How Data Governance Principles Influence Participation in Biodiversity Science

Authors:

Beckett Sterner
School of Life Sciences
Arizona State University
bsterne1@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. The management of these repositories, often called biodiversity data portals, can serve either to redress or to perpetuate the colonial history of biodiversity science and current inequities. Both researchers and Indigenous Peoples are implementing new strategies to influence whom biodiversity data portals recognise as salient participants in data management and use. Two notable efforts are the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective benefit, Authority, Responsibility, Ethics) Data Principles. Actors use these principles to influence the governance of biodiversity data portals. 'Fit-for-use' data is a social status provided by groups of actors who approve whether the data meets specific purposes. Advocates for the FAIR and CARE Principles use them in a similar way to institutionalise the authority of different groups of actors. However, the FAIR Principles prioritise the ability of machine agents to understand the meanings of data, while the CARE Principles prioritise Indigenous Peoples and their data sovereignty. Together, FAIR and CARE illustrate a broader emerging strategy for institutionalising international norms for digital repositories about who they should recognise as having a formal role in determinations of the fitness-for-use of data.

Introduction

The 2022 Kunming-Montreal agreement marks an important moment for the Convention on Biological Diversity (CBD), which governs the international rights of scientists, companies, nations, and Indigenous Peoples to benefit from genetic information about biodiversity. One of the CBD's new targets for 2030 highlights the need to 'ensure that the best available data, information and knowledge, are accessible to decision makers, practitioners and the public to guide effective and equitable governance." The target also emphasises that 'traditional knowledge, innovations, practices and technologies of [I]ndigenous peoples and local communities should only be accessed with their free, prior and informed consent, in accordance with national legislation' (CBD, 2022).

 Data principles provide emerging international norms for biodiversity data governance addressing the CBD's 2030 targets. 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 extracting specimens from Indigenous lands, renaming places and organisms known to Indigenous Peoples,[i] and omitting Indigenous interests and contributions to knowledge. In the months leading up to the Kungmin-Montreal agreement, the CBD Secretariat highlighted two sets of principles as especially relevant to its aims (Open-Ended Working Group, 2022): the FAIR Principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016) and the CARE Principles (Collective benefit, Authority, Responsibility, Ethics) (Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019).[ii]

Two different sets of actors have developed and promoted the FAIR and CARE principles. The FORCE11 working group of scientific researchers and publishers published the FAIR Principles in 2016. The original publication (Wilkinson et al., 2016) has now been cited over 8,900 times as of February 2023. The European Union has invested substantially in developing the FAIR principles into standardised criteria and indicators that can adopted as official policy (European Commission Expert Group on FAIR Data, 2018). The Indigenous Data Interest Group initially drafted the CARE principles in 2018 at the plenary meeting of the Research Data Alliance in Botswana (Carroll et al., 2020a). The group was composed of thirteen international academic scholars and was co-led by Indigenous scholars, Stephanie Russo Carroll and Maui Hudson. The CARE Principles are grounded in the United Nation Declaration on the Rights of Indigenous Peoples (UNDRIP). CARE is gaining rapid international interest and adoption, for example in the recent United Nations Educational, Scientific and Cultural Organisation Recommendation on Open Science (UNESCO, 2021).

Both sets of principles seek to institutionalise general norms for whose contributions to data should be explicitly recognised and who has authority over access rights and appropriate uses. Senior 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 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 in making decisions about the use of data related to themselves and their lands and waters (Carroll et al., 2020a).

What strategies are advocates of FAIR and CARE principles using in prioritising different groups of actors (i.e., machine agents and Indigenous Peoples) in formulating and

implementing data principles? How are they similar or different? One analytical approach to this question is to characterise the social effects of scientific norms in terms of their effects on who scientists treat as legitimate participants in their projects (Díaz-Reviriego et al., 2019). A common approach in the past has been to recognise people as participants if they 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, though, have argued this 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). Who counts as a participant also has further implications for the criteria scientists adopt to track the impacts of their projects, which often favor specific groups of stakeholders while excluding others. Some projects now embrace expanded understandings of participation that empower non-professional scientists to guide the aims of research and the terms of data collection and use (Pareja et al., 2018; Turreira-García et al., 2018; Thompson et al., 2020).

Following this analytical approach, we characterise how the growing use of the FAIR and CARE Principles by researchers, academic institutions, national governments, and Indigenous Peoples marks an important development in the governance of biodiversity data. In Section 2, we introduce an analytical framework using the concepts of knowledge infrastructure and data governance, and we use this framework to articulate the social and technical dimensions of fitness-for-use. In Section 3, we show how advocates for the FAIR and CARE Principles establish different but sometimes compatible classes of participants—machine agents and Indigenous Peoples, respectively—as authoritative in judging the fitness-for-use of data held in digital repositories. Next, we address current and prospective uses 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. We then discuss how the CARE Principles apply to three examples of longstanding practices of publishing data related to Indigenous cultures, lands, and economic activity.

Analytical Perspectives

We use three analytical concepts: knowledge infrastructures, data governance, and fitness-for-use. STS scholars often use 'knowledge infrastructure' to characterise 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). Researchers and governments increasingly recognise knowledge infrastructures as a class of organisations that operate outside individual state or academic institutions. Teams of scientists employed by multiple universities, for example, often collaborate to run large scientific data repositories. Additionally, while many data repositories receive government funding, they typically are not subject in their daily operations to any particular government agency or authority. We treat the sense of knowledge that is relevant to a particular knowledge infrastructure as constructed and revisable over time as actors exert influence over the development of the infrastructure.

As part of generating, sharing, and maintaining knowledge, knowledge infrastructures contend with establishing their products or services as authoritative and legitimate (Strasser et al., 2019). Christine Hine, for example, analysed several knowledge infrastructures involving citizen science participation and noted two general kinds of knowledge infrastructures: those

characterised 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). Her typology, however, does not address the process by which participants in knowledge infrastructures acquire authority and legitimacy for their contributions by way of external relationships with stakeholders, including funders, users, universities, and other infrastructures (Mitchell et al., 1997). There is an opportunity, then, to investigate how a broader range of parties seek to influence knowledge infrastructures through negotiating what counts as knowledge for a particular project and who has the authority to decide.

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. In this respect, portals are more than web interfaces for databases; they are also social organisations whose collective activities constitute a spatially and socially distributed network of relationships among people and places. Some of these portals directly facilitate citizen science activities, such as iNaturalist, while others, such as the Global Biodiversity Information Facility (GBIF), focus on providing a single global point of access for all biodiversity data sources. Regardless, these portals establish their authoritative status as producers of knowledge in relation to serving the aims and values of their users and stakeholders more broadly. Put differently, the professional scientists who run the portals constitute only one group among many who matter in characterising the value and authority of the portals' data resources or set portal norms (Frischmann et al., 2014, Leonelli, 2016).

Norms and policies for data governance are therefore of critical importance for the status of biodiversity data portals as knowledge infrastructures (Frischmann et al., 2014). We define data governance as the assemblage of institutional mechanisms, norms, and policies that regulate the rights and responsibilities of people or organisations to produce and use data. Data governance norms can be implicit or explicit in the activities of infrastructure personnel and frequently vary from one context to the next (Leonelli and Tempini, 2020). Community standards and policies also dynamically evolve with time in response to funding and other institutional incentives. Scientists working with the Ecological Metadata Language, for example, sought to refashion and renegotiate their collective narratives of progress in response to difficulties experienced at different project sites (Millerand et al., 2013). Data principles such as FAIR and CARE are examples of emerging cross-organisational norms for data governance, and the implementation of FAIR and CARE through new community standards and policies has the potential to change the structures, functions, and practices of knowledge infrastructures to improve their utility and benefits for broad classes of actors.

The concept of fitness-for-use is an important focus for data governance as users and stakeholders seek to influence the norms, aims, and functions of knowledge infrastructures. We borrow the term fitness-for-use 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 to be 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 the legitimacy of using the portals' data 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 the sampling design of a dataset provides meaningful evidence for a modelling question—we use the concept to analyse how claims of usability also reflect social relationships and status. We therefore treat being fit-for-use as a status people attribute to a body of data, analogous to how a third party is required designate a person as fit-for-work or fit-for-office. In each of these examples, 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 Indigenous persons, lands, and waters. Alternatively, many scientists and businesses want data infrastructures to provide sufficient background information so that their computational proxy agents can make authoritative judgments about fitness-for-use on their behalf.

Having recognised fitness-for-use as a socially attributed status, we can analyse data portals to determine who has the authority to make fitness-for-use judgments about data held in the repository. Additionally, we can analyse 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 recognised 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. Oversight: 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 organisational oversight role in the data portal
2. Purpose-based: 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. Material contribution: Actors whose contributions to a data resource materially affect judgements about its fitness-for-use, e.g., because the contribution entails legal obligations for other parties or because of community norms about data ownership

These sets of actors and their characteristics may also intersect and overlap. An example of overlap would be cases in which scientists must document permissions for collecting data in particular locations in order to meet ethical research standards for their research project, which combines the oversight and material contribution types above.

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

FAIR Principles

Researchers developed the FAIR Principles for scientific data with the explicit aim of using machine agents to automate 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' centres 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 also 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 the need to designate some license for the data, but they do not specifically require open licensing (Higman et al., 2019). The 'interoperability' principle focuses on using shared, standardised vocabularies for describing information in datasets, making sure these standards are themselves FAIR and ensuring 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 of 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 (Wilkinson et al., 2016), 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 (Wilkinson et al., 2016: 4)

'To be Findable:
    F1. (Meta)data are assigned a globally unique and persistent identifier
    F2. Data are described with rich metadata (defined by R1below)
    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 standardised 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 common sense 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 directly 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. Researchers' widespread support of FAIR by researchers partly relates to 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 the principles' advocates garner endorsements from researchers and stakeholders before understanding how communities will operationalise these concepts in local situations. Further

research and policy development, led especially by the European Union, has focused on developing 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, recognising the need for community-level deliberation on the standards and metrics appropriate to the domain. However, the advocates of the principles abstractly 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.

CARE Principles

As noted in Section 3.1, the FAIR Principles' promoters explicitly decline to comprehensively 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 organisations offers an exemplary model for institutionalising compliance with emerging data principles. Indeed, advocates for other sets of data principles have positioned themselves as articulating complementary rather than opposing priorities to FAIR. Our conception of fitness-for-use as a social status that actors attribute to a dataset is especially helpful in characterising how different sets of data principles interact and influence the aims and legitimacy of knowledge infrastructures.

Indigenous data sovereignty, for example, addresses the importance of Indigenous Peoples' participation in and authority over 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). UNDRIP provides an international framework that formally acknowledges the sovereignty and cultural and intellectual property rights of Indigenous Peoples, which the CBD's 2011 Nagoya Protocol further specifies to include genetic resources (United Nations, 2007; 2011; Davis, 2016). UNDRIP complements more specific national policies and laws that address Indigenous data rights and sovereignty in, e.g., Canada and New Zealand.

Indigenous Peoples and scholars developed concepts of data sovereignty to characterise Indigenous rights and interests in light of historical trends (Carroll et al., 2020a). These trends include power imbalances favouring non-Indigenous over Indigenous Peoples; the collection and use of data about Indigenous Peoples without their consent, including to portray them as solely deficient in some characteristic; the use of these data and narratives to further the aims and values of non-Indigenous People and governments; and the lack of control over and access to Indigenous data by Indigenous Peoples.

The CARE Principles, listed in Box 2, build on the right of sovereignty to articulate principles specifically addressing data from or about Indigenous Peoples, including about individual Indigenous persons, groups, lands, and waters. Between 2017 and 2019, the CARE Indigenous Data Interest 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 (Carroll et al., 2020a).

As minimal norms for open data and metadata, the FAIR Principles do not explicitly encourage the perpetuation of oppression of and harm to Indigenous Peoples. However, they also do not prioritise redressing these issues. As demonstrated in Section 3.1, the FAIR Principles focus on properties of data and metadata as objects, abstracted from any effects the collection and use of that data and metadata may have on people. This abstraction enables the presentation of a generic machine agent perspective on scientific data as objectively measurable and separated from human subjectivity, obscuring how the fitness-for-use of data is inherently related to the aims and situations of specific actors. While abstraction can serve the goal of generalisation across many actors and situations, socially dominant groups can also use the elision of concrete details about who benefits most to foster ignorance about how seemingly raceless contemporary institutions perpetuate privileged power and access to resources (Mills, 2015). In response, scholars of Indigenous data argue that pooled data resources, and the infrastructures used to store

---

Box 2: The CARE Principles for Indigenous Data Governance (Carroll et al., 2020, 5)

'**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 recognised 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. Recognising 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'

---

and transmit them, are inextricable from their social and historical context (Chan et al., 2019).

Further, the use of these data differently impacts those with unequal power relations (Kukatai and Taylor, 2016b, Walter and Andersen, 2016).

By comparison, 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 using technical components like unique identifiers. In joint meetings with FAIR working group members, for example, leaders of the CARE Principles have noted the importance for Indigenous sovereignty of making data about Indigenous Peoples and their lands more findable and accessible when these data are held by colonial governments or organisations (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, categorised, distributed, and used. Theorising and institutionalising Indigenous data sovereignty, therefore, begins to move Indigenous Peoples from obscured to formally recognised authorities on fitness-for-use on multiple levels in data infrastructures. For instance, the idea of access in FAIR is unmarked with respect to any particular stakeholder—rhetorically it is presented as accessible for anyone. In contrast, 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 (Carrol 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 marginalisation. The following sections further examine how FAIR and CARE prioritise the authoritativeness of distinct groups of actors in the governance of biodiversity data infrastructures.

Use of FAIR to harness efficiencies by prioritising 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, governments and private companies hold a substantial proportion of data describing the locations and traits of species and do not make them available 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 the Convention on International Trade in Endangered Species of Wild Fauna and Flora). Given these facts, any group will confront challenges when seeking 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). The work of biodiversity data collection and sharing continues to be highly decentralised and

spatially distributed. The largest international data portal, GBIF, aggregates about 2.3 billion data records as of January 2023 by combining 81,000 datasets sourced from 1,950 data publishers. These publishers are predominantly scientific organisations, 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). Nonetheless, after decades of work, the transaction costs of finding, harmonising, and reusing biodiversity data are still prohibitive obstacles for many researchers. Biodiversity researchers also recognise that making data open in the sense of accessible freely online with few or no legal restrictions on re-use is still insufficient for their purposes. Such openness does not guarantee that users of the data will be able to identify redundant records, that data providers will follow standardised formats, or that it will be possible to integrate information about the geolocations and taxonomic identities of observed organisms across sources.

Researchers and data portal leaders are using FAIR to set norms and rules that prioritise the ability of computational agents to access and evaluate biodiversity datasets, reducing the need for human labour. The principle of Findability, for example, addresses basic challenges of 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). 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). The Darwin Core format, for example, imposes minimum required metadata information for observations of species occurrences, such as providing a scientific name for the observed organism (Wieczorek et al., 2012). Darwin Core does not designate a standardised biological taxonomy, however, 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 optimises fitness-for-use for all stakeholders (Garnett and Christidis, 2017, Franz and Sterner, 2018, Sterner et al., 2020). Another set of relevant data standards is 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 as motivating its efforts to coordinate global interoperability for trait data (Gallagher et al., 2020).

These examples show how researchers use FAIR to prioritise 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. by exercising the oversight type of authority we identified in Section 2). To the extent they make independent judgments that are not further shared or made available to others, they exercise the purpose-based type of authority we

identified. The ultimate equitability and desirability of these practices, however, are controversial.

Use of CARE to prioritise Indigenous Peoples in biodiversity data governance

As leaders of CARE move toward implementing technical standards and criteria for compliance, they follow a similar strategy to FAIR by seeking to translate initial widespread stakeholder support (e.g., from organisational 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—including scientists, conservationists, and businesses—continue to value the collection and study of organisms on Indigenous lands. A recent global study found that Indigenous Peoples manage or have tenure rights over about 38 million square kilometers, which represents 'a quarter of the world's land surface, and intersects about 40% of all terrestrial protected areas and ecologically intact landscapes' (Garnett et al., 2018). In principle this fact could provide the foundation for a powerful alliance among Indigenous peoples, conservationists, and researchers, but external actors still regularly overlook or override the interests and right to self-determination of Indigenous groups (Rimmer, 2015; Gilbert and Lennox, 2019).

This section examines several examples of how Indigenous sovereignty can apply to biodiversity data commonly stored in data portals. These examples serve to illustrate how Indigenous Peoples assert data sovereignty by influencing who gets included in the three sets of actors having authority over fitness-for-use that we introduced in Section 2. The CARE principles help managers and stakeholders of biodiversity data portals legitimise and justify attributing authoritative status to a broader group of actors than data portals have historically recognised.

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 categorisation 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 quantify the precise proportion because the required provenance information is typically lacking. Historically, however, scientists have not prioritised Indigenous people, knowledge, and language as part of the natural history information network they sought to assemble (Foster, 2017). Furthermore, many biodiversity data portals as organisations have not formally recognised Indigenous peoples, knowledges, and languages as authorities over and contributors to their data.

The CARE Principles can address this lack of recognition by establishing Indigenous Peoples as authorities in accessing and making judgements about the fitness-for-use of data. While the FAIR Principles lack 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). In other words, the CARE Principles warrant

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.

A current goal for many Indigenous communities is to have Western science serve to strengthen rather than erase Indigenous knowledge and cultures, especially names for organisms and places. For example, ecological researchers working in Aotearoa (New Zealand) have largely failed to acknowledge or discuss the Māori knowledge system, mātauranga Māori, in their research on species (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).

Relatedly, United States biodiversity data portals rarely provide metadata linking scientific species names to Native American names for the corresponding organisms, making it difficult to search for data using to culturally meaningful categories.

Researchers are recognising this gap as an important opportunity for biodiversity researchers to advance the preservation and continued use of Indigenous names by incorporating them into metadata, publications, and data collection (Wehi et al. 2019). For biodiversity data science, the names scientists attribute to organisms are essential metadata for realising all four pillars of FAIR. However, prioritising this scientific terminology and language can also actively erase or perpetuate barriers to access for Indigenous communities. As a result, the CARE Principles provide a significant complement to FAIR by setting out a systematic basis for recognising Indigenous Peoples as actors whose categories and language should be prioritised in accessing and using data from Indigenous territories.

Indigenous lands

We use examples of current and past governance of biodiversity data about Indigenous lands to illustrate how the CARE Principles establish and legitimate the authority of Indigenous Peoples over whose input is required or legitimate in making determinations on fitness-for-use for other actors. Open access to biodiversity data collected from Indigenous peoples' lands and waters can aid or hinder their rights to self-determination depending, for example, on whether systemic inequalities prevent the Indigenous groups from using the information. Similarly, data in the portals may make it easier for businesses, states, or conservation organisations to more effectively exploit resources on Indigenous lands or establish new land exclusions. In this respect, one cannot assume that the collection and sharing of biological data — including specimens or other vouchered observations — generally has a neutral effect on Indigenous peoples' claims and access to natural resources.

Indigenous communities, for example, 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 also increasingly inform how national governments represent their territories. In Indonesia, for example, the international REDD+ program (Reducing Emissions from Deforestation and forest Degradation, plus enhancing forest carbon stocks and conservation) has centralised forest governance in the country's federated system. This centralised 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, 2274; see also Mulyani and Jepson, 2017). In combination with Indonesian recognition of some Indigenous land rights, the initiative 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).

Similarly, activities such as participatory mapping can establish new claims to rights for Indigenous Peoples by factually documenting 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 may bring previously isolated peoples and natural areas into greater contact with economic and political forces that do not prioritise collective benefits with Indigenous groups or biodiversity. The formalisation 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.

Considering the risks as well as benefits of sharing digital information about Indigenous lands, simply increasing the FAIR-ness of biodiversity data is not sufficient to establish 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 United States 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).

While biodiversity researchers commonly exert their authority as data managers to restrict who can access high-resolution spatial data about protected species, they have not developed parallel standards for restricting access to Indigenous data in light of potential harms to Indigenous Peoples. In addition, the CARE Principles articulate more robust expectations than many national legal and ethical standards, such as U.S. copyright law or informed consent protocols (United Nations General Council, 2007; Davis, 2016). Nor have they addressed the language of 'free, prior and informed consent, in accordance with national legislation' used in the Kunming-Montreal target we mentioned in the introduction (CBD, 2022).

UNDRIP provides a useful reference in this respect, as it establishes a general right to self-determination and many more specific supporting rights for Indigenous peoples. Article 31.1, for example, declares that

> Indigenous peoples have the right to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions, as well as the manifestations of their sciences, technologies and cultures, including human and genetic resources, seeds, medicines, knowledge of the properties of

fauna and flora, oral traditions, literatures, designs, sports and traditional games and visual and performing arts. They also have the right to maintain, control, protect and develop their intellectual property over such cultural heritage, traditional knowledge, and traditional cultural expressions (United Nations General Council 2007, 9).

To address this gap between scientific and Indigenous governance norms, leaders of CARE have co-authored the Traditional Knowledge and Biocultural Labels as a means to articulate conditions under which Indigenous data may be viewed and used. They are currently collaborating with data repositories to implement the labels at scale (Liggins et al., 2021). The conditions described in the labels are customisable to particular Indigenous groups and repositories. They may, for example, restrict access occur during certain times of year or require researchers to contact and establish relationships with relevant Indigenous groups. The labels illustrate how compliance with CARE is likely to regularise and document Indigenous participation in the governance of knowledge infrastructures—see especially CARE Principles C1 and E3 in Box 2.

Commercialisation

Bioprospecting provides an internationally significant example of how the CARE Principles establish authority through documenting the provenance of data collected on or about Indigenous lands, waters, and Peoples. 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. Companies have developed many successful commercial products based on studies of the properties or behaviours of biological species, including products such as biofertilizers, nutritional supplements, industrial chemicals, and medicinal drugs or treatments (Efferth et al., 2016). Since 1993, the CBD has governed the legal extraction of genetic resources from a country for the purpose of research and commercialisation.

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. While in rare cases researchers may discover useful properties of species fortuitously without any prior guidance, it is more common for researchers to focus on collecting and investigating species based on local guidance, which may incorporate Indigenous knowledge.[iii] Most biodiversity data portals, however, do not generally mandate or collect information about who gave permission for data collection or who provided background knowledge leading to the observation, e.g., about the importance of the species or likely locations where it could be found. 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. These practices are changing under guidance from the CBD, however, and an emerging best practice is for researchers to link digitised collection permits with individual data records in biodiversity databases (Zimkus, 2021). By normalising requirements to document Indigenous permission, the CARE Principles also institutionalise recognition of Indigenous Peoples as participants in the collection, storage, publication, and use of biodiversity data collected from their lands.

Two examples of bioprospecting illustrate how the presence or absence of metadata recognising contributions of Indigenous Peoples can influence their authority over the process

and results of bioprospecting. Our first example concerns commercialisation 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.*

Our second example shows how gaps in provenance can fuel disputes over biopiracy. In 1993, the International Cooperative Biodiversity Group (ICBG) funded a team from Washington University in the United States, 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).

A leader of an Indigenous organisation, Evaristo Nugkuag, contests this account from Lewis, however. Nugkuag is the founder and president of Consejo Aguaruna y Huambisa, which had entered into a collaborative agreement with Washington University and the research team. As quoted by Greene (2004, 215), Nugkuag said

> that the ICBG researchers made a critical mistake in choosing to work with the Ministry of Agriculture and that "without having authorisation 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."

Here, a 'discrete' sampling strategy means targeted rather than comprehensive. This dispute contributed to Consejo's eventual withdrawal from the partnership and a battle over permissions and contracts, ultimately undermining the potential for equitable benefits-sharing. The norms that CARE articulates for provenance documentation, in conjunction with international agreements such as the CBD, therefore have concrete effects on the recognition of Indigenous contributions and authority over biodiversity data.

Conclusion

This paper inquired about the similarities and differences of FAIR and CARE in formulating and implementing data principles. We addressed how proponents of the FAIR and CARE Data Principles use them similarly to drive institutional change in the authority of different groups of actors to participate in setting the fitness-for-use of biodiversity data: machine agents and Indigenous Peoples, respectively. Our analytical framework was grounded in the study of knowledge infrastructures, and our empirical study focused in particular on data principles as an emerging approach to setting new norms and standards in the data governance of biodiversity data portals. We analysed the concept of fitness-for-use as not just technological but also involving inherent social relations, i.e. that people and purpose are inherently involved in

attributing fitness-for-use as a relational status of data records. On this basis, we characterised how advocates of FAIR and CARE use data principles to influence the participation of three distinct sets of actors in determining fitness-for-use.

Our analysis of FAIR and CARE showed that advocates for both sets of principles are using them to influence which groups are recognised by data portals as having the authority to make judgments about the fitness-for-use of the data they hold. We showed that advocates use the FAIR Principles to institutionalise new authority for computational agents in biodiversity knowledge infrastructures by ensuring these agents have access to sufficient metadata that they can 'know' what the data mean. In particular, FAIR empowers computational agents as actors whose input is required or legitimate in determining the fitness-for-use of data by others, and as actors with the authority to judge fitness-for-use for their given purposes as agents. We then argued that the CARE Principles address all three forms of authority in biodiversity knolwedge infrastructures on behalf of Indigenous Peoples, by citing examples relating to Indigenous cultures, lands, and commercialisation of biological resources. In addition to regulating the sovereignty of Indigenous Peoples over determining access and use of Indigenous data, CARE also expands the set of actors that biodiversity data portals typically recognize as having authority over fitness-for-use because of material contributions they made to producing or maintaining the data. 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.

More broadly, we have illustrated how data principles are an emerging means of governing which groups of actors are formally recognised 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 organisation, 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 to drive implementation of those principles in particular data infrastructures. Our analytical approach may be applicable to study the use of data principles beyond FAIR and CARE and in fields other than biodiversity.

Biographical Notes

Beckett Sterner is an Assistant Professor in the School of Life Sciences at Arizona State University. His research investigates what knowledge is needed for people to work together while differing in fundamental ways, including in beliefs, aims, and values. He examines this driving question using knowledge infrastructures for global biodiversity loss as a case study, including data commons, modeling methods, and international governance. He has an interdisciplinary background in history, philosophy, and social studies of science as well as statistics and biology.

Steve Elliott is a Research Affiliate at Arizona State University's Center for Biology and Society, for which he studies the economics, social organization, and reasoning strategies of

research. Some of his recent work examines science diaspora networks and proof-of-concept research.

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, Oscar Figueroa-Rodríguez, Jarita Holbrook, et al. 2020a. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19: 1–12.

Carroll, Stephanie Russo, Maui Hudson, Jarita Holbrook, Simeon Materechera, 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, Maui Hudson, Keith Russellet, et al. 2021. "Operationalising the CARE and FAIR Principles for Indigenous Data Futures." *Scientific Data* 8: 108.

Convention on Biological Diversity. 2022. "COP15: Final Text of Kunming-Montreal Global Biodiversity Framework." Accessed January 24, 2023. https://www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222.

Chan, Leslie, Angela Okune, Rebecca Hillyer, Denisse Albornoz, 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.

Edwards, Paul N. 2010. *A Vast Machine*. Cambridge: MIT Press.

Efferth, Thomas, Mita Banerjee, Norbert W. Paul , Sara Abdelfatah, 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, Katherine Strandburg, and Michael 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.

Garnett, Stephen, Neil Burgess, John Fa, Álvaro Fernández-Llamazares, et al. 2018. "A Spatial Overview of the Global Importance of Indigenous Lands for Conservation." Nature Sustainability 1: 369–74.

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, Margaret O'Brien , Kristin Vanderbilt, 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, Donat Agosti, Enrique Alonso García, 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, Melissa George, Catherine Robinson, 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, Francesco de Bello, José Alexandre Diniz-Filho, Thomas Lewinsohn 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, Carolina Behe, Finn Danielsen, 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, Benedict Wheeler, Lora Fleming, 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, Shaun Ogilvie, Rauru Kirikiri, et al. 2020. "Kua Takoto Te Mānuka: Mātauranga Māori in New Zealand Ecology." *New Zealand Journal of Ecology* 43 (3).

Millerand, Florence, David Ribes, Karen Baker and Geoffrey Bowker. 2013. "Making an Issue out of a Standard." *Science, Technology, & Human Values* 38: 7–43.

Mills, Charles. 2015. "Global White Ignorance." In *Routledge International Handbook of Ignorance Studies*, edited by Matthias Gross and Linsey McGoey, 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, and Annika Jacobsen. 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.

Open-Ended Working Group on the Post-2020 Global Biodiversity Framework. 2022. "Co-Leads' Report on the Work of the Informal Co-Chairs' Advisory Group on Digital Sequence Information on Genetic Resources Since the Fourth Meeting of the Open-Ended Working Group on the Post-2020 Global Biodiversity Framework." Convention on Biological Diversity, CBD/WG2020/5/INF/1, Appendix II.

Pareja, Claudio, Jordi Honey-Rosés, Nadja Kunz, Jocelyn Fraser, 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, Viktor Senderov, Georgi Zhelezov, et al. 2019. "OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science." *Publications* 7: 38.

Raymond-Yakoubian, Julie, Peter Pulsifer, Fraser Taylor, Camilla Brattland, 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/5da9f4479ecab221c e848fb2/1571419335217/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.

Strasser, Bruno, Jérôme Baudry, Dana Mahr, 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, Pablo Domínguez, Elena Carrillo-Anglés, 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, Thomas Buckley, Mana Cracknell, 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, Te Waiarani Harawira, Gerard Fitzgerald, 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, Robert Guralnick, Stan Blum, et al. 2012. "Darwin Core: An Evolving Community-Developed Biodiversity Data Standard." *PLoS ONE* 7: e29715.

Wilkinson, Mark, Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 1–9.

Wilkinson, Mark, Michel Dumontier, Susanna-assunta Sansone, Luiz Olavo Bonino da Silva Santos, 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.