International Journal of Digital Earth

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Harmonizing quality measures of FAIRness assessment towards machine-actionable quality information

Ge Peng, Gary Berg-Cross, Mingfang Wu, Robert R. Downs, Sudhir R. Shrestha, Lesley Wyborn, Nancy Ritchey, Hampapuram K. Ramapriyan, S. Jeanette Clark, Jenny Wood, Zhong Liu & Abdelhak Marouane

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2024, VOL. 17, NO. 1, 2390431
https://doi.org/10.1080/17538947.2024.2390431

Harmonizing quality measures of FAIRness assessment towards machine-actionable quality information

Ge Peng [a], Gary Berg-Cross [b], Mingfang Wu [c], Robert R. Downs [d], Sudhir R. Shrestha [e], Lesley Wyborn [f], Nancy Ritchey [g], Hampapuram K. Ramapriyan [h,i], S. Jeanette Clark [j], Jenny Wood [a], Zhong Liu [k] and Abdelhak Marouane [a]

[a]Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, USA; [b]Ontolog Forum Board, Potomac, MD, USA; [c]Australian Research Data Commons, Melbourne, Australia; [d]Center for International Earth Science Information Network (CIESIN), Columbia Climate School, Columbia University, Palisades, NY, USA; [e]Office of Water Prediction (OWP), National Weather Service, National Oceanic and Atmospheric Administration (NOAA), Silver Spring, MD, USA; [f]National Computational Infrastructure, Australian National University, Canberra, Australia; [g]NOAA's National Centers for Environmental Information (NCEI), Asheville, NC, USA; [h]Science Systems and Applications, Inc., Lanham, MD, USA; [i]NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA; [j]National Center for Ecological Analysis and Synthesis (NCEAS), University of California Santa Barbara, Santa Barbara, CA, USA; [k]Center for Spatial Information Science and Systems (CSISS), George Mason University and NASA Goddard Earth Sciences Data (GES) and Information Services Center (DISC) NASA/GSFC, Greenbelt. MD, USA

ABSTRACT

FAIR Principles are a set of high-level guidelines for sharing digital resources. The growing global adoption of the FAIR Principles by policymakers, funders, and organizations compels data professionals, projects, and repositories to demonstrate the level of FAIR-compliance (referred to as FAIRness) of their digital data, metadata, and infrastructures. Because the FAIR Principles offer general objectives rather than specific implementation instructions, discrepancies exist due to different interpretations, domain-specific requirements, and intended applications. These discrepancies hinder direct comparisons and integration of assessment outcomes. To address this issue, we propose a novel framework, including a consolidated FAIR vocabulary. This framework establishes quality measures upfront in FAIRness assessment workflows to surpass the intricacies arising from the aforementioned dependencies. The established quality measures encapsulate the distinctive core concepts inherent in individual FAIR principles and can serve as common, fundamental pillars of holistic FAIRness assessment workflows. Building upon this fundamental set of the quality measures, we introduce a FAIRness quality maturity matrix (FAIR-QMM) as a structured, tiered, and progressive approach for evaluating and reporting the degree of FAIR-compliance. The FAIR-QMM can be used as a FAIRness assessment tool independently and/or as a translator between other FAIRness assessment tools or models.

ARTICLE HISTORY
Received 22 January 2024
Accepted 5 August 2024

KEYWORDS
FAIR; quality measure; FAIRness assessment; scientific data; FAIRness quality maturity matrix

CONTACT Ge Peng gpeng93@gmail.com

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17538947.2024.2390431
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

1. Introduction

The FAIR Principles lay out a set of high-level data management and stewardship behaviors that digital data objects should exhibit to be findable, accessible, interoperable, and reusable (i.e. FAIR; Wilkinson et al. 2016, 2022). FAIR aims to support open science by optimizing data sharing in a machine-friendly environment across systems, disciplines, and regional boundaries for both human and machine end-users (Mons et al. 2017). The FAIR Principles were intentionally designed to be domain- and technology-agnostic (Mons et al. 2017; Wilkinson et al. 2022). Moreover, Wittenburg and Strawn (2018, 11) emphasized: 'experts widely agree that the FAIR principles express policy goals, not blueprints for data infrastructure building.'

With rapidly increased acceptance and adoption of the FAIR Principles by policymakers, funders, and organizations worldwide for promoting data sharing and enabling open science, the requirements have increasingly arisen for data professionals, publicly funded projects, and repositories to demonstrate the FAIR-compliance, i.e. FAIRness, of their digital data objects and supporting infrastructures. To meet these requirements, various organizations have attempted to operationalize ways of testing compliance with the FAIR Principles (e.g. Devaraju et al. 2022; Wilkinson et al. 2019).

The characteristics of specific communities for which data products were created significantly influence the framing of some FAIR principles such as data accessibility and data usage license requirements. (In this paper, the term 'FAIR principles', with 'principles' in lowercase, is used to refer to individual principles, e.g. F1, F2, etc.) Additionally, these characteristics provide context for standards related to descriptive metadata, communication protocols, vocabularies, knowledge representation languages, and more (Berg-Cross and Arbor 2022; Bernasconi et al. 2023). Consequently, these implementations and practices have led to the emergence of a range of FAIRness assessment models and indicators. While beneficial in implementing the FAIR Principles and/or assessing FAIRness, large disparities exist among those assessment models and metrics due to different subjective interpretations of the FAIR Principles, domain-dependent specifications, and the targeted applications (e.g. Bahim et al. 2020; Candela, Mangione, and Pavone 2024; Jacobsen et al. 2020; Peters-von Gehlen et al. 2022; Wittenburg and Strawn 2018).

Unique application priorities can lead to assessment models that address only a subset of the FAIR Principles or implementation from a particular perspective. While examining compliance with a few specific FAIR principles may be important for a particular use case, such examination lacks a holistic view of FAIR-compliance. In addition, the lack of consistency among different approaches poses several challenges. Firstly, it makes comparing assessment results across tools and systems difficult (Peng 2023a). Secondly, it hinders the implementation of machine readability and interpretability of data and associated information, which requires far more rigorous specifications and standardization than is needed for human understandability of data. Peng (2023a), therefore, called for a common set of harmonized FAIRness indicators that adhere to the original definitions of the FAIR principles. A method was introduced in Peng (2023a) for systematically breaking down definitions into key categories, category-specific requirements, core concepts, and focus elements. The method adopted the basic structure of the subject-predicate-object part of the principles, separating the subject data, metadata, and infrastructure entities. The resulting complete mappings of those entities are captured in Peng (2023b).

In this paper, we enhance the formalization of the mapping processes and present a novel framework that includes the establishment of FAIRness quality measures and a common vocabulary. This framework helps rise above the complexity that stems from subjective interpretations and the domain and application dependencies. The established quality measures represent the unique core concepts captured in individual FAIR principles and can serve as common, fundamental pillars of holistic FAIRness assessment workflows.

It has been recognized that the degree to which digital data adhere to the FAIR Principles can vary along a continuum, and the process of making data FAIR involves incremental steps and continuous improvement (e.g. Mons et al. 2017). Built on the established quality measures, a FAIR-compliance maturity matrix is constructed. The matrix provides a structured, tiered, and progressive approach to evaluating and reporting the FAIRness of individual digital data objects and associated infrastructure. Improved structuring of FAIRness maturity will also assist in moving towards machine-actionable quality information.

The remainder of this paper is organized as follows. The next section provides an overview of existing models, indicators, and metrics for assessing FAIRness. Section 3 provides description of the characteristics of the FAIR Principles and methodologies employed to identify the core concepts associated with these principles. Subsequently, we introduce a framework for FAIRness quality measures in Section 4. This framework focuses specifically on the core concepts of the FAIR principles. We then establish a comprehensive set of quality measures and craft a maturity matrix framed around these FAIRness quality measures. This matrix offers a systematic and incremental approach to evaluating and documenting FAIRness. In Section 5, we examine various issues related to FAIRness assessment, such as variations in implementing specific quality measures within particular domains and applications. Finally, in Section 6, we present our conclusions and discuss potential future activities.

2. Review of existing FAIR implementations, assessment models, and metrics

There are numerous existing efforts to develop models, indicators, and metrics for implementing the FAIR Principles and assessing the FAIRness of digital data. Without being exhaustive, outcomes from some of those efforts are highlighted in this section.

(1) The Research Data Alliance (RDA) FAIR Data Maturity Model Working Group (2020) defined 41 FAIR data maturity indicators (RDA FAIR-DMIs) by applying 15 FAIR principles to data and metadata separately with a consideration for modifiers. For example, four indicators are defined for F1 – *(Meta)data are assigned a globally unique and persistent identifier*:
- Metadata is identified by a persistent identifier,
- Metadata is identified by a globally unique identifier,
- Data is identified by a persistent identifier, and
- Data is identified by a globally unique identifier.

One might also note the subtle differences between the verb in the original F1 definition (i.e. assigned) and that in the indicators (i.e. identified). The RDA FAIR-DMIs are framed around verbs and the subjects of the indicators are always either data or metadata, for example, A1.2 – *the protocol allows for an authentication and authorization procedure, where necessary*, is represented by 'Data is accessible through an access protocol that supports authentication and authorisation' (RDA FAIR Data Maturity Model Working Group 2020, 11). Each indicator is assigned with three levels of priorities: Essential, Important, and Useful. Among the 41 FAIR-DMIs, 20 are considered as Essential, 14 as Important, and 7 as Useful. Moreover, each indicator is evaluated against five levels of implementation: 0 (not applicable), 1 (not being considered this year), 2 (under consideration or in planning stage), 3 (in implementation phase), and 4 (fully implemented). This RDA FAIR data maturity model can be used by data providers and publishers for a self-assessment that identifies where to concentrate efforts to make data and metadata more FAIR, including explicit consideration for human users.

(2) FAIR Implementation Profiles (FIPs) developed by GO-FAIR (Schultes et al. 2020) are based on the two principal concepts: FAIR Implementation Community (a self-identified organization that seeks to create and share FAIR data and services) and FAIR-Enabling Digital Resources (FEDR), such as datasets, metadata, code, protocols, data policies, identifier mechanisms, and standards. Each FEDR is a digital object, which in itself requires compliance with the FAIR Principles, including the assignment of a Globally Unique Persistent Resolvable

Identifier (GUPRI) (Schultes et al. 2022; Wittenburg et al. 2019). FIP Questionnaires are designed with 21 questions, each of which can have multiple associated FAIR Enabling Resource (FER) types. They also separate data and metadata entities, while still retaining infrastructure-related entities. The last FAIR principle, i.e. R1.3, is not explicitly addressed.

For example, two questions are included on F1:

- What globally unique, persistent, resolvable identifiers do you use for metadata records?
- What globally unique, persistent, resolvable identifiers do you use for datasets?

With another two on A1.2:

- Which authentication & authorization technique do you use for metadata records?
- Which authentication & authorization technique do you use for datasets?

Each community can publish an FIP at the granularity of a dataset, collection or repository. The FIP explicitly declares the choices made by a given community on the specific resource types used for each of the FAIR principles, such as considering Digital Object Identifier (DOI) as an identifier type in response to the questions on F1. Each FIP itself can be version-controlled, time-stamped, and published with a DOI that can also be linked to the dataset/data collection/repository (Wyborn et al. 2023). As more communities publish these questionnaires, a convergence emerges. This convergence leads to aggregated insights, revealing the most commonly adopted resource types by communities or identifying challenges they encounter in their development endeavors (Schultes et al. 2020, 2022). FIPs (and their related FERs) allow at least two types of comparison between individual datasets. Firstly, it is possible to use knowledge graphs to map across the communities that are using specific FERs to enhance machine-to-machine interoperability. Where FERs are not the same, it is possible to develop crosswalks, e.g. using the Simple Standard for Sharing Ontological Mappings (SSSOM) of Matentzoglu et al. (2022). Secondly, each FIP can be used to assess the FAIRness of a particular dataset and determine whether a higher level of maturity is required for any of the FAIR principles for more effective machine-to-machine interoperability.

(3) Providing a framework for evaluating implementations, Schultes (2023) defined 5 layers of FAIR implementations. The framework separated the FAIR Principles into two classes: information technology-related and data content-related; and in addition, grouped the FAIR implementation activities into two distinct phases: FAIRification for data to be FAIR-ready and FAIR Orchestration to allow data integration and sharing (Schultes 2023). Their framework in some way denotes the readiness for data sharing, integration and analysis. However, that framework is not specified at the level of individual FAIR principles.
(4) The FAIR Metrics established by FAIRsFAIR defined 14 core and universal metrics (Wilkinson et al. 2018) and subsequently developed 15 maturity indicators (MIs) with 22 associated compliance tests for all FAIR principles except R1.2 and R1.3 (Wilkinson et al. 2019). The information on implementation maturity levels is conveyed in MIs as either 'loose' or 'strict' for two specific cases: the use of knowledge representation language and the use of FAIR vocabulary.
(5) The self-evaluation FAIR assessment tool from the Australian Research Data Commons (ARDC) contains 12 assessment questions with a dropdown menu (ARDC 2022). The options include progressive criteria for attaining greater maturity for most, if not all, FAIR principles.
(6) The DataONE suite of FAIR checks consists of 52 checks spread across all the dimensions of FAIR (Jones, Slaughter, and Habermann 2019). Checks were developed by community consensus at the Earth Science Information Partners (ESIP) 2019 summer meeting session. The checks have been implemented programmatically across multiple metadata dialects including

DataCite, Dryad, EML, and ISO-19115/19115-2, and have been deployed to assess the metadata of all datasets in DataONE (Slaughter et al. 2023). Each check is classified as either required, optional, or informational, and returns either a pass or fail depending on metadata content. An overall FAIR metric can be calculated to evaluate FAIRness along a continuum. This approach rewards passing both required and optional checks, while only penalizing failures in required checks.

(7) The FAIR rubric of the United States Geological Survey (USGS; Hutchison et al. 2024) consists of 62 questions with options of 'Yes', 'No', and 'Not Applicable'. Questions are separated for data and metadata. Some of the questions imply a progressive level of maturity. For example, questions for data identifier are:
- Is an identifier assigned for the data release and documented in the data release's metadata record?
- Is the assigned identifier persistent?
- Is the assigned identifier unique (i.e. has a unique value)?

The USGS questions are categorized into three levels: essential, intermediate, and advanced, based on their importance for the FAIR principles. However, some questions are beyond the original definitions of the FAIR principles and include implementation practices that are tailored specifically to the USGS.

(8) Adhering to the original definitions of the FAIR principles, Peng (2023a, 2023b) decomposed the principles into a set of core indicators using the basic structure of the subject-predicate-object part of the principles, separating the subject data, metadata, and infrastructure entities. A set of supplemental indicators is constructed with objective nouns, modifiers and characteristics to provide a comprehensive set of harmonized indicators (Table 1; from Peng 2023b). For example, for *F1. (meta)data are assigned a globally unique and eternally persistent identifier,* the subject-predicate-object part of the principle is '*(meta)data – are assigned – identifier*' and '*identifier*' is the core concept of F1. '*Globally unique*' and '*eternally persistent*' are modifiers for '*identifier*', referred to as focus elements. The resultant indicators for F1 are:
- F1-01D: Data are assigned an identifier,
- F1-01M: Metadata are assigned an identifier,
- F1-02D: Data identifier is globally unique,
- F2-02M: Metadata identifier is globally unique,
- F3-03D: Data identifier is eternally persistent,
- F3-03M: Metadata identifier is eternally persistent.

Where F1-01D and F1-01M are core indicators, formulated with a category and a core concept. The other four indicators are supplemental indicators, constructed with a category-specific core concept and a focus element.

Building upon these existing efforts, we have developed a more systematic way that abides by the definitions of the FAIR principles to promote data sharing in machine-to-machine environments. The methodology for it is described in the following section.

3. Methodology

As a part of our methodology, we first identify key entities contained in the FAIR principles, then utilize a concept-mapping approach for our analysis while drawing upon the design principles of Semantic Web and Web Ontology frameworks, such as the Resource Description Framework (RDF) developed by the World Wide Web Consortium (W3C). Additionally, we incorporate concepts from the FAIR ontology specification created by Kuhn and Dumontier (n.d.). We

Table 1. List of harmonized indicators derived from FAIR principles.

Harmonized Indicators (HI) For the FAIR Guiding Principles			
Principle ID	**HI Type**	**HI ID**	**Description**
F1	**Core**	**F1-01D**	**Data are assigned an identifier**
	Core	**F1-01M**	**Metadata are assigned an identifier**
	Modifier	F1-02D	Data identifier is globally unique
	Modifier	F1-02M	Metadata identifier is globally unique
	Modifier	F1-03D	Data identifier is eternally persistent
	Modifier	F1-03M	Metadata identifier is eternally persistent
F2	**Core**	**F2-01D**	**data are described with rich metadata**
	Modifier	F2-01M	Metadata include data usage licence
	Modifier	F2-02M	Metadata include data provenance
	Extension	F2-03M	Metadata include attributes for discovery
	Extension	F2-04M	Metadata include attributes for data retrieval
	Extension	F2-05M	Metadata include attributes for data access
F3	**Core**	**F3-01M**	**Metadata include the identifier of the data it describes**
	Modifier	F3-02M	Metadata clearly include the data identifier
	Modifier	F3-03M	Metadata explicitly include the data identifier
F4	**Core**	**F4-01D**	**Data are registered or indexed in a resource**
	Core	**F4-01M**	**Metadata are registered or indexed in a resource**
	Modifier	F4-02IS	Resource for data is searchable
	Modifier	F4-03IS	Resource for metadata is searchable
A1	**Core**	**A1-01D**	**Data identifier resolves to the digital data object**
	Core	**A1-01M**	**Metadata identifier resolves to the digital metadata object**
	Modifier	A1-01IS	The communication protocol for data retrieval is standardised
	Modifier	A1-02IS	The communication protocol for metadata retrieval is standardised
A1.1	Modifier	A1.1-01IS	The protocol is open
	Modifier	A1.1-02IS	The protocol is free
	Modifier	A1.1-03IS	The protocol is universally implementable
A1.2	**Core**	**A1.2-01IS**	**The protocol allows for procedure, where necessary**
	Modifier	A1.2-02IS	The procedure includes that for authentication
	Modifier	A1.2-03IS	The procedure includes that for authorization
A2	**Core**	**A2-01M**	**Metadata are accessible permanently**
I1	**Core**	**I1-01D**	**Data use language for knowledge representation**
	Core	**I1-01M**	**Metadata use language for knowledge representation**
	Modifier	I1-02D	The data language is formal
	Modifier	I1-02M	The metadata language is formal
	Modifier	I1-03D	The data language is accessible
	Modifier	I1-03M	The metadata language is accessible
	Modifier	I1-04D	The data language is shared
	Modifier	I1-04M	The metadata language is shared
	Modifier	I1-05D	The data language is broadly applicable
	Modifier	I1-05M	The metadata language is broadly applicable
I2	**Core**	**I2-01D**	**Data use vocabulary**
	Core	**I2-01M**	**Metadata use vocabulary**
	Modifier	I2-02D	The data vocabulary is findable
	Modifier	I2-02M	The metadata vocabulary is findable
	Modifier	I2-03D	The data vocabulary is accessible
	Modifier	I2-03M	The metadata vocabulary is accessible
	Modifier	I2-04D	The data vocabulary is interoperable
	Modifier	I2-04M	The metadata vocabulary is interoperable
	Modifier	I2-05D	The data vocabulary is reusable
	Modifier	I2-05M	The metadata vocabulary is reusable
I3	**Core**	**I3-01D**	**Data include references**
	Core	**I3-01M**	**Metadata include references**
	Modifier	I3-02D	The data references are qualified
	Modifier	I3-02M	The metadata references are qualified
	Modifier	I3-03D	References included in data are to other data
	Modifier	I3-03M	References included in metadata are to other data
	Modifier	I3-04D	References included in data are to other metadata
	Modifier	I3-04M	References included in metadata are to other metadata
R1	**Core**	**R1-01D**	**Data are described with attributes**
	Core	**R1-01M**	**Metadata are described with attributes**
	Modifier	R1-02D	Data attributes are plural
	Modifier	R1-02M	Metadata attributes are plural
	Modifier	R1-03D	Data attributes are accurate
	Modifier	R1-03M	Metadata attributes are accurate
	Modifier	R1-04D	Data attributes are relevant
	Modifier	R1-04M	Metadata attributes are relevant
	Extension	R1-05M	Metadata include attributes for data use
	Extension	R1-06M	Metadata include attributes for data contextual information
	Extension	R1-07M	Metadata include attributes for data processing information
R1.1	**Core**	**R1.1-01D**	**Data are released with a data usage licence**
	Core	**R1.1-01M**	**Metadata are released with a data usage licence**
	Modifier	R1.1-02D	Data licence is clear
	Modifier	R1.1-02M	Metadata licence is clear
	Modifier	R1.1-03D	Data licence is accessible
	Modifier	R1.1-03M	Metadata licence is accessible
R1.2	**Core**	**R1.2-01D**	**Data are associated with provenance**
	Core	**R1.2-01M**	**Metadata are associated with provenance**
	Modifier	R1.2-02D	Data provenance is detailed
	Modifier	R1.2-02M	Metadata provenance is detailed
R1.3	**Core**	**R1.3-01D**	**Data meet standards**
	Core	**R1.3-01M**	**Metadata meet standards**
	Modifier	R1.3-02D	Data standards are domain-relevant
	Modifier	R1.3-02M	Metadata standards are domain-relevant
	Modifier	R1.3-03D	Data standards are community standards
	Modifier	R1.3-03M	Metadata standards are community standards

Notes: The naming convention for indicators is: {FAIR-ID}–0{n}{Category ID}, where {FAIR-ID}consists of [F1, F2, F3, F4; A1, A1.1, A1.2, A2; I1, I2, I3; R1, R1.1, R1.2, R1.3], denoting an individual FAIR principle; {Category ID} consists of [D, M, IS], denoting the Data, Metadata, or Infrastructure category, respectively. {n} can take on any value from the set [1, 2, … , 7], denoting the nth indicator associated with the given principle and category, respectively. The core indicators are denoted in bold – they are fundamental indicators formulated with only a category and a core concept, while those in gray are extended indicators defining additional category-specific focus elements that are not explicitly defined in the original FAIR Principles. The rest are supplemental indicators which further modify the related core concepts. From Peng (2023b).

aim to establish a common vocabulary for systematically and consistently describing the multi-dimensional, multi-layered, and multi-faceted nature of the FAIR principles. The goal is to provide a structural framework without delving into axiom scheme, structural specifications, and annotations, which may be addressed at a later stage when implementing the FAIRness quality measures and assessment specifications.

3.1. Identification of key entities in the FAIR principles

Wilkinson et al. (2016) designed high-level guiding principles under four nonfunctional requirements – Findability, Accessibility, Interoperability, and Reusability. In this paper, we subsequently refer to these four nonfunctional requirements as four FAIR quality dimensions, following Peng (2023a).

Wilkinson et al. (2016) also developed 15 FAIR principles to establish specific functional requirements for digital objects along the four FAIR quality dimensions in a computational environment. These 15 FAIR principles cover three unique high-level entities, or categories: data, metadata, and infrastructure. The principles also entail essential properties or specific functional requirements for each of the three categories. For example, the F1 principle requires that data and metadata should each be assigned a property: *identifier*; the identifier is further qualified to be globally unique and eternally persistent. After mapping the 15 principles onto the three categories, 28 category-specific requirements are identified for data, metadata, and infrastructure (Table 2; adapted from Peng 2023b).

In contrast to its simple acronym, the FAIR Principles are highly complex even before accounting for domain and application dependencies. The FAIR principles are inherently multi-dimensional. They are also multi-layered and multi-faceted, revolving around core concepts with multiple elements, as depicted in Figure 1 and outlined in Figure 2. The approaches for identifying core concepts are described next.

3.2. Conceptual mapping of the FAIR principles

Concept mapping is a visual and systematic technique to organize and graphically represent knowledge, ideas, concepts, or relationships between concepts. It is well-established and widely used to clarify complex topics and facilitate understanding. This approach is utilized in this study to identify the core concepts associated with the definitions of the FAIR principles.

Our concept mapping process starts with each category entity (denoted by ovals in Figure 1; based on Peng 2023b). The definitions of individual principles are first decomposed into the subject-predicate-object structure, similar to the design principle of RDF triplets that are commonly utilized in semantic web and web ontology frameworks. In this structure, the ‘subject’ is a category entity (e.g. Data), and ‘object’ is an objective noun (e.g. Identifier for the F1 principle). The ‘predicate’ denotes the connection (relationship) between the subject and the object (e.g. are assigned). This process yields a total of 21 category-specific core concepts associated with the objective nouns (denoted by rounded rectangles in Figure 1), excluding duplicates such as ‘Data Identifier’ and ‘Metadata Identifier’ from the A1 principle for the Data and Metadata categories, respectively. At the end of this process, the following 12 unique core concepts emerge: Identifier, Rich Metadata, (Searchable) Resource, (Communication) Protocol, Language, Vocabulary, Reference, Procedure, Attribute, License, Provenance, and Standard, which together are referred to as FAIRness quality measures. Parentheses are used to specify the scope of the underpinning concept. While ‘Rich’ is also a modifier, ‘Rich Metadata’ is treated as a single term to distinguish it from ‘Metadata’ and focuses on measuring metadata richness.

Utilizing the same concept mapping approach, Peng (2023b) also derived 48 additional elements consisting of modifiers and characteristics to the core concepts, which are utilized for a set of harmonized indicators for the FAIR Principles (recaptured in Table 1).

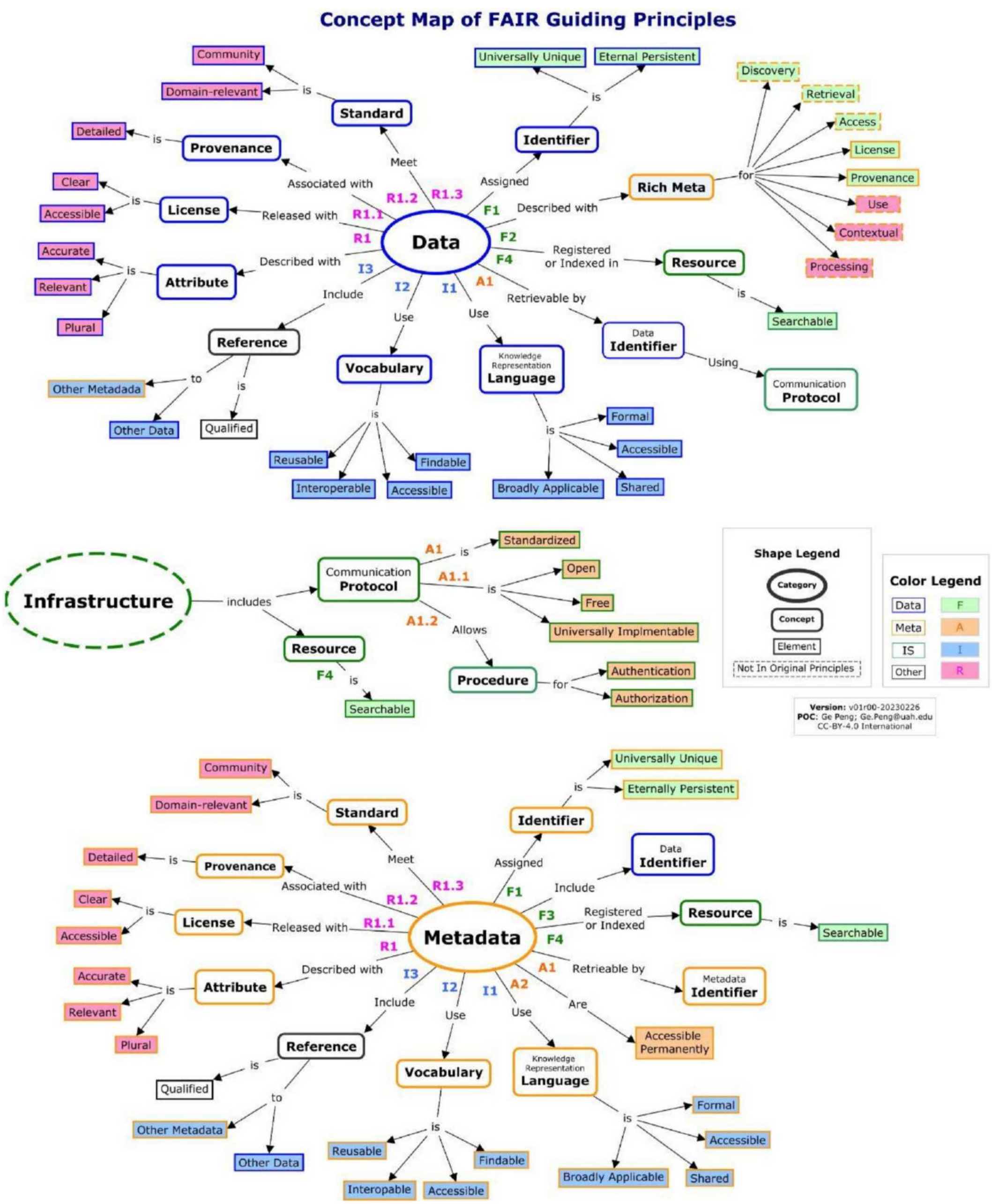

Figure 1. A conceptual map illustrates the key categories and associated concepts and elements that have been mapped from the definitions of the FAIR principles using a concept mapping approach (explained in Section 3). {F, A, I, R}n.m denotes the association to the individual principle, e.g. R1.3. Shape and color legends are on the diagram. Created using Cmap (IHMC n.d.). Source: Peng et al. (2023). Version: v01r00-20230226.

3.3. Semantic ontology perspective of FAIR principles

Semantic models and ontologies have been used to organize the identified concepts and to identify relation between concepts. The approach is to leverage well-defined models and community standards to help understand details of, and supporting relationships between, some of the core concepts. This understanding contributes incrementally to the development of a formal, consistent,

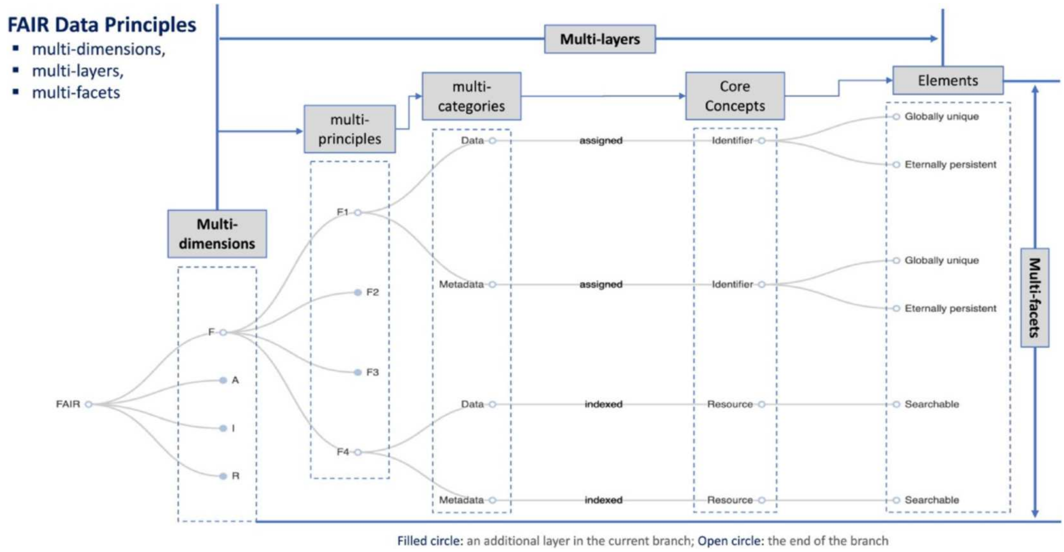

Figure 2. A schematic diagram shows the complexity of the FAIR Principles in terms of multi-dimensions, multi-layers, and multi-facets. A filled circle indicates an additional layer in the current branch while an open circle signifies the end of the branch.

and well-defined terminology, aiding in the comprehension of the underlying entities and their relationships (Berg-Cross and Arbor 2022).

While a full ontology implementation of the FAIRness quality measures is beyond the scope of this paper, here we provide an example of viewing the FAIR principles from a semantic ontology perspective.

In semantic web and web ontology frameworks, an RDF triple in the form of a subject–predicate–object expression is well-used to represent a semantic relationship. We adapt this RDF triplet approach to analyze the FAIR principles. For example, the *F1 principle*: '*meta(data) are assigned a globally unique and eternally persistent identifier*', indicates a basic semantic triplet relation: metadata or data is the subject, identifier is the object, and 'are assigned' specifies a relationship between the subject and the object. The identifier for 'meta(data)' is further assigned the qualities of being globally unique and eternally persistent, similar to a subtype of 'hasCharacteristic' from the Information Artifact Ontology (IAO; Ruttenberg et al. 2022). Therefore, the F1 principle simply ensures that a globally unique and persistent identifier be used to identify a data element, while also ensuring an associated link can be resolved, as specified by the A1 principle.

In our derived FAIRness quality measure framework, subjects or objects are categories and core concepts, while the semantic relations are executable actions for making connections between the two.

The outcomes of the simplified semantic and ontology analysis and the systematic concept mapping of the FAIR principles are a set of formally defined FAIRness quality measures, as described next in Section 4. These measures establish a strong foundational framework for a holistic FAIRness assessment workflow, enabling machine-actionable reporting of FAIRness assessment and outcomes, which is presented next.

4. FAIRness quality measure framework

4.1. Common vocabulary for the FAIR principles

The concept mapping and the semantic approaches as described in Section 3 are applied to the definitions of individual FAIR principles to identify and organize core concepts. The visual depiction of the results is captured in Figure 3 and outlined below.

We categorize the top level of the FAIR Principles as Class (the most inner circle, dark-shaded in Figure 3), following Kuhn and Dumontier n.d.:

Table 2. Category-specific requirements mapped from FAIR principles.

FAIR ID	REQ ID	Description
F1		(meta)data are assigned a globally unique and eternally persistent identifier
	F1-REQ-D	data are assigned a globally unique and eternally persistent identifier
	F1-REQ-M	metadata are assigned a globally unique and eternally persistent identifier
F2		data are described with rich metadata (defined by R1 below)
	F2-REQ-D	data are described with rich metadata
	F2-REQ-M	rich metadata are defined by R1 below
F3		metadata clearly and explicitly include the identifier of the data it describes
	F3-REQ-M	metadata clearly and explicitly include the PID-D
F4		(meta)data are registered or indexed in a searchable resource
	F4-REQ-D	data are registered or indexed
	F4-REQ-M	metadata are registered or indexed
	F4-REQ-IS	The registering or indexing resource is searchable
A1		(meta)data are retrievable by their identifier using a standardized communications protocol
	A1-REQ-D	data are retrievable by their identifier
	A1-REQ-M	metadata are retrievable by their identifier
	A1-REQ-IS	The retrieving communications protocol is standardized
A1.1		the protocol is open, free, and universally implementable
	A1.1-REQ-IS	the protocol is open, free, and universally implementable
A1.2		the protocol allows for an authentication and authorization procedure, where necessary
	A1.2-REQ-IS	the protocol allows for an authentication and authorization procedure, where necessary
A2		metadata are accessible, even when the data are no longer available
	A2-REQ-M	metadata are accessible permanently
I1		(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
	I1-REQ-D	data use a formal, accessible, shared, and broadly applicable language for knowledge representation
	I1-REQ-M	metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2		(meta)data use vocabularies that follow FAIR principles
	I2-REQ-D	data use vocabularies that follow FAIR principles
	I2-REQ-M	metadata use vocabularies that follow FAIR principles
I3		(meta)data include qualified references to other (meta)data
	I3-REQ-D	data include qualified references to other (meta)data
	I3-REQ-M	metadata include qualified references to other (meta)data
R1		(meta)data are richly described with a plurality of accurate and relevant attributes
	R1-REQ-D	data are richly described with a plurality of accurate and relevant attributes
	R1-REQ-M	metadata 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
	R1.1-REQ-D	data are released with a clear and accessible data usage licence
	R1.1-REQ-M	metadata are released with a clear and accessible data usage licence
R1.2		(meta)data are associated with detailed provenance
	R1.2-REQ-D	data are associated with detailed provenance
	R1.2-REQ-M	metadata are associated with detailed provenance
R1.3		(meta)data meet domain-relevant community standards
	R1.3-REQ-D	data meet domain-relevant community standards
	R1.3-REQ-M	metadata meet domain-relevant community standards

Notes: Descriptions of individual FAIR principles are in blue. The naming convention for requirement identifiers (REQ IDs) is: {FAIR ID}-REQ-{Category ID}, where, {FAIR ID} = [F1, F2, F3, F4; A1, A1.2, A1.2, A2; I1, I2, I3; R1, R1.1, R1.2, R1.3], denoting individual FAIR principles, respectively; {Category ID} = [D; M; IS], denoting the Data, Metadata, and Infrastructure category, respectively. Based on Peng (2023b): F3-REQ-D has been removed as deemed to be redundant with F3-REQ-M.

Class:

- FAIR Principles

The FAIR Principles encompass four quality dimensions by design:

Dimensions:

- Findability (F),
- Accessibility (A),

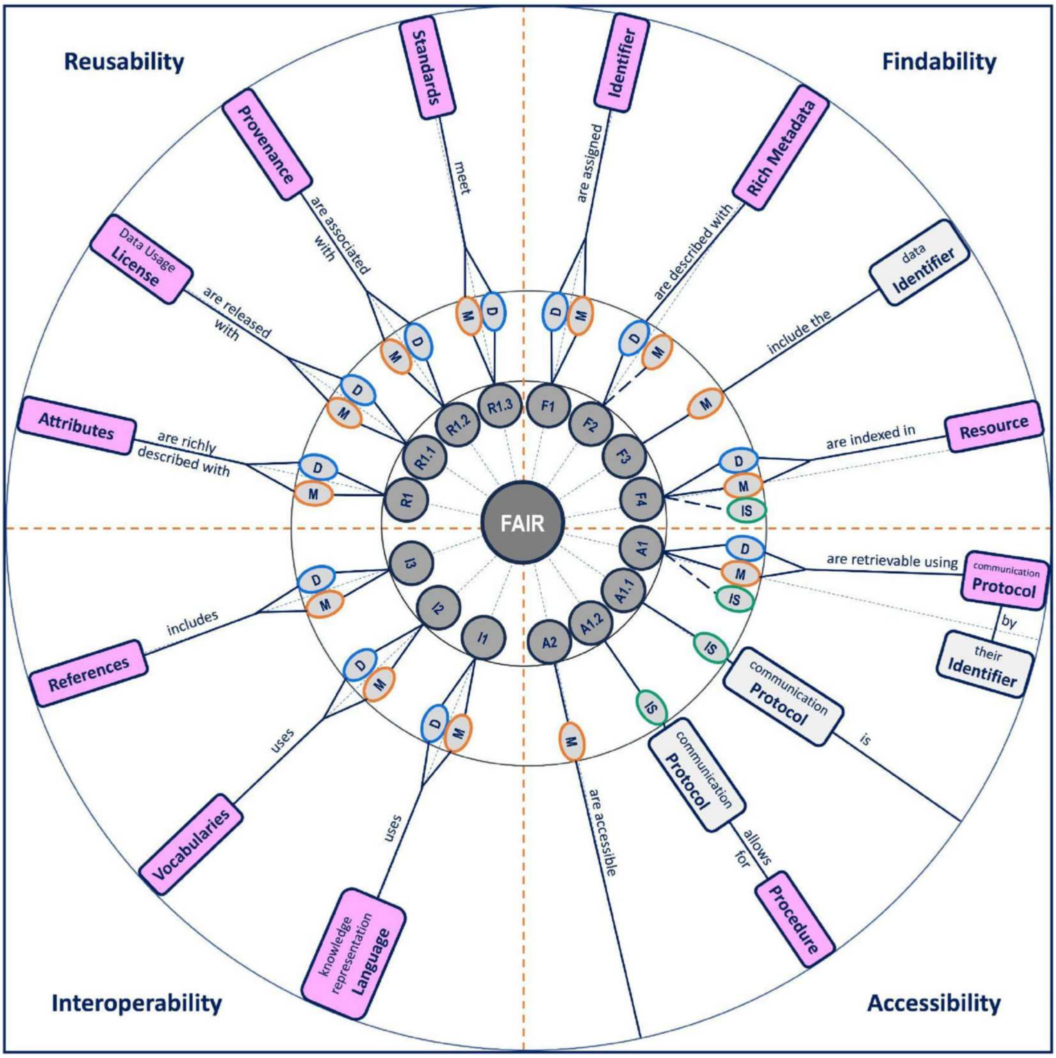

Figure 3. Visual depiction of the FAIR dimensions represented by each quadrant; class and its subclasses represented by dark-shaded circles; associated categories in light-shaded ovals; and identified core concepts in rounded rectangles. Color-coded oval outlines denote three key categories: Blue – Data (D); Orange – Metadata (M); and Green – Infrastructure (IS), respectively. Dashed lines extending from FAIR subclass entities (F2, F4, A1) to the respective categories (M, IS, IS) indicate the category's presence without it being the subject of the sub-principle. Pink-filled rounded rectangles are used to highlight the unique core concepts – to be established as FAIRness quality measures. Additional elements or modifiers are not included in the diagram for simplicity.

- Interoperability (I),
- Reusability (R)

The FAIR Principles are composed of 15 sub-principles (second inner circle, dark-shaded circles in Figure 3), representing as Sub-Classes; adopted from Kuhn and Dumontier (n.d.):

Sub-Classes:

- FAIR Sub-Principles: {F1, F2, F3, F4, A1, A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, R1.3}

(In this paper, both terms ‘FAIR sub-principle’ or ‘FAIR principle’, with ‘principle’ in lowercase, may be used interchangeably to denote one of the above 15 individual sub-principles.)

The key entities in the FAIR sub-principles can be further categorized into three categories (light-shaded ovals in Figure 3):

Categories:

- Data (D),
- Metadata (M),
- Infrastructure (IS)

Applying the concept mapping approach to the subject–predicate–object portion of all 15 FAIR sub-principles yields 12 unique core concepts (pink-filled rounded rectangles in Figure 3).

Core Concepts:

(Descriptions consolidate existing community FAIR interpretations including those from GO-FAIR Foundation (n.d.))

- *Identifier*, a unique and persistent identifier for data and/or metadata;
- *Rich Metadata*, a set of descriptors of data including those minimally required for search and discovery, as well as those needed for understanding and reuse. This core concept is closely related to ‘Attribute’ in the context of (re)use.
- *Resource*, infrastructure such as search engines like Google that users can perform searches to find relevant data;
- *Protocol*, a computational agent such as HTTPS (Secure Hypertext Transfer Protocol) or FTP (File Transfer Protocol) that facilitates efficient information retrieval;
- *Procedure*, a set of defined and implemented specified rules and roles in the data search and retrieval infrastructure for user authentication (e.g. Single Sign-on with Two-Factor-Authentication) and access control (user permission/profile);
- *Language*, a formal system used to express the context of data and/or metadata in a format such as XML (eXtensible Markup Language) and RDF (Resource Description Framework) that can be utilized by machines;
- *Vocabulary*, a standardized set of terms and their meanings or definitions; A data vocabulary may consist of terms and definitions that describe the types of data, their characteristics, and the relationships between different data elements. A metadata vocabulary includes terms and definitions used to describe metadata attributes, properties, and relationships.
- *Reference*, a reference to another resource that provides additional relevant and useful information, including references to published documents (journal articles, reports, conference proceeding papers, etc.) on data product algorithms and validation, (meta)data standard specifications, as well as that to previous metadata records, if appropriate.
- *Attribute*, closely related to Rich Metadata, a set of attributes that focus on providing information for use suitability and conditions of the discovered data;
- *License*, a license that describes under which conditions the discovered data can be used;
- *Provenance*, information on when and how the data or metadata was created and modified, and by whom, its sources and ownership throughout its lifecycle;
- *Standard*, an established or agreed-upon set of guidelines, rules, specifications, or criteria used as a reference or norms.

4.1.1. FAIR-compliance quality measures

These unique core concepts are fundamental building blocks of the FAIR principles and FAIRness assessment and are therefore designated as *Quality Measures* to facilitate a holistic FAIR-compliance assessment. They encompass most of the entities within the FAIRsFAIR metrics as well as most of the FER types defined in the GO FAIR FIPs.

4.2. FAIR-Compliance maturity matrix (FAIR-QMM)

Beyond the variability of FAIRness across individual datasets, the level of FAIR-compliance requirements may be influenced by available resources and the potential impact of the datasets. Highly influential data include those that inform important policy and decision-making processes such as data used by government agencies for climate assessments and resilience in formulating policies (Ramapriyan et al. 2016; Tilmes et al. 2015). Climate assessments and related policies help reduce vulnerability and enhance the ability to withstand and adapt to climate-related shocks and stressors that affect humans and their lives. Highly influential data should have a higher degree of FAIR-compliance compared to other general research data, due to the enhanced level of requirements on transparency and reproducibility. To this end, a multi-level maturity matrix with pre-defined, progressive behaviors in discrete stages can help evaluate and document a dataset's current state of being FAIR. An organization can begin by defining its desired level of maturity based on its own objectives for providing a data product. It can then use a maturity matrix to determine the FAIRness requirements, identify gaps, and develop a roadmap for meeting the requirements, as depicted in Figure 4. This process is similar to the approach described by Peng et al. (2015) for scientific data stewardship. The approach has been shown to be effective in ensuring the data stewardship maturity and demonstrating the trustworthiness of organizations' core datasets (e.g. Dunn et al. 2021; Peng et al. 2019).

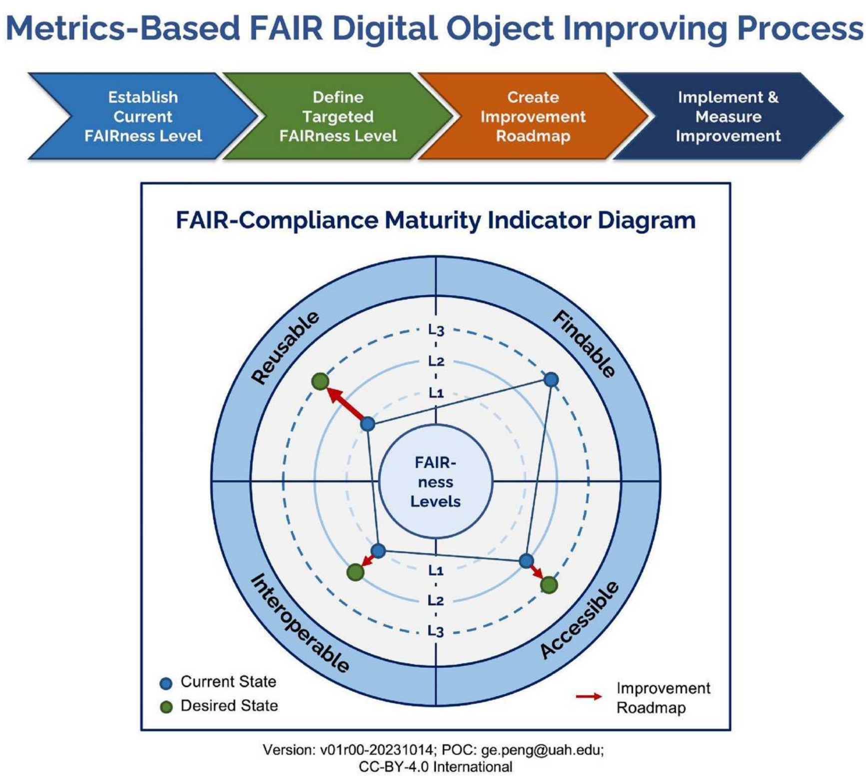

Figure 4. Visual depiction of a metrics-based FAIRness improvement process, incorporating FAIR-compliance maturity indicators. The current/desired states of FAIR-compliance maturity matrix indicators are represented by solid blue/green dots, respectively. The red arrows denote the direction of the improvement needed to reach the desired state.

Building on the aforementioned quality measures and elements deduced from the FAIR principles, a FAIRness quality maturity matrix (referred to FAIR-QMM hereafter) is developed, leveraging existing data management and FAIR-enabling practices and information technologies. Three maturity levels are defined, depicting the stages of minimal (Level 1), intermediate (Level 2), and optimal (Level 3) compliance with the FAIR principles, respectively. When reporting the maturity evaluation results, one may use Level 0 to denote the state of 'None or No Information Available'. It is recommended to include evidence or justification on the achieved maturity level to provide traceability and transparency of the FAIRness assessment. Additionally, it is important to describe the targeted domain and community, especially with regard to standards.

The definitions of the FAIR-QMM are documented in Appendix A (included in the supplementary material), accompanied by explanations and examples illustrating potential realizations of various maturity levels. For demonstration purposes, we will describe the maturity levels for the F1 and I2 principles, focusing on their associated quality measures, namely Identifier and Vocabulary. An in-depth discussion on F2 in terms of Rich Metadata and domain-specific use cases can be found in the Discussion Section.

4.2.1. F1-Identifier

The maturity levels of F1-compliance in terms of (meta)data identifiers are measured by the levels of uniqueness and persistence, which are defined as follows:

- Level 1 (L1): Data and metadata are assigned identifiers;
- Level 2 (L2): L1 + identifiers are either: (i) eternally persistent but not globally unique; (ii) globally unique but not eternally persistent; or (iii) unique within a particular scientific discipline;
- Level 3 (L3): L1 + identifiers are globally unique and eternally persistent.

At Level 1 – a minimal stage, (meta)data are assigned identifiers, usually just locally unique within an internal database or (meta)data management system. Examples include concept IDs for data products and associated data files in the Common Metadata Repository (CMR) of the National Aeronautics and Space Administration (NASA). CMR assigns a collection concept ID (like C1996881146-POCLOUD) to each data product. The collection concept ID is associated with the assigned DOI (like https://doi.org/10.5067/GHGMR-4FJ04) and resolves to the data product landing page (like https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1). A granule concept ID (like G2743390617-POCLOUD) is assigned to each associated data file, which is internal to CMR.

At Level 2 – an intermediate stage. Identifiers assigned can be eternally persistent but not globally unique. Or they may be unique within a particular domain/group such as International Virtual Observatory Identifier (VOID). On the other hand, identifiers assigned may be globally unique but not eternally persistent such as UUID or Uniform Resource Locator (URL) – both can be persistent if managed well. In some cases, identifiers may be valid only for the duration of a project. Identifiers may or may not be resolvable.

At Level 3 – an optimal stage, identifiers are globally unique and eternally persistent, such as Digital Object Identifiers (DOIs), Persistent Uniform Resource Locator (PURL), or Archival Resource Keys (ARK). These types of identifiers are by design globally unique and eternally persistent. Being eternally persistent implies persistent binding and identifier resolution which is explicitly addressed in A1. In practice, identifier persistence in terms of identifier resolution, as measured by the A1 principle, depends largely on the sustainability of the identifier hosting services and resolution infrastructures. Utilizing a GUPRI ensures the compliance to both F1 and A1.

4.2.2. I2-Vocabulary

The I2 principle requires that vocabularies used by (meta)data follow the FAIR principles, especially to be machine actionable. The maturity levels of I2-compliance in terms of vocabularies are defined as:

- L1: Data and metadata use vocabularies;
- L2: L1 + vocabulary is managed, findable and accessible; with a GUPRI assigned to the vocabulary itself;
- L3: L2 + vocabulary is governed, version controlled, interoperable and reusable; with a GUPRI assigned to each term within the vocabulary.

At Level 1, (meta)data use internal vocabularies or glossaries, often created and utilized by an individual project or institution, available only internally to associated entities, such as a project glossary in a shared Google Document.

At Level 2, vocabularies are managed and available online in HTML or PDF format (not machine-actionable); Being managed implies that there is a recognized entity that oversees the updating (adding, removing terms, modifying definitions, etc.). GUPRIs should also be assigned for the vocabularies themselves. An example includes NASA Atmospheric Composition Variable Standard Names (Silverman et al. 2023).

At Level 3, vocabularies are governed, version-controlled, interoperable and reusable. Controlled vocabularies are formally defined and version controlled with an established formal process to manage the vocabularies. Definition of each term is accompanied by a comprehensive description of its source and interconnection to other terms and/or sources as appropriate. A GUPRI is assigned to each of its terms, such as the practice used by Global Change Master Directory (GCMD) keywords (GCMD 2023). Ideally, the vocabularies are also available in multiple languages for enhanced reusability.

4.2.3. Importance of repositories and data service providers

At the core of the FAIR principles, i.e. subject–predicate–object expression parts, the definitions are the actionable terms such as ‘are assigned’, ‘are described by’, ‘are indexed’, ‘allow for’, ‘are released with’, as captured in the definitions of Level 1 stages in the FAIR-QMM (Table A2, which is included in the supplementary material). They underpin workflows that can be established by repositories. From this perspective, repositories can play a central role in offering FAIR-enabling capabilities, which reinforces previous opinions (Stall 2020; d’Aquin et al. 2023).

Standardizing those workflows offers an opportunity to systematically create, publish, and access datasets. It may involve how the Rich Metadata are defined, what Identifier to utilize, and which format/schema to adopt, leading to enhanced data and metadata interoperability.

By providing a centralized resource/protocol established by repositories and/or data service providers for discovery and retrieval of the data, it helps ensure that the technologies are open, secure and widely implemented. This leads to a higher FAIR-compliance of individual datasets.

The most important aspect of such an approach is to alleviate burdens on individual scientists of having to learn about the FAIR Principles and determining how to implement them, which could involve knowledge and expertise from other domains (data management and technology). The approach will not only help to enhance FAIRness of all of the repository’s data holdings but also make sharing data more inclusive and equitable.

5. Discussion

The most challenging aspect of implementing the FAIR Principles and evaluating the FAIRness of individual data products is the variation in FAIR practices and enabling resources across organizations, science disciplines, knowledge domains, communities, topical areas, and applications. Known as dependencies, these variations often occur in intertwined and nested ways, affecting the standardization of FAIR practices and FAIRness assessments. Shown below is an example scenario:

(1) Across different science disciplines (e.g. physics, biology, astrology, etc.),

(a) within one organization such as Earth science and Planetary science in NASA;
  (i) Across different topic areas, within one science discipline, such as physical oceanography and atmospheric chemistry in NASA Earth science;
    (i.1) Across different applications, such as satellite measures for weather prediction and climate monitoring.
(b) Across different organizations,
  (ii) within the same science discipline, topic area, and application, such as NASA and USGS in Earth sciences for drought monitoring;

(2) Across different knowledge domains (e.g. science, data management, and technology) and communities (e.g. Environmental science, Biodiversity, Genomics).

In FAIR implementation and FAIRness assessments, *Rich Metadata* and *Vocabulary* are two of the quality measures that exhibit various dependencies. In this section, the discussion begins by exploring the organization-dependency of *Rich Metadata* through implementation examples of established community metadata standards by different Earth science organizations. It then proceeds to the benefits of utilizing the FAIR-QMM to assess and document the FAIRness of individual data products, including mitigating the discipline dependency of *Vocabulary*. Moreover, the discussion addresses how the combination of data and metadata, as well as the relevance of standards, affects the FAIRness assessment. This Discussion section concludes with examples of community effort towards the convergence of FAIR implementations, which is essential for achieving optimal interoperability across disciplines and systems.

5.1. Domain dependency of rich metadata

'Rich' depends on its intended use. Fenner et al. (2019) proposed a minimum set of descriptive metadata properties (i.e. creator, title, publisher, publication date, summary, keywords, identifier, resource type) that support data discovery and citation, representing a good intersection among sets of metadata elements used by domain-agnostic metadata standards: Dublin Core, DCAT-2, Schema.org, and DataCite metadata schema. (As these may be insufficient from a data reuse perspective, rich metadata should also include provenance, standard vocabularies for variables, measures, etc.) Crosswalks from 15 metadata standards/schemas to that of Schema.org carried out by the RDA Metadata Interest Group (Wu et al. 2023; Wu, Hagan et al. 2021) clearly show variation of how individual metadata fields are encoded in them.

Table 3 captures the minimum set of metadata fields from four U.S. Earth science organizations. The first three are federal agencies with NASA focusing on space exploration, NOAA (National Oceanic and Atmospheric Administration) on the Earth's atmosphere, oceans, climate, and weather, and USGS on the Earth's natural resources, natural hazards, geology, and ecosystems. The fourth one, DataONE, is a federation of earth and environmental science repositories. Additional fields such as 'Spatial Coverage', 'Temporal Coverage', 'Topic Category' are added as they are important in searching for geo-spatiotemporal science data – also indicated by the metadata schema of Science on Schema.org (SOSO; Shepherd et al. 2022). It is intended to show if and how the fields suggested by Fenner et al. (2019) are implemented in various Earth science organizations.

The results demonstrate that the metadata schemas employed by the four organizations encompass most of the essential metadata fields. Notably, the 'Resource Type' field is not commonly included in these schemas since they are designed for data, making the resource type implicit. However, explicitly including this field could enhance machine understanding.

The terms used or the ways they are implemented can be quite different. For interoperability within an organization, crosswalks among different metadata schemas, including the use of 'synonyms'', have been shown to be useful. For example, NASA's Unified Metadata Model (UMM) is a crosswalk for several NASA supported metadata standards such as NASA's Earth Observing System

Table 3. Mapping of essential rich metadata fields suggested by Fenner et al. (2019) for data discovery and citation based on collection-level metadata schemas from different U.S. Earth science organizations.

Rich Metadata (Fenner et al. 2019)	NASA-ESDIS (Reiter and Stevens 2023)	NOAA-NCEI (NCEI Metadata Working Group 2019)	USGS (Data Catalog Vocabulary, DCAT US 1.1[a])	DataONE (DataONE 2020)
Title	Entry Title [M]	Title [M]	title (M)	title [M]
Resource Type	CollectionDataType [O]	–	–	–
Creator	CollectionCitations/ Creator [O], ContactPersons/Role = Metadata Author [O]	Citation [M]	creator* (M) *not DCAT-US but collected by USGS	origin/investigator [M]
Summary / Description	Abstract [M]	Abstract [M]	description (M)	abstract [R]
Identifier	DOI [M]	DOI [M]	identifier (type DOI) (R)	identifier [M]
Version	Version [M]	hierarchyLevel [M]	–	–* * A version chain is established via fields obsoletes/obsoletedBy that are populated by identifiers
Keywords	AncillaryKeywords [O], LocationKeywords [O], ScienceKeywords [M], TemporalKeywords [O]* * Also keywords for platform, instrument, data center, and project	descriptiveKeywords [M], theme [M], Datacenter [M], Place [M], Project [R], platform [R], instrument [R], resolution [R]	keywords (M) Locationkeywords, ScienceKeywords	keywords [R]
Publisher	CollectionCitations/ Publisher [O]	citedResponsibleParty [M]	publisher (M)	originMemberNode [M]
Publication Date	DataDate/Type = CREATE [O], MetadataDate/Type = CREATE [O]	dateStamp [M]	publication date (M)	datePublished [M]
License	UseConstraints [M]	resourceConstraints [M]	rights (M)	licenseDescription [R]* * not yet released
Related Dataset(s)	Associated DOIs/DOI [O]	MD_AggregateInformation [M] for data if applicable	–	–
Related Publication (s)	Publication References [O]	MD_AggregateInformation [M] for publications if applicable	related publication (O)	–
Data Repository /Archive	DataCenters [M]	citedResponsibleParty [M]	harvest source* * this is an internal USGS field	originMemberNode [M]
Extra Fields				
Spatial Coverage	SpatialExtent [M]	geographicElement [M]	Coordinates	eastBoundCoord, westBoundCoord, northBoundCoord, southBoundCoord [R]
Temporal Coverage	TemporalExtents [M]	temporalElement [M]	temporal information (start date and end date)	beginDate, endDate [R]
Topic Category	ISOTopicCategory [O]	topicCategory [M]	Theme	–

Notes: [M/R/O] denotes 'Mandatory' or 'Recommended' or 'Optional', respectively.
[a]https://resources.data.gov/resources/dcat-us/ (accessed: 2023-11-27).

(EOS) Clearinghouse (ECHO) metadata model, the Directory Interchange Format (DIF) content metadata, and ISO 19115-2 (Reiter and Stevens 2023). The DataONE terms are themselves a harmonization of seven commonly used metadata standards including the Ecological Markup Language (EML), DublinCore, and ISO 19115. The same is expected to be useful across organizations and disciplines, reconfirming the philosophy behind the Metadata Schema and Crosswalk Registry (MSCR) by FAIRCORE4EOSC (n.d.).

5.2. Benefits of utilizing the FAIR-QMM

If the FAIR-QMM is applied, the quality maturity can be compared at the same level across disciplines. Taking the *Vocabulary* quality measure as an example, there is no specific vocabulary that is stipulated in each level, but each level specifies what the status of a vocabulary should be. At Level 1, a vocabulary should be adopted for data and metadata. For Level 2, no matter which vocabulary is adopted, it should be managed, findable and accessible, as well as assigned a GUPRI. For Level 3, a chosen vocabulary should be governed, version controlled, interoperable and reusable, and moreover, a GUPRI should be assigned to each term of the vocabulary.

The concept, category and qualifier of the FAIR-QMM are at the atom level, systematic, and comprehensive. They are derived by adopting the Semantic Web Ontology framework and concept mapping approach in analyzing each of the FAIR principles. The FAIR-QMM could be regarded as an abstract model, which other FAIR implementation and maturity models could be mapped to and thus enable interoperability among FAIR maturity models. For example, in the RDA FAIR – DMIs by the RDA FAIR Data Maturity Model Working Group (2020), RDA-F1-01M (Metadata is identified by a persistent identifier) and RDA-F1-02M (Metadata is identified by a globally unique identifier) can be mapped to FAIR-QMM identifier – level 2 and level 3 respectively. In the FAIR-QMM, this identifier quality maturity levels can be applied to any resources – i.e. data, metadata and IS. The questions in the FIP mini-questionnaire can be more rigorously measured by the FAIR-QMM model. For example, the question 'What globally unique, persistent, resolvable identifiers do you use for metadata records', instead of proving identifier type(s) as an answer which may indicate binary assessment of the implementation, FAIR-QMM provides three levels of maturity that can be consistently applied to any identifier and could potentially be coded in a way that is easier for machines to understand and interpret.

5.3. Inter-dependencies and potential impact of whether data and metadata are combined

In practice, requirements of individual FAIR principles may not be interdependent for data and metadata. As a result, meeting FAIR requirements is not entirely autonomous but rather contingent on whether data and metadata are integrated or separate.

When combined, data and metadata are treated as a single digital entity – meeting a data-specific requirement automatically fulfills the corresponding metadata-specific requirement and vice versa.

When they are separated, however, implementation and assessment of quality measures such as assigning identifiers, providing licenses, and capturing provenances, need to consider both data – and metadata-specific requirements, although different workflows may potentially impact this.

We use *Identifier* to illustrate the last point. The assignment of a persistent identifier and the level at which it is assigned often depend on the institutional data management workflow. For example, NOAA workflow consists of assigning a DOI at the dataset-level and an internally unique identifier to the associated dataset-level metadata record that is indexed in a search engine and drives the dataset/DOI landing page. NASA and USGS have a similar workflow. This type of workflow appears to be sufficient to satisfy both F1 and A1 principles, even though F1-REQ-M is satisfied at a lower maturity level.

Conversely, it is also possible for data and metadata to be integrated. In this situation, an assigned and resolvable DOI will satisfy both F1-REQ-D and F1-REQ-M as well as A1-REQ-D and A1-REQ-M. It also satisfies F3-REQ-M if the DOI is captured in the data file metadata.

5.4. Relevance of standard

While '*Standard*' is designated as a specific quality measure for R1.3, modifying *Attribute* (R1), it underpins many of the other FAIR principles and associated quality measures such as:

- F1: Standard for establishing *Identifier* that is globally unique and eternally persistent;
- A1: Standard for communication *Protocol* for retrieval;
- A1.2: Standard for security *Procedure*;
- I1: Standard for knowledge representation *Language*;
- I2: Standard for *Vocabulary*;
- I3: Standard for linking *Reference;*
- R.1: Standard for providing *License*;
- R1.2: Standard for capturing *Provenance*.

Therefore, some FAIRness assessment tools have indicated that the assessment of R1.3 is embedded in that of other FAIR principles (Devaraju et al. 2022; Wilkinson et al. 2019). We agree with their reasoning but recommend to indicate the level of maturity of the standard utilized in each aspect, using the levels of maturity for *Standard* as defined in the FAIR-QMM (Table A2 in the supplementary material).

5.5. Examples of community effort on convergence of FAIR implementations

To overcome the large disparity in FAIR implementations, many coordinated community efforts have been carried out or are underway, seeking to enable the gathering and convergence of FAIR implementations both within and across various disciplines. For example, FAIRsharing services as a general registry of diverse and often manually curated terminology artifacts, models/formats, reporting guidelines, identifier schema, and metrics; and its Subject Resource Application Ontology (SRAO) provides a formal, hierarchical structure and richer semantics of the collected contents, leveraging other domain and community ontologies (FAIRsharing n.d.). The Go FAIR Implementation Profile (FIP) collects and synthesizes community implementation choices (GO FAIR n.d.), recognizing that some are primarily associated with technical implementation while others are associated with content- and domain-specific standards and practices (Schultes 2023). The WorldFAIR project strives to harmonize FIPs within eleven specific disciplines and to develop a cross-domain interoperability framework to foster interdisciplinary research across all these disciplines (Gregory and Hodson 2022; Hodson and Gregory 2023; WorldFAIR n.d.). To improve metadata interoperability under the European Open Science Cloud (EOSC) and beyond, the Horizon Europe projects FAIR-IMPACT and FAIRCORE4EOSC have teamed up to establish a Metadata Schema and Crosswalk Registry (MSCR) (FAIRCORE4EOSC n.d.).

6. Conclusion

This paper consolidates the FAIR vocabulary which is essential in harmonizing FAIRness assessment and reporting. The paper demonstrates the nature of multi-dimensions, multi-layers, and multi-facets of the FAIR principles and therefore the complexity of implementing them across disciplines.

This paper designates a set of quality measures upstream of the FAIRness assessment workflow, striving to overcome challenges related to subjective interpretations and domain-specific dependencies. The established quality measures align with unique core concepts in individual FAIR principles, serving as fundamental pillars for a holistic FAIRness assessment workflow.

Building upon this foundational set of quality measures, a FAIR-compliance maturity matrix is formulated, offering a structured and tiered approach to quantify and report progressive FAIR-compliance levels, referred to as FAIR-QMM. The criteria presented in the FAIR-QMM not only facilitate the assessment of current FAIRness status but also aid in defining the desired level of FAIRness and formulating a strategy to achieve that goal. To ensure that sharing the results of assessments is also in line with the FAIR Principles, we recommend following the FAIR quality information guidelines established by Peng et al. (2022) for documenting and reporting the assessed quality measures, assessment methods, approaches, and outcomes.

We envision that the integration of the FAIRness quality measures, the FAIR-QMM, harmonized FAIRness indicators, and the quality information guidelines, will play a role in fostering the creation of comprehensive FAIRness assessment workflows and the generation and integration of machine-actionable FAIRness quality information. Case studies on climate data, including both observational and modeling data, are currently being developed in collaboration with the World Meteorological Organization (WMO) and the European Centre for Medium-Range Weather Forecasts (ECMWF). The results of these efforts will be reported in the future.

Acknowledgments

We thank Leslie Hsu for the information on the USGS FAIR Rubric and Vivian Hutchison for the information on USGS collection-level metadata. Discussions with Rob Atkinson, Vivian Hutchison, Mark Parsons, and Matthew Jones were beneficial. We thank Janet Fennema for creating the list of Acronyms for Appendix A in the supplementary material and the two anonymous reviewers for their constructive comments that have helped enhance the clarity and quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

No additional data was generated for this work.

ORCID

Ge Peng http://orcid.org/0000-0002-1986-9115
Gary Berg-Cross http://orcid.org/0000-0002-2282-7215
Mingfang Wu http://orcid.org/0000-0003-1206-3431
Robert R. Downs http://orcid.org/0000-0002-8595-5134
Sudhir R. Shrestha http://orcid.org/0000-0001-6820-5879
Lesley Wyborn http://orcid.org/0000-0001-5976-4943
Nancy Ritchey http://orcid.org/0000-0003-3939-6287
Hampapuram K. Ramapriyan http://orcid.org/0000-0002-8425-8943
S. Jeanette Clark http://orcid.org/0000-0003-4703-1974
Jenny Wood http://orcid.org/0009-0007-9932-4716
Zhong Liu http://orcid.org/0000-0001-8150-7556
Abdelhak Marouane http://orcid.org/0009-0002-2794-1313

References

ARDC (Australian Research Data Commons). 2022. "FAIR Data Self Assessment Tool." Australian Research Data Commons. Accessed December 26, 2023. https://ardc.edu.au/resource/fair-data-self-assessment-tool/.

Bahim, C., C. Casorrán-Amilburu, M. Dekkers, E. Herczog, N. Loozen, K. Repanas, K. Russell, and S. Stall. 2020. "The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments." *Data Science Journal* 19:41. https://doi.org/10.5334/dsj-2020-041.

Berg-Cross, G., and S. Arbor. 2022. "Beyond Simple FAIR Principles for Ontologies and Semantic Resources Grounding Rich, Meaningful Metadata." *Journal of the Washington Academy of Sciences* 108:1–26.

Bernasconi, A., A. S. Garcia, G. Guizzardi, L. O. B. Da Silva Santos, and V. C. Storey. 2023. "Ontological Representation of FAIR Principles: A Blueprint for FAIRer Data Sources." In *Advanced Information Systems Engineering. CAiSE 2023. Lecture Notes in Computer Science 13901*, edited by M. Indulska, I. Reinhartz-Berger, C. Cetina, and O. Pastor, 262–277. Cham: Springer. https://doi.org/10.1007/978-3-031-34560-9_16.

Candela, L., D. Mangione, and G. Pavone. 2024. "The FAIR Assessment Conundrum: Reflections on Tools and Metrics." *Data Science Journal* 23:33. https://doi.org/10.5334/dsj-2024-033.

d'Aquin, M., F. Kirstein, D. Oliveira, S. Schimmler, and S. Urbanek. 2023. "FAIREST: A Framework for Assessing Research Repositories." *Data Intelligence* 5:202–241. https://doi.org/10.1162/dint_a_00159.

DataONE. 2020. "DataONE Content Indexer Documentation". *DataONE*. Accessed January 10, 2024. https://indexer-documentation.readthedocs.io/en/latest/.

Devaraju, A., R. Huber, M. Mokrane, P. Herterich, L. Cepinskas, J. de Vries, H. L'Hours, J. Davidson, and W. Angus. 2022. "FAIRsFAIR Data Object Assessment Metrics (0.5)." *Zenodo*. https://doi.org/10.5281/zenodo.6461229.

Dunn, R., C. Lief, G. Peng, W. Wright, O. Baddour, M. Donat, B. Dubuisson, et al. 2021. "Stewardship Maturity Assessment Tools for Modernization of Climate Data Management." *Data Science Journal* 20:7. https://doi.org/10.5334/dsj-2021-007.

FAIRCORE4EOSC. n.d. "Metadata Schema and Crosswalk Registry (MSCR)." FAIRCORE4EOSC. Accessed January 2, 2024. https://faircore4eosc.eu/eosc-core-components/metadata-cschema-and-crosswalk-registry-mscr.

FAIRsharing. n.d. "Subject Resource Application Ontology." FAIRsharing. Accessed September 25, 2023. https://fairsharing.org/FAIRsharing.b1xD9f.

Fenner, M., M. Crosas, J. S. Grethe, D. Kennedy, H. Hermjakob, P. Rocca-Serra, G. Durand, et al. 2019. "A Data Citation Roadmap for Scholarly Data Repositories." *Scientific Data* 6:28. https://doi.org/10.1038/s41597-019-0031-8.

GCMD (Global Change Master Directory). 2023. "GCMD Keywords, Version 17.5." Earth Science Data and Information System, Earth Science Projects Division, Goddard Space Flight Center, NASA. Accessed August 5, 2023. https://forum.earthdata.nasa.gov/app.php/tag/GCMD(Keywords.

GO FAIR Foundation. n.d. "Interpreting FAIR." GO FAIR Foundation. Accessed November 12, 2023. https://www.gofair.foundation/interpretation.

GO FAIR. n.d. "FAIR Implementation Profile." GO FAIR. Accessed September 25, 2023. https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/.

Gregory, A., and S. Hodson. 2022. "WorldFAIR Project (D2.1) 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' (1.0)." *Zenodo*. https://doi.org/10.5281/zenodo.7378109.

Hodson, S., and A. Gregory. 2023. "WorldFAIR Project (D1.3) First Policy Brief (Version 1)." *Zenodo*. https://doi.org/10.5281/zenodo.7853170.

Hutchison, V. B., T. Norkin, L. S. Zolly, and L. Hsu. 2024. "State of the Data: Assessing the FAIRness of US Geological Survey Data." *Data Science Journal* 23: 22. https://doi.org/10.5334/dsj-2024-022.

IHMC (Institute for Human & Machine Cognition). n.d. "Cmap". Florida Institute for Human & Machine Cognition. Accessed March 26, 2023. https://cmap.ihmc.us.

Jacobsen, A., R. M. de Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, et al. 2020. "FAIR Principles: Interpretations and Implementation Considerations." *Data Intelligence* 2:10–29. https://doi.org/10.1162/dint_r_00024.

Jones, M. B., P. Slaughter, and T. Habermann. 2019. "Quantifying FAIR: Automated Metadata Improvement and Guidance in the DataONE Repository Network." *Zenodo*. https://doi.org/10.5281/zenodo.3408466.

Kuhn, T., and M. Dumontier. n.d. "FAIR Vocabulary". Accessed October 7, 2023. https://peta-pico.github.io/FAIR-nanopubs/principles/index-en.html.

Matentzoglu, N., J. P. Balhoff, S. M. Bello, C. Bizon, M. Brush, T. J. Gallahan, C. G. Chute, et al. 2022. "A Simple Standard for Sharing Ontological Mappings (SSSOM)." *Database* 2022:baac035. https://doi.org/10.1093/database/baac035.

Mons, B., C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, and M. A. Wilkinson. 2017. "Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud." *Information Services & Use* 37:49–56. https://doi.org/10.3233/ISU-170824.

NCEI (National Centers for Environmental Information) Metadata Working Group. 2019. "Guidance for the NCEI Collection Level Metadata Template v1.2." Accessed August 1, 2023. https://www.ncei.noaa.gov/sites/default/files/2022-05/AB-GUID-02823_R1_Guidance%20for%20The%20NCEI%20Collection%20Level%20Metadata%20Template%20v1.2.pdf.

Peng, G. 2023a. "Finding Harmony in FAIRness." *Eos* 104: 22–25. https://doi.org/10.1029/2023EO230216.

Peng, G. 2023b. "Dissecting the FAIR Guiding Principles - Key Categories, Core Concepts, Focus Elements, and Harmonized Indicators (v01r00-20230225) [Data set]." *Zenodo*. https://doi.org/10.5281/zenodo.7896948.

Peng, G., R. R. Downs, H. K. Ramapriyan, M. A. Parsons, D. F. Moroni, Z. Liu, S. J. S. Khalsa, C. Mears, Y. Wei, B. Ramachandran, S. Smith, and NASA O'FAIR WG. 2023. *An Overview of Community FAIR Practices – NASA O'FAIR WG Inception Report*. Document ID: NASA-OFAIR-ESDSWG-DOC-0001. Version: v01r00-20230508. CC-BY 4.0 International. https://doi.org/10.5067/DOC/ESCO/ESDSWG-0001V1.
Peng, G., C. Lacagnina, R. R. Downs, A. Ganske, H. Ramapriyan, I. Ivánová, L. Wyborn, et al. 2022. "Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets." *Data Science Journal* 21:8. https://doi.org/10.5334/dsj-2022-008.
Peng, G., A. Milan, N. Ritchey, R. P. Partee II, S. Zinn, P. E. McQuinn, P. Lemieux III, et al. 2019. "Practical Application of a Stewardship Maturity Matrix for the NOAA OneStop Program." *Data Science Journal* 18:41. https://doi.org/10.5334/dsj-2019-041.
Peng, G., J. L. Privette, E. J. Kearns, N. A. Ritchey, and S. Ansari. 2015. "A Unified Framework for Measuring Stewardship Practices Applied to Digital Environmental Datasets." *Data Science Journal* 13:231–253. https://doi.org/10.2481/dsj.14-049.
Peters-von Gehlen, K., H. Höck, A. Fast, D. Heydebreck, A. Lammert, and H. Thiemann. 2022. "Recommendations for Discipline-Specific FAIRness Evaluation Derived from Applying an Ensemble of Evaluation Tools." *Data Science Journal* 21:7. https://doi.org/10.5334/dsj-2022-007.
Ramapriyan, H. K., J. G. Goldstein, H. Hua, and R. E. Wolfe. 2016. "Tracking and Establishing Provenance of Earth Science Datasets: A NASA-Based Example." In *Provenance and Annotation of Data and Processes. IPAW 2016. Lecture Notes in Computer Science*, edited by M. Mattoso, and B. Glavic, Vol. 9672, 226–229. Cham: Springer. https://doi.org/10.1007/978-3-319-40593-3_27.
RDA FAIR Data Maturity Model Working Group. 2020. "FAIR Data Maturity Model: Specification and Guidelines." Research Data Alliance. https://doi.org/10.15497/RDA00050.
Reiter, E., and T. Stevens. 2023. "Earthdata Wiki: UMM-C Schema Representation. Atlassian Confluence." https://wiki.earthdata.nasa.gov/display/CMR/UMM-C(Schema(Representation.
Ruttenberg, A., M. Courtot, J. A. Overton, J. Zheng, M. Brochhausen, J. M. Horton, J. Malone, et al. 2022. "Information Artifact Ontology." https://github.com/information-artifact-ontology/IAO/.
Schultes, E. 2023. "The FAIR Hourglass: A Framework for FAIR Implementation." *FAIR Connect* 1:13–17. https://doi.org/10.3233/FC-221514.
Schultes, E., B. Magagna, K. M. Hettne, R. Pergl, M. Suchánek, and T. Kuhn. 2020. "Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence." In *Advances in Conceptual Modeling, Vol. 12584*, edited by G. Grossmann, and S. Ram, 138–147. New York: Springer International Publishing. https://doi.org/10.1007/978-3-030-65847-2_13.
Schultes, E., M. Roos, L. O. Bonino da Silva Santos, G. Guizzardi, J. Bouwman, T. Hankemeier, A. Baak, and B. Mons. 2022. "FAIR Digital Twins for Data-Intensive Research." *Frontiers in Big Data* 5:883341. https://doi.org/10.3389/fdata.2022.883341.
Shepherd, A., M. Jones, S. Richard, N. Jarboe, D. Vieglais, R. Duerr, D. Fils, M. Minch, et al. 2022. "Science on Schema.org. v1.3.1." Earth Science Information Partners. Accessed September 19, 2023. https://github.com/ESIPFed/science-on-schema.org.
Silverman, M., M. Shook, R. Hornbrook, L. Ziemba, S. Hall, K. Ullmann, J. Crounse, et al. 2023. "Atmospheric Composition Variable Standard Name Convention version 1." NASA Earth Science Data and Information System Standards Coordination Office. Accessed August 5, 2023. https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/acvsnc.
Slaughter, P., B. Leinfelder, J. Clark, B. Mecum, S. Gordon, M. Jones, D. Mullen, J. Tao, and M. B. Artntek. 2023. "MetaDig Engine: Multi-Dialect Metadata Assessment Engine." Version: 2.4.1. Release date: 6 July 2023. https://github.com/NCEAS/metadig-engine.
Stall, S. 2020. "Enabling FAIR Data – The Importance of Our Scientific Repositories." EGU General Assembly. Online, 4–8 May 2020. https://doi.org/10.5194/egusphere-egu2020-17993.
Tilmes, C., R. E. Wolfe, B. Duggan, S. Aulenbach, J. C. Goldstein, X. Ma, and S. Zednik. 2015. "Supporting Trust with Provenance of the Findings of the National Climate Assessment." METHOD 2015: The 4th Intl. Workshop on Methods for Establishing Trust of (Open) Data. 11 October 2015, Bethlehem, PA, USA. http://www.few.vu.nl/~dceolin/method2015/papers/METHOD_2015_paper_2.pdf.
Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3:160018. https://doi.org/10.1038/sdata.2016.18.
Wilkinson, M. D., M. Dumontier, S. A. Sansone, L. O. B. da Silva Santos, M. Prieto, D. Batista, P. McQuilton, et al. 2019. "Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework." *Scientific Data* 6:174. https://doi.org/10.1038/s41597-019-0184-5.
Wilkinson, M. D., S. A. Sansone, E. Méndez, R. David, R. Dennis, D. Hecker, M. Kleemola, C. Lacagnina, A. Nikiforova, and L. J. Castro. 2022. "Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion [Version 1]." *Open Research Europe* 2. https://doi.org/10.12688/openreseurope.15364.1.

Wilkinson, M. D., S. A. Sansone, E. Schultes, P. Doorn, L. O. B. de Silva Santos, and M. Dumontier. 2018. "A Design Framework and Exemplar Metrics for FAIRness." *Scientific Data* 5:180118. https://doi.org/10.1038/sdata.2018.118.

Wittenburg, P., and G. Strawn. 2018. "Common Patterns in Revolutionary Infrastructures and Data." B2SHARE. https://doi.org/10.23728/B2SHARE.4E8AC36C0DD343DA81FD9E83E72805A0.

Wittenburg, P., G. Strawn, B. Mons, B. Boninho, and E. Schultes. 2019. "Digital Objects as Drivers towards Convergence in Data Infrastructures." B2SHARE. https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11.

WorldFAIR. n.d. "The WorldFAIR Project." Accessed September 25, 2023. https://worldfair-project.eu/.

Wu, M., P. Hagan, B. Cecconi, S. Richard, C. Verhey, and RDA Research Metadata Schemas WG. 2021. "A Collection of Crosswalks from Fifteen Research Data Schemas to Schema.org." Research Data Alliance. https://doi.org/10.15497/RDA00069.

Wu, M., S. M. Richard, C. Verhey, L. J. Castro, B. Cecconi, and N. Juty. 2023. "An Analysis of Crosswalks from Research Data Schemas to Schema.org." *Data Intelligence* 2023:100–121. https://doi.org/10.1162/dint_a_00186.

Wyborn, L., A. Prent, J. Croucher, N. Rees, and R. Farrington. 2023. "Using FAIR Implementation Profiles (FIPs) and FAIR Enabling Resources (FERs) to Accelerate Machine-to-machine Interoperability of Geoscience Datasets Within and Across Repositories, Communities and Other Domains." AGU 2023 Fall Meeting Abstract, https://agu.confex.com/agu/fm23/meetingapp.cgi/Paper/1440456.