Open data ownership and sharing: Challenges and opportunities for application of FAIR principles and a checklist for data managers

Albert I. Ugochukwu [*], Peter W.B. Phillips

Centre for the Study of Science and Innovation Policy, Johnson Shoyama Graduate School of Public Policy, University of Saskatchewan, 101 Diefenbaker Place Saskatoon, Saskatchewan, S7N 5B8, Canada

A B S T R A C T

The amount of data generated across various disciplines has been steadily increasing and is projected to experience exponential growth in the foreseeable future. This underscores the pressing need for proficient and streamlined data management. Data has proven to be a crucial tool in addressing complex societal challenges on a global scale. However, the challenge of producing and openly disseminating data that are easily discoverable, accessible, interoperable, and reusable (FAIR) has emerged as a significant concern for policymakers. The potential for data to be repurposed for advancing scientific research and innovation across different disciplines is contingent on its willingness to be shared. This paper employs a systematic literature review to investigate the motivating factors, advantages, and obstacles associated with open data sharing. Additionally, it explores governance frameworks that can create unique opportunities for implementing FAIR principles in real-time scientific research.

1. Introduction

In recent years, data-driven research has been on the increase globally, producing digital data that are relevant for development across disciplines. This has prompted calls from governments, international development agencies, and multilateral organizations for digital data to be treated as a public good and publicly made available to facilitate research and scientific process, and to supplement national-level data to provide evidence-based interventions [1].

Given that science underscores and operates on the principle of openness, government agencies often want the results, including data from the research they funded to be made open and publicly accessible. In addition, most publishers now require researchers to share their data after publication via public platforms [2,3]. This sharing policy is gaining ground across disciplines, thereby resulting in an increasing rate of data deposition in publicly accessible repositories [4]. Going by these requirements and from the economics perspective, one could consider data generated through publicly funded research as a 'public good'. This public good theory suggests that open data is *non-rivalrous* and *non--excludable*, implying that there are no barriers preventing anyone from accessing the data, and the use of the data by anyone does not exclude someone else from using it.

Making data openly and freely available has not gained traction in the research space due to limited appropriation of gains by owners for sharing data and their inability to control access, reuse their data, and recover costs associated with data collection, preparation, and unauthorized use [5]. The imbalance in risks and rewards serves as a disincentive to open data sharing, limits data release, and has prompted data protection and/or exclusion, particularly for private research activities, through some mechanisms ranging from climatic/geographic restrictions to legal protection, and licensing [6]. In principle, researchers will be incentivized to make their data publicly available if there are some expectations of extracting returns on their investment.

In agriculture, vast amounts of data have been collected in recent years, especially in breeding research. Unfortunately, the greater percentage of these data are neither made public nor shared in the research space. In response, a group of scholars developed and published the FAIR Guiding Principles to advance more structured and purposeful data management and stewardship [7]. The group offered "four foundational principles—Findability, Accessibility, Interoperability, and Reusability"—to guide data producers and publishers to maximize value from the algorithms, tools, and workflows that led to data.

To promote the practice of FAIR, research publishers, funders and institutions have stepped up their demands on data creators to open up

their research data for reuse. The GO FAIR Initiative, the sponsor and promoter of the FAIR principles, notes that the idea is being taken up and used widely. For example, leading journals like Nature, Science, and PLOS One have required authors to post their data in supplementary appendices or public repositories to enable replication and reuse of their work. Similarly, universities and public research organizations have long practice and experience in requiring and supporting open data. Most if not all public universities will either adhere to or assert alignment to some or all of the FAIR principles. Sustainability of the GO-FAIR Initiative, therefore, requires a structured template for assessing an institution's or community's efforts in pursuit of an open data outcome.

Digital data sharing occurs mostly through interpersonal exchange by trusted colleagues [8–10]. This suggests that scientific data sharing has been mostly private rather than public. Although some journals have policies that require authors to submit articles with data, this has not improved data sharing [11]. Data reuse is usually contingent on the applicability (appropriateness) of the data to the current study, the quality of data documentation, and the ability to interpret the data [12]. Some researchers do not follow best practices around data structure presentation, metadata, and licensing [13] and some open data does not fully comply with the FAIR Principles, often making existing data inaccessible, less interoperable, and reusable [14]. Efficient data management planning would optimally define the type of data to be collected, the associated costs and benefits, and how data will be stored and shared for reuse by others.

Against this backdrop, the paper aims to provide in-depth insight on data ownership, sharing, governance, and a multitude of factors influencing researchers' motivation for sharing and re-using open data, including the FAIR principles and strategies that enhance data sharing and interoperability. The paper also proposes a framework for achieving FAIR datasets that can incentivize open data sharing for the advancement of science and innovation.

2. Methodology

The study uses a systematic review approach to explore the literature on and practice of data sharing and reuse, including motivations, benefits and barriers, governance, and FAIR principles. Numerous studies in the existing literature concentrate on data ownership and sharing within particular fields. In contrast, this review centers on three interconnected streams: the drivers, benefits, and barriers to open data sharing and reuse; the governance of data and its application; and the implementation of the FAIR principle. While various discipline-specific investigations exist regarding data sharing and reuse practices, these are shaped by the distinct attributes and obstacles within each discipline. While each study delves into discipline-specific concerns, there are overarching insights, such as shared motivations, barriers, and opportunities that transcend disciplinary boundaries.

A systematic review to address the research question, "what are the drivers, incentives, and barriers to data sharing?" was conducted by electronically searching for original quantitative and qualitative articles across various disciplines. This review followed methodologies outlined by several authors in different disciplines, including [15–17], and [18]; and adhered to the PRISMA statement guidelines for systematic reviews and meta-analyses [19].

Databases searched included PubMed, Medline, Embase, Scopus, ProQuest, Cochrane Library, CINAHL, PROSPERO Database for Systematic Reviews, and PsycINFO. The search encompassed papers from 1997 to 2023 without limitations on publication dates. Key search terms (i.e. search strategy/query string) used to extract papers from the databases include: a) data sharing among researchers, b) data reuse, c) motivations for data sharing, d) benefits of data sharing e) barriers to data sharing, f) data governance, g) FAIR principles, h) data management and stewardship i) a and b, and j) c and d.

The review utilized specific criteria, as outlined by Ref. [18], to determine the eligibility of papers. Papers were considered eligible for inclusion if related to data sharing in research, published in English and a peer-reviewed journal, reported original research, and were qualitative or quantitative with any study design. Reference lists and hand searching were undertaken to identify additional papers (e.g., Refs. [6, 15–17,41,79,81]) classified as 'others'. Papers that focused on opinion pieces, personal letters, reviews, editorials, or non-peer-reviewed papers, as well as theses from master's or doctoral research were deemed ineligible and excluded. Additionally, duplicate papers were dropped. One author screened papers for eligibility and the other was involved in the full-text review process; conflicts were resolved by consensus.

The title and structured summary (abstract) of each eligible paper were extracted and compiled. These were used to group the eligible papers into eight themes, including data sharing and reuse, barriers, incentives, benefits, data governance, data management and stewardship, FAIR principles, and others (i.e., papers deemed relevant for the study). Out of the initial pool of one hundred and fifteen papers covering the key themes that were identified and extracted from technical reports, conferences, and journals for the review, twenty-nine were deemed ineligible based on the exclusion criteria. Subsequently, eighty-six papers met the eligibility criteria, were analyzed and included in the review process. The study makes a theoretical contribution to the body of literature.

The pictorial representation of the key themes and the number of articles identified and used under each theme are shown in Fig. 1.

3. Open data sharing: drivers, benefits, and barriers

In recent years, scientific research has become more collaborative and data-intensive owing to advances in technological development, such as the automated digital data capturing devices used in collecting plant phenotyping data [20–22]. Efficient research collaboration requires information (e.g., data) sharing and provision of access for use and reuse. Despite the increasing volume of research data in all disciplines, especially agriculture, the entire research community is yet to embrace the idea of sharing data openly for reuse by others. While open data and/or sharing practice is not completely absent in the research space, it is almost exclusively carried out in a few disciplines where these practices are embedded in research design from the onset [23]. Open data sharing and reuse by researchers has not become a common practice in most disciplines, largely due to discipline-specific challenges and the complex interaction of factors influencing data sharing and re-use.

3.1. Drivers and benefits of open data sharing and reuse

Although data sharing has been promoted as an increasingly essential part of research and the scientific process, there are mixed motivations and opportunities for sharing and reusing open research data, especially data that meets the FAIR data principles: findable, accessible, interoperable, and reusable [7]. Planning and policy choices might be informed by this open research data and would enhance the scope, quality, speed, and value of scientific research [24,25]. Often researchers are incentivized to share their data if it will increase the citation and visibility of their work [26] while combining data from multiple sources to generate a novel dataset (big data in some cases) can be an incentive for researchers to reuse open data [27]. Big data, most times, gives more robust insights than restricted data from a single source [28].

Data sharing and reuse offer scientists the opportunity for reanalysis and results verification [29], enhance science and innovation [20,30], and reduce duplication of efforts [31–34]. Data reanalysis can improve scientific research. According to Ref. [35], open access to data will encourage more carefulness and greater scrutiny of data and results by authors with the anticipation that reusers will carry out thorough checks of their data. He argued that this would engender trust in published results. Governments also use open data for data-informed policy decision-making and to maximize returns on public investment in
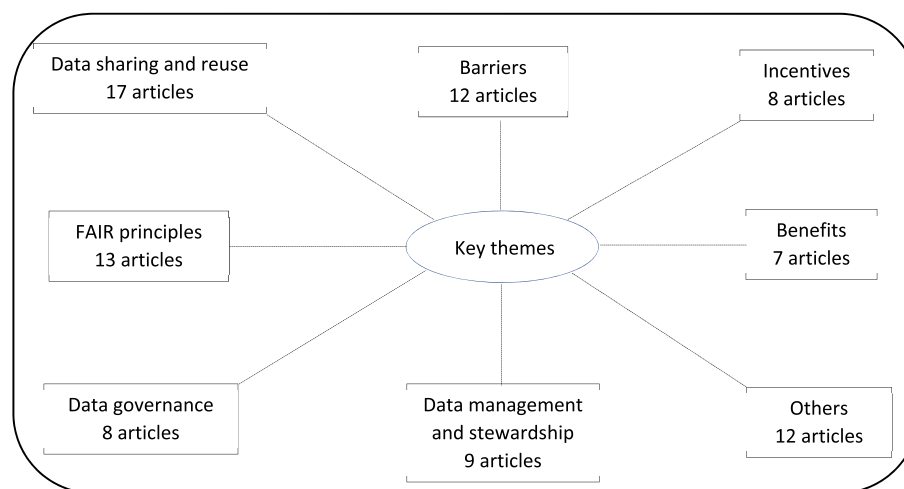
Fig. 1. Pictorial representation of key themes and number of articles identified and used under each theme.

research [36]. Some scholars (e.g. Ref. [37]) argue that making data accessible for reuse can reduce costs by minimizing duplication of data collection, reducing data falsification and fabrication, maintaining data integrity, and enhancing scientific progress through interdisciplinary research that draws different interpretations from the same data sets.

Making data openly and publicly accessible also creates the opportunity to enhance efficiency, transparency, and accountability in public service delivery, as well as economic value for data [38]. Making data publicly available could enhance collaboration within the broader scientific community and give researchers the opportunity to build on the work of others. It can also stimulate meta-analysis.

Specifically for food, the implementation of FAIR could also bring about increased innovation and help in solving food security and climate change challenges in the case of agriculture and nutrition [39]. While sharing data publicly could serve as an incentive for researchers to effectively manage data and ensure quality, it will also increase transparency and recognition of researchers and funders [40]. In addition, open data sharing can enhance trust and transparency among stakeholders in the food system [41] and effectively accelerate research discovery [42].

3.2. Barriers to open data sharing and reuse

Making data open and available for reuse is associated with some challenges. Lack of requisite digital infrastructure, the significant effort required in data sharing [43], fear of losing publication opportunities from own-generated data sets [11,44], receipt of appropriate credit [24], and protection of proprietary interests in the case of academic-industry relations [45] all can work to deter researchers from sharing and making their data openly available. Open data sharing also requires institutional investments in infrastructure and data stewardship, which can be a challenge.

Some scholars hesitate to share their research data due to privacy and confidentiality concerns [46]. This underscores the importance of ownership structures and data governance.

Ownership rights are key to accessing and reusing open data, particularly in agriculture. Challenges include inadequate understanding and incomplete legal frameworks that regulate open data ownership, as well as a lack of standardization and data governance strategies. Further, insufficient awareness regarding adherence to standards, a deficiency in data science proficiency, a focus on data collection rather than its subsequent reuse, and a constrained allocation of funds and standards management are contributing factors [47]. In most cases, legal rights are owned by intermediaries – those who make significant investments in data collection, arrangement, and compilation of

databases–rather than by individuals who produce or use the data [38]. The involvement of several actors (e.g. creator, compiler, funder, consumer, purchaser, licensor) further exacerbates the complexity of data ownership. In plant phenotyping, in addition to these challenges, storage infrastructure, data sharing, and access are problems affecting data exchange and reuse.

Administrative bottlenecks also can discourage sharing data openly, particularly when two or more organizations are involved in data collection. In this case, who makes decisions about data sharing becomes an issue as no single organization can claim data ownership. For some individuals or small organizations, the high costs of acquiring facilities for storing and managing data, especially in the absence of any returns on investment, serve as a disincentive to open data sharing [48].

Barriers to open data sharing have also been examined from social, political, institutional, legal, and operational perspectives [48]. While cultural differences and social barriers to interaction among users are common barriers, lack of support about and knowledge of open data constitute small political barriers. The absence of open data policy, guidelines, and standards, and sometimes difficulty in handling user requirements, create institutional barriers to open data sharing. A number of operational barriers can also inhibit sharing, including data fragmentation, absence of metadata and information on data quality, and changing and/or absence of clear semantics.[1] Legal barriers can arise from policy differences and diverging licensing practices [49–51].

The summary of drivers (motivations), benefits, and barriers with action points to open data sharing and reuse is presented in Table 1.

4. Data governance

Data governance has been viewed by many as a potential solution to the challenges and/or barriers associated with open data sharing by public and private organizations. Data governance has been viewed from different lenses by many scholars and organizations, differentiated based on objectives and priorities that, oftentimes, vary from one organization to another. Central to these definitions is that behaviours (e. g. decisions on data) must be in line with the performance objectives, priorities, and needs of the organization.

---

[1] Effective open data sharing and reuse require a clear and common understanding of the semantics (database description) used by the sharer and potential users (stakeholders). It is important that all stakeholders within a domain (e.g. plant phenotyping community) understand the terminologies and coding systems used for the data set. This requires a coordinated approach and use of common guidelines (data management and governance principles) that would enhance interoperability and socio-economic value of data.

Table 1

Summary of the drivers (motivations), benefits, barriers, and action points to open data sharing and reuse.

Item	Description	
Drivers	*What could motivate a researcher to share data*	
	a. Publicly accessible research data can be beneficial for planning and policy development, while also improving the breadth, caliber, pace, and significance of scientific inquiry	
	b. Sharing data boosts the citation rate and visibility of research endeavours.	
	c. Creating a big data by combining data from several sources could encourage reuse.	
Benefits	*Advantages linked with sharing data*	
	a. Policymakers, practitioners, and stakeholders can utilize shared data to make evidence-based policy decisions, thereby maximizing the returns on research investment.	
	b. The opportunity to reuse openly shared data allows researchers to broaden their knowledge base, delve deeper into insights, and improve scientific research through reanalysis and result verification.	
	c. Data sharing reduces research costs and prevents duplication of efforts in data collection.	
	d. The likelihood of thorough scrutiny and checks by data reusers creates an incentive to handle data more carefully, ultimately enhancing data integrity and fostering trust in published results.	
	e. Promotes meta-analysis, increases collaboration, and provides opportunities to expand on the work of others.	
	f. Data sharing provides the opportunity for streamlined data management and promotes transparency, quality, and trust among stakeholders.	
Barriers	*Challenges to data sharing*	*Recommended actions*
	a. Inadequate institutional infrastructure.	Public and institutional investment in necessary infrastructure is crucial.
	b. Concerns about losing opportunities for publication arising from datasets generated independently.	Systems of acknowledgment and reward should be in place for citation, stewardship, and data sharing.
	c. Researchers who generate data may not receive appropriate credit or returns on their investment.	
	d. Privacy and confidentiality concerns underscore the importance of ownership structures and data governance.	Provision of well-defined legal frameworks and policies that regulate open data ownership rights and reuse.
	e. The absence of standardization and data governance strategies poses a challenge.	Implementing standardization and data governance, along with licensing strategies, which could be either common or discipline-specific, is important.
	f. Limited data stewardship and the substantial effort required are obstacles to data sharing.	Investing in capacity building, particularly in training related to data science, stewardship, and management, is crucial.

Data governance refers to "the collective set of decision-making processes for the use and value-maximization of an organization's data assets" [52, p.7]. Information Technology (IT) governance has been generally described as those decisions that enhance effective management and use of data, who the decisions are made for, and the management involved in the actual decision-making and implementation [53]. Scholars (e.g., Refs. [54–57]) refer to data governance as the specification of accountability frameworks (policies, guidelines, standards, and practices) and decision rights that shape people's behaviours about the use of data. Data governance creates standards and guidelines for data quality management, defines roles, and allocates responsibilities for decisions [54]. Effective data management and sharing practices require that data should be properly governed using appropriate institutional frameworks.

Although some scholars advocate for a universal data governance approach (model) for all organizations [58], identify different institutional models of data governance, including centralized, replicated, federated, collaborative, and decentralized models. They argue that the choice of a model by an individual organization is contingent on the goals of the organization, the reason for data collection, and what works for the organization. A more centralized data governance model may be attractive for accountability but may delay responses to the needs of data users. A decentralized model, in contrast, would work best where control over data is distributed between several groups [58]. A Federated data governance framework would work well for multiple organizations that work together with common governance policies, standards, and procedures without any individual organization having control over all data and infrastructure. The replicated model works best when different organizations adopt a common data governance model. The collaborative data governance model, which provides an efficient solution to problems of interoperability and should be more effective and flexible in the multi-stakeholder setting, would appear to be appropriate for plant phenotyping data.

Data governance focuses on data quality, management, policies, decision and access rights, accountability (data stewardship), and Metadata [54]. Of all these, producing high-quality data is the most challenging; it requires appropriate governance principles [59]. Quality attributes and/or dimensions of a dataset include data format, accessibility, conciseness, completeness, reputation, believability, semantics, timeliness, accuracy, validity, and uniqueness [60,61]. Deficiencies in any of these attributes in a data set could create technical barriers to data sharing. For example, some data sets are published in formats that are not machine-readable and, therefore, could make transformation configuration and processing difficult [50,51].

5. FAIR principles and applications

The FAIR Principles – representing Findable, Accessible, Interoperable, and Reusable – were created by Ref. [7] to enhance the reuse of scholarly data. They identify preconditions and important attributes for data to enhance its propensity for reuse by humans and machines (third parties) and to create value [62,63]. FAIR guiding principles consist of 15 principles divided among the four headings of Findable, Accessible, Interoperable, and Reuse, and have operated in some academic institutions in Europe for a while; since 2016 they have been endorsed by a number of prominent organizations globally, including G7[2] and a range of science funding organizations and national governments.[3] The principles guide data stewards and managers in defining and providing quality criteria that data must meet before being deposited in any digital repository, to enhance reproducibility and reusability.

The principles apply to all digital resources – data, software, protocols, images, repositories, web services, and other resources or products arising from research, which must be associated with a unique and persistent identifier and rich metadata (i.e., contextual and supporting information) to enhance discovery, accessibility, and reuse [64].

5.1. The key question is: how can FAIR data be generated in reality? This requires a strategic framework and best practices

Following the increasing volume and variety of data produced in various disciplines, enhancing research reproducibility,[4] collaboration, and transparency through data sharing for reuse requires significant improvement and strengthening of the capacity to find, access, interoperate, and reuse digital resources, especially data [65]. In principle, the process of achieving FAIR research data starts at the beginning of any project, especially in the data management plan that describes the steps, characteristics, and behaviours that can facilitate the achievement of

---

[2] G7 Expert Group on Open Science (http://.g8.utoronto.ca/science/2017-annex4-open-science.html).

[3] http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464.

[4] Ability of human and machines to find and use digital assets.

FAIRness and how research data can be preserved and reused.

It is important to state that while FAIR principles should be applicable to algorithms, tools, and workflows that help in generating data [7], metadata (i.e., the proper description and information about the data) plays a crucial role in finding, managing, and reusing data. Therefore, making data FAIR requires a systemic change in the way research processes are being designed and outputs disseminated using digital tools. We examine critical ways of achieving the four cardinal FAIR principles as follows:

5.2. Findable

Making data findable implies that potential reusers can locate the data online. To find and/or discover data,[5] a globally unique and persistent identifier - e.g., Digital Object Identifier (DOI), Uniform Resource Locator – (URL), or Archival Resource Key (ARK) must be attached to the data. In addition, data should be described with rich metadata that specifies the unique and persistent identifier (e.g., DOI), cataloged or registered in a searchable resource (e.g., Google or MEDLINE database for health research), and published in a credible and rightful repository. Appropriate metadata provides contextual information about the data. This would make data easily citable, findable, and reuse. Citing data makes it easier to identify and access a data set used to support a research result or claim, thereby facilitating the reanalysis and reproduction of a study.

A data set should be stored and preserved in a secured location (repository) until it serves its usefulness. The repository could be general (e. g., zenodo), subject-specific, institutional, national, or domain-specific. The persistent identifier must be referenced in the research output. The findability principle ensures that data can be found unambiguously when a common search resource is used.

5.3. Accessible

Data should be accessible by both humans and machines from a repository with some conditions/restrictions attached. Conditions or restrictions may manifest in various forms, such as roles, terms of use, permission, and authentication procedures. Furthermore, data access might be limited due to confidentiality or proprietary concerns. Certain data may only be accessible through designated specific restricted-use agreements. Nevertheless, adhering to FAIR principles necessitates transparently documenting such conditions or restrictions within the metadata. Once data is assigned a persistent identifier by the repository, data retrieval has to be done through an open and free standardized communication protocol (e.g., HTTP), which allows for authorization and authentication where necessary. However, metadata should be accessible even when the data or original resource is no longer accessible. In addition, there has to be a form of authentication before accessing data.

5.4. Interoperable

This seems to be the most challenging of the four high-level FAIR guiding principles due to the complexity of some data [66]. Literarily interoperability, according to Ref. [67], is the "ability of two or more systems or components to exchange information and to use the information that has been exchanged". From a research data perspective, it represents the technical capability of different systems (organizations) to exchange data via a common format and protocols with every stakeholder having a common understanding of the data [68]. This defines the ability of data from multiple sources to be integrated (merged) while retaining its meaning. A prime example is the integration of phenotyping and genotyping data.

A major attribute of good quality data is interoperability, which creates knowledge and value, fulfills the objective for which the data will be used, and stimulates collaboration. A key challenge to achieving data interoperability is a lack of common standards and protocols arising from divergent needs, competence, and interpretations by various data users [58,69]. Integrating a data set with other files from diverse sources may not be possible if files are created using proprietary software that is often protected. Such software is prone to restriction, thereby making it difficult to merge data sets from different sources.

Being more of a data management and governance problem and less of technology, achieving data interoperability requires institutional (legal and regulatory) frameworks and metadata practices. The regulatory framework can specify standards for data security and protection, data that can or cannot be shared and integrated, and how data can be shared with other organizations within the ambient of the law – potentially through Memoranda of Understanding (MOU), licensing or sharing agreements. A common understanding of the semantic metadata by all stakeholders involved is very important. Data producers and reusers must understand how data is structured and how the various components relate to each other and to the components of other datasets to facilitate sharing or exchange. Interoperability is best taken into consideration at every stage in the life cycle (value chain) of the data – from collection to reuse.

An interoperability framework is important, especially in interdisciplinary research and data development that crosses disciplines. The use of community-accepted and recognized specification (schemas) standards for metadata, vocabularies, and keywords that precisely define concepts and qualities is key to enhancing data interoperability. Standard metadata schema describes elements (e.g., title, description, date of issue or modification, publisher, license, keywords, etc.) that should accompany a dataset within a domain of application. Data sets and metadata should be prepared using appropriate common, standardized, and recognized open file formats for easy integration. The establishment and adoption of standard controlled vocabularies and classifications at the design stage of data collection, processing, and sharing eliminate ambiguities and strengthen semantic interoperability and exchange of knowledge.

5.5. Reuse

New research often builds on existing studies. Data reuse is contingent on several factors, including its discoverability, accuracy, clarity, accessibility, and availability of appropriate documentation and usage license indicating the type of reuse permitted [70]. These imply that the reuse potential of data is determined by the way (file format, metadata, resolution) it is shared[14].

For easy discovery and reuse, good metadata should: indicate where, when, what, and how the data was collected; be clear for easy interpretation; be well organized; disclose instruments used in collecting data and standards/calibrations used; define the table/row column headings, units of measurement and codes used, especially if they are not provided in the publication; explain how to locate and access the data; provide pointers (e.g. number of columns and rows) to check if the entire data set was imported; author's contact details for clarification; and report strengths, weaknesses, and limitations of the data [71–74]. Information on data provenance (e.g., survey protocol, instruments, experimental process, methodology) is vital for reusability. This also requires a "clear and accessible data usage license" [75, p.20] to enable third parties to understand the terms and allowable types of reuse permitted. Ultimately, this is facilitated if data is housed through suitable repositories.

---

[5] 'Finding' and 'discovering' data represent a prevalent distinction often referenced in research data discussions. Finding data entails searching for literature, repositories, or datasets indexes, employing criteria such as topics, research questions, or geographical regions. Conversely, data discovery involves identifying the precise storage location and processing methods of a dataset, typically facilitated by tools like Digital Object Identifier (DOI) that are associated to the data.

Advancing scientific research and enhancing collaboration among different communities and research disciplines requires the production of quality data with FAIR attributes through efficient data management, stewardship, and synergies in data structures and standards. This could serve as a checklist and/or framework for data managers to create a sustainable data ecosystem that facilitates knowledge and value creation through optimal reuse of data. The desire to achieve FAIR data should ideally start at the early stage of every project, preferably at the proposal stage, and be constantly revisited all through the cycle of the project for possible adjustments. Achieving FAIR data requires a good understanding of the different stages of data collection, publication, uptake (reuse), and impact (e.g., breeding and policy decisions) as well as critical activities within each stage along the data value chain. Identifying the different stages and associated activities at the early stage will guide in allocating resources and defining who will be responsible for each activity. Other important considerations that could form a preliminary checklist for data managers wanting to produce FAIR data include identification of the presence or absence of data-driven policies within the institution; intellectual, ownership, and access rights especially if there are other collaborating institutions; choice of data repository; sharing and reuse.[6]

Data stewardship starts with a data management plan that clearly identifies and outlines data management and documentation strategies that will enhance the propensity of the data to be reused by others to support research findings [76–78]. The challenges related to the implementation of FAIR principles are summarized in Table 2.

Although some challenges outlined in Table 2 can be addressed by applying FAIR principles, the endeavour to achieve FAIR can also present new obstacles of its own. The principles of Accessibility (A), Interoperability (I), and Reusability (R) within FAIR largely depend on the 'Findability' principle. According to Ref. [79], the feasibility of data accessibility, interoperability, and reusability is contingent upon its location. The Digital Object Identifier (DOI), Archival Resource Key (ARK), Persistent Uniform Resource Locators (PURIs), and Identifiers.org are examples of common FAIR identifiers. There may be difficulties when choosing an identification type. Although there is a consensus that identifiers should be globally unique, each form has pros and cons and is supported by Registration Agencies that are domain-centric [79]. For example, the California Digital Library supports the ARK, whereas DataCite supports the DOI (https://datacite.org) [80].

Challenges may also emerge when publishing in appropriate repositories. Digital scholarly communication relies on four primary types of repositories: subject-based repositories (such as RePEc, SSRN), research repositories (like GenBank for genomics data), national repository systems (e.g., Dutch DAREnet, French HAL system), and institutional repositories (such as the Astrophysics Data System, eScholarship Repository). However, certain repositories impose strict policies that pose significant challenges to researchers who publish through them. For instance, while some institutional repositories enforce policies like restricted open access, which reduces the visibility and reusability of research outputs, others impose eligibility restrictions on deposits, raise copyright concerns, and require funding for maintenance. Additionally, subject-based repositories often encounter scope-related issues; certain repositories have limited scopes that fall short of meeting the needs of researchers [81]. Furthermore, certain repositories lack the necessary technologies and collaborative standards for users and academics to monitor and track citation statistics [82].

Gaining access to restricted resources, including data that is not easily accessible, poses a challenge for researchers. While certain data may be restricted due to ethical, legal, or commercial reasons and only made available and accessible through specific requests [83], studies (such as those by Refs. [84–86]) indicate that the quality of such data

---

Table 2
Challenges related to the implementation of the FAIR principles.

FAIR Principle	Description	Challenges related to implementation
Findable	The capacity for reusers (third parties) to find data on the internet, is typically made easier by employing a distinct and enduring identifier linked to the data.	a. Certain data lacks a unique identifier (such as a DOI) that would make it easy to locate. b. Some researchers publish data with inadequate or nonexistent metadata, which complicates the process of finding, citing, and reusing. c. Certain data are not published in the rightful (e.g., discipline-specific) and credible repository. d. Data can occasionally be shared in different file formats and on different platforms, which poses a challenge to finding and reusing it.
Accessible	The ability to retrieve data from a repository typically with the help of an identifier that is linked to it through a standardized communication protocol.	a. Some data is subject to access restrictions, including permission requirements, usage agreements, and authentication procedures, which hinder third-party reuse. It can be difficult to adopt FAIR data and open sharing when access is restricted due to concerns about confidentiality, privacy, or proprietary rights.
Interoperable	The capacity to combine data from multiple sources while maintaining its significance.	a. Absence of a common standard due to differing needs, expertise, and interpretations among diverse data users limit data integration. b. While some data created using proprietary software are often protected, others are created in incompatible formats and vocabularies. These non-standardized data models complicate the merging of FAIR data from multiple sources.
Reuse	The possibility of others reusing a dataset by others.	a. Certain data require a license to permit reuse. Identifying the most appropriate license may be a challenge and limit data reuse. b. Some data are not stored in rightful (i.e. discipline-specific) repositories (databases), thereby limiting access and reuse.

Effective management of any chosen persistent identifier is crucial for enabling Findability within the FAIR data ecosystem [79].

---

when accessed often falls short and may not have the necessary supporting documentation needed for efficient reuse. Moreover, users of research outputs or resources often lack awareness of the existence of restricted resources, such as datasets, that could be used for research purposes. To make restricted resources easier to find, more accessible and reusable, data infrastructure that supports restricted resources must be improved to ensure efficient management and access, training programs for custodians of restricted data focusing on documentation for effective reuse should be implemented, and metadata standards tailored specifically to accommodate restricted data and facilitate easy discovery and understanding should be established [83].

Research findings indicate a significant correlation between metadata and the discoverability of a research output, which pertains to how

easily the content can be located within a database or search engine. While processed by machines for human comprehension, metadata holds equal importance to research data as it establishes a connection between research data and publication, particularly from the FAIR standpoint. Hence, accurate descriptive metadata components, such as persistent identifiers (e.g., ORCHIDs) and their corresponding Digital Object Identifiers (DOIs), play a crucial role in enabling access to the full text of open-access contents.

Some research outputs, especially those of privately funded research, have limited user rights and conditions for reuse. To propel scientific research forward, it is imperative to reevaluate data to ensure its quality and explore its potential for reuse in fresh analyses or new studies. Hence, choosing the right repository that offers user access and sharing rights is crucial for enhancing reusability.

6. Policy implications and conclusions

The increasing volume of data across disciplines has spurred innovative ways of conducting scientific research and underscores the need for efficient data management and investment of resources in requisite data infrastructure to facilitate data sharing. Despite the benefits associated with data sharing, many scholars consistently feel reluctant in making their data openly available for reuse by others due to some concerns.

Efficient research collaboration and data reuse and/or integration to answer new questions is contingent on data sharing and the ability to integrate data from different sources. This can only be made possible if we create digital data one can easily find, access, interoperate and synthesize with other data and can be readily reused for new advancements in research and innovation.

To promote and strengthen FAIR digital resources, there is need for new infrastructure, human capital development (e.g., training of people on data stewardship), and institutional policies that will incentivize data sharing and promote reproducible research. Governments and institutions need to invest in human capital building. To achieve FAIR and open data eco-system, researchers need adequate training in data collection, management (curation and stewardship) and analysis in order to manage data at every stage in the life cycle. Meanwhile, data managers need to have a good understanding of the different stages of data life cycle and how various research activities and policies can be geared towards producing FAIR data. However, the required skills and expertise depend on the domain (discipline). Therefore, identifying specific skill gaps in any domain should be a matter of policy concern.

Policies that promote open science should be adopted at different levels. The ongoing efforts by many journals and funding agencies in making data openly available for reuse is a fruitful strategy that will modernize stewardship. Recognition and incentive systems for data sharing, stewardship, and citation should be created in various institutions and incorporated as part of an assessment for researchers. One of the key challenges in data sharing is the lack of research data infrastructure. Provision of adequate funding by the public sector for appropriate data infrastructure will incentivize private investment. Intellectual Property, copyright, and licensing policies should be strengthened and well-defined to ensure adequate protection and reduce tensions surrounding data ownership between individuals, private institutions, and the public sector. Above all, a multi-pronged and coordinated strategy, including well-defined data governance principles, technical standards, and practices, are essential for FAIR data and in promoting data sharing.

CRediT authorship contribution statement

Albert I. Ugochukwu: Writing – original draft, Methodology, Formal analysis, Data curation. **Peter W.B. Phillips:** Writing – review & editing, Supervision, Conceptualization, Funding acquisition, Project administration.

Declaration of competing interest

The authors declare no conflict of interest in the research, authorship, and publication of this article.

Data availability

No data was used for the research described in the article.

Acknowledgements

References

[1] L. Taylor, The ethics of big data as a public good: which public? whose good? Phil, Trans. R. Soc A 374 (2016) 2060126, https://doi.org/10.1098/rsta.2016.0126.

[2] D.G. Roche, Open data: policies need policing, PMID: 27708293, Nature 538 (7623) (2016) 41, https://doi.org/10.1093/jamia/ocz037, 10.1038/538041c Journal of the American Medical Informatics Association: JAMIA 26(89), 825–834 (2019.

[3] S. Renaut, A.E. Budden, D. Gravel, T. Poisot, P. Peres-Neto, Management, archiving, and sharing for biologists and the role of research institutions in the technology-oriented age, Bioscience 68 (2018) 400–411. https://doi:10.1093/biosci/biy038.

[4] T.H. Vines, R.L. Andrew, D.G. Bock, M.T. Franklin, K.J. Gilbert, N.C. Kane, et al., Mandated data archiving greatly improves access to research data, Faseb. J. 27 (4) (2013) 1304–1308. PMID: 23288929, https://doi:10.1096/fj.12-218164.

[5] F. Ritchie, R. Welpton, Sharing risks, sharing benefits: data as a public good, in: Paper Prepared for Conference of European Statisticians and Presented at the Joint Session on Statistical Data Confidentiality, Tarragona, Spain, 26-28 October, 2011.

[6] P.W.B. Phillips, G.G. Khachatourians, The Biotechnology Revolution in Global Agriculture: Invention, Innovation and Investment in the Canola Sector, 2001, 376P, CABI.

[7] M.D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, B. Mons, The FAIR guiding principles for scientific data management and stewardship, Nature 3 (160018) (2016) 1–9, https://doi.org/10.1038/sdata.2016.18.

[8] J.C. Wallis, E. Rolando, C.L. Borgman, If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology, PLoS One 8 (7) (2013) e67332. https://doi:10.1371/journal.pone.0067332.

[9] L. Awada, P.W.B. Phillips, Governance and stewardship for research data and information sharing: issues and prospective solutions in the Transdisciplinary Plant Phenotyping and Imaging Research Center Network, Plants, People, Planet 4 (2021) 84–95, https://doi.org/10.1002/ppp3.10238.

[10] A.I. Ugochukwu, P.W.B. Phillips, Data sharing in plant phenotyping research: perceptions, practices, enablers, barriers and implications for science policy on data management, Plant Phenome J. 5 (1) (2022) e20056, https://doi.org/10.1002/ppj2.20056.

[11] C.J. Savage, A.J. Vickers, Empirical study of data sharing by authors publishing. In plos journals, PLoS One 4 (9) (2009) e7078, https://doi.org/10.1371/journal.pone.0007078.

[12] I.M. Faniel, T.E. Jacobsen, Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data, Comput. Support. Coop. Work 19 (2010) 355–375, https://doi.org/10.1007/s10606-010-9117-8.

[13] M.B. Jones, M.P. Schildhauer, O.J. Reichman, S. Bowers, The new bioinformatics: integrating ecological data from the gene to the biosphere, Annu. Rev. Ecol. Evol. Syst. 37 (2006) 519–544, https://doi.org/10.1146/annurev.ecolsys.37.091305.110031.

[14] D.G. Roche, M. Granados, C.C. Austin, S. Wilson, G.M. Mitchell, P.A. Smith, S. J. Cooke, J.R. Bennett, Open government data and environmental science: a federal Canadian perspective, FACETS 5 (2020) 942–962. https://doi:10.1139/facets-2020-0008.

[15] B.A. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, EBSC Technical Report 1 (2–3) (2007) 1–65. Available at: https://www.cs.auckland.ac.nz/~norsaremah/2007%20Guidelines%20for%20performing%20SLR%20in%20SE%20v2.3.pdf. (Accessed 7 February 2024).

[16] O.P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, J. Syst. Software 80 (4) (2007) 571–583, https://doi.org/10.1016/j.jss.2006.07.009.

[17] B.A. Kitchenham, O.P. Brereton, A systematic review of systematic review process research in software engineering, J. Inf. Software Technol. 55 (2013) 2049–2075, https://doi.org/10.1016/j.infsof.2013.07.010.

[18] E. Hutchings, M. Loomes, P. Butow, F.M. Boyle, A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data, Syst. Rev. 9 (2020) 240, https://doi.org/10.1186/s13643-020-01485-5.

[19] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097, https://doi.org/10.1136/bmj.b2535, 2009.

[20] T. Hey, S. Tansley, K. Tolle, 287, in: The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington: Microsoft Research, 2009. Available: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf. (Accessed 15 March 2022).

[21] National Academies of Science, in: Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age, 2009. Available: http://www.nap.edu/catalog.php?record_id=12615. (Accessed 14 May 2022).

[22] National Science Foundation Press Release 10-077, Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans (2010). Available: http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928. Accessed on March 15, 2022.

[23] CWTS, Centre for Science and Technology Studies, Report of the Leiden University's Centre for Science and Technology Studies, in: Open Data: the Researcher Perspective, 2017. Available at: https://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf. (Accessed 23 May 2022).

[24] J.C. Molloy, The open knowledge foundation: open data means better science, PLoS Biol. 9 (12) (2011) e1001195, https://doi.org/10.1371/journal.pbio.1001195.

[25] C.L. Borgman, in: Why Are the Attribution and Citation of Scientific Data Important?, 2012. Available: https://escholarship.org/uc/item/65b51130. (Accessed 18 March 2022).

[26] R. Viseur, Open science: practical issues in open research data, in: Proceedings of 4ᵗʰ International Conference on Data Management Technologies and Applications – DATA, 2015, pp. 201–206, 2015, Colmar, Alsace, France.

[27] A. Zuiderwijk, Open Data Infrastructures: the Design of an Infrastructure to Enhance the Coordination of Open Data Use, 2015, https://doi.org/10.4233/uuid:9b9e60bc-1edd-449a-84c6-7485d9bde012 s-Hertogenbosch: Uitgeverij BOX Press.

[28] D.G. Roche, R. Lanfear, S.A. Binning, T. Haff, L.E. Schwanz, K.E. Cain, et al., Troubleshooting public data archiving: suggestions to increase participation, 2014, PLoS Biol. 12 (1) (2014) e1001779, https://doi:10.1371/journal.pbio.1001779.

[29] V. Stodden, Enabling reproducible research: open licensing for scientific innovation, Int. J. Commun. Law Pol. 13 (2009) 1–25. Available at: https://www.stodden.net/papers/ijclp-STODDEN-2009.pdf.

[30] C.L. Borgman, The conundrum of sharing research data, J. Am. Soc. Inf. Sci. Technol. 63 (2012) 1059–1078, https://doi.org/10.1002/asi.22634.

[31] B. Hanson, A. Sugden, B. Alberts, Making data maximally available, Science 331 (6018) (2011) 649, https://doi.org/10.1126/science.1203354.

[32] S.D. Kahn, On the future of genomic data, Science 331 (2011) 728–729, https://doi.org/10.1126/science.1197891.

[33] L. Lyon, Dealing with Data: Roles, Rights, Responsibilities, and Relationships, UKOLN, 2007, p. 65. Available at: https://purehost.bath.ac.uk/ws/portalfiles/portal/419529/dealing_with_data_report-final.pdf. (Accessed 4 August 2023).

[34] O.J. Reichman, M.B. Jones, M.P. Schildhauer, Challenges and opportunities of open data in ecology, Science 331 (1618) (2011) 703–705, https://doi.org/10.1126/science.1197962.

[35] A. Cox, The benefits of data sharing, in: National Academics of Sciences, Engineering, and Medicine 2016. Principles and Obstacles for Sharing Data from Environmental Health Research: Workshop Summary, The National Academies Press, Washington, DC, 2016, https://doi.org/10.17226/21703.

[36] C. Tenopir, S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Frame, Data sharing by scientists: practices and perceptions, PLoS One 6 (6) (2011) e21101, https://doi.org/10.1371/journal.pone.0021101. The 4th International Conference on Data Management Technologies and Applications.

[37] D. Cowan, P. Alencar, F. McGarry, Perspectives on open data: issues and opportunities, in: Technical Report CS-2014-01 Published in: 2014 IEEE International Conference on Software Science, Technology and Engineering, Ramat Gan, Israel, 2014, https://doi.org/10.1109/SWSTE.2014.18.

[38] J. De Beer, Ownership of open data: governance options for agriculture and nutrition (2016), in: (Wallingford: Global Open Data for Agriculture and Nutrition) [Report], 2016. Available at, https://ssrn.com/abstract=3015958. (Accessed 18 April 2022). SSRN.

[39] A. Zuiderwijk, Analysing open data in virtual research environments: new collaboration opportunities to improve policy making, Int. J. Electron. Govern. Res. 13 (4) (2018) 76–92, https://doi.org/10.4018/IJEGR.2017100105.

[40] B. Nelson, Data sharing: empty archives, 2009, Nature 461 (7261) (2009 Sep 10) 160–163, https://doi.org/10.1038/461160a.

[41] Food and Agriculture Organization, Realizing the Potential of Digitalization to Improve the Agri-Food System. Proposing a New International Digital Council for Food and Agriculture, A Concept Note, 2020 (Rome, Italy).

[42] C. Graf, How and Why We're Making Research Data More Open, Wiley Publication, 2018. November 7, Available at: https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-we-re-making-research-data-more-open. (Accessed 3 February 2022).

[43] S. Childs, J. McLeod, E. Lomas, G. Cook, Opening research data: issues and opportunities, Record Manag. J. 24 (2014) 142–162, https://doi.org/10.1108/RMJ-01-2014-0005.

[44] H. Mooney, M.P. Newton, The anatomy of a data citation: discovery, reuse, and credit, J. Librarianship Scholarly Commun. 1 (1) (2012) eP1035. Available at: http://jlsc-pub.org/jlsc/vol1/iss1/6.

[45] D. Blumenthal, M. Gluck, K.S. Louis, M.A. Stoto, D. Wise, University-Industry research relationship in biotechnology: implications for the University, Science 232 (4756) (1986) 1361–1366, https://doi.org/10.1126/science.3715452.

[46] C. Bizer, P. Boncz Peter, M.L. Brodie, O. Erling, The meaningful of big data: four perspectives – four challenges, SIGMOD Record 40 (4) (2011) 56–60, https://doi.org/10.1145/2094114.2094129.

[47] M. Devare, E. Arnaud, E. Antezana, B. King, Governing agricultural data: challenges and recommendations, in: H.F. Willamson, S. Leonelli (Eds.), Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development, 2023, https://doi.org/10.1007/798-3-031-13276-6-11.

[48] A.S. Figueiredo, Data sharing: convert challenges into opportunities, Front. Public Health 5 (2017) 327. https://doi:10.3389/fpubh.2017.00327.

[49] M. Janssen, Y. Charalabidis, A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, Inf. Syst. Manag. 29 (4) (2012) 258–268, https://doi.org/10.1080/10580530.2012.716740.

[50] E. Barry, F. Bannister, Barriers to open data release: A view from the top. Inf. Polity 19, (1,2), 129-152. http://doi.org/10.3233/ip-140327.

[51] A. Zuiderwijk, N. Helbig, J. Gil-Garcia, M. Janssen, Guest editors' introduction. innovation through open data: a review of the state-of-the-art and an emerging research agenda, J. Theoretical Appl. Electronic Commerce Res. 9 (2) (2014), https://doi.org/10.4067/S0718-18762014000200001.

[52] E. Pierce, W.S. Dismute, C.L. Yonke, The state of information and data governance - understanding how organizations govern their information and data assets. Industry Report, International Association for information and data quality, Baltimore, MD, 2016. https://www.dataversity.net/iaidq-publishes-new-report-state-information-quality-data-governance/. (Accessed 7 September 2023).

[53] P. Weill, J.W. Ross, IT governance on one page, SSRN, https://ssrn.com/abstract=664612, November 2004.

[54] K. Weber, B. Otto, H. Österle, One size does not fit all - a contingency approach to data governance, 1, in: ACM Journal of Data and Information Quality, vol. 1, 2009, pp. 1–27, https://doi.org/10.1145/1515693.1515696.

[55] V. Khatri, C.V. Brown, Designing data governance, 1, in: Communications of the ACM, vol. 53, 2010, p. 148, https://doi.org/10.1145/1629175.1629210.

[56] B. Otto, Data governance, 4, in: Business and Information Systems Engineering, vol. 3, 2011, pp. 241–244, https://doi.org/10.1007/s12599-011-0162-8.

[57] P. Brous, M. Janssen, R. Vilminko-Heikkinen, Coordinating decision-making in data management activities: a systematic review of data governance principles, Electron. Govern. 9820 (2016) 115–125, https://doi.org/10.1007/978-3-319-44421-5_9.

[58] L.G. Morales, T. Orrell, Data interoperability: a practitioner's guide to joining up data in the development sector, Available at: https://www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner%E2%80%99s%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf, 2018. (Accessed 28 April 2022).

[59] S. Eckartz, W. Hofman, A. Veenstra, A decision model for data sharing, Sep. 2014. 13th International Conference on Electronic Government (EGOV), 2014, pp. 253–264. Dublin, Ireland, https://doi:10.1007/978-3-662-44426-9_21.

[60] P. Nousak, R. Phelps, A Scorecard Approach to Improving Data Quality, P158–27, PWC Consulting, Chapel Hill, NC, 2002. Available at: https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p158-27.pdf. (Accessed 23 February 2022).

[61] S. Knight, J. Burn, Developing a framework for assessing information quality on the world wide web, Inf. Sci. 8 (2005) 159–172, https://doi.org/10.28945/493.

[62] European Expert Group on FAIR Data, Turning FAIR into reality, Final report and action plan from the European Commission Expert Group on FAIR Data (2018). November, Available at: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. (Accessed 10 February 2022).

[63] A. Lamprecht, L. Garcia, M. Kuzak, et al., Towards FAIR principles for research software, Data Sci. 3 (1) (2020) 37–59, https://doi.org/10.3233/DS-190026.

[64] M. Dumontier, K. Wesley, Advancing discovery science with FAIR data stewardship: findable, accessible, interoperable, reuse, Ser. Libr. 74 (1–4) (2018) 39–48, https://doi.org/10.1080/0361526x.2018.1443651.

[65] The Turing Way Community, The FAIR Principles, 2020. Available at: https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-fair.html. (Accessed 15 March 2022).

[66] R. Allen, D. Hartland, FAIR in Practice – JISC Report on the Findable Accessible Interoperable and Reuseable Data Principles, zenodo, 2018, https://doi.org/10.5281/zenodo.1245564.

[67] IEEE, Standard Glossary of Software Engineering Terminology, vol. 610, IEEE Std, 1990, pp. 12–1990, https://doi.org/10.1109/IEEESTD.1990.101064.

[68] S. Heiler, Semantic interoperability, ACM Comput. Surv. 27 (2) (1995) 271–273, https://doi.org/10.1145/210376.210392.

[69] Joined-up data standards project, the frontiers of data interoperability for sustainable development, Available at: https://www.publishwhatyoufund.org/wp-content/uploads/2017/11/JUDS_Report_Web_061117.pdf, 2016. (Accessed 8 February 2022).

[70] OpenAIRE, FAIR – in depth (n. d). Available at: https://www.openaire.eu/how-to-make-your-data-fair. Accessed on March 23, 2022.

[71] W.K. Michener, J.W. Brunt, J.J. Helly, T.B. Kirchner, S.G. Stafford, Nongeospatial metadata for the ecological sciences, Ecol. Appl. 7 (1997) 330–342, https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2.

[72] A.S. Zimmerman, Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists, The University of Michigan, 2003. PhD dissertation.

[73] A.S. Zimmerman, Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse, Int. J. Digit. Libr. 7 (2007) 5–16, https://doi.org/10.1007/s00799-007-0015-8.

[74] C.A. Strasser, R. Cook, W.K. Michener, A. Budden, Primer on data management: what you always wanted to know. UC Office of the President: California Digital

Library. http://escholarship.org/uc/item/7tf5q7n3, 2012. (Accessed 2 February 2022).

[75] European Expert Group on FAIR Data, Turning FAIR into reality, Final report and action plan from the European Commission Expert Group on FAIR Data (2018). November, Available at: https://ec.europa.eu/info/sites/info/files/turning_fair _into_reality_1.pdf. (Accessed 10 February 2022).

[76] E.T. Borer, E.W. Seabloom, M.B. Jones, M. Schildhauer, Some simple guidelines for effective data management, Bull. Ecol. Soc. Am. 90 (2) (2009) 205–214, https:// doi.org/10.1890/0012-9623-90.2.205.

[77] C.C. Austin, Guidelines for the Deposit and Preservation of Research Data in Canada, Research Data Canada, Ottawa, Ontario, 2015. Available at: https://www. rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Cana da-2015.pdf. (Accessed 19 August 2022).

[78] W.K. Michener, Ten simple rules for creating a good data management plan, PLoS Comput. Biol. 11 (10) (2015) e1004525, https://doi.org/10.1371/journal. pcbi.1004525.

[79] N. Juty, S.M. Wimalaratne, S. Soiland-Reyes, J. Kunze, G.A. Goble, T. Clark, Unique, persistent, resolvable: identifiers as the foundation of FAIR, Data Intelligence 2 (2020) 30–39, https://doi.org/10.1162/dint_a_00025.

[80] J. Kunze, R. Rogers, The ARK identifier scheme, Available at: https://escholarship. org/uc/item/9p9863nc, 2008. (Accessed 17 March 2024).

[81] T. Morrell, The Opportunities and challenges of research data and software for libraries and institutional repositories. Against the Grain, November 2019,

Available at: chrome- extension://efaidnbmnnnibpcajpcglclefindmkaj/https://doc s.lib.purdue.edu/cgi/viewcontent.cgi?article=8419&context=atg. (Accessed 19 March 2024).

[82] C. Armbruster, L. Romary, Comparing repository types: challenges and barriers for subject-based repositories, research repositories, national repositories systems and institutional repositories in serving scholarly communication, Int. J. Digit. Libr. Syst. 1 (4) (2010) 61–73, https://doi.org/10.4018/jdls.2010100104.

[83] K.B. Read, G.A. Gibson, A. Leahey, L. Peterson, S. Rutley, J. Shi, K. Stathis, Understanding the challenges associated with finding and accessing restricted data in Canada: a mixed methods study. https://doi.org/10.31219/osf.io/pa5fx, 2022, August 18.

[84] B. Pongiglione, A. Torbica, H. Blommestein, S. de Groot, O. Ciani, S. Walker, F. Dams, R. Blankart, M. Mollenkamp, S. Kovacs, R. Tarricone, M. Drummond, Do existing real-world data sources generate suitable evidence for the HTA of medical devices in Europe? Mapping and critical appraisal, Int. J. Technol. Assess. Health Care 37 (1) (2021) e62, https://doi.org/10.1017/S0266462321000301.

[85] J.B. Byrd, A.C. Greene, D.V. Prasad, X. Jiang, C.S. Greene, Responsible, practical genomic data sharing that accelerates research, Nat. Rev. Genet. 21 (10) (2020) 615–629, https://doi.org/10.1038/s41576-020-0257-5.

[86] B. Bekemeier, S. Park, U. Backonja, I. Ornelas, A.M. Turner, Data, Capacity-Building, and Training Needs to Address Rural Health Inequalities in the Northwest United States: a Qualitative Study.