documenting when Indigenous Peoples or persons are the subjects or contributors to biodiversity data. Although FAIR includes requirement R1.2, '(meta)data are associated with detailed provenance,' current community practices among biodiversity data collectors and managers are rarely sufficient to ensure that provenance 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 the organism's taxonomic group. Information about who gave permission for data collection, or to whom the species relate, isn't 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 any prior guidance, so they more commonly collect and investigate species based on local guidance, which

may include Indigenous knowledge.<sup>ii</sup> 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 species' appetite suppressant properties, which they 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 the species' existence.

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, 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 nonIndigenous 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,"' i.e. according to a targeted rather than a comprehensive sampling strategy (Greene, 2004, 215). This dispute fueled the eventual withdrawal of Consejo from the partnership and a period of battles over permissions and contracts, ultimately undermining the potential for equitable benefits.

#### **4. Discussion and Conclusion**

We've shown how two sets of international principles for data governance, FAIR and CARE, are shifting decisions about who counts as a participant in biodiversity research. Historically the scope of participation has been determined 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 scientific data infrastructures as a general class of organization, however, there is an opportunity to influence many projects by establishing general norms and rules for their governance. We focused on norms and rules for making judgments about the fitness-for-use of data, especially who has authority, whose access and information needs are prioritized, and whose contributions should be formally documented. We conceived of fitness-for-use in this respect as a social as well as technical status that actors attribute to a body of data that is similar to judging someone fit-for-office. Our analytical approach could be used to study data governance principles beyond FAIR and CARE and in fields other than biodiversity.

Our analysis showed that those who developed and implement FAIR and CARE seek to institutionalize different categories of actors as legitimate participants, thereby prioritizing different criteria for fitness-for-use. Given the perceived value of general data governance

principles for the aims of biodiversity science, people have developed a strategy to influence the operations, outcomes, and authority of data portals. This strategy is distinct from the top-down and bottom-up strategies detailed by Hine, and it may be used in conjunction with either of them. The strategy 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 to implement those principles in particular data infrastructures. Notably, the two sets of data principles we considered both advocate for particular classes of agents. Proponents for FAIR, for example, advocate most directly for machine agents, and indirectly for those who have access to such machine agents. Proponents for CARE advocate most directly for Indigenous Peoples, and indirectly for those with less access to machine agents. When a particular portal implements FAIR or CARE, it develops criteria for fitness-of-use that advance the aims of those who developed the principles and those agents assumed therein as relevant participants. Thus, by adopting different or multiple sets of principles, portals influence their particular conceptions of who counts as a participant.

## 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.
- Effert, 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.
- Stephen Garnett 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 Lehtälä, 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](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.
- Sterner, Beckett, Edward Gilbert, and Nico Franz. 2020. “Decentralized but Globally Coordinated Biodiversity Data.” *Frontiers in Big Data* 3: 519133.
- Bruno Strasser 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.
- 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](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.

---

<sup>i</sup> 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.).

<sup>ii</sup> 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.