Open Free Findable Accessible Interoperable Reusable

INCEPTION REPORT

AN OVERVIEW OF COMMUNITY FAIR PRACTICES

Prepared by

NASA Earth Science Data Systems Working Groups (ESDSWG)

Open, Free & FAIR Working Group (O'FAIR WG)

May 8, 2023

AUTHOR LIST

Ge Peng

Earth System Science Center (ESSC)/Interagency Implementation and Advanced Concepts Team (IMPACT) of NASA's Marshall Space Flight Center project (MSFC), University of Alabama in Huntsville (UAH), Huntsville, AL, USA
ORCID: 0000-0002-1986-9115

Robert R. Downs

Center for International Earth Science Information Network (CIESIN), Columbia Climate School, Columbia University, Palisades, NY, USA
ORCID: 0000-0002-8595-5134

Hampapuram K. Ramapriyan

Science Systems and Applications, Inc. (SSAI), Lanham, MD, USA and
Earth Science Data and Information System Project (ESDIS), NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA
ORCID: 0000-0002-8425-8943

Mark A. Parsons

UAH/NASA MSFC IMPACT & Chief Science Data Office, Huntsville, AL, USA
ORCID: 0000-0002-7723-0950

David F. Moroni

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
ORCID: 0000-0003-2994-557X

Zhong Liu

George Mason University/GES DISC, Greenbelt, Maryland, USA
ORCID: 0000-0001-8150-7556

Siri Jodha S. Khalsa
University of Colorado, Boulder, Colorado, USA
ORCID: 0000-0001-9217-5550

Carl Mears
Remote Sensing Systems, Santa Rosa, CA, USA
ORCID: 0000-0002-6020-9354

Yaxing Wei
ORNL.DAAC, Oak Ridge National Laboratory, Oak Ridge, TN, USA
ORCID: 0000-0001-6924-0078

Bhaskar Ramachandran
LAADS DAAC, GSFC/SSAI, Greenbelt, Maryland, USA
ORCID: 0000-0002-1684-8888

Shawn R. Smith
Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, USA
ORCID: 0000-0003-1392-3077

DISCLAIMER

The opinions expressed in this document do not reflect NASA's concurrence, approval, or indicate steps to implementation.

CHANGE HISTORY LOG

Version	Author/Contributor [Reviewer]	Description
v00r01-20221019	Ge Peng	Initial document draft - document architecture.
v00r02-20221109	Ge Peng	Incomplete draft for WG review and contribution - focusing on content. Utilized the contents contributed by Ge Peng, Hampapuram Ramapriyan, David Moroni, Mark Parsons, Zhong Liu, Robert Downs, and other members of the WG. Sections impacted: All.
v00r03-20230208	Ge Peng, Robert Downs, Hampapuram Ramapriyan, Mark Parsons, David Moroni, Zhong Liu, Carl Mears, Siri Jodha Khalsa/Shawn Smith, Rudi Gens, Keith Bryant, [Stephanie Wingo, Sara Lubkin]	First complete draft for WG review and proofread. Sections impacted: All.
v00r04-20230301	[R. Downs, M. Parsons, H. Ramapriyan, B. Ramachandran, S. Smith, Y. Wei, S. Wingo, etc.]	Revised based on review comments from the WG members. Section impact: All.
v01r00-20230508	Peng et al. (2023)	First baselined document.

RECOMMENDED CITATION

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. CC-BY-4.0 International. Version: v01r00-20230508. DOI: https://doi.org/10.5067/DOC/ESCO/ESDSWG-0001V1

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Earth Science Division (ESD) of the NASA Science Mission Directorate (SMD) has been promoting open data since the 1990s. The latest NASA SMD data and information policy (i.e., SPD-41a) requires all NASA-funded data and information to be freely available and openly shared, including such Earth science data products that are managed by NASA data repositories. SPD-41a also explicitly recommends all SMD-funded scientific data and information to follow the FAIR Guiding Principles.

Meeting the FAIR data requirements specified in SPD-41a will undoubtedly impact how data producers generate their datasets and how data repositories manage their data stewardship workflows. Currently, there is no consistent and practical guidance on how to best follow the FAIR Principles but such guidance is expected to be in high demand.

To help address this challenge, the Earth Science Data System Working Group on Making NASA SMD-funded Earth Science Data Open, Free and FAIR (O'FAIR) has been created. This document, the Inception Report, is the first of two deliverables to be produced by the working group. The O'FAIR Inception Report describes the FAIR Principles and provides examples for interpreting the FAIR Principles in practice. It identifies and lists FAIR practices and describes approaches that are being applied, currently.

The O'FAIR Inception Report also serves as a centralized place to hold consolidated materials and information that have been collected by the working group to prepare for the development of a guidance document. The Guide Document, the second deliverable, will be developed to provide principle-by-principle guidance on how to apply exemplary practices that enable NASA SMD-funded Earth science data and information to be findable, accessible, interoperable, and (re)usable (aka, FAIR) in addition to ensuring that they will be open and freely available for use. The Guide Document will leverage community FAIR data practices that pertain to NASA data holdings. It will be produced with input and consensus from various relevant NASA stakeholders and offering benefits to all NASA ESD data repositories and data producers. It is envisioned that the Guide Document can be used by future mission project managers to define data management and stewardship requirements.

1. BACKGROUND

1.1 About O'FAIR WG

The Open, Free & FAIR Working Group (O'FAIR WG) was established in May 2022 as one of NASA's Earth Science Data System Working Groups (ESDSWG), with a proposed term of two years. The purpose of the O'FAIR WG is to explore community practices and develop a guidance document on what existing practices can be utilized to ensure or enable NASA Earth science data and information that are funded by the NASA SMD to be open and free as well as findable, accessible, interoperable, and (re)usable (aka, FAIR), following the FAIR Guiding Principles as defined by Wilkinson et al. (2016). The working group membership is open to anyone who is interested in participation, especially those in the NASA Earth Science Division (ESD) and its affiliated organizations. Additional information can be found at:

https://wiki.earthdata.nasa.gov/display/ESDSWG/Making+NASA+SMD-funded+Earth+Science+Data+FAIR+Working+Group (NASA ESDSWG wiki access may be required.)

1.2 Motivation

NASA's Science Mission Directorate (SMD) seeks to expand human knowledge through new scientific discoveries in order to understand the Sun, Earth, Solar System, and Universe. As one of five SMD divisions, the ESD supports the collection/acquisition, processing, stewardship, and applications of data and information from NASA's Earth observations. These activities are conducted to facilitate understanding and monitoring of the current state of the Earth system through its affiliated Science Investigator-led Processing Systems (SIPS) (Fig. 1.1) and Distributed Active Archive Centers (DAACs) located throughout the United States (Fig. 1.2), comprising the Earth Observing System Data and Information System (EOSDIS). At the end of FY2022, there were about 15,360 unique datasets with a total archive volume of about 72 PB, including about 20 PB in the Cloud, with an average archive growth of about 49 TB per day. In FY2022 (October 1, 2021 – September 30, 2022), EOSDIS served more than 3.64 million distinct users world-wide with an average end user distribution volume of 281 TB per day (Wanchoo et al. 2022).

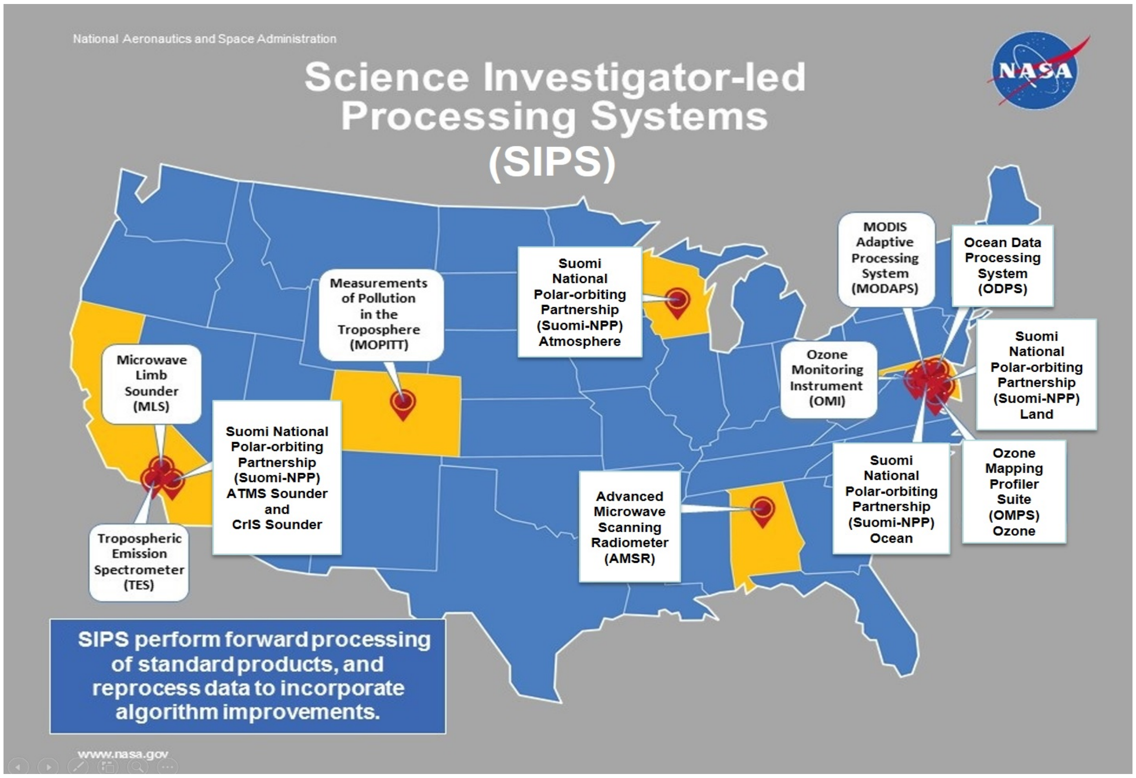

Figure 1.1. EOSDIS Science Investigator-led Processing Systems (SIPS)

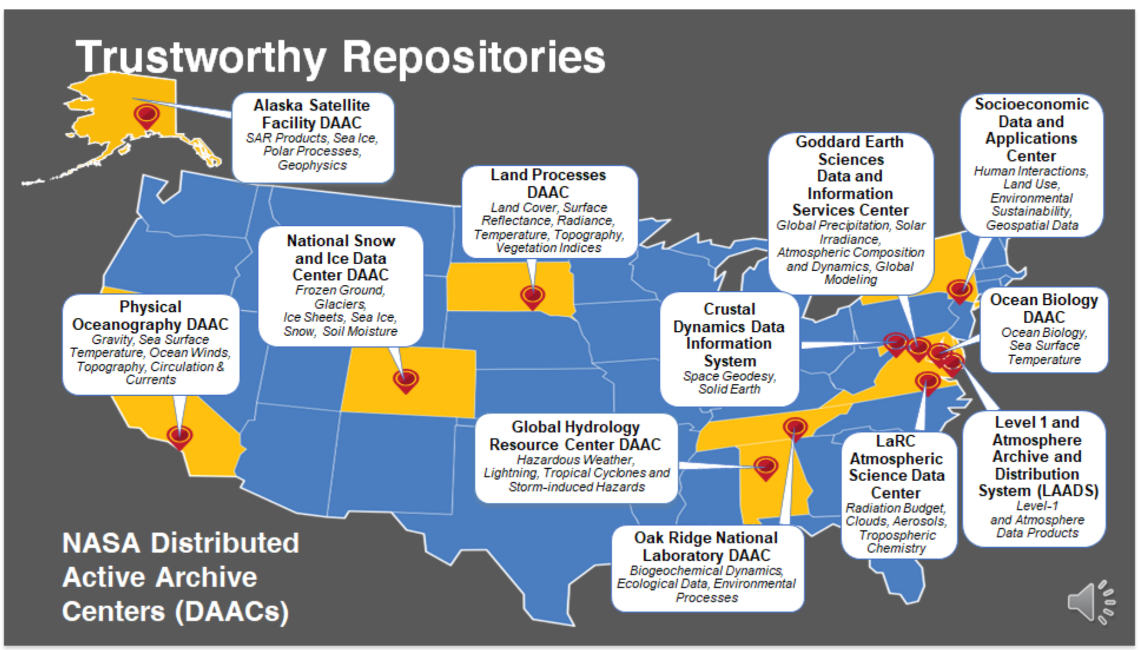

Figure 1.2. EOSDIS Distributed Active Archive Centers (DAACs)

To maximize the value of NASA data and information, NASA SMD released a new draft data and information policy in 2021 for public comment. The final version was released in September 2022 (SPD-41a; NASA SMD 2022a) followed by a more detailed guidance document in December 2022 (NASA SMD 2022b). The policy mandates all NASA-funded data and information that are not subject to specific laws, regulations, or policies, be freely available and openly shared, including Earth science data products that are managed by NASA DAACs. SPD-41a defines "Data" as "scientific information that can be stored digitally and accessed electronically." In the event a variance is granted to the free distribution, SMD will charge no more than the cost of dissemination for the distribution of data (NASA SMD 2022a).

The SPD-41a policy also explicitly recommends all SMD-funded data follow the FAIR Guiding Principles along with other specific requirements. This will undoubtedly impact how data producers generate their datasets and how DAACs manage their data stewardship workflows and disseminate data products and services. Guidance or guidelines on how to best follow these principles is expected to be in high demand by data producers and data managers whose current and future practices can benefit from the experiences of others. In particular, a guide that builds on leading community FAIR data practices, but pertains to NASA data holdings with input and consensus from various relevant NASA stakeholders, is expected to be beneficial to all DAACs and data producers such as SIPS, missions and campaigns, and projects that constitute the Making Earth System Data Records for Use in Research Environment (MEaSUREs) Program. It is also useful for future program and project managers in defining data management and stewardship requirements.

1.3 Mission Statement

The O'FAIR WG aims to provide principle-by-principle guidance on what leading practices should be used to ensure that NASA SMD-funded Earth science open and free data/information are also findable, accessible, interoperable, and (re)usable (FAIR).

1.4 About This Document

This document provides a high-level overview of the current landscape of community FAIR practices as the inception report of the O'FAIR WG. The methodology is described in Section 2, followed by descriptions/interpretations of the FAIR Guiding Principles (Section 3), FAIR assessment models and implementation examples (Section 4), the state of NASA FAIR assessments (Section 5), and a summary in Section 6.

Given the limitation on the time available to the members of this working group to dedicate to this effort, the community FAIR practices captured are not necessarily exhaustive or comprehensive but provide a starting point to gauge the current community FAIR practices landscape and to leverage this knowledge to develop a FAIR-practices guide document, pertaining to NASA-funded Earth science data and information.

2. METHODOLOGY

Upon receiving approval to commence, the O'FAIR WG started its activities by familiarizing the group members with NASA policies and the U.S. FAIR initiatives. Familiarization activities included presentations from invited speakers who possess relevant expertise on these issues, followed by discussions with each of the speakers and subsequent discussions among group members. The members also made a significant effort to review and gather information from their network and current literature that is relevant to the FAIR Principles and to applying the FAIR Principles in practice. In addition, a crosswalk was developed to match and represent conceptual relationships between the stated data requirements in NASA's SPD-41a and the FAIR principles to identify any synergies and distinctions in either of them as compared to the other. Furthermore, a diagram (not shown) was developed to represent the relationships between the relevant concepts that have been identified and discussed among the members of the Working Group.

2.1 Invited Speaker Presentations

Recognizing the need to obtain current information about national and NASA policies and about emerging technology and practices, both nationally and internationally, that are relevant to the objectives of the O'FAIR WG, speakers were invited to present during the monthly meetings of the O'FAIR WG on pertinent topics.

On June 22, 2022, the speaker was Dr. Steven M. Crawford, the Science Data Officer for the NASA SMD. Dr. Crawford presented, "Status Update on SPD-41: Scientific Information Policy," describing the goals of the SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024, the current policy, and proposed changes in the SPD-41a draft for software, data, and publications, in light of the comments received from the community. Several topics were discussed with Dr. Crawford during the question and discussion portion of the presentation. Of particular note, the O'FAIR WG raised several questions that led to discussions about the SPD-41a terminology, target audience, strategic objectives, vision, efforts within and across divisions of the SMD, and planned WG activities.

On July 27, 2022, the speaker was Dr. Melissa Cragin, former Chief Strategist for Data Initiatives in the Research Data Services division at the San Diego Supercomputer Center (SDSC) at the University of California San Diego (UCSD) and current Associate Vice President for Information Technology at Rice University. Dr. Cragin presented, "A brief look at the FAIR Principles and opportunities for the Earth sciences." During the questions and discussion portion of the presentation, a variety of issues were raised, including FAIR community membership within the United States (U.S.), types of FAIR initiatives within the U.S., support for such activities, and current challenges that are being faced within the community.

2.2 FAIR Practices Data Collection

Recognizing the diversity of approaches in which the FAIR practices have been implemented within various organizations and how identifying such approaches can inform the efforts of the Earth science community, the O'FAIR WG has identified and described practices that address the FAIR Principles.

2.3 SPD-41a and FAIR Principles Matching

Recognizing that there are similarities and differences in the concepts described in SPD-41a and the FAIR Principles, and understanding that these relationships could help improve how we benefit from the potential synergies between their objectives, the O'FAIR WG assessed and represented these relationships in a crosswalk between the terms used in both documents (Fig. 2.1). As shown in Fig. 2.1, the SPD-41a data requirements map well to principles in findability and reusability. As somewhat expected, the SPD-41a data requirements do not cover infrastructure-related principles such as A1.1 and A1.2, which, however, may be addressed by different NASA policies. Detailed discussion of the crosswalk will be provided in the Guide document (in preparation).

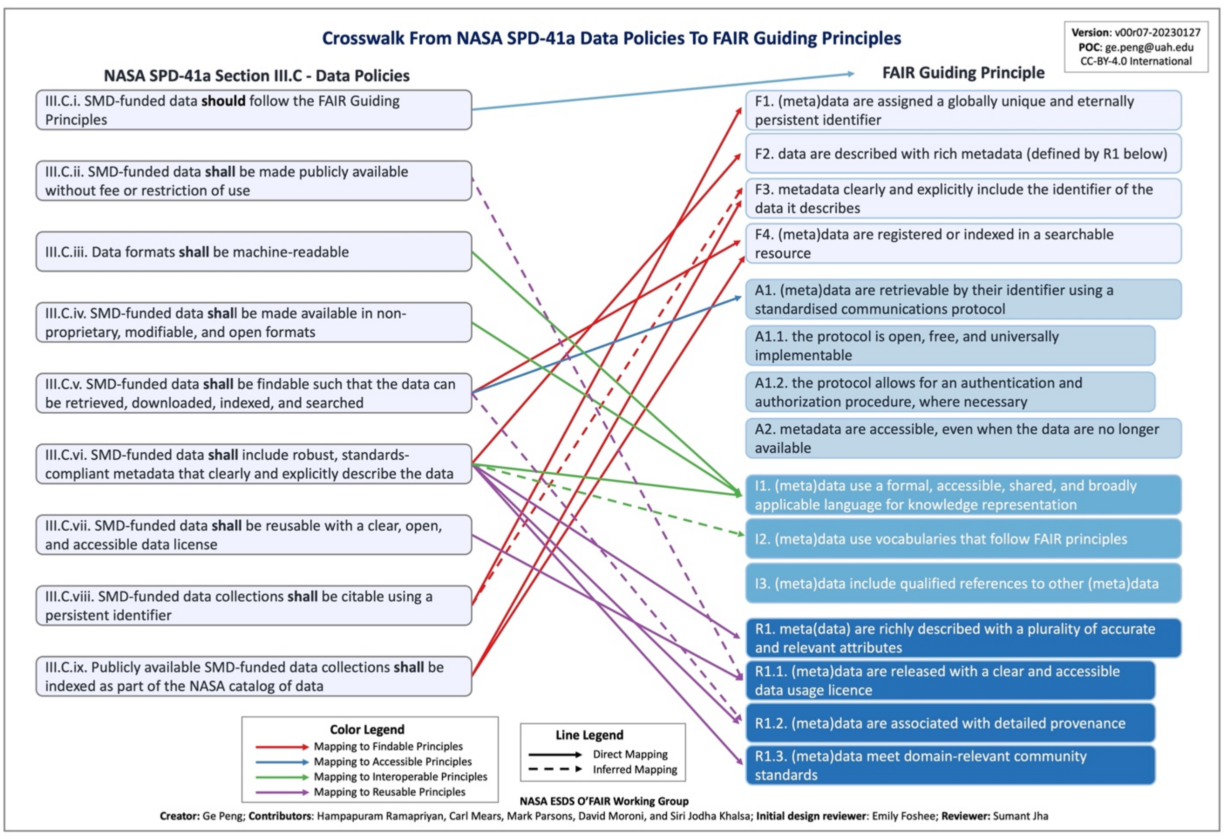

Figure 2.1. Diagram shows the crosswalk from the data requirements defined in SPD-41a (NASA SMD 2022a) to the FAIR Principles defined in Wilkinson et al. (2016).

3. DESCRIPTIONS OF FAIR DATA GUIDING PRINCIPLES

3.1 FAIR Data Guiding Principles

The FAIR Guiding Principles (‘FAIR Data Principles’, ‘FAIR Principles’ or simply ‘FAIR’) were initiated at the 2014 Lorentz Workshop ‘Jointly Designing a Data FAIRport’ [1] and formally defined by stakeholders from academia, industry, funding agencies, and scholarly publishers to encourage data sharing across systems, disciplines, and regional boundaries in 2016 (Wilkinson et al. 2016). The FAIR Principles have quickly gained popularity in the global data management and stewardship community, have been endorsed by many international organizations and countries, and have had a major impact in prompting data sharing and reuse worldwide (Peng et al. 2022).

The description of the FAIR Principles is captured in Table 3.1. To facilitate the FAIRness assessment process, each principle can be further mapped into requirements that are associated with Data (D), Metadata (M), and Infrastructure (IS) categories (Peng 2023). For example, the A1 principle can be decomposed into the following category-specific requirements (Fig. 3.1):

- A1-REQ-D: data are retrievable by their identifier;
- A1-REQ-M: metadata are retrievable by their identifier;
- A1-REQ-IS: communication protocol is standardized.

The category-breakdowns of the total FAIR requirements are displayed in a pie-chart (Fig. 3.2). The largest percentage of the FAIR requirements pertains to metadata (44.8%; 13 metadata requirements) with the close second for data (41.4%; 12 data requirements), while less than 14% pertaining to infrastructure (13.8%; 4 infrastructure related requirements). See Peng (2023) for more details.

[1] https://www.openaire.eu/how-to-make-your-data-fair

Table 3.1. The description of the FAIR Data Guiding Principles from Wilkinson et al (2016) and their mappings onto requirements in three key categories: Data (D), Metadata (M), and Infrastructure (IS). From: Peng (2023).

FAIR Data Guiding Principles			Key Categories		
#	ID	Description	Data (D)	Metadata (M)	Infrastructure (IS)
1	F1	(meta)data are assigned a globally unique and eternally persistent identifier	1	1	
2	F2	data are described with rich metadata (defined by R1 below)	1	1	
3	F3	metadata clearly and explicitly include the identifier of the data it describes	1	1	
4	F4	(meta)data are registered or indexed in a searchable resource	1	1	1
5	A1	(meta)data are retrievable by their identifier using a standardised communications protocol	1	1	1
6	A1.1	the protocol is open, free, and universally implementable			1
7	A1.2	the protocol allows for an authentication and authorization procedure, where necessary			1
8	A2	metadata are accessible, even when the data are no longer available		1	
9	I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	1	1	
10	I2	(meta)data use vocabularies that follow FAIR principles	1	1	
11	I3	(meta)data include qualified references to other (meta)data	1	1	
12	R1	(meta)data are richly described with a plurality of accurate and relevant attributes	1	1	
13	R1.1	(meta)data are released with a clear and accessible data usage licence	1	1	
14	R1.2	(meta)data are associated with detailed provenance	1	1	
15	R1.3	(meta)data meet domain-relevant community standards	1	1	

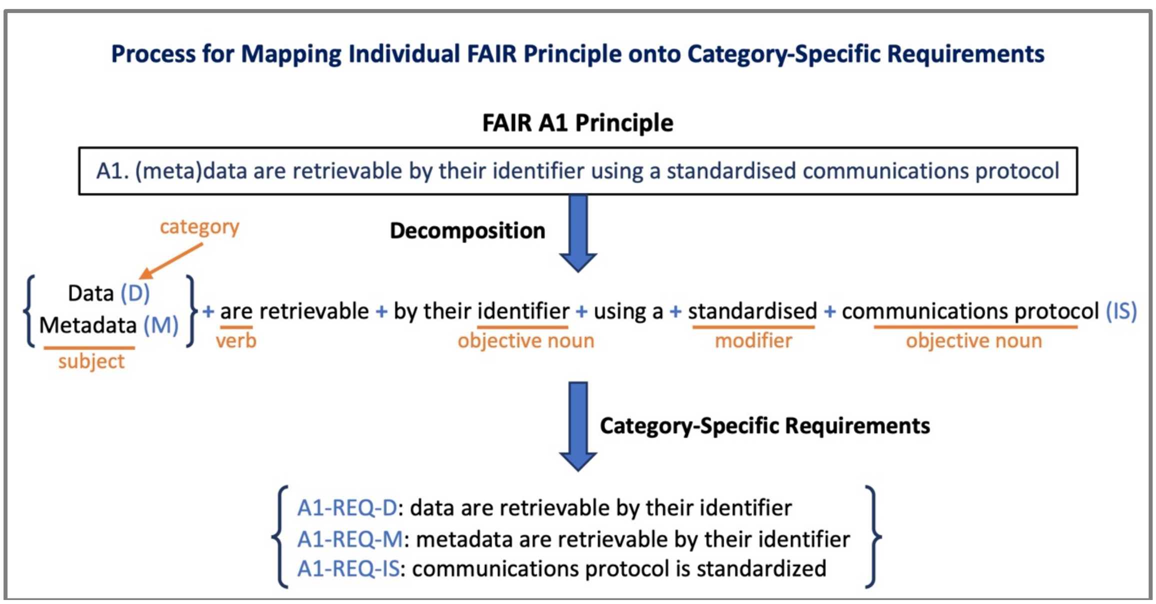

Figure 3.1. Diagram shows the process of decomposing the FAIR-A1 principle into category-specific requirements. Based on: Peng (2023).

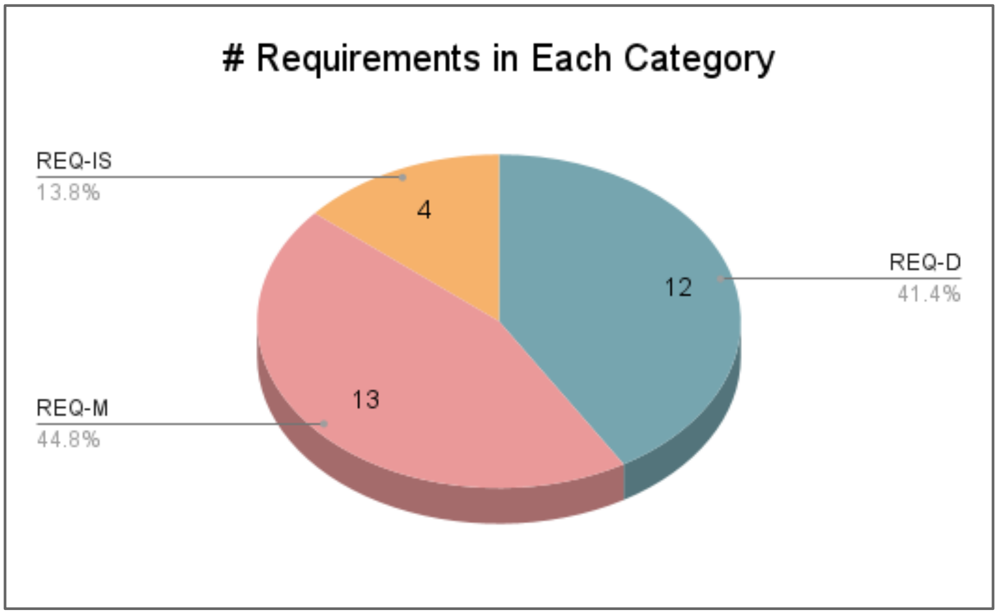

Figure 3.2. The breakdown of category-specific requirements from the FAIR Principles. Based on: Peng (2023)

3.2 Organizational Explanations of the FAIR Principles

There are a number of explanations or interpretations of the FAIR Principles available online. Table 3.2 provides examples from four well-established organizations that represent different perspectives.

Table 3.2. What do the FAIR Guiding Principles mean?

Organization ->	DataOne	Go FAIR	IPCC	Swiss National Science Foundation
Source	https://www.dataone.org/fair	https://www.go-fair.org/fair-principles/	Pirani et al. (2022)	http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf
Findable	Metadata and data should be easy to find for both humans and computers.	The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.	Rich metadata describing the resource held in an easily searchable platform.	Data and metadata should be easy to find by both humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.
Accessible	Once someone	Once the user finds the	Adoption of	Data and metadata should

	finds the required data, they need to know how the data can be accessed.	required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.	collaborative open-access platforms supporting the export of structured metadata.	be stored for the long term such that they can be easily accessed and downloaded or locally used by machines and humans using standard communication protocols.
Interoperable	The data needs to be easily integrated with other data for analysis, storage, and processing.	The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.	Adoption of standards (for data, metadata, and software) and defined vocabularies, common workflow and development protocols.	Data should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.
Reusable	Data should be well-described so they can be reused and replicated in different settings.	The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.	All products will have documentation and metadata that integrates different sources of information about authorship, formats, and lineage, for full provenance, traceability, and reproducibility (e.g. input data, metadata, diagnostics, tool version) and implement relevant standards for file formats.	Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

NASA SMD Open -Source Science Guidance (NASA SMD 2022b) have provided the following interpretation of the FAIR Principles:

- Findable - consistent and persistent descriptions make scientific data easy to find by both humans and computers;
- Accessible - use of standard, open protocols ensure data and metadata can be accessed by all;
- Interoperable - formal, accessible, and widely adopted semantics and vocabularies are used to expand data usability across systems and communities;
- Reusable - data are richly described according to standards to ensure they can be combined or replicated, and usage rights are clarified.

4. THE STATE OF COMMUNITY FAIR PRACTICES

The twenty-nine high-level requirements among the data (D), metadata (M), and infrastructure (IS) categories can be further subdivided into key concepts with associated elements (Fig. 4.1).

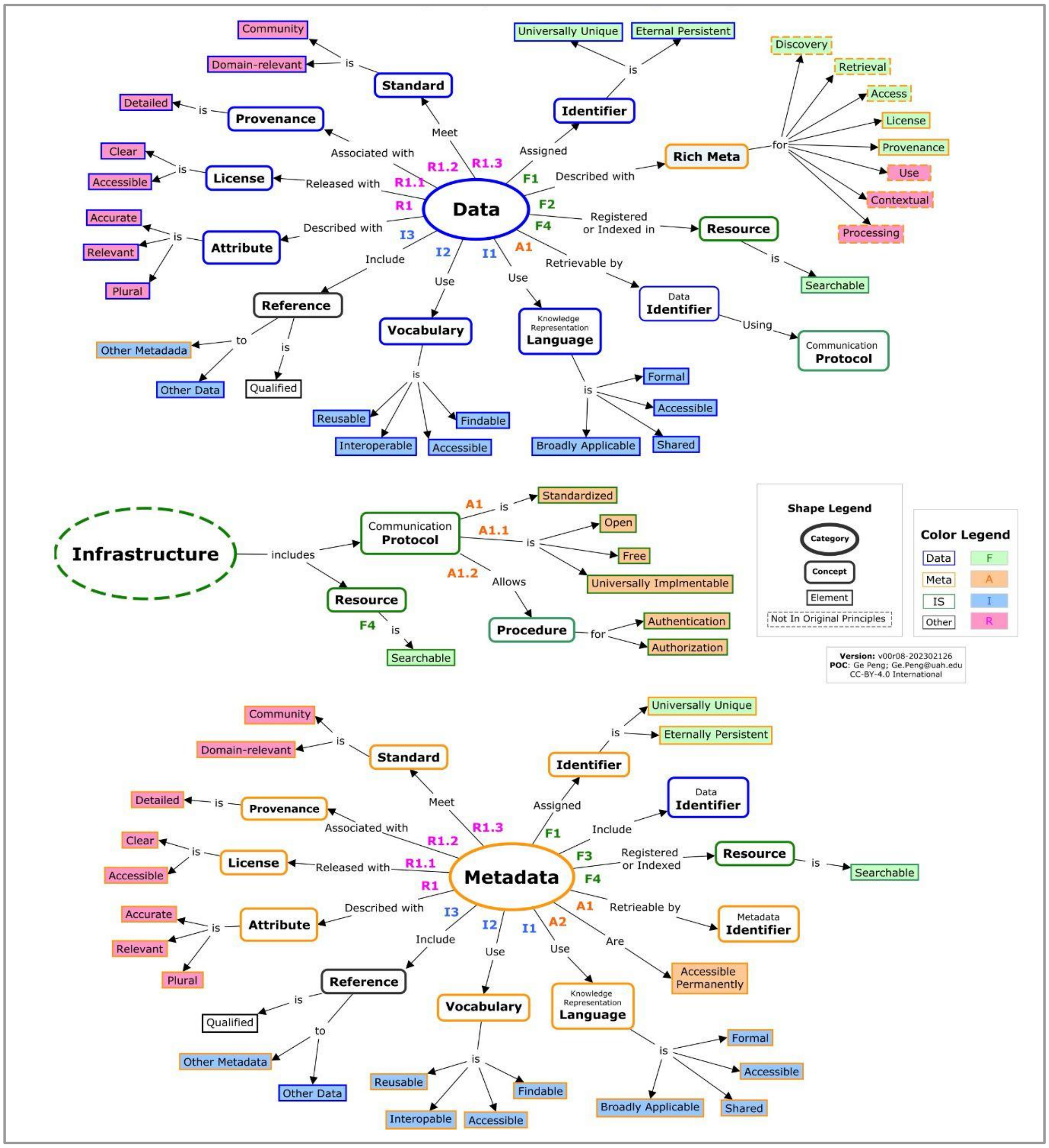

Figure 4.1. Diagram shows key categories and associated concepts and elements based on the FAIR Guiding Principles. {F, A, I, R}n.m denotes the association to the individual principle, e.g., R1.3. Created using Cmap.[2] Based on: Peng (2023). Version: v00r08-20230226. Creator: Ge Peng. Contributor: Siri Jodha Khalsa.

---

[2] https://cmap.ihmc.us/

Practices for identified concepts, including the application of Persistent Identifiers (PID), Rich Metadata, Knowledge-Representation Languages, Vocabularies, Attributes, and Standards, are often domain specific and digital-object-type dependent. The question of what entails rich metadata is one of the pain-points for ensuring the compliance of FAIR. So are FAIR vocabularies, as well as key elements such as data usage licenses and provenance. It is beyond the scope of this document to thoroughly discuss those issues. Instead, in this section, we will only outline some of the community practices in the areas of persistent identifiers, rich metadata, FAIR vocabularies, detailed provenance, data usage licenses, and communication protocols.

4.1 Persistent Identifiers

The persistent, unambiguous identification and location of (meta)data objects is now recognized as a necessary and central component of modern data infrastructure, especially in an interdisciplinary context (Klump et al. 2015; Klump et al. 2017; Wittenburg and Strawn 2018). Science builds on past work, and data available today from specific organizations in specific ways may still be very important many years from now when transitory details of ownership, data formats, and access protocols have all changed. Assigning and maintaining Persistent Identifiers (PIDs) allow us to add layers of abstraction that survive rapid technical change and changing access details and allow us to reconstruct past scientific workflows.

To that end, PIDs are a major component of the FAIR principles. They are explicitly mentioned in three of the principles and their use is implied in several other principles. The use of PIDs is also required by SPD-41a and the 2022 OSTP Memo[3] on "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research". The application of PIDs to enhance data sharing and reuse has also been a major focus of the Research Data Alliance (RDA) (Lannom 2014), but unfortunately, it has been more difficult than anticipated (Klump et al. 2017; Wittenburg and Strawn 2018).

Oddly, PIDs are not explicitly defined in the original Wilkinson et al. (2016) description of the FAIR principles. In this document we adopt the National Science and Technology Council Definition - "A persistent identifier is a digital identifier that is globally unique, machine resolvable and processable, and with an associated metadata schema"[4] that identifies an entity (e.g., individual researcher, digital research object) in perpetuity and is used to disambiguate as well as build associations between entities.

It turns out, however, that identity is a complex, intersectional concept. A review of this complexity, including an overview of different types of PIDs can be found in a related essay "Persistent Identifiers in FAIR-Data Infrastructure" (Parsons 2023).

[3] https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf
[4] https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf

Different types of PID systems may be utilized for different digital objects and for different disciplines and applications (e.g., Hakala 2010; Parsons et al. 2019; Parsons 2023). Common examples include:

- Handles, managed and globally resolved by the DONA Foundation[5].
- Digital Object Identifiers (DOIs), which are a form of Handle broadly used in scientific publishing and citation and resolved by the International DOI Foundation (IDF).[6]
- The Archive Resource Key (ARK), a highly-distributed, open system, which supports many research institutions, archives, and museums and provides some unique resolution and metadata features.[7] The service, n2t.net, is a global resolver for ARKs and other PIDs but ARK creators are free to create their own local resolvers.
- 'Persistent' URLs (PURLs) which are currently managed by the Internet Archive.[8]
- Identifiers.org is another resolver for registered URIs or Compact URIs broadly used in the life sciences community.

The use of PIDs for data is probably most mature in the context of data citation. NASA Earth Science Division has well-established guidelines[9] for this based on the Earth Science Information Partners Data Citation Guidelines (ESIP Data Preservation and Stewardship Committee 2019), which have also been adopted by major Earth science publishers such as the American Geophysical Union[10]. The USGS requires scientific data approved for release to be assigned a PID (USGS 2017). In its Citation Procedural Directive, NOAA has established the requirements for obtaining DOIs as PIDs for data and publications (NOAA EDMC 2021).

The NASA SMD Standards Working Group currently recommends that NASA repositories register DOIs through DataCite to enable data citation and provide guidelines on how to do so in compliance with SPD-41a[11].

Finally, it is worth noting that PIDs can be used for many things besides data, including organizations (Research Organizations Registry—ROR[12]), people (Open Researcher and Contributor IDs—ORCIDs[13]), and instruments (DOIs or Handles, see Stocker et al. (2020)). The use of these may contribute to data FAIRness, but they are not required at the same level as PIDs for data.

---

[5] http://handle.net/
[6] https://www.doi.org/
[7] https://arks.org/
[8] https://purl.archive.org/
[9] https://www.earthdata.nasa.gov/engage/doi-process
[10] https://data.agu.org/resources/agu-data-software-sharing-guidance
[11] https://github.com/nasa/smd-open-science-guidelines/blob/main/guidance/guideline001_doi_registration.md
[12] https://ror.org/
[13] https://orcid.org/

4.2 Rich Metadata

Rich metadata comes into play in the FAIR-F2 principle for describing data, which is constrained by the R1 principle by requiring "a plurality of accurate and relevant attributes" (Wilkinson et al. 2016; see Fig. 4.2). The FAIR principles do not define what constitutes rich metadata or what attributes, apart from specifically mentioning data usage license and provenance. As implied in Fig. 4.2, one can argue that rich metadata should include sufficient metadata elements for search and discovery, retrieval and access, near- and long-term understanding and use, and provide contextual and processing information, in addition to license and provenance. Rich description of metadata is necessary for machine interpretability, and supports infrastructure. However, what constitutes rich metadata is still evolving and tends to be domain specific, which does not necessarily support interdisciplinary use.

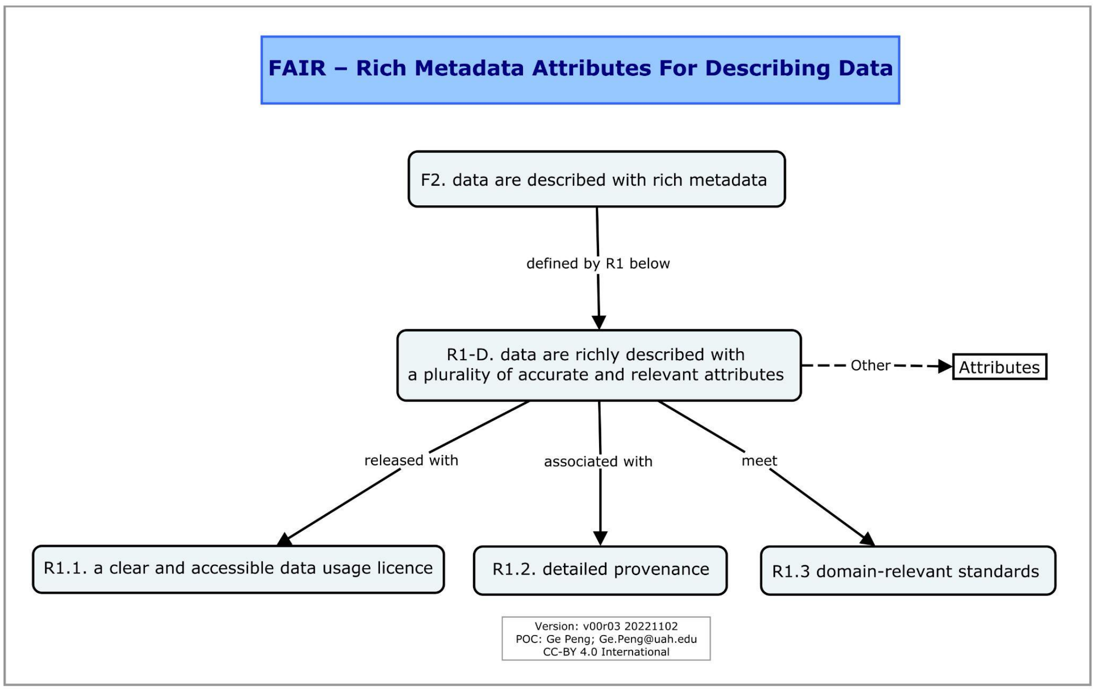

Figure 4.2. Mapping of rich metadata for describing data based on the FAIR Principles. Dashed arrow denotes potential additional attributes that are not explicitly defined in the FAIR Principles. Created using Cmap.

There is a rich collection of community metadata profiles and schemas. Among them are metadata elements defined by the Metadata Interest Group of Research Data Alliance (RDA) (Table A1), a metadata profile for Earth System Science data by GeoKUR (Henzen et al. 2021), core metadata profiles of the World Meteorological Organization (WMO) for search and discovery (Table A2), and the collection-level metadata elements of NASA's Common Metadata Repository (CMR) for NASA EOSDIS system (Table A3).

Examples of rich metadata for use include ISO 19165-2 standard (ISO 19165-2 2020) based on NASA's Preservation Content Specification (PCS - initially published in 2011 and updated by

Ramapriyan et al. 2022) and ESA/CEOS Earth Observation Preserved Data Set Content (ESA and CEOS WGISS 2015).

4.3 FAIR Vocabularies

The FAIR-I2 principle states: "(meta)data use vocabularies that follow FAIR principles" (Wilkinson et al. 2016), which recommends following FAIR Principles when creating and publishing a new vocabulary resource or updating an existing vocabulary. Cox et al. (2021) summarized FAIR Principles that applied to a vocabulary (Table 4.5) and proposed 10 simple rules for making a vocabulary FAIR (Table 4.6).

Table 4.5. Summary of FAIR Principles Applied to a Vocabulary (From Cox et al. 2021; Table 1)

FAIR Aspect	FAIR Principles Applied to a Vocabulary
F	● Each vocabulary is denoted by a persistent unique web identifier ● Each term is denoted by a persistent unique web identifier ● It is possible to search for a term or vocabulary and get a web identifier for it ● The vocabulary is available from at least one repository recognised by the community
A	● When the vocabulary or term identifier is de-referenced, a machine- or human-readable representation is returned, as requested
I	● At least one representation conforms to a community standard for vocabularies ● The vocabulary includes mapping relations to other vocabularies
R	● The license for use of the vocabulary is clear and accessible ● Enough metadata at vocabulary and term-level is provided, including provenance and maintenance information ● The definitions are sufficient for a user to understand what each term means

Table 4.6. Ten Simple Rules for Making a Vocabulary FAIR (From Cox et al. 2021; Table 3)

# Rule	Rule Description
1	Determine the governance arrangements and custodian of the legacy vocabulary
2	Verify that the legacy-vocabulary license allows repurposing, and agree on the license for the FAIR vocabulary
3	Check term and definition completeness and consistency in the legacy vocabulary
4	Establish a traceable maintenance-environment for the FAIR vocabulary content
5	Assign a unique identifier to (a) the vocabulary and (b) each term in the vocabulary
6	Create machine readable representations of the vocabulary terms
7	Add vocabulary metadata

8	Register the vocabulary
9	Make the vocabulary accessible for humans and machines
10	Implement a process for publishing revisions of the FAIR vocabulary

Examples of controlled vocabularies are listed in Table 4.7. Among them, ENVO, I-ADOPT, SOSA, and SWEET are ontologies. NERC and RVA are services for hosting community vocabularies. Note: These controlled vocabularies have not yet been evaluated to determine compliance with the FAIR Principles.

Table 4.7. Examples of Controlled Community Vocabularies and Ontologies

Acronym	Description	Version	Source
CF	Climate and Forecast (CF) Metadata Conventions include controlled vocabularies of standard names, area types, and standardized regions for NetCDF files.[14] The CF data model and reference software can be found in Hassell et al. (2017).	Version 1.10, 31 August 2022	Eaton et al. (2022); https://cfconventions.org/index.html
ENVO	The Environmental Ontology (ENVO) is a FAIR-compliant community ontology for the concise, controlled description of environments.	Version: 14 May 2021	Buttigieg et al. (2016). https://sites.google.com/site/environmentontology/
GCMD	NASA Global Change Master Directory (GCMD) keywords are a hierarchical set of controlled Earth science vocabularies, including platforms, instruments, measurement names, etc.	Version 14.9	GCMD (2022); https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords
I-ADOPT	RDA InteroperAble Descriptions of Observable Property Terminology (I-ADOPT) is an ontology as an interoperable framework for representing observable properties.	Version 0.0.1; 6 December 2021	https://i-adopt.github.io/
NERC Vocabularies Server	Natural Environment Research Council (NERC) Vocabularies Server (NVS) hosts a list of vocabularies used by the geoscience community. NVS includes multiple elements of the GCMD.	On-going	https://vocab.nerc.ac.uk/collection/
RVA	Research Vocabularies Australia (RVA) is a service - both an aggregator and a publisher of vocabularies - and contains a copy of the GCMD vocabularies.	On-going	https://vocabs.ardc.edu.au/
SOSA	SOSA is a formal but lightweight specification for Sensors, Observations, Samples, and Actuators (SOSA)	2015	Janowicz et al. (2015); https://www.w3.org/2015

[14] Additional NetCDF conventions can be found: https://www.unidata.ucar.edu/software/netcdf/conventions.html

			/spatial/wiki/SOSA_Ontology
SWEET	ESIP Semantic Web for Earth and Environmental Terminology (SWEET) is a highly modular ontology suite with ~6000 concepts in ~200 separate ontologies covering Earth system science.	Version 3.5.0, 13 July 2022	https://github.com/ESIPFed/sweet
UDUNITS	UDUNITS is a UNIDATA software package supporting units of physical quantities.	Version 2.2.28; 2 December 2020	https://www.unidata.ucar.edu/software/udunits/

Some glossaries have been developed for the research data management community, such as those maintained by the Committee on Data of the International Science Council (CODATA) [15], Data Management Association (DAMA) UK [16], the Data Foundation and Terminology (DFT) Interest Group of Research Data Alliance (RDA) [17], Consortia Advancing Standards in Research (CASRAI) [18], and ISO/TC 211 [19] (see Peng et al. 2021, Appendix A for additional discussion on terms and definitions.)

Additionally, EOSDIS maintains a list of acronyms and symbols [20] and a glossary.[21] An individual program, DAAC, or project may also maintain a list of relevant vocabularies, for example, the glossary for NASA Earth Observatory [22], standard names for atmospheric composition variables (under ESCO review), and the glossary for NASA Catalog of Archived Suborbital Earth Science Investigations (CASEI).[23]

4.4 Detailed Provenance

Provenance captures the information about what was changed; when, why, and how the change was made; and who made or was responsible for the change. It may also describe whence and how the current digital object was derived. The detailed provenance in FAIR refers to that of both data and metadata.

A systematic review of data quality in Earth data in the last decade has noted that the provenance or lineage information was rarely included (Yang et al. 2013). An overview of geoscience data

[15] https://codata.org/rdm-glossary/
[16] https://www.dama-uk.org/Glossary
[17] https://smw-rda.esc.rzg.mpg.de/dft-2.0.html
[18] https://casrai.org/rdm-glossary/
[19] https://isotc211.geolexica.org/
[20] https://www.earthdata.nasa.gov/learn/acronym-list
[21] https://www.earthdata.nasa.gov/learn/glossary
[22] https://earthobservatory.nasa.gov/glossary/all
[23] https://impact.earthdata.nasa.gov/casei/glossary/

provenance frameworks can be found at Di et al. (2013). Significant attention has been paid to provenance in recent years as evidenced by a series of workshops on provenance and annotations, for example, see Mattoso and Glavic (2016). An example of tracing of provenance as completely as possible for the figures in the Third National Climate Assessment (NCA3) that predominantly use NASA's data is given in Ramapriyan et al. (2016). The DataONE project has developed provenance systems that "enable reproducible research and facilitate proper attribution of scientific results transitively across generations of derived data products." (Cao et al, 2016). Compliance with the ISO 19165-2 standard (ISO 19165-2 2020) will ensure that the provenance and context are captured for Earth science data products.

Two somewhat exclusive frameworks are starting to be used in the Earth science community. First, The World Wide Web Consortium (W3C), an international standards-setting organization for the web, established the W3C provenance (PROV) family of specifications to capture machine-interoperable provenance information.[24] Second, the ISO 19115 metadata standard has recently been extended to include a model to record all inputs necessary to execute processing steps, as well as to describe the process itself (ISO 19115-2)[25]. Efforts have been made to reconcile these two models, e.g., Jiang et al. (2018), with varying success. In addition, automated tools have been developed to capture and/or visualize provenance information as processing is performed (e.g., provo: https://github.com/GeoinformationSystems/provo).

4.5 Data Usage License

The FAIR-R1.1 principle is about including a clear and accessible data usage license when releasing (meta)data in order for users to know whether they have permission to use the data, and if so, under what license conditions and who should be cited if the data are re-used (Wilkinson et al 2016). U.S. federally funded scientific data are released under agency-specific policies. For example, NASA's Earth science program has a long-standing data and information policy for "open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available." [26] The statement is clear and publicly accessible and may be included as a free text in NASA collection-level metadata records as the part of the *Use Constraints* element of UMM-C, but not necessarily with a machine-interoperable format at this time.

The SPD-41a policy recommends releasing scientific data that have no other restrictions with a Creative Commons Zero (CC0) license (NASA SMD 2022a). Individual NASA DAACs, such as SEDAC, have adopted the Creative Commons Attribution 4.0 International (CC BY) license[27], which enables data producers from various disciplines to contribute data relevant to human

[24] https://www.w3.org/TR/prov-dm/
[25] https://www.iso.org/standard/67039.html
[26] https://www.earthdata.nasa.gov/data-and-information-policy
[27] https://creativecommons.org/licenses/

interactions in the environment without restrictions on how the data are used. Offering the CC BY license as a simple way for data producers to share their data provides teams of data producers, who often receive funding from a variety of sources, to agree on an open data sharing approach that also encourages data users to cite the data. For example, on behalf of a team of data producers, a member of that team can allow their data to be used by anyone for any purpose, by indicating that the data are licensed under the CC BY license. Furthermore, when publishing data that have been licensed under the CC BY license, SEDAC provides a simple rights declaration statement within the metadata so that potential users are able to understand their rights for data use, without having to wade through legal jargon. Providing a link to the CC BY license within the metadata also enables users to read the summary describing the license and to access the actual license if they are interested in reading the legal language of the license. An example of a dataset that contains a simple rights declaration statement within the metadata can be seen by viewing the SEDAC dataset, Country Trends in Major Air Pollutants.[28]

The Software Package Data Exchange (SPDX) specification, an international open standard (ISO/IEC 6692:2021), provides a common and machine-interoperable format for licenses at: https://spdx.org/licenses/

4.6 Communication Protocol

A communications protocol is a set of formal rules describing how to transmit or exchange data, especially across a network. A standardized communications protocol is one that has been codified as a standard, such as the Hypertext Transfer Protocol (HTTP).
Examples of standardized communication protocols for making data available include:[29]

- From a web server via HTTP using a browser,
- From a file server via File Transfer Protocol (FTP) using an FTP client application,
- Through a well-documented Application Programming Interface (API).

There are various community data servers developed building on open web APIs, such as the Thematic Real-time Environmental Distributed Data Services (THREDDS – data servers are primarily built on OPeNDAP which is a widely used, subsetting data access method extending the HTTP protocol [30]), as well as organizational search platforms such as NASA Earthdata Search [31] and NOAA *OneStop* Data Portal. [32]

[28] https://doi.org/10.7927/et1q-jj80
[29] https://ardc.edu.au/resource/standardised-communications-protocols/
[30] https://www.unidata.ucar.edu/software/tds/current/
[31] https://search.earthdata.nasa.gov/search
[32] https://data.noaa.gov/onestop/

Within Amazon Web Services™ (AWS)-based cloud services and NASA's Earthdata Cloud platform [33], one can utilize a URL which specifies the location of an Amazon S3-based [34] object-storage bucket to securely access data on the cloud. Note: The Amazon S3™ protocol, while unique, is proprietary to AWS as a specific object-storage and access protocol.

4.7 FAIR Implementation Profiles and Convergence Matrix

Aiming to accelerate community convergence on FAIR implementation options, the GO FAIR community launched and hosted the development of machine-actionable FAIR implementation profiles (FIP) by collecting community implementation choices.[35] These community-specific, comprehensive FIP collections are made openly available to other communities. One can also create their own FAIR implementation profile using the FIP mini-questionnaire that is available on the https://bit.ly/yourFIP webpage (the questions are captured in Fig. 4.3).

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

Figure 4.3. Screenshot of the FAIR Implementation Profile (FIP) questionnaire template located at https://bit.ly/yourFIP.

The FAIR Convergence Matrix is an online platform, managed by the FAIR Convergence Matrix Working Group[36], that composes mature and trustworthy FIPs (Schultes et al. 2020). The FAIR Convergence Matrix is also used "to track the evolving landscape of FAIR implementations" and to guide self-identified communities in developing their own FAIR implementations and practices.[37]

[33] https://www.earthdata.nasa.gov/eosdis/cloud-evolution
[34] Amazon Simple Storage Service (Amazon S3) is an object storage service, https://docs.aws.amazon.com/s3/index.html
[35] https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/
[36] https://osf.io/n7uwp/
[37] https://github.com/go-fair-ins/GO-FAIR-Ontology/tree/master/Models/FIP

5. THE STATE OF COMMUNITY FAIR ASSESSMENTS AND IMPLEMENTATIONS

The FAIR Principles were not originally designed to be an assessment approach per se (Mons et al. 2017), but many organizations are already using the FAIR Principles in this way (Peng et al. 2022). To this end, unfortunately, the FAIR Principles are subject to different interpretations, resulting in a wide range of implementations for assessing data FAIRness (RDA FAIR Data Maturity Model Working Group 2018). For example, the RDA FAIR Data Maturity Model Working Group (2020) has developed a set of core assessment criteria for FAIRness, also known as RDA FAIR data maturity indicators (DMIs), to aim for direct comparison of assessment results. Although it is not straightforward to measure the FAIRness of data objects, based on a subset of the RDA FAIR DMIs, Devaraju and Huber (2021) derived a list of practical tests as core metrics with underlying contextual assumptions on repository and domain requirements, data types, and formats, etc., to develop an online tool, aka, F-UJI Tool,[38] to systematically and automatically assess the FAIR compliance of scientific data. Peters-von Gehlen (2022) evaluated five FAIRness assessment tools, which are recaptured here in Table 5.1, and concluded that, even with their individual strengths, none of the five approaches is fully fit for evaluating (discipline-specific) FAIRness. This evaluation is also calling for a hybrid approach because manual approaches can capture "the contextual aspects of FAIRness relevant for reuse", whereas automated approaches need to build on "the strictly standardized aspects of machine actionability." (Peters-von Gehlen 2022).

Table 5.1. Examples of FAIRness Assessment Tools
(From Peters-von Gehlen 2022)

Tool	Method	Covered FAIR Dimensions	Reference	Evaluation URL
Checklist for Evaluation of Dataset Fitness for Use	manual	n/a *	Austin et al. (2019)	https://docs.google.com/forms/d/1cqvWAC5BJ8aiJ0m4YXN8eGCijcxxekyWz73SpnEMICc/viewform?edit_requested=true
FAIR Maturity Evaluation Service	automated	F: 8, A: 5, I: 7, R: 2	Wilkinson et al. (2019; updated 2020)	https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/collections/new/evaluate
FAIRshake [dataset rubric]	hybrid	F: 3, A: 1, I: 0, R: 5	Clarke et al. (2019)	https://fairshake.cloud/rubric/8/
F-UJI	automated	F: 7, A: 3, I: 4, R: 10	Devaraju et al. (2021)	https://www.f-uji.net/
RDA FAIR Data Maturity Indicators	manual	F: 13, A: 12, I: 10, R: 10	Bahim et al. (2020)	N/A

[38] https://www.f-uji.net/?action=test

* A questionnaire consists of twenty questions covering aspects of dataset identification, state of the repository's certification, data curation, metadata completeness, accessibility, data completeness and correctness as well as findability and interoperability. The topics covered by the questions map very well on to the FAIR principles (Peters-von Gehlen 2022).

Additional examples of FAIR data and metadata assessment models can be found in Peng et al. (2021, updated 2022).

Australia Research Data Commons (ARDC) created an online FAIR Data Self Assessment Tool, aiming to help users to check the 'FAIRness' of a dataset with tips on how to enhance its FAIRness.[39]

A list of community FAIR implementation examples is captured in Table 5.2 and details of those implementations can be found in a supplementary Google Spreadsheet (NASA ESDSWG Wiki access may be required).[40] This Google Spreadsheet also captures other relevant information and will be treated as a living document.

Table 5.2. Examples of Community FAIR Implementations

Reference	Description	DOI/URL
ARDC FAIR Data Guidelines for Project Output	Guidelines and resources for ARDC projects on how data outputs involved in the project will be made more FAIR	https://ardc.edu.au/resource/fair-data-guidelines-for-project-data-outputs/
Cox et al. (2021)	Ten simple rules for making a vocabulary FAIR	https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009041
Jones and Slaughter (2019)	DataONE FAIR Checks	
Data Sharing Registry	A registry of terminology artifacts, models/formats, reporting guidelines, and identifier schemas	https://fairsharing.org/search?fairsharingRegistry=Standard
Devaraju and Huber (2021)	A set of actionable core metrics for assessing the FAIR compliance of scientific data	https://doi.org/10.1016/j.patter.2021.100370
FAIR Implementation Profiles	A repository that captures the comprehensive set of implementation choices made at the discretion of individual communities of practice	https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/
OpenAIRE	A set of guidelines for researchers on how to make your data FAIR	https://www.openaire.eu/how-to-make-your-data-fair

[39] https://ardc.edu.au/resource/fair-data-self-assessment-tool/

[40] https://wiki.earthdata.nasa.gov/display/ESDSWG/O%27FAIR+ESDSWG+-+FAIR+Practices+%28Descriptions_Metrics_Assessments_Tools%29+Collection

O'Toole and Tocknell (2022)	Implementation of FAIR of All-Sky Virtual Observatory	https://arxiv.org/pdf/2203.10710.pdf
Peng et al. (2021)	Guidelines and practical examples of making quality information FAIR	https://doi.org/10.31219/osf.io/xsu4p
Pirani et al. (2022)	Guidance on making IPCC Process FAIR	https://doi.org/10.5281/zenodo.6504469
Wilkinson et al. (2019)	FAIR maturity evaluation framework	https://doi.org/10.1038/s41597-019-0184-5

6. THE STATE OF NASA FAIR ASSESSMENTS AND IMPLEMENTATION

NASA ESD promotes the full and open sharing of all data with the research and applications communities, private industry, academia, and the general public.[41] To this end, standards and practices are developed to ensure and/or enhance the management and stewardship of SMD-funded research data.[42]

A subjective evaluation of EOSDIS with respect to the FAIR principles was carried out by members of the ESDIS Project in collaboration with the 12 EOSDIS DAACs (Ramapriyan and Lynnes 2019; Ramapriyan and Behnke 2020). Each of the DAACs performed a self-assessment as well as EOSDIS as a whole. Rather than assess each of the over 10,000 datasets for compliance with FAIR principles, this assessment evaluates how the overall tools and approaches used in the EOSDIS and the DAACs satisfy the FAIR principles. For each of the 15 principles, a verbal description is provided of how the principle is addressed from the perspectives of human actionability (HA) and machine actionability (MA). Also, a score between 1 and 10 is assigned subjectively for each of the 15 principles under HA and MA, where 1 indicates the lowest compliance and 10, the highest. The scores averaged over all the sub-principles under each of F, A, I, and R are shown in Table 6.1 below.

Table 6.1. Summary of FAIRness assessment of NASA Earth Science Data and Information Systems (ESDIS). From: Ramapriyan and Behnke 2020

FAIR Dimension	HA	MA	Comments
Findability	7	6	More work needed
Accessibility	10	8	Excellent in HA; MA needs some improvements

[41] https://www.earthdata.nasa.gov/data-and-information-policy

[42] https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices

Interoperability	8	8	Doing well here; some improvements needed
Reusability	9	7	Excellent in HA; MA needs work

The detailed evaluation can be found at the supplementary Google Spreadsheet [43] (NASA ESDSWG Wiki access may be required).

The Data Product Development Guide Working Group, has developed a document titled Data Product Development Guide (DPDG) for Data Producers (Ramapriyan and Leonard 2020). The appendices D and E of this document provide tables of metadata attributes recommended to be used in the data products submitted to the EOSDIS DAACs for archival and distribution. A total of 107 attributes are listed including global attributes and variable-level attributes. The global attributes are further categorized into those needed for Interpretability, Discovery, Geolocation, Temporal Location, Usability, Provenance (General, Attribution, and Lineage). Each of these attributes have been characterized with a justification for use, and labeled with one or more of the letters F, A, I, and R to show how they support FAIR. The numbers and percentages of attributes that have been tagged with each of the four letters are shown in Table 6.2.

Table 6.2. Summary of a breakdown of recommended metadata attributes in data product development guide (DPDG) for data producers (Ramapriyan and Leonard 2020) for supporting FAIR. (The total number of the recommended metadata attributes is 107.)

FAIR Dimension	# of Attributes	Percentage of Attributes Tagged
F	46	43%
A	35	33%
I	48	45%
R	52	49%

The NASA ESDIS Metrics System (EMS) (NASA EOSDIS 2023) establishes requirements and methods for each DAAC to collect data activity and usage metrics, which are routinely analyzed and provided to NASA management to inform the best allocation of resources for the scientific user community and also to improve data services. User satisfaction metrics provide a key measurement for "fitness for use" in NASA EOSDIS data services. In particular, the American

[43] https://wiki.earthdata.nasa.gov/display/ESDSWG/O'FAIR+WG+-+FAIR+Practices+(Descriptions_Metrics_Assessments_Tools)+Collection

Customer Satisfaction Index (ACSI) survey (NASA EOSDIS 2023) of users of DAACs has provided uniform, consistent and uninterrupted measurements of data services for each of the twelve DAACs since 2004. The ACSI, along with other metrics, can be used to identify barriers in data services and benchmark progress.

Using GES DISC, a multidisciplinary data center, as an example, Liu et al. (2022) described current NASA metrics and recommended that the metrics could be integrated in conjunction with the FAIR Guiding Principles.

Additional examples of NASA FAIR assessments or implementation are listed in Table 6.3.

Table 6.3. Additional Examples of NASA FAIR Assessments or Implementation

Assessment Type	Method	Reference
Enhancing the FAIRness of Carbon Monitoring System Data	Implementing FAIR Principles	Singh et al. (2019)
Applying FAIR and TRUST Principles to Support the Earth Science Community at the Global Hydrometeorology Resource Center DAAC	Adopting FAIR & TRUST Principles	Ellett et al. (2021)
NASA's Catalog of Archived Suborbital Earth Science Investigations (CASEI)	Online platform - Adopting FAIR Principles	Wingo et al. (2022)
The Use of Atmospheric Composition Variable Standard Names in Airborne and Field Data Products	Controlled vocabulary	Silverman et al. (2022)

7. SUMMARY AND FUTURE WORK

The O'FAIR WG was established in May 2022 as one of NASA's Earth Science Data System Working Groups (ESDSWG). This WG has been exploring community practices with the goal of developing a guide document to ensure that Earth science data and information funded by NASA are open and free as well as FAIR. The activities of this WG are motivated by the extensive, diverse, globally used Earth science data that have been open and free for nearly three decades, as well as the recently published Science Mission Directorate Program Directive (SPD-41a) that explicitly recommends that all SMD-funded data should follow the FAIR Guiding Principles. Guidelines on how to best follow these principles are expected to be very useful to data producers and data managers whose current and future practices can benefit from the experiences of others. Future program and project managers should find such guidelines to be useful in defining data management and stewardship requirements.

This report is a first step in developing principle-by-principle guidance for making the NASA Earth science data more FAIR. It provides a high-level overview of the current landscape of community FAIR practices. While the report is not exhaustive in the coverage of the extensive community practices that have come into being over the last few years, it provides a starting point to gauge the current community FAIR practices landscape and to help develop a guide document pertaining to NASA-funded Earth science data and information.

The O'FAIR WG started its activities by familiarizing the group members with NASA policies and the U.S. FAIR initiatives. Expert presenters were invited to the WG meetings to brief the group and discuss policy implications and issues. The WG members reviewed and gathered information from their network and current literature that is relevant to the FAIR Principles and to applying the FAIR Principles in practice. A crosswalk was developed to represent conceptual relationships between the stated data requirements in NASA's SPD-41a and the FAIR principles.

Various organizations have interpreted the FAIR principles in slightly different ways to suit their respective needs. This report shows four representative interpretations – DataOne, GO FAIR, IPCC and the Swiss National Science Foundation. It also includes the interpretation from NASA SMD Open -Source Guidance.

To facilitate the FAIRness assessment process, each principle has been mapped into one of three categories: Data (D), Metadata (M), and Infrastructure (IS). As expected, the largest percentage of the FAIR requirements pertains to metadata, data coming a close second, and infrastructure coming last. A diagram is included that shows key categories and associated concepts and elements based on the FAIR Guiding Principles as well as the relations among them. Some of the community practices in the areas of persistent identifiers, rich metadata, FAIR vocabularies, detailed provenance, data usage licenses, and communication protocols are further discussed in this report.

Even though the FAIR Principles were not originally designed to be an assessment approach per se, many organizations are already using the FAIR Principles in this way. To this end, the FAIR Principles are subject to different interpretations, resulting in a wide range of implementations for assessing data FAIRness. This has also led to a variety of assessment tools, some of which are discussed in this report. Also, several examples are shown of community FAIR implementations. The ESDIS Project and some of the DAACs have worked on FAIRness assessment and implementation of FAIR-compliant datasets. Some of these are identified in this report.

Equipped with the information collected and knowledge gained from reviewing and consolidating the information, the O'FAIR WG will next focus on developing a Guide Document to provide practical guidance to stakeholders on being compliant with the FAIR Principles.

ACKNOWLEDGEMENT

Support from NASA ESDSWG, affiliated projects and DAACs is appreciated for the O'FAIR WG to be established and for members to participate in the working group activities. We thank Stephen (Steve) Olding and Sara Lubkin for serving as liaisons between the working group and the ESDIS leadership; Together with many working group members as the point-of-contact for individual DAACs, they have brought different perspectives and discipline-specific knowledge that are beneficial to the development of this document.

This document is the result of participation and contributions from the O'FAIR WG members (See Table below for their name, affiliation, and role/subject area.) Special THANK-YOU goes to Steve Olding for providing beneficial guidance during the planning phase, for creating and maintaining the O'FAIR mailing list and wiki website, as well as for supporting the monthly telecons.

We thank Steven Crawford and Melissa Cragin for sharing their knowledge on the SPD-41a and the FAIR Principles, Lesley Wyborn, Australian National University/Australian Research Data Commons (ARDC), for the information on SPDX interoperable licenses, and Juha Hakala, Library of Finland, for the information on PIDs.

We thank the members of SMD Standards Working Group and UAH/MSFC IMPACT project for reviewing Fig. 2.1, especially Sumant Jha, Emily Foshee, Dan Berrios, Brian Thomas, and Graham Berriman for beneficial comments.

Name, Affiliation, and Role/Subject Area of NASA O'FAIR WG Members.
(People whose names are in bold contributed to or reviewed the proposal
for establishing the working group.)

Name (First Last)	**Affiliation**	**Role; Subject Area**
Ge Peng	UA Huntsville/MSFC IMPACT	Chair; Scientific data stewardship
Robert R. Downs	SEDAC; Columbia University; ESCO	Co-Chair
Hampapuram Ramapriyan	SSAI/NASA GSFC ESDIS	NASA Data Systems, data preservation and stewardship, information quality
David Moroni	JPL; MEaSUREs, formerly PO.DAAC	Data Manager & Producer
Yaxing Wei	ORNL.DAAC	
Mark A. Parsons	NASA Chief Science Data Office	Persistent Identifiers, SMD Standards and Guidelines, Data Stewardship
Manil Maskey	ESDS HQ on ML	
Keith Bryant	GES DISC	UI/UX/Usability
Zhong Liu	George Mason University/GES DISC	Data curator and user needs
Carl Mears	Remote Sensing Systems	Data producer
Leigh Sinclair	GHRC.DAAC	

Stephanie Wingo	UA Huntsville/MSFC IMPACT; ADMG	Airborne
Rudi Gens	ASF.DAAC	
Jeoffrey Stano	GHRC.DAAC	
Deborah Smith	UA Huntsville/IMPACT; ADMG	
Douglas Rao	NCICS/NCAI	Observer; AI/ML
Michele Thornton	ORNL.DAAC	
Bhaskar Ramachandran	LAADS.DAAC, GSFC	
Stephen (Steve) Olding	GSFC, ESDSWG	
Sara Lubkin	GSFC, ESDIS; ESCO	
Francis Lindsay	GSFC	
Siri Jodha S Khalsa	NSIDC.DAAC; ESCO	Observer; Standards
Shannon Leslie	NSIDC.DAAC	Observer;
Karen Yuen	JPL/TOPS	Observer;
Tammy Walker	ORNL	Observer;
Shawn R. Smith	COAPS/FSU; NASA ACCESS	In-situ marine data management; Interoperability; Vocabulary

REFERENCES

Austin, C., H. Cousjin, M. Diepenbroek, J. Petters, M. Soares e Silva, and World Data System-Research Data Alliance, 2019: WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations. https://doi.org/10.15497/rda00034

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, https://doi.org/10.5334/dsj-2020-041

Buttigieg, P.L., E. Pafilis, S.E. Lewis, M.P. Schildhauer, R.L. Walls and C.J. Mungall, 2016: The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *J Biomed Semant,* 7, https://doi.org/10.1186/s13326-016-0097-6

Cao, Y., C. Jones, V. Cuevas-Vicenttin, M. B. Jones, and others, 2016: DataONE: A Data Federation with Provenance Support. In: Mattoso, M., Glavic, B. (eds) Provenance and Annotation of Data and Processes. IPAW 2016. *Lecture Notes in Computer Science*, 9672. Springer, Cham. https://doi.org/10.1007/978-3-319-40593-3_28

Clarke D.J.B., L. Wang, A. Jones, M.L. Wojciechowicz and others, 2019: FAIRshake: Toolkit to Evaluate the FAIRness of Research Digital Resources. *Cell Syst.* 9. https://doi.org/10.1016/j.cels.2019.09.011

Cox, S.J.D., A.N. Gonzalez-Beltran, B. Magagna, and M.-C. Marinescu, 2021: Ten simple rules for making a vocabulary FAIR. *PLoS Comput. Biol.* 17, https://doi.org/10.1371/journal.pcbi.1009041

Devaraju, A. and R. Hurt, 2021: An automated solution for measuring the progress toward FAIR research data. *Patterns*, 2, https://doi.org/10.1016/j.patter.2021.100370

Di, L., Y. Peng, H.K. Ramapriyan, R.L. King, 2013: Geoscience Data Provenance: An Overview. *IEEE Transactions on Geoscience and Remote Sensing*, 51. https://dx.doi.org/10.1109/TGRS.2013.2242478

Eaton, B., J. Gregory, B. Drach, K. Taylor, S. Hankin, and others, 2022: NetCDF Climate and Forecast Metadata Conventions. Version: 1.10, 31 August 2022. Available at: https://cfconventions.org/Data/cf-conventions/cf-conventions-1.10/cf-conventions.pdf

Ellett, W., G. Stano, L. Sinclair, S. J. Graves, M. Maskey, and D. Wright, 2021: Applying FAIR and TRUST Principles to Support the Earth Science Community at the Global Hydrometeorology Resource Center DAAC. Abstract IN42A-08. AGU 2021 Fall Meeting. Abstract available at: https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/977351

ESIP Data Preservation and Stewardship Committee, 2019: Data Citation Guidelines for Earth Science Data. Ver. 2. *Earth Science Information Partners*. https://doi.org/10.6084/m9.figshare.8441816.

ESA and CEOS WGISS, 2015: Long Term Preservation of Earth Observation Space Data - Earth Observation Preserved Data Set Content (PDSC). Version: 1.0 September 2015. Available at: https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf

GCMD (Global Change Master Directory), 2022: GCMD Keywords, Version 14.9, Greenbelt, MD: Earth Science Data and Information System, Earth Science Projects Division, Goddard Space Flight Center (GSFC), National Aeronautics and Space Administration (NASA). URL (GCMD Keyword Forum Page): https://forum.earthdata.nasa.gov/app.php/tag/GCMD+Keywords

Hakala, J., 2010: Persistent identifiers - an overview. *National Library of Finland*. Available at: http://www.persid.org/downloads/PI-intro-2010-09-22.pdf

Hassell, D., J. Gregory, J. Blower, B. N. Lawrence, and K. E. Taylor, 2017: A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1). *Geosci. Model Dev.*, 10, https://doi.org/10.5194/gmd-10-4619-2017

Henzen, C., A. Rümmler, and M. Wagner, 2021: Recommendations for Developing a Metadata Profile for Earth System Science Data. *Zenodo*. https://doi.org/10.5281/zenodo.4916698

ISO 19165-2, 2020: Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products.

Janowicz, K., A. Phil, and S. Simmons, 2015: Sensor Observation Sampling Actuator ontology. https://www.w3.org/2015/spatial/wiki/SOSA_Ontology

Jiang, L., P. Yue, W. Kuhn, C. Zhang, C. Yu, and X. Guo, 2018: Advancing interoperability of geospatial data provenance on the web: Gap analysis and strategies. *Computers & Geosciences*, 117. https://doi.org/10.1016/j.cageo.2018.05.001

Jones, M. and P. Slaughter, 2019: Quantifying FAIR: metadata improvement and guidance in the DataONE repository network. *DataONE Webinar*, May 14, 2019. Slides are available at: https://www.dataone.org/uploads/dataonewebinar_jonesslaughter_fairmetadata_190514.pdf

Klump, J., R. Huber, and M. Diepenbroek, 2015: DOI for geoscience data-how early practices shape present perceptions, *Earth Science Informatics*, pp. 1-14. https://doi.org/10.1007/s12145-015-0231-5

Klump, J., F. Murphy, T. Weigel, M/A/ Parsons, 2017: 20 Years of Persistent Identifiers--Applications and Future Directions, *Data Science Journal*, 16. http://doi.org/10.5334/dsj-2017-052

Lannom, L., 2014, Special Issue on the Research Data Alliance, D-Lib, 20(1/2), http://dx.doi.org/10.1045/january2014-editorial

Liu, Z., C.-L. Shie, A.J. Ritrivi, G.-D. Lei, G.T. Alcott, M. Greene, J. Acker, J.C. Wei, D.J. Meyer, A. Li, and A.F. Al-Jazrawi, 2022: Developing Metrics for NASA Earth Science Interdisciplinary Data Products and Services. *Data Science Journal,* 21, http://doi.org/10.5334/dsj-2022-005

Mattoso, M. and B. Glavic (eds.), 2016: Provenance and Annotation of Data and Processes, 6th International Provenance and Annotation Workshop, IPAW 2016, McLean, VA, USA, June 7–8, 2016, Proceedings. https://doi.org/10.1007/978-3-030-80960-7

Mons, B., C. Neylon, J. Velterop, M. Dumointier, L. Santos, and M. D. Wilkinson, 2017: Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, 37. https://doi.org/10.3233/ISU-170824

NASA, 2011: Earth Science Data Preservation Content Specification (PCS). Updated: Revision C, May 3, 2022. Available at: https://www.earthdata.nasa.gov/s3fs-public/2022-06/NASA_ESD_Preservation_Spec_C.pdf

NASA EOSDIS, 2023, American Customer Satisfaction Index (ACSI) Reports. Available at: https://earthdata.nasa.gov/eosdis/system-performance/acsi-reports (Last accessed on February 6, 2023)

NASA SMD, 2022a: Scientific Information Policy for the Science Mission Directorate. SMD Policy Document SPD-41a. Version: September 26, 2022. Available at: https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD-information-policy-SPD-41a.pdf

NASA SMD, 2022b: Open-Source Science Guidance. Version: December 8, 2022. Available at: https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD%20Open-Source%20Science%20Guidance%20v1%2020221208.pdf

NOAA EDMC (Environmental Data Management Committee), 2021: NOAA Data and Publication Citation Procedural Directive. Version: v.2.0. Available at: https://nosc.noaa.gov/EDMC/documents/EDMC-PD-DataPublicationCitation_2.0-accessible.pdf

O'Toole, S. and J. Tocknell, 2022: FAIR standards for astronomical data. Available at" https://arxiv.org/pdf/2203.10710.pdf

Parsons, M.A., R. Duerr, and M. Jones, 2019: The History and Future of Data Citation in Practice. *Data Science Journal*, 18, https://doi.org/10.5334/dsj-2019-052

Parsons, M.A., 2023: Persistent Identifiers in FAIR-Data Infrastructure. *Zenodo*, https://doi.org/10.5281/zenodo.8007232

Peng, G, C. Lacagnina, I. Ivánová, R.R. Downs, H.K. Ramapriyan, A. Ganske, and others, 2021: International Community Guidelines for Sharing and Reusing Quality Information of Individual Earth Science Datasets. Updated on March 26, 2022. *Open Science Framework,* https://doi.org/10.31219/osf.io/xsu4p

Peng, G., W.S. Gross, and R. Edmunds, 2022: Crosswalks among stewardship maturity assessment approaches promoting trustworthy FAIR data and repositories. *Sci Data* 9, 576, https://doi.org/10.1038/s41597-022-01683-x

Peng, G., 2023: Finding Harmony in FAIRness. *Eos,* https://doi.org/10.1029/2023EO230216

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, https://doi.org/10.5334/dsj-2022-007

Pirani, A., A. Alegria, A. Al Khourdajie, W. Gunawan, J. M. Gutiérrez, K. Holsman, D. Huard, M. Juckes, M. Kawamiya, N. Klutse, V. Krey, R. Matthews, A. Milward, C. Pascoe, G. van der Shrier, A. Spinuso, M. Stockhause and X. Xing, 2022: The implementation of FAIR data principles in the IPCC AR6 assessment process: A TG-Data Guidance Document, Supporting

Material, Intergovernmental Panel on Climate Change Task Group on Data Support for Climate Change Assessments (TG-Data), 14 pp, https://doi.org/10.5281/zenodo.6504469

Ramapriyan, H.K., J.C. Goldstein, H. Hua, and R.E. Wolfe, 2016: Tracking and Establishing Provenance of Earth Science Datasets: A NASA-Based Example. In: Mattoso, M., Glavic, B. (eds) Provenance and Annotation of Data and Processes. IPAW 2016. *Lecture Notes in Computer Science*, 9672. Springer, Cham. https://doi.org/10.1007/978-3-319-40593-3_27

Ramapriyan, H.K. and C. Lynnes, 2019: Reusability in NASA's Earth Observation System Data and Information System (EOSDIS). Abstract IN21A-09. AGU 2019 Fall Meeting. Abstract available at https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/497316

Ramapriyan, H.K. and J. Behnke, 2020: NASA's Earth Observing System Data and Information System (EOSDIS) and FAIR – A Self-Assessment. Abstract IN044-08. AGU 2020 Fall Meeting. Abstract available at: https://agu.confex.com/agu/fm20/meetingapp.cgi/Paper/683318. The presentation recording is available at: https://wiki.earthdata.nasa.gov/display/ESDSWG/References+to+Assessments (Earthdata sign-on credential may be required)

Ramapriyan, H.K., and P. J. T. Leonard, 2020: Data Product Development Guide (DPDG) for Data Producers version1. NASA Earth Science Data and Information System Standards Office. Version: July 2020. Publish date: 21 October 2021. https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1

Ramapriyan, H.K., J.F. Moses, and D. Smith, 2022: NASA Earth Science Data Preservation Content Specification, Revision C. NASA Earth Science Data and Information System Standards Coordination Office, 3 May 2022. doi:10.5067/DOC/ESO/423-SPEC-001

RDA FAIR Data Maturity Model Working Group, 2018: Case statement for a new RDA Working Group - FAIR Data Maturity Model: core criteria to assess the implementation level of the FAIR data principles. Version: 1.7. Available at: https://www.rd-alliance.org/sites/default/files/case_statement/Case%20Statement-FAIR%20Data%20Maturity%20Model-v1.7.pdf

RDA FAIR Data Maturity Model Working Group, 2020: FAIR Data Maturity Model: specification and guidelines. *Research Data Alliance.* https://doi.org/10.15497/RDA00045

Schultes, E., B. Magagna, K.M. Hettne, R. Pergl, M. Suchanek, and T. Kuhn, 2020: Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann, G., Ram, S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer, Cham. https://doi.org/10.1007/978-3-030-65847-2_13

Silverman, M. L., G. Chen, M. Shook, and others, 2022: The Use of Atmospheric Composition Variable Standard Names in Airborne and Field Data Products. Abstract IN32C-0399. AGU 2022 Fall Meeting. Abstract available at: https://agu.confex.com/agu/fm22/meetingapp.cgi/Paper/1088790

Singh, D., A. Boyer, M. Donnovan, L. Hook, J. J. McNels, and others, 2019: Enhancing the FAIRness of Carbon Monitoring System Data. Abstract B13F-2445. AGU 2019 Fall

Meeting. Abstract available at: https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/596337

Stocker, M., L. Darroch, R. Krahl, T. Habermann, A. Devaraju, U. Schwardmann, C. D'Onofrio, and I. Häggström, 2020: Persistent Identification of Instruments. *Data Science Journal*, 19. http://doi.org/10.5334/dsj-2020-018

USGS, 2017: Fundamental Science Practices: Review and Approval of Scientific Data for Release. Chapter in Survey Manual. Version: January 13, 2017. Available at: https://www.usgs.gov/survey-manual/5028-fundamental-science-practices-review-and-approval-scientific-data-release

Wanchoo, L., J. Ludema, and D. Turicios, 2022: EOSDIS FY2022 Annual Metrics Report. Version: November 2022. Available at: https://www.earthdata.nasa.gov/eosdis/system-performance-and-metrics/eosdis-annual-metrics-reports

Wilkinson, M.D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, and others, 2016: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., M. Dumontier, S.A. Sansone, and others, 2019: Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci Data 6, https://doi.org/10.1038/s41597-019-0184-5

Wingo, S. M., D. K. Smith, C. Davis, S. Bagwell, C. Woods, and others, 2022: An Overview of NASA's Catalog of Archived Suborbital Earth Science Investigations (CASEI): Supporting FAIR and Open Access to Airborne and Field Data. Abstract IN35A-01. AGU 2022 Fall Meeting. Abstract available at: https://agu.confex.com/agu/fm22/meetingapp.cgi/Paper/1115591

Wittenburg, P. and G. Strawn, 2018: Common Patterns in Revolutionary Infrastructures and Data. http://dx.doi.org/10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0.

Yang, X., J.D. Blower, L. Bastin, V. Lush, A. Zabala, J. Masó, D. Cornford, P. Díaz, and J. Lumsden, 2013: An integrated view of data quality in Earth observation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371, https://doi.org/10.1098/rsta.2012.0072

APPENDIX A. EXAMPLES OF RICH METADATA

As community examples, Tables A1-3 list metadata elements defined by the Metadata Interest Group of Research Data Alliance (RDA), core metadata profiles of the World Meteorological Organization (WMO) for search and discovery, and the collection-level metadata elements of NASA's Common Metadata Repository (CMR) for NASA's Earth Observing System Data and Information System (EOSDIS) system, respectively.

Table A1. A set of metadata elements defined by the RDA Metadata Interest Group (Based on: https://www.rd-alliance.org/groups/metadata-ig.html)

Element	URL
Unique Identifier (for later use including citation)	http://bit.ly/2ryRr12
Location (URL)	http://bit.ly/2rujALv
Description	https://bit.ly/2X11UFq
Temporal coordinates	http://bit.ly/2se44QX
Spatial coordinates	http://bit.ly/2ru6kGt
Originator (organisation(s) / person(s))	http://bit.ly/2ruFCgZ
Project	http://bit.ly/2rukIid
Facility / equipment	http://bit.ly/2sdEj3h
Quality	http://bit.ly/2svs0Cc
Availability (license, persistence)	http://bit.ly/2t56LEy
Provenance	http://bit.ly/2se59Z1
Citations	http://bit.ly/2se9efQ
Related publications	http://bit.ly/2rjHFR5
Related software	http://bit.ly/2rutPzn
Schema	http://bit.ly/2srMUl3
Medium / format	http://bit.ly/2svtEEe

Table A2. WMO Core Metadata Profile - Search and Discovery Metadata
(https://github.com/wmo-im/wcmp2/tree/main)

Element	Description
Type	Type of resources such as dataset or services.
Title	A human-readable name for a given resource collection.
Description	Free-text summary description of the resource.
Identifiers	
Keywords	Keywords, tags, key phrases, or classification code such as weather, real-time.
Themes and Topic Hierarchy	A knowledge organization system used to classify the data that the record is describing (Earth systems codelist: https://github.com/wmo-im/wcmp2-codelists/blob/main/codelists/earth-system-domain.csv)
Geospatial Extent	Bounding geometries for a given dataset collection
Temporal Extent	Temporal extents as time instants or time periods.
Providers	These elements provide contact information based on the role of the provider.
Version	Versioning of the resource such as 1.0.1
Digital Object Identifier	Provides a Digital Object Identifier (DOI) as a means to cite research or resource identification using the DOI framework.
Record Creation Date	Describes the date that the record was created.
Record Update Date	Describes the date that the record was changed.
Distribution Information	Provides information regarding how to access and retrieve data and products.

Table A3. Elements in NASA Collection-Level Unified Metadata Model (UMM-C)
(https://wiki.earthdata.nasa.gov/display/CMR/UMM-C+Schema+Representation; Last update: February 17, 2023; Last accessed: May 4, 2023)

Element	Definition	Required
Abstract	This element provides a brief description of the dataset the metadata represents.	Yes
Access Constraints	This element describes any restrictions imposed on data access. Access Constraints can be described in a free text field with the option to provide an access control list (ACL) value.	No
Additional Attributes	This element stores the data's distinctive attributes (i.e. attributes used to describe the unique characteristics of the resource which extend beyond those defined in this mapping).	No

Ancillary Keywords	This element allows metadata authors to provide words or phrases beyond the controlled Science Keyword vocabulary to further describe the collection.	No
Archive and Distribution Information	This element and all of its sub-elements allow a data provider to provide archive and distribution file information upfront to an end user, to help them decide if they can use the product. The file information includes AverageFileSize - typically used for granules as well as file formats and other file information.	No
Associated DOIs	This element stores DOIs that identify associated collections or other associated items to this collection.	No
Collection Citation	This element provides the information required to properly cite the collection in professional scientific literature.	No
Collection Data Type	This element is used to identify the collection as a Science Quality Collection or as a non-science-quality collection such as a Near Real Time (NRT) collection.	No
Collection Progress	This element describes the production status of the dataset.	Yes
Contact Groups	This element is used to provide contact information for a group associated with the dataset.	No
Contact Person	This element is used to provide contact information for an individual associated with the dataset.	No
Data Center	This element is used to identify and provide contact information for the organization responsible for originating, processing, archiving, and/or distributing the dataset being described in the metadata.	Yes
Data Dates	This element is used to identify dates when the *data* or *resource itself* changed in some way.	No
Data Language	This element describes the language used in the preparation, storage, and description of the collection. It is the language of the collection data itself. It does not refer to the language used in the metadata record (although this may be the same language)	No
Direct Distribution Information	The direct distribution information main element allows data providers to provide users information on getting direct access to data products that are stored in the Amazon Web Service (AWS) S3 buckets when they are initially looking at a collection. The end users get information such as the S3 credentials end point, a credential documentation URL, as well as bucket prefix names, and an AWS region.	No
Directory Names	This element has been used historically by the GCMD internally to identify association, responsibility and/or ownership of the dataset, service or supplemental information. Note: This field only occurs in the DIF. When a DIF record is retrieved in the ECHO10 or ISO 19115 formats, this element will not be translated.	No

DOI	This element stores the DOI (Digital Object Identifier) that identifies the dataset.	Yes
Entry Title	This element describes the title of the dataset described by the metadata.	Yes
Instrument	This element is used to register the device that measured or recorded the data, including direct human observation.	No
ISO Topic Category	This element identifies the topic category (or categories) from the EN ISO 19115 Topic Category Code List that pertain to a collection.	No
Location Keywords	This element contains keywords that characterize the study area/region where data was collected.	No
Metadata Association	This element is used to identify other metadata resources that are dependent on or related to the data described by the metadata.	No
Metadata Dates	This element is used to identify dates when the *metadata* changed in some way. This element is made of two sub-elements, Type and Date.	No
Metadata Language	This element specifies the language used in the metadata record (i.e. English, French, Chinese, etc.).	No
Metadata Specification	The Metadata Specification element requires the user to add in schema information into every collection record. It includes the schema's name, version, and URL location.	Yes
Paleo Temporal Coverage	This element defines the time period for geologic and/or paleoclimate data. The element is predominantly used for data samples that originated prior to 01-01-0001.	No
Platform	This element describes the relevant platforms used to acquire the data.	Yes
Processing Level	This element describes an identifier indicating the level at which the data in the collection are processed, ranging from level 0 (raw instrument data at full resolution) to level 4 (model output or analysis results).	Yes
Project	This element describes the scientific endeavor(s) with which the collection is associated.	No
Publication References	This element describes key bibliographic citations pertaining to the collection.	No
Purpose	This element contains suggested usage for the data and/or a description of why the resource exists.	No
Quality	This element describes the quality of the dataset.	No
Related URL	This element describes any resource-related URLs that include project home pages, resource information pages, services, related data,	Yes

	archives/servers, metadata extensions, direct links to online software packages, web mapping services, links to images, documents, or other data.	
Science Keywords	This element enables the specification of Earth Science keywords.	Yes
Short Name	This element identifies the dataset's short name.	Yes
Spatial Extent	This element describes the geographic coverage of the data.	Yes
Spatial Information	This element stores information about the reference frame from which horizontal and vertical spatial domains are measured. The horizontal reference frame includes fields for Geodetic Model, Geographic Coordinates, and Local Coordinates. The Vertical reference frame includes fields for altitudes (elevations) and depths.	No
StandardProduct	This element is reserved for NASA records only. A Standard Product is a product that has been vetted to ensure that they are complete, consistent, maintain integrity, and satisfies the goals of the Earth Observing System mission. The NASA product owners have also committed to archiving and maintaining the data products. More information can be found here: https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-standard-products.	
Temporal Extents	This element describes when data were acquired or collected.	Yes
Temporal Keywords	This element specifies a word or phrase which serves to summarize the temporal characteristics of a dataset.	No
Tiling Identification System	This element defines a named two-dimensional tiling system related to the collection.	No
Use Constraints	This element defines how data may or may not be used to assure the protection of privacy or intellectual property. This includes license information, or any special restrictions, legal prerequisites, terms and conditions, and/or limitations on using the dataset.	No
Version	This element identifies the dataset version.	Yes
Version Description	This element describes the version of the dataset.	No

APPENDIX B. ACRONYMS

ACCESS	Advancing Collaborative Connections for Earth System Science (Program; NASA)
ADMG	Airborne Data Management Group
AGU	American Geophysical Union
API	Application Programming Interface
ARDC	Australian Research Data Commons
ARK	Archival Resource Key
ASF	Alaska Satellite Facility
AWS	Amazon Web Services
CASEI	Catalog of Archived Suborbital Earth Science Investigations (NASA)
CASRAI	Consortia Advancing Standards in Research
CC	Creative Commons
CEOS	Committee on Earth Observation Satellites
CF	Climate and Forecast (Metadata Conventions)
CIESIN	Center for International Earth Science Information Network
COAPS	Center for Ocean-Atmospheric Prediction Studies (Florida State University)
CMR	Common Metadata Repository
CODATA	Committee on Data of the International Science Council
DAAC	Distributed Active Archive Center
DAMA	Data Management Association (UK)
DFT	Data Foundation and Terminology (Interest Group of RDA)
DIF	Directory Interchange Format
DMI	Data Maturity Indicator
DOI	Digital Object Identifier
EDMC	(NOAA) Environmental Data Management Committee
ENVO	Environmental Ontology
EOSDIS	Earth Observing System Data and Information System
ESA	European Space Agency
ESCO	ESDIS Standards Coordination Office
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System (Project)
ESDS	Earth Science Data Systems (Program)
ESDSWG	Earth Sciecne Data System Working Groups(s)

ESIP	Earth Science Information Partners
ESSC	Earth System Science Center (at NASA MSFC)
FAIR	Findable, Accessible, Interoperable, and Reusable
FIP	FAIR Implementaiton Profile(s)
FTP	File Transfer Protocol
F-UJI	FAIR UJI (UJI is not an acronym; it is Malay word meaning test)
GCMD	NASA Global Change Master Directory
GES DISC	Goddard Earth Sciences Data and Information Services Center
GHRC	Global Hydrometeorology Resource Center
GSFC	Goddard Space Flight Center
HTTP	Hypertext Transfer Protocol
I-ADOPT	InteroperAble Descriptions of Observable Property Terminology
IDF	International DOI Foundation
IMPACT	Interagency Implementation and Advanced Concepts Team
IPCC	Intergovernmental Panel on Climate Change
ISBN	International Standard. Books Number. A unique identifier for books.
ISO	International Organization for Standardization
ISO/TC 211	ISO Technical Committe - Geographic information/Geomatics
ISSN	International Standard Serial Number. A unique number for serials, such as periodicals.
JPL	Jet Propulsion Laboratory
LAADS	Level 1 and Atmosphere Archive and Distribution System
MEaSUREs	Making Earth System Data Records for Use in Research Environments
MSFC	Marshall Space Flight Center
NASA	National Aeronautics and Space Administration
NBN	National Bibliography Number. A country-specific identification format for some national libraries
NCICS	North Carolina Institute for Climate Studies
NCAI	NOAA Center for Artificial Intelligence
NERC	Natural Environment Research Council (Vocabularies Server)
NOAA	National Oceanic and Atmospheric Administration
NRT	Near-Real Time
NSIDC	National Snow and Ice Data Center
NVS	NERC Vocabularies Server

O'FAIR	Open, Free, and FAIR
OPeNDAP	Open-source Project for a Network Data Access Protocol
ORCID	Open Researcher and Contributor ID
ORNL	Oak Ridge National Laboratory
PB	Petabyte(s)
PDSC	Preserved Data Set Content
PID	Persistent Identifier
PO.DAAC	Physical Oceanography Distributed Active Archive Center
PROV	(W3C) Provenance
RDA	Research Data Alliance
RVA	Research Vocabularies Australia
SDSC	San Diego Supercomputer Center
SEDAC	Socioeconomic Data and Applications Center
SIPS	Science Investigator-led Processing System
SMD	Science Missions Directorate
SNSF	Swiss National Science Foundation
SOSA	Sensors, Observations, Samples, and Actuators
SPD	SMD Program Directive
SPDX	Software Package Data Exchange
SSAI	Science Systems and Applications, Inc.
SWEET	Semantic Web for Earth and Environmental Terminology
S3	Amazon Simple Storage Service
TB	Terabyte(s)
THREDDS	Thematic Real-time Environmental Distributed Data Services
TOPS	Transform to Open Science
TRUST	Transparency, Responsibility, User focus, Sustainability and Technology (principles)
UCSD	University of California, San Diego
UMM-C	Unified Metadata Model-Collections
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
UUID	Universally Unique Identifier. A unique information identifier within a computer system.
USGS	United States Geological Survey

W3C	World-Wide Web Consortium
WG	Working Group
WGISS	(CEOS) Working Group on Information Systems and Services
WMO	World Meteorological Organization
W3C	World Wide Web Consortium