

NASA ESDIS Standards Coordination Office

# A Practical Guide for Open, Free & FAIR NASA Earth Science Data Products

Version 1.0, 11 August 2025

## STATUS OF THIS MEMO

This memo summarizes the outcomes of the ESDIS Working Group focused on making NASA Science Mission Directorate (SMD) Earth science data products open, free & FAIR (O'FAIR) for the benefit of the NASA Earth science data community. This memo does not specify an ESDIS standard of any kind. Distribution is unrestricted.

## CHANGE EXPLANATION

This is the original version of the document.

## COPYRIGHT NOTICE

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. All other materials from the literature maintain their own original usage licenses and copyrights.

## SUGGESTED CITATION

Peng, G., R. R. Downs, H. K. Ramapriyan, Y. Wei, B. Ramachandran, M. Parsons, Z. Liu, and NASA ESDIS O'FAIR WG, 2025: A Practical Guide for Open, Free & FAIR NASA Earth Science Data Products. Version 1. *NASA ESDIS Standards Coordination Office*.  
<https://doi.org/10.5067/DOC/ESCO/ESDS-RFC-057V1>.

## ABSTRACT

The O'FAIR Earth science data products guide was developed by the O'FAIR Working Group, formerly part of the Earth Science Data Systems Working Groups (ESDSWG) which is now known as ESDIS Working Groups. The guide highlights the alignment of existing NASA resources with the FAIR Principles and provides practical guidance for ensuring and enhancing the level of FAIR-compliance of NASA-funded open and free Earth science data products.

## TABLE OF CONTENTS

STATUS OF THIS MEMO .....	2
CHANGE EXPLANATION .....	2
COPYRIGHT NOTICE .....	2
ABSTRACT .....	2
<b>1 EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>2 INTRODUCTION .....</b>	<b>7</b>
2.1 O'FAIR WORKING GROUP AND ABOUT THIS DOCUMENT .....	7
2.2 ORGANIZATION OF THIS DOCUMENT .....	7
2.3 OPEN AND FREE DATA .....	8
2.4 FAIR DATA .....	8
2.5 OPEN SCIENCE AND OPEN-SOURCE SCIENCE .....	8
<b>3 BACKGROUND .....</b>	<b>8</b>
3.1 DRIVERS .....	8
3.2 GOAL AND ENVISIONED OUTCOME .....	10
3.3 INTENDED AUDIENCE AND STAKEHOLDERS .....	10
3.4 BUILDING COMMUNITY .....	11
3.5 APPLICABILITY .....	11
3.6 LIMITATIONS .....	11
<b>4 FAIR PRINCIPLES, SPD-41A, AND THE CROSSWALK .....</b>	<b>12</b>
4.1 FAIR PRINCIPLES .....	12
4.2 COMPLEXITY OF THE FAIR PRINCIPLES .....	13
4.3 WHAT DOES FAIR-COMPLIANCE MEAN? .....	14
4.4 NASA OPEN DATA AND INFORMATION POLICY (SPD-41A) .....	15
4.5 WHAT DO NASA'S OPEN SCIENCE AND OPEN-SOURCE SCIENCE MEAN? .....	16
4.6 CROSSWALK BETWEEN SPD-41A DATA REQUIREMENTS AND FAIR PRINCIPLES .....	19
<b>5 MAKING NASA EARTH SCIENCE DATA FAIR .....</b>	<b>20</b>
5.1 BEING FINDABLE .....	21
5.2 BEING ACCESSIBLE .....	28
5.3 BEING INTEROPERABLE .....	32
5.4 BEING REUSABLE .....	45
5.5 ROLES AND RESPONSIBILITIES OF STAKEHOLDERS .....	54
<b>6 DISCUSSION .....</b>	<b>56</b>
6.1 HOW DO WE DEAL WITH THIS FOR NEW DATASETS UNDER DEVELOPMENT? .....	56
6.2 HOW DO WE "RETROFIT" EXISTING DATASETS TO MAKE THEM COMPLIANT? .....	56
6.3 WHAT ARE THE RECOMMENDED OPEN LICENSES FOR NASA-HOSTED DATA? .....	57
6.4 WHAT DOES "COMMUNITY STANDARDS" MEAN? .....	57
6.5 BEYOND FAIR .....	58
6.6 HOW TO ADDRESS DYNAMICALLY SUBSETTED OR AGGREGATED TIME SERIES? .....	59
6.7 HOW ABOUT SOFTWARE? .....	59
6.8 HOW ABOUT COMMERCIAL DATA PRODUCTS? .....	60

<b>7</b>	<b>CONCLUSION AND PATH FORWARD .....</b>	<b>60</b>
<b>8</b>	<b>ACKNOWLEDGEMENT .....</b>	<b>63</b>
<b>9</b>	<b>REFERENCES.....</b>	<b>65</b>
	<b>APPENDIX A. KEY TERMS AND DEFINITIONS .....</b>	<b>70</b>
	<b>APPENDIX B. CATEGORY-SPECIFIC FAIR REQUIREMENTS.....</b>	<b>72</b>
	<b>APPENDIX C. CROSSWALKS BETWEEN FAIR AND CMR UMM-C FIELDS .....</b>	<b>75</b>
	<b>APPENDIX D. NASA EARTH SCIENCE GLOSSARIES AND VOCABULARIES.....</b>	<b>77</b>
	<b>LIST OF FIGURES .....</b>	<b>78</b>
	<b>LIST OF TABLES .....</b>	<b>79</b>
	<b>AUTHOR LIST.....</b>	<b>80</b>
	<b>CONTRIBUTING MEMBERS OF THE O'FAIR WORKING GROUP.....</b>	<b>81</b>

## 1 EXECUTIVE SUMMARY

NASA's Science Mission Directorate (SMD) seeks to advance human knowledge through new scientific discoveries to understand the Sun, Earth, Solar System, and Universe. The Earth Science Division (ESD), one of five SMD science divisions, supports the collection or acquisition, processing, stewardship, archiving, distribution, and applications of high-quality data and information from NASA's Earth observations. These activities, overseen by the NASA's Earth Science Data Systems (ESDS) Program, facilitate exploring, monitoring, and understanding of the current state of the Earth system through its affiliated Science Investigator-led Processing Systems (SIPs) and discipline-specific Distributed Active Archive Centers (DAACs).

SIPs and DAACs are distributed across the United States as part of the Earth Science Data and Information System (ESDIS) Project within the ESDS Program. At the end of FY2024, NASA archived a total of 128.56 PB data, including 66.69 PB in the Cloud, with an average archive growth of 147.26 TB per day. In FY2024 (October 1, 2023 - September 30, 2024), ESDIS served more than 8 million distinct users world-wide with an average end-user distribution volume of about 450.8 TB per day (Wanchoo et al. 2024).

NASA Earth science data support activities by the U.S. and international organizations and governments, disaster-relief agencies worldwide, scientists from academia and the private sector, business, the military, and the general public across Earth science disciplines.<sup>1</sup>

ESD has been promoting the open and free sharing of scientific data for over three decades. For example, NASA Earth science data has been free and open to all users since 1994.<sup>2</sup> The latest SMD data and information policy, aka, SPD-41a, codifies and requires that all NASA-funded scientific data and information be freely available and openly shared (NASA SMD 2022). For the first time, NASA explicitly recommends that SMD-funded scientific data *follow the Findable, Accessible, Interoperable, and Reusable (FAIR) Principles* defined in Wilkinson et al. (2016).

Specifically, SPD-41a states that SMD-funded scientific data products shall be:

- Described by standards-compliant metadata;
- Assigned a persistent identifier and released with clear, open, and accessible license;
- Indexed in a NASA catalog and searchable; and,
- Made available and retrievable in an open and machine-readable format.<sup>3</sup>

---

<sup>1</sup> <https://science.nasa.gov/earth-science/earth-data> (accessed: 2023-09-01)

<sup>2</sup> <https://www.earthdata.nasa.gov/esds> (accessed: 2023-08-17)

<sup>3</sup> Definitions of key terms and their sources are captured in Appendix A.

These SPD-41a data requirements closely align with the FAIR Principles regarding data and metadata and apply to Earth science data products generated by NASA SMD-funded researchers and managed by NASA DAACs.

In response to the need for technical guidance on implementing SPD-41a, the NASA ESDS Working Group on Open, Free & FAIR Data (O'FAIR WG) has developed this document for the NASA Earth science community. This comprehensive guide, intended to help NASA's open and free data adhere to the FAIR Principles, identifies NASA FAIR-enabling resources to stakeholders such as data producers, and offers practical and actionable recommendations to DAACs and SIPs, to bolster compliance of their data products with these principles.

Advancing the level of FAIR-compliance (i.e., FAIRness) is an ongoing journey characterized by incremental progress. Interpretation of the FAIR Principles can be subjective, and their implementation can vary considerably across different domains and applications. To navigate this complexity, the guidance outlined in this document primarily focuses on the data product (aka collection) level, not the granule or variable level. Nonetheless, this effort represents a solid step towards systematically ensuring or improving the FAIR-compliance of NASA SMD-funded Earth science data products. Our approach leverages existing NASA FAIR-enabling resources and formulates recommendations by integrating community standards and best practices.

The identified NASA FAIR-enabling resources underscore the crucial role of existing NASA enterprise processes, standards, tools, and infrastructures in adhering to the FAIR Principles. High compliance is evident in findability and accessibility, though there is room for improvement across all four FAIR aspects, particularly in interoperability. Coordinated efforts are essential for improving (meta)data interoperability to facilitate seamless data and information sharing, integration, and exchanges across NASA tools and systems, other agencies managing Earth science data products, and various disciplines.

We advocate for the collaborative development of additional variable-level and domain-specific guidance in partnership with domain experts, and for active participation in developing domain-specific standards and best practices, including those relevant to Cloud environments. In the interim, we encourage data producers and repositories to leverage existing variable-level standards and community best practices, both within and across domains and disciplines, to the fullest extent possible.

The guidance in this document primarily targets digital scientific data products from NASA SMD-funded scientific missions and projects starting with the 2023 Research Opportunities in Space and Earth Science (ROSES) solicitations. However, the FAIR-compliance practices, standards, and technologies can also benefit data products from existing missions and investigations. To support

quick reference and practical use, the WG also developed a companion cheat sheet summarizing key points from this document (Peng et al. 2024b).

Given the rapid evolution of community FAIR practices and technologies, this guide should be periodically reviewed and updated.

## 2 INTRODUCTION

### 2.1 O'FAIR Working Group and About This Document

The O'FAIR WG was established in late May 2022. The O'FAIR WG relies mostly on in-kind support from organizations and projects within the ESDS Program and SMD-funded projects. The objectives are to explore community practices and develop practical guidance for NASA and the scientific community to enhance findability, accessibility, interoperability, and reusability (i.e., FAIR) of Earth science data products and their associated metadata, while upholding their openness and free access. A data product in this document is defined as a set of data files that can contain multiple parameters and that compose a logically meaningful group of related data (Ramapriyan et al. 2024).

This document provides practical and actionable guidance on making NASA open and free data more FAIR. It represents the second and final deliverable produced by the O'FAIR WG for this term.

### 2.2 Organization of This Document

Following the Executive Summary (Chapter 1), we describe what the following concepts mean within the context of this document in this Introduction (Chapter 2):

- Open and free data,
- FAIR data,
- Open science and open-source science.

Chapter 3 provides background, describing the drivers behind our effort, the primary goal and envisioned outcomes, the intended audience and stakeholders, as well as the document's applicability and limitations.

Additional information on the FAIR Principles, SPD-41a, and crosswalk between the two are provided in Chapter 4.

If one prefers to bypass all these contextual information chapters, one can go directly to Chapter 5 for identified NASA resources and practical guidance on FAIR NASA Earth science data.

A selection of questions and corresponding answers are captured in the Discussion chapter (Chapter 6), followed by the Conclusion and Path Forward chapter (Chapter 7).

## 2.3 Open and Free Data

Open and free data in this document denotes scientific data and associated metadata that are publicly available in a non-proprietary format, both comprehensible to humans and interpretable by machines, and is obtainable without any charges or usage restrictions (in rare cases where it is necessary to charge for data, the fees should not exceed the cost of data dissemination). Open and freely accessible data are critical in lowering barriers to access and maximizing the utilization of federally funded scientific data.

## 2.4 FAIR Data

FAIR data in this document denote scientific data that are uniquely identified, comprehensively and consistently documented with standard-based metadata for discovery and reuse, and can be securely retrieved using open protocols and understood for permissible use. FAIR data are essential in supporting open science by making data and relevant information ready for sharing and reuse. More detailed descriptions of the FAIR Principles are provided in Section 4.1.

## 2.5 Open Science and Open-Source Science

Open Science is the principle and practice of making research products and processes available to all, while respecting diverse cultures and maintaining security and privacy. Open science fosters collaborations, reproducibility, and equity.<sup>4</sup> In addition, open-source science involves use and/or sharing of open-source software, data, and collaborative methodologies in scientific investigations. Open science can minimize duplication in collecting, creating and reusing scientific artifacts and lowers the barriers to public research. Additional information can be found in Section 4.5.

# 3 BACKGROUND

## 3.1 Drivers

Bringing together current federal mandates and NASA policies and guidance, (e.g., U.S. Law 115-254 2018; U.S. Law 115-435 2019; OSTP 2022; NASA ESD data and information policy<sup>5</sup> and ESDS open source policy<sup>6</sup>), NASA SMD (2022) released its latest Scientific Information Policy for the

---

<sup>4</sup> <https://open.science.gov/> (accessed: 2023-07-26)

<sup>5</sup> <https://www.earthdata.nasa.gov/engage/open-data-services-software-policies/data-information-guidance> (accessed: 2025-06-19)

<sup>6</sup> <https://www.earthdata.nasa.gov/engage/open-data-services-and-software/esds-open-source-policy> (accessed: 2023-04-19)

Science Mission Directorate (SMD), also known as SPD-41a. SPD-41a requires all SMD-funded scientific information, i.e., data, software, and publications, to "be made publicly available to the extent legally permitted", which "applies to all SMD-funded scientific activities to the extent possible, regardless of the funding vehicle" (NASA SMD 2022). SPD-41a addresses three types of scientific information: 1. Mission scientific information produced by SMD-funded flight missions; 2. Research scientific information produced by a science investigation funded via research award(s); and 3. Scientific information produced by all other SMD-funded scientific activities, such as, but not limited to, experiments, investigations using sub-orbital platforms, airplanes, field campaigns, or citizen science projects.

A few key policy statements from SPD-41a are quoted below:

Applicable to all SMD-funded scientific information:

- "Scientific data underlying peer reviewed manuscripts resulting from SMD-funded scientific activities shall become publicly available, to the extent allowed by applicable law and existing NASA policies, no later than the publication of the peer-reviewed article that describes it.
- "SMD-funded data should follow the FAIR Guiding Principles for scientific data management and stewardship. This means data should be findable, accessible, interoperable, and reusable (FAIR)."

Applicable to mission information:

- "There shall be no period of exclusive access to Mission data. A period after the data have been obtained may be allowed for activities such as calibration and validation of the data. This period shall be as short as possible and shall not exceed six months."

Applicable to research information (and other non-mission information):

- Scientifically useful data from models and simulations developed using SMD funding shall be made available at the time of publication of the peer reviewed manuscript that describes the scientific results.

Grants from ROSES-2023 onwards are expected to fully comply with SPD-41a. For existing missions and investigations, adherence is recommended if feasible within available resources.

As noted above, SPD-41a explicitly recommends all SMD-funded scientific data and information to follow the FAIR Principles. Consistent and practical guidance tailored to NASA Earth science data products is needed for adhering to SPD-41a, especially on how to best follow the FAIR

principles. Such guidance will be beneficial to stakeholders such as data producers, data managers, and stewards.

### 3.2 Goal and Envisioned Outcome

The overarching objective behind the creation of this Guide document is to align the requirements of NASA SPD-41a data policy with the implementation requirements of stakeholders. Our focus is on enhancing the machine-actionability of NASA Earth science data by providing practical and actionable FAIR-compliance guidance at the level of individual data products.

By doing so, we also envision additional impacts through the following:

- Supporting open science and open-source science by enabling sharing and maximizing the value of NASA Earth science data and information;
- Reaching new users by facilitating access to and (re)use of NASA data and information with minimized barriers to entry.

We have no intention of duplicating efforts and are instead making extensive use of the work that has already been accomplished. The landscape of the FAIR practices and FAIRness assessment frameworks was explored in the first phase of the O'FAIR WG. Those practices and frameworks are captured in an Excel workbook<sup>7</sup> and a high-level summary was provided in the O'FAIR WG Inception Report (Peng et al. 2023). They serve as the foundation for this document, complemented by the continuous collection and synthesis of additional FAIR practices and implementation approaches.

The guidance is structured around four FAIR aspects in Chapter 5. Each aspect begins with an overview of the key focus areas of associated FAIR principles. NASA's current practices and standards are then highlighted, identifying areas of alignment and gaps with specific FAIR-compliance requirements. In cases where gaps are identified or additional improvement can help enhance the level of compliance, recommendations will be formulated by drawing upon community standards and best practices, tailored to NASA SMD-funded Earth science data products.

### 3.3 Intended Audience and Stakeholders

Intended audiences include data and information producers and managers, and service providers such as those from the ESDIS Project, DAACs, and SIPs, as well as ROSES-funded principal investigators and mission science team members, who can benefit directly from this guide.

---

<sup>7</sup> <https://wiki.earthdata.nasa.gov/display/ESDSWG/O%27FAIR+WG+-+FAIR+Practices+%28Descriptions+Metrics+Assessments+Tools%29+Collection> (version: 2023-03-01)

New NASA mission managers can benefit from this guide by defining stewardship requirements for mission standard data products, leveraging FAIR practices described in this guide.

Decision-makers at ESDS, ESDIS, and the SMD Office of the Chief Science Data Officer (OCSDO) can leverage insights from this guide to assess the current state of FAIR-enabling resources and make informed decisions about implementing the O'FAIR WG's recommendations.

### 3.4 Building Community

The O'FAIR WG consists of domain experts spanning science, data management, and technology fields, with representation from ESDIS enterprise tools/services, eight domain-specific DAACs, as well as data producers and managers from NASA-funded projects. Their diverse expertise and perspectives in producing and managing a variety of Earth science data products contribute to the comprehensiveness of outcomes produced by the working group. Moreover, their active involvement ensures that insights gained from working group activities are shared with their organizations or projects, fostering a lasting impact beyond the group's existence.

### 3.5 Applicability

This document focuses on digital scientific data and associated metadata. They are defined by SPD-41a as scientifically or technically relevant information that can be stored digitally, available openly, and accessed electronically. The practices and recommendations described in this document should be applied to scientific data from NASA SMD-funded Earth science missions starting from the 2023 ROSES solicitations. They are also beneficial to scientific data from existing missions and investigations.

Scientific data in this document are assumed to be publicly available without any restrictions. Any data that are subjected to specific laws, regulations, and policies, e.g., Export Administration Regulations (EAR) or International Traffic in Arms Regulations (ITAR), are not considered - see Section II-C of SPD-41a (NASA SMD 2022) for a list of laws, regulations, and policies that generate exceptions to data sharing requirements. In addition, SPD-41a does not require the sharing of preliminary results.<sup>8</sup>

### 3.6 Limitations

The practices and recommendations outlined in this document are generally broad and focus on the product or collection level for both data and metadata. (Within the context of this document, "collection-level" corresponds to "product-level" while "granule-level" pertains to "file-level". These terms may be used interchangeably.) This approach aims to limit the diversity and

---

<sup>8</sup> [https://science.nasa.gov/researchers/science-data/science-information-policy\\_faq](https://science.nasa.gov/researchers/science-data/science-information-policy_faq) (accessed: 2023-05-15)

complexity of practices that can arise from domain-specific or variable-dependent variations, which are often less standardized.

Some file-level or specific data types related discussions, such as those for in-situ measurements, are provided but by no means are exhaustive. We acknowledge that certain cases may not be addressed and could require additional attention from a specialized task team.

This document primarily addresses the technical aspects of FAIR-enabling resources and does not explicitly cover organizational aspects. Technical aspects focus on infrastructure, standards, and tools—such as persistent identifiers (PIDs), metadata standards, controlled vocabularies, and protocols—that enable data to meet FAIR principles. In contrast, organizational aspects address governance, policy, social, and administrative considerations that facilitate secure and timely data sharing and utilization.

## 4 FAIR PRINCIPLES, SPD-41a, AND THE CROSSWALK

### 4.1 FAIR Principles

The FAIR Principles provide a framework for data management and stewardship. They aim at enabling digital resources, especially data, to be findable, accessible, interoperable, and reusable (FAIR) in a machine-friendly environment for both human and machine end users (Wilkinson et al. 2016; recaptured in Table B1 of Appendix B). The FAIR Principles were designed to ensure digital data to be machine-actionable with no or minimal human intervention and prescribed expected behaviors - but not how - to remain domain and technology agnostic (Mons et al. 2017; Wilkinson et al. 2019; 2022). They offer overarching concepts and objectives that can be applied across various disciplines and adapted to different contexts. As a result, however, the implementation of the FAIR Principles can be subject to subjective interpretation and domain-specific requirements. We also recognize that some standards and best practices required to implement certain elements of the FAIR Principles are still under development.

The FAIR Principles consist of data enabling behaviors in four aspects, i.e., findability, accessibility, interoperability, and reusability, referred to as FAIR dimensions hereafter. They touch on data, metadata, and infrastructure, referred to as FAIR categories hereafter. From the FAIR Principles, Peng (2023) derived 28 category-specific requirements (updated and recapitulated in Table B1). In this document, we will use "FAIR Principles" to denote the overarching framework, while using "FAIR principles" in lowercase to indicate individual principles (e.g., F1, F2, ...).

Since their inception, various explanations and interpretations of the FAIR Principles have been developed by different organizations (e.g., Swiss National Science Foundation<sup>9</sup>; GO FAIR<sup>10</sup>; GO FAIR Foundation<sup>11</sup>). We have consolidated and summarized them below to provide high-level information on what they are and how to achieve them along each of the four FAIR dimensions in the content of digital data.

The *Findability* (F) principles focus on making data and digital resources discoverable. To achieve this, data should be assigned globally unique and persistent identifiers (such as Digital Object Identifiers), cataloged, and described using rich metadata for search and discovery.

The *Accessibility* (A) principles emphasize that data and digital resources should be accessible with persistent metadata through an open, free, standardized communication protocol that allows for authentication and authorization procedures when appropriate.

The *Interoperability* (I) principles aim to facilitate seamless integration and exchange of data and digital resources across different systems, tools, services, and research domains. To achieve this, it suggests that data and digital resources should be structured using standardized, widely accepted data models, formats, and vocabularies, including references to other relevant data and metadata, for example, a reference to metadata standard that was utilized.

The *Reusability* (R) principles promote the creation of data and digital resources that can be readily understood, interpreted, and (re)used. This involves adhering to domain-relevant standards and providing comprehensive data documentation, including clear data usage license and detailed provenance.

## 4.2 Complexity of the FAIR Principles

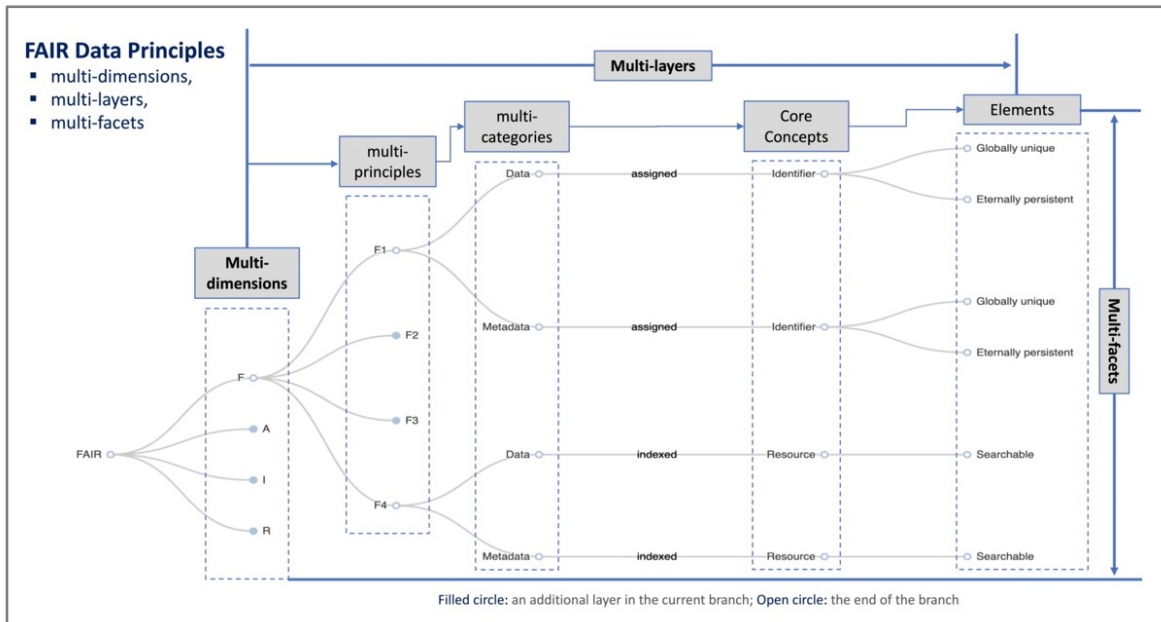
The FAIR Principles may seem straightforward due to their simple acronym, but they are quite complex, even without considering domain and application dependencies. These principles are inherently multi-dimensional, layered, and multifaceted, encompassing core concepts with numerous elements, as depicted in Peng et al. (2024a; shown here as Figure 4.1). The complexity of the FAIR principles underscores the need for guidance.

---

<sup>9</sup> [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_principles\\_translation\\_SNSF\\_logo.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf) (accessed: 2023-08-30)

<sup>10</sup> <https://www.go-fair.org/fair-principles> (accessed: 2023-08-30)

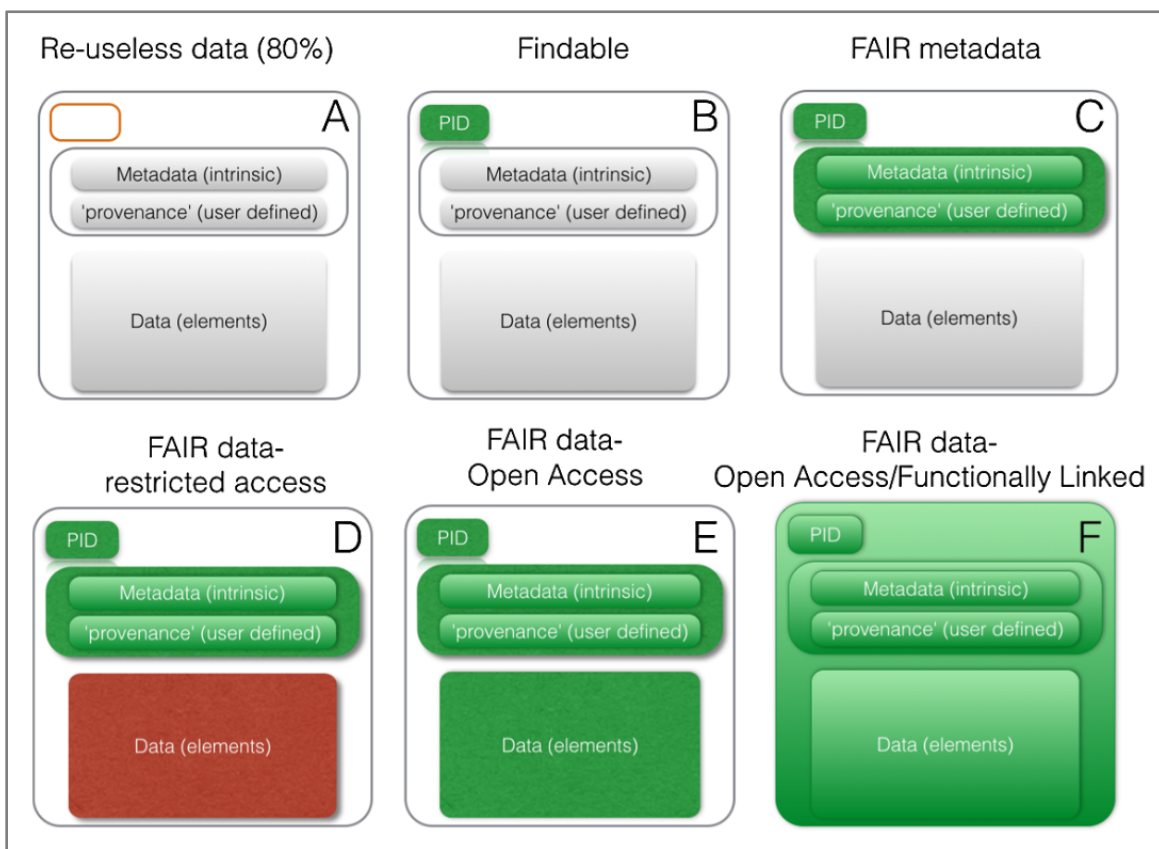
<sup>11</sup> <https://www.gofair.foundation/interpretation> (accessed: 2023-08-30)



**Figure 4.1.** A schematic diagram depicts the complexity of the FAIR principles in terms of multi-dimensions, multi-layers, and multi-facets. Source: Peng et al. (2024a). License: CC-BY 4.0

### 4.3 What Does FAIR-Compliance Mean?

Improving FAIR-compliance is a continuing journey with incremental steps – one example is shown in Mons et al. (2017; captured here as Figure 4.2).



**Figure 4.2.** A schematic diagram shows varying degrees of FAIRness - Data as increasingly FAIR digital objects. Source: Mons et al. (2017), which one should turn to for a detailed description of and examples for various degrees of FAIRness (i.e., boxes A-F). Colored elements in green are FAIR and open, colored elements in red are FAIR and closed. License: CC-BY 4.0

#### 4.4 NASA Open Data and Information Policy (SPD-41a)

The NASA Science Information Policy, aka SPD-41a, provides guidance on the open sharing of publications, data, and software created in the pursuit of scientific knowledge. The SPD-41a codifies emerging data requirements from federal and NASA policies. They are applicable to data and information products produced either from NASA flight missions or from research investigations that have been supported by the SMD. For both cases, as described in SPD-41a (NASA SMD 2022), data and information that have been produced with NASA support should follow the FAIR Principles to ensure their findability, accessibility, interoperability, and usability. Mission data and information include that from both NASA-supported missions and NASA partner missions. Research data and information includes awards from contracts, cooperative agreements, and grants as well as from direct funding from NASA. Data products from one individual mission, program, or project can cover multiple Earth science disciplines and are managed by their associated discipline specific DAACs.

NASA has been promoting data sharing, enhancing capabilities, and establishing data management practices to improve the user experience in discovering and utilizing ESDIS data products for both human and machine end users. Ramapriyan and Behnke (2020) have manually evaluated the FAIRness of ESDIS data systems, which is summarized in the O'FAIR WG Inception Report (Peng et al. 2023, Table 6.1). The results have shown good FAIR-compliance at the data system level. Across all four FAIR dimensions, FAIRness is comparable or slightly higher for human data users than machine end users, highlighting opportunities to enhance machine-readability and interoperability.

The FAIR Principles are inherently designed to facilitate data sharing in a computational environment with no or minimal human intervention for both humans and machine end users. Despite the significant presence of human end users within the NASA data user community, our approach will align with the original scope of the FAIR Principles. We will emphasize practices and capabilities that enhance FAIRness within computational environments, thereby streamlining the guidance to manage complexity.

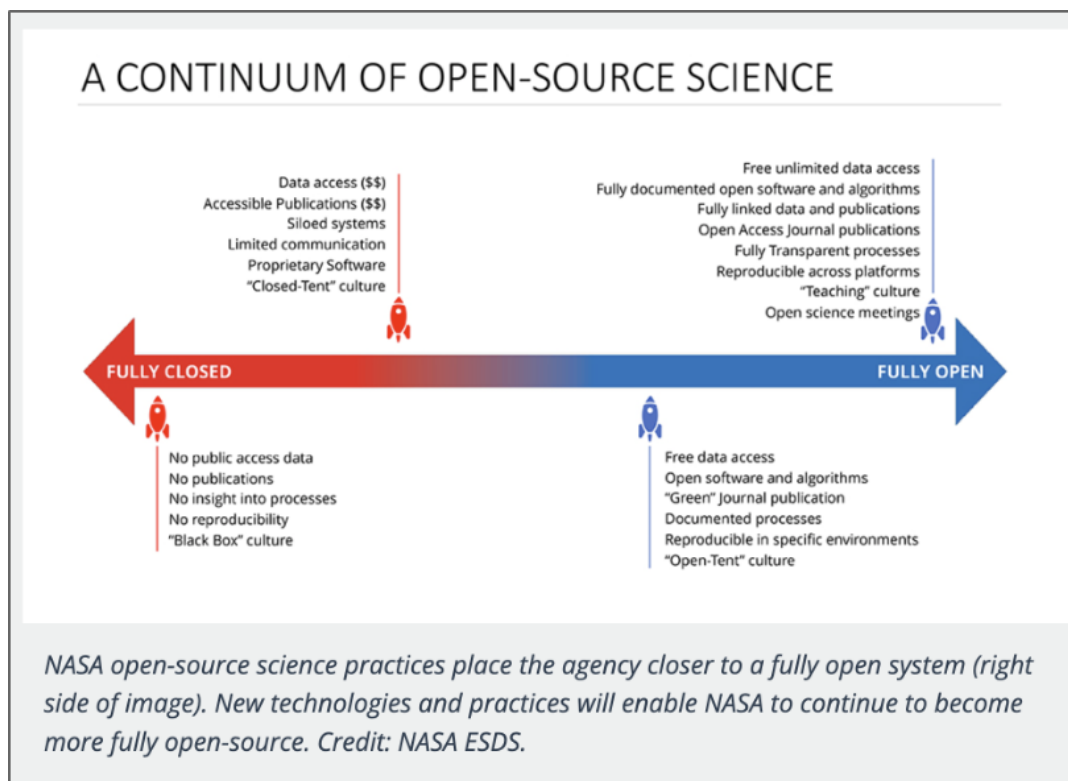
In addition, although the FAIR Principles have been intentionally formulated to be domain- and technology-agnostic, it is necessary to address the domain- and technology-specific aspects when implementing the FAIR Principles. The NASA Earth science data holding encompasses a wide variety of data products that serve a wide spectrum of data users and applications. Therefore, the requirements for the levels of FAIRness can vary, which makes enabling data FAIR a very challenging task.

Lastly, NASA Earth science data products are categorized with different levels of service requirements (NASA ESDS 2020). Data management and stewardship practices applied to different data product categories can vary. However, going forward, all NASA SMD-funded data products must comply with the SPD-41a requirements and align with the FAIR Principles.

By exploring existing NASA standards, practices, capabilities and workflows, this Guide document aims to assist stakeholders in initiating enhanced data management and stewardship through the entire data product lifecycle, especially for metadata collection and data curation. By taking incremental steps to make continuous improvements, we can collectively move forward to make NASA Earth science data more open and FAIR.

#### **4.5 What Do NASA's Open Science and Open-Source Science Mean?**

Open-source science is a continuum (Figure 4.3). Open, free & FAIR NASA data are one of the cornerstones of NASA's open science and support reproducibility. Promoting and enabling FAIR data supports NASA's Open-Source Science Initiative.



**Figure 4.3. Continuum of Open-Source Science.** Source: <https://www.earthdata.nasa.gov/esds/open-science> (accessed: 2023-04-19)

#### 4.5.1. Why open science?

Open science encompasses the open sharing of research data, methodologies, software, and scientific outcomes. It allows for transparency and removes barriers to access, and therefore potentially accelerates the discovery and advancement of current scientific findings through broad collaborations. Additionally, it may increase efficiency and quality of research, and fosters a culture of collaboration (e.g., Allen and Mehler 2019).

The memo by the U. S. Office of Science and Technology Policy (OSTP 2022) stated: "openness in science is fundamental, security is essential, and freedom and integrity are crucial." <sup>12</sup>

NASA has developed Open Science 101<sup>13</sup>, a free five-module training course designed to introduce the principles and best practices of open science.

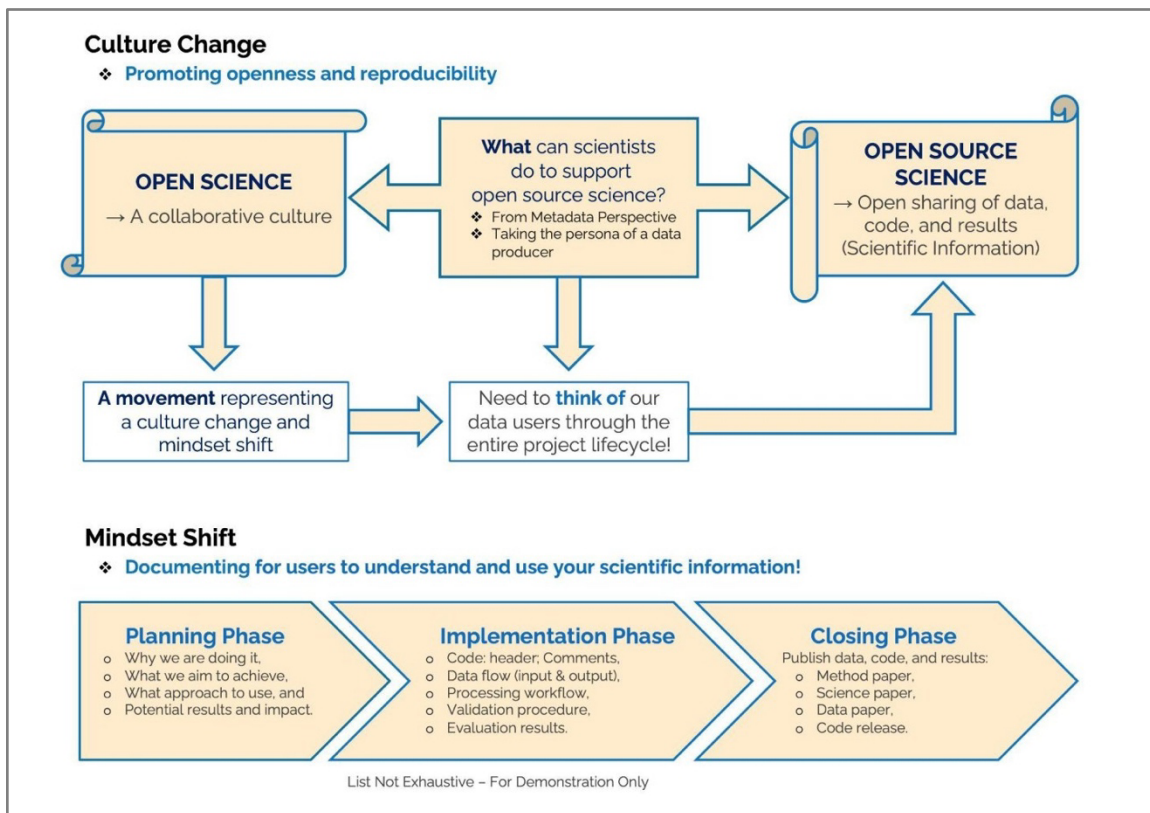
<sup>12</sup> See: <https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf> (accessed: 2023-05-23; link updated: 2025-07-31)

<sup>13</sup> <https://science.nasa.gov/open-science/os101> (accessed: 2025-07-31)

#### 4.5.2. When is an appropriate stage to be open?

Open-source science requires a culture and mindset change, which takes time and resources. Ultimately, new NASA-funded scientific data, code, and results should be open for "the entirety of the scientific process, from start to finish" (Albers 2023).

At a minimum, scientific data resulting from SMD-funded scientific activities (excluding intermediate data) shall be made publicly available, to the extent permitted by applicable laws and existing NASA policies - no later than the publication date of the corresponding peer-reviewed article (NASA SMD 2023). In addition, we encourage publishing documentation on the method/algorithm used to generate the data product, if not already included in the peer-reviewed article (Figure 4.4).



**Figure 4.4.** An illustrative depiction of the actions that scientists and data producers can take to support open-source science, including pertinent information to be documented and disseminated for data users across the project lifecycle. A method paper may be published during the planning phase. Based on: Nosek (2021) and Peng (2022).

The SPD-41a requires that scientific and technically relevant information such as algorithms be stored digitally and accessed electronically (NASA SMD 2022). Presently, description of algorithms for data retrieval and/or processing tends to be captured in an Algorithm Theoretical Basis Documents (ATBD). ATBDs are generally written by highly specialized scientists for their peers. Essentially, the content of ATBDs is technical and detailed, reflecting the advanced knowledge and specialization of both the authors and the readers. ATBD usually contains specific and in-depth information that may not be included in typical journal publications. Examples of ATBDs are available at: <https://eosps.nasa.gov/content/algorithm-theoretical-basis-documents>. To support the modernization and standardization of ATBD curation and approval process, an Algorithm Publication Tool (APT) has been developed to facilitate the standardized end-to-end publication process.<sup>14</sup>

Journals that offer open access for Earth science data papers include:

- Copernicus Earth System Science Data<sup>15</sup>,
- AGU Earth and Space Science<sup>16</sup>,
- Big Earth Data.<sup>17</sup>

#### 4.6 Crosswalk Between SPD-41a Data Requirements and FAIR Principles

To understand potential synergies that could be leveraged to meet the objectives of SPD-41a and the FAIR Principles, the O'FAIR WG assessed these relationships in a crosswalk between the terms used in both (Figure 4.5). The figure demonstrates a clear alignment between the SPD-41a data requirements and the FAIR principles for both data and metadata. A1.1 and A1.2 principles do not crosswalk to any explicit data requirements in the SPD-41a as they are pertained to infrastructure capabilities which were discussed in SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024 (NASA SMD 2019).

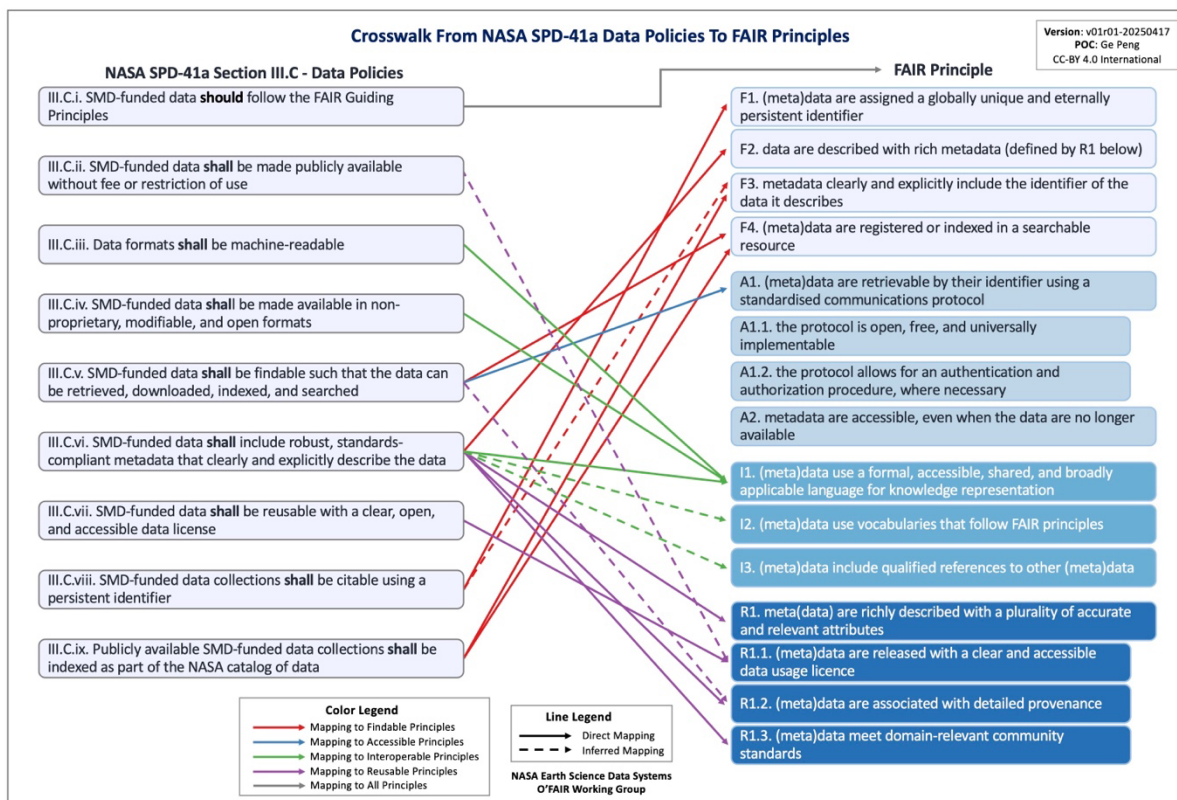
---

<sup>14</sup> <https://www.earthdata.nasa.gov/apt/> (accessed: 2024-04-30)

<sup>15</sup> <https://www.earth-system-science-data.net/> (accessed: 2023-05-30)

<sup>16</sup> <https://agupubs.onlinelibrary.wiley.com/journal/23335084> (accessed: 2023-11-15)

<sup>17</sup> <https://www.tandfonline.com/journals/tbed20> (accessed: 2023-11-15)



**Figure 4.5.** Crosswalk from the data requirements that are defined in the SPD-41a (NASA SMD 2022) to the FAIR principles that are defined in Wilkinson et al. (2016). Direct mappings from SPD-41a to FAIR are denoted by solid lines, while the dashed lines denote the mappings may be inferred or not always hold. Mappings to Findable/Accessible/Interoperable/Reusable principles are denoted by red/blue/green/purple lines, respectively.

## 5 MAKING NASA EARTH SCIENCE DATA FAIR

In this chapter, we describe the focus areas of the principles within each of the four FAIR dimensions. We highlight existing NASA FAIR-enabling standards, practices, workflows, and tools/services for Earth science data products. Additionally, we evaluate and discuss their alignment with the category-specific requirements of each FAIR principle. The category-specific requirements are adapted from Peng (2023) and summarized in Appendix B. The compliance levels are categorized as *satisfying* all key characteristic requirements, *meeting* some but not all requirements, or *supporting* but not directly meeting any requirements, to indicate high, intermediate, or minimal level of compliance, respectively. Table 5.1 summarizes three commonly used identifier types and their alignment with the FAIR F1 principle. In this context, the key characteristics evaluated are global uniqueness and long-term persistence. While a local identifier is generally neither globally unique nor eternally persistent, it still qualifies as an

assigned identifier, thus supporting the F1 principle at a minimal level. One can, however, use a global unique identifier such as Universally Unique Identifier (UUID) for local identifiers, and therefore increase their FAIR compliance to “Meeting” for F1.

**Table 5.1.** *Characteristics of a Digital Object Identifier (DOI), Universally Unique Identifier (UUID), and local identifier (ID), as well as their compliance level to the F1 principle (i.e., (meta)data are assigned a globally unique and eternally persistent identifier)*

Identifier Type	DOI	UUID	Local ID
<b>What it is</b>	A globally unique, persistent identifier used to cite and access digital resources (e.g., datasets, publications)	A 128-bit number used to uniquely identify information in distributed systems	A unique identifier used within a specific organization, repository, or system
<b>Example</b>	10.5067/GHGMR-4FJ04	8012fda7-3ea4-4ef2-bb4e-0f66d4d9e850	C1996881146
<b>Globally Unique</b>	Yes	Yes	No
<b>Eternally Persistent</b>	Yes	No	No
<b>F1 Compliance Level</b>	Satisfying	Meeting	Supporting

To address any gaps or improve compliance, we provide recommendations specific to each FAIR dimension, presented in encapsulated boxes. In some instances, we also offer practical examples, leveraging leading community practices.

The recommendations are classified as *low, medium, and high* for NASA based on their criticality in facilitating FAIR compliance. The estimated level of effort and resource requirement for implementing a recommendation is indicated as *Easy, Moderate, Challenging, or Resources Intensive*. Additionally, when appropriate, a responsible party or target audience may be specified.

### 5.1 Being Findable

The FAIR *Findability* principles aim to enhance the discoverability of digital data. This entails assigning data with globally unique, persistent, and resolvable identifiers, such as DOIs. Moreover, data should be cataloged and described with comprehensive metadata, including linked data identifiers, to facilitate search and discovery.

Over the past several decades, NASA has invested substantial efforts to ensure Earth science data are discoverable and accessible to both human and machine users. These efforts have involved establishing standards, developing processes, and building infrastructure capacity. The current

NASA ESDS standards, practices and capabilities have been shown to strongly support the principles in the Findability dimension at the data system level (Ramapriyan and Behnke 2020). Our findings align with these conclusions at the data product level. We also offer recommendations for further improvement, primarily for machine users.

### 5.1.1. NASA ESDS enabling resources

The resources within the ESDS Program for enabling the Findability principles are highlighted in the context of a typical workflow (Steps 1 - 3 as discussed below) for curating and publishing NASA Earth science data products. The compliance level with the category-specific requirements defined in Table B1 is described and explained. The category includes data (D), metadata (M), and infrastructure (IS). When gaps are found, recommendations are provided to improve compliance. The results are summarized in Table 5.2.

**Table 5.2.** A list of category-specific requirement IDs associated with the Findability principles (described in Table B1), NASA enabling resources and their compliance levels, and O'FAIR working group (WG) recommendation identifiers (IDs) along with their level of priority and effort required to implement.

FAIR-F REQ ID	NASA Enabling Resources	Compliance Level (Satisfying/Meeting/Supporting)	O'FAIR WG Recommendation ID (Priority/Effort)
F1-REQ-D	Data Product DOI; NASA DOI Process	Satisfying	
F1-REQ-M	CMR UMM-C ConceptID	Supporting	F-1 (Low/Easy); F-2 (Medium/Moderate)
F2-REQ-D	CMR UMM-C Required metadata elements	Meeting	R-2 (Medium/Moderate)
F2-REQ-M	CMR UMM-C fields	Supporting	F-3 (Low/Moderate); A-2 (Medium/Easy); R-4 (High/Easy)
F3-REQ-M	CMR UMM-C required DOI element;  Linked DOI (landing page website; data files)	Satisfying;  Compliance level varies by DAACs	I-2 (Medium/Moderate); R-5 (High/Easy)
F4-REQ-D	CMR; Earthdata Search	Satisfying	
F4-REQ-M	CMR; Earthdata Search	Satisfying	

<b>F4-REQ-IS</b>	CMR API; Earthdata Search Portal	Satisfying	
------------------	----------------------------------	------------	--

*Step 1. Obtain a DOI for the data product through a DAAC, using the ESDIS DOI Registrar (currently DataCite) and following the defined ESDS DOI process<sup>18</sup>*

DOI is a globally unique and permanent persistent identifier that is currently recommended by NASA SMD for use in identifying data for citation in the scientific literature.<sup>19</sup> This step satisfies F1-REQ-D.

The ESDIS DOI Process includes data and metadata requirements, the DOI assignment and submission process<sup>20</sup>, and the recommendations for the DOI landing page content. (One can and should reserve DOIs prior to the delivery of actual data to a DAAC. Detailed steps for assigning/registering/reserving DOIs within ESDIS as well as updating a DOI are described as a part of the ESDIS DOI Process.)

There are additional nuances in providing a data product DOI to users. Currently the decision on the following matters is made at the discretion of individual DAACs and/or projects:

- Naming conventions and selection of optional metadata elements when assigning and minting DOIs for individual data products by DataCite, and
- The arrangement of the data product landing page<sup>21</sup>, formed from the Common Metadata Repository (CMR) Table 5.1<sup>22</sup> collection-level metadata record. The DOI may directly appear on or be linked to the landing page (see Subsection 5.3.4.4. for discussion and examples of linking data product DOI to the landing page and data file metadata). This practice induces differences in the user experience even if they all follow the same DOI process and are based on the CMR collection-level metadata record, which is discussed in more detail in the next subsection.

*Step 2. Curate data product-level metadata record, utilizing the Unified Metadata Model's for Collections (UMM-C) schema of NASA CMR<sup>23</sup>*

<sup>18</sup> <https://www.earthdata.nasa.gov/engage/doi-process> (accessed: 2023-06-16)

<sup>19</sup> [https://github.com/chutneyboy/smd-open-science-guidelines/blob/main/guidance/guideline001\\_doi\\_registration.md](https://github.com/chutneyboy/smd-open-science-guidelines/blob/main/guidance/guideline001_doi_registration.md) (accessed: 2023-06-16)

<sup>20</sup> <https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/ESDIS+DOI+Process> (accessed: 2023-06-30)

<sup>21</sup> Moving forward, data product landing pages will be more consistent with the Earthdata Web Unification effort.

<sup>22</sup> <https://www.earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr> (accessed: 2024-05-22)

<sup>23</sup> <https://wiki.earthdata.nasa.gov/display/CMR/UMM+Documents> (accessed: 2023-06-16)

The term "rich metadata" was used by Wilkinson et al. (2016) in reference to the F1 principle, which states that data should be described with rich metadata. This concept is further elaborated under the Reusability principles, where R1 calls for "*a plurality of accurate and relevant attributes.*" While Wilkinson et al. (2016) did not specify exactly which attributes qualify, they emphasized that metadata should at a minimum include a data usage license (R1.1) and provenance information (R1.2). Thus, the "rich metadata" are essential for search and discovery in Findability and (re)use in Reusability. The introduction of this term serves to make the FAIR principles domain and application agnostic while the implementation of these principles will need to be domain and application dependent. Optimality is served when enough metadata exists to facilitate discovery and understanding of the basics of the data including its usage and utility without having to consult its data producers. For Findability, "rich" means that there is sufficient information for non-expert (machine) discovery.

At the barebone of the FAIR principles, "rich" refers to data usage license and detailed provenance. In practice, however, additional metadata elements are necessary to effectively search and use the underpinning data products. Fenner et al. (2019) proposed a minimum set of descriptive metadata properties (i.e., creator, title, publisher, publication date, summary, keywords, identifier, resource type) that support data discovery and citation.

For data products, the resource type is explicit, and UMM-C uses an optional element (Collection Data Type) to denote different types of data (e.g., Near\_Real\_Time, Low\_Latency, Expedited, Science\_Quality).

The required CMR UMM-C elements include Dataset Title, Short Name, Version, Abstract, DOI, Metadata Specification, Collection Progress, Processing Level, Data Center, Science Keywords, Temporal Extent, Spatial Extent, and Platform (NASA ESDIS CMO 2023). These UMM-C elements are generally sufficient for geospatial data discovery (meeting F2-REQ-D, together with the UseConstraints element which often includes the license information).

UMM-C uses an optional element (Data Date) for capturing the information of when the data product was created/published, last updated, planned review, and deletion dates. Complete crosswalks of UMM-C fields to the FAIR principles are captured in Appendix C.

Some basic versioning and provenance information for data and very limited information for metadata are captured in UMM-C elements, supporting F2-REQ-M. See additional discussions on data usage license and provenance in Section 5.4. The need for cross referencing different versions of metadata records at the data product level is discussed in Section 5.2.

Data product DOI is mandatory to be included in the CMR UMM-C metadata schema (satisfying F3-REQ-M).

*Step 3. Integrate the data product-level metadata record into CMR for the metadata to be searchable and discoverable via the NASA Earthdata Search Portal*

One can use CMR API or the Metadata Management Tool<sup>24</sup> to enter collection records in the CMR User Acceptance Test (UAT) environment.<sup>25</sup> The metadata records are then ingested into CMR through an internal workflow which indexes metadata records with CMR tags for search and discovery.

CMR is the authoritative management system for all ESDIS metadata records. The existing CMR holdings for all the ESDIS providers can be found at:

<https://cmr.earthdata.nasa.gov/search/site/collections/directory/eosdis>.

In CMR, each collection-level metadata record is assigned an internally unique identifier (Concept ID, e.g., C2036882456-POCLOUD), which is at the minimum level of meeting F1-REQ-M.

Publicly available NASA data products are searchable in common search engines such as Google. Through the DOI minting process, data product metadata records are indexed in and searchable on DataCite. Additionally, all data product metadata records are indexed in CMR with tags and integrated into Earthdata Search Portal for search and discovery (satisfying F4-REQ-D; F4-REQ-M; F4-REQ-IS). Earthdata Search is a standards-based web application developed by NASA ESDIS to enable data discovery, search, comparison, visualization, and access across ESDIS' Earth Science data holdings. It builds upon several public-facing services provided by ESDIS, including the CMR for data discovery and access, ESDIS User Registration System (URS) authentication, the Global Imagery Browse Services (GIBS) for visualization, and a number of OPeNDAP services hosted by data providers.<sup>26</sup>

CMR maintains an endpoint for data.gov and data.nasa.gov for them to read and synchronize collection metadata records. Thus, NASA Earth science data products are searchable in these registries. The data products are also searchable via the restful CMR Search API (e.g., using curl and wget), utilizing OpenSearch.

The workflow outlined by steps 1-3 is used by DAACs, which appears to be generally sufficient to enable NASA Earth science data products to be discoverable, meeting all the *Findability* requirements. It is expected that NASA will continue to enhance the Findability for all data

---

<sup>24</sup> <https://wiki.earthdata.nasa.gov/display/CMR/Metadata+Management+Tool+%28MMT%29+User%27s+Guide> (accessed: 2023-08-10)

<sup>25</sup> <https://wiki.earthdata.nasa.gov/display/CMR/Onboarding+Process+for+a+new+CMR+provider> (accessed: 2023-08-10)

<sup>26</sup> <https://github.com/nasa/earthdata-search> (accessed: 2024-0530)

products, such as utilizing additional variable-level metadata for more effective discovery, especially for in-situ measurements and field campaign observations.

Although the CMR metadata identifiers (i.e., collectionConceptID) are not globally unique and eternally persistent, having the data product DOI linked in the data product metadata record and indexed with the Earthdata Search Portal appears to be sufficient.

To help fulfill F1-REQ-M at a higher maturity level, we recommend that all CMR Concept IDs be assigned either Universally Unique Identifiers (UUIDs) or Archival Resource Keys (ARKs), which are globally unique and persistent when managed properly for practical purposes (Recommendation F-1).

**RECOMMENDATION F-1 [Low/Easy: CMR]**

Assign globally unique identifiers, such as Universally Unique Identifiers (UUIDs) or Archival Resource Keys (ARKs), to each of the CMR ConceptIDs, at least for all the collection ConceptIDs. These identifiers should be managed for their uniqueness and persistence within CMR.

Both UUIDs and ARKs are free and open. They can be generated at scale, but both require being managed by the entity that uses them.

Top 10 online UUIDs generators are described by Geekflare<sup>27</sup>, including:

- <https://www.uuidgenerator.net/> (The website will enable generation of up to 500 UUIDs simultaneously.)
- <https://domsignal.com/uuid-generator> (This generator can generate up to 1000 UUIDs.)

Assigning and managing ARKs are less straightforward. For getting started on using ARKs, one can find relevant information at:

<https://arks.org/about/getting-started-implementing-arks/>

### *5.1.2. Persistent identifiers for other entities*

One of the decision points involves determining the appropriate granularity for assigning a DOI (Parsons 2023). A common approach is to assign a DOI to each of the individual data products even if they may be based on the measurements from the same satellite mission. On the other hand, airborne campaigns tend to assign one DOI to represent the whole collection of data products related to a common theme or category, like trace gas, aerosol properties, or even an

---

<sup>27</sup> <https://geekflare.com/best-uuid-generators/> (accessed: 2023-08-30)

instrument, from the campaign and additional DOIs to individual data products – like a book and individual chapters.

NASA maintains a list of the missions without assigning PIDs.<sup>28</sup> A unique and persistent identifier for each mission would offer several advantages. It ensures permanent identification and enhances discovery of the mission. Users can reliably locate the mission through a stable link. This identifier also facilitates citation, attribution, and tracking of impacts of the mission. We therefore recommend that a PID such as DOI be assigned to each of satellite missions or field campaigns (Recommendation F-2). The PID should resolve to a permanent landing page. The mission/campaign PID should be explicitly embedded as structured metadata on the landing page, using linked data standards such as JSON-LD or RDFa, for enhanced search engine visibility and semantic understanding (brief description of those standards and examples of linked data can be found in Subsection 5.3.4.4). We also encourage that data product PIDs be linked to the mission/campaign PID and vice versa – the latter can be considered as a qualified reference (I3). The information on how to manage this interconnected system and complex citation is much needed but it is beyond the scope of this working group.

**RECOMMENDATION F-2 [*Medium/Moderate: Mission or Field Campaign Leads*]**

Assign and maintain a globally unique and persistent identifier (PID) to the satellite mission or field campaign and ensure that the PID is resolved to a permanent mission/campaign landing page and explicitly embedded as structured, machine-actionable metadata, utilizing standards for linked data on the web.

Beyond data products, it is also beneficial to uniquely identify other entities—such as the scientists who produced the data products, metadata specialists who curated the metadata records, and the data centers that host the data products—to ensure proper credit and capture detailed provenance.

Community persistent identifier systems have been developed to uniquely identify individual researchers, e.g., Open Researcher and Contributor IDs (ORCID)<sup>29</sup> and organizations, e.g., Research Organizations Registry (ROR)<sup>30</sup> and Registry of Research Data Repositories (re3data).<sup>31</sup> We recommend individual data producers and organizations to register with appropriate persistent identifier systems if they have not yet done so, and use those PIDs in both metadata and documents when appropriate to denote uniquely the individual data producers and

---

<sup>28</sup> <https://www.nasa.gov/a-to-z-of-nasa-missions/> (accessed: 2024-01-05)

<sup>29</sup> <https://orcid.org/> (accessed: 2023-06-16)

<sup>30</sup> <https://ror.org/> (accessed: 2023-06-16)

<sup>31</sup> <https://www.re3data.org/> (accessed: 2024-05-16)

organizations (Recommendation F-3). This practice will help ensure that contributions are credited correctly, support data reproducibility, and establish provenance.

All the DAACs have already been assigned a UUID within the Global Change Master Directory's (GCMD) Keyword Management System (KMS). The current list can be downloaded at: [https://gcmd.earthdata.nasa.gov/kms/concepts/concept\\_scheme/providers?format=csv](https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/providers?format=csv)

**RECOMMENDATION F-3 [Low/Moderate: Data Producers/Curators/SIPs/DAACs]**

Register with appropriate unique and persistent identifier (PID) systems (e.g., data producers/curators with ORCID.org and organizations with ROR.org or re3data.org). Those PIDs should be used and included in relevant metadata and documents to uniquely denote individual data producers and organizations.

Other PIDs, like instrument PIDs, are under development by different external communities, for example, Research Data Alliance (RDA) Persistent Identification of Instruments-PIDINST.<sup>32</sup> For other artifacts, one may utilize Research Resource Identifiers-RRIDs.<sup>33</sup> It would be good to monitor and/or contribute to the development activities and introduce the applicable ones to the NASA community. We also encourage documenting use cases for specific needs of additional PIDs or inadequacies, if any, in the existing PIDs.

### 5.1.3. DOIs for data products from external entities

For data products from projects not assigned to a DAAC, non-ESDIS datasets, or where getting a DOI through a DAAC is not possible, it is recommended that the data products be deposited into discipline-specific, community-recognized repositories when possible, or in generalist repositories if suitable community resources are not available.<sup>34</sup> One can utilize [www.re3data.org](http://www.re3data.org) to search for a relevant domain repository. Commonly used generalist repositories include: Zenodo (<https://zenodo.org>); Open Science Framework (<https://osf.io>); and Figshare (<https://figshare.com>). One can turn to NASA SMD (2023) for additional repositories and a list of criteria to consider when selecting an external repository.

## 5.2 Being Accessible

The FAIR *Accessibility* principles underscore the importance of ensuring that the assigned data and metadata are resolvable with persistent metadata. This accessibility should be facilitated through an open, freely available, widely implementable, and standardized communication protocol, capable of accommodating authentication and authorization procedures as needed.

<sup>32</sup> <https://www.pidinst.org/> (access: 2024-05-16)

<sup>33</sup> <https://www.rrids.org/> (accessed: 2024-05-16)

<sup>34</sup> <https://www.earthdata.nasa.gov/engage/doi-process> (accessed: 2023-06-16)

Over the past three decades NASA has developed and continually refined its infrastructure to effectively provide open and free Earth science data to users worldwide. Generally speaking, the existing NASA ESDS processes, practices, tools and capabilities support the A1 principles and provide the infrastructure for enhanced access to data to better address user needs and support modern research and applications. The findings are summarized in Table 5.3.

**Table 5.3.** A list of category-specific requirement IDs associated with the Accessibility principles (description in Table B1), NASA enabling resources and their compliance levels, and O'FAIR WG recommendation IDs along with their level of priority and effort required to implement.

<b>FAIR-A REQ ID</b>	<b>NASA Enabling Resources</b>	<b>Compliance Level</b> (Satisfying/Meeting/Supporting)	<b>O'FAIR WG Recommendation ID</b>
<b>A1-REQ-D</b>	Data product DOI landing page; Worldview; Harmony	Satisfying	
<b>A1-REQ-M</b>	CMR/Earthdata Search	Satisfying	
<b>A1-REQ-IS</b>	HTTPS; OPeNDAP; CMR API/Earthdata Search Portal; Worldview; Harmony;	Satisfying	
<b>A1.1-REQ-IS</b>	HTTPS; OPeNDAP; CMR/Earthdata Search; Harmony;	Satisfying	
<b>A1.2-REQ-IS</b>	Earthdata Login System	Satisfying	
<b>A2-REQ-M</b>	Version management - Vary by DAACs	Meeting	A-1 (Low/Moderate); A-2 (Medium/Easy)

Currently, a data product DOI (e.g., 10.5067/GHGMR-4FJ04) is assigned as a part of a DAAC's archival process, registered with DataCite, and resolved to a landing page that is normally built on the CMR collection-level metadata record and managed by individual DAACs. The landing page, aka, data product landing page, often contains detailed information of access method(s) and tool(s) to get data if appropriate (e.g., <https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1>; selecting the "Data Access" tab results in a list of methods for accessing data). This practice enables the dataset to be retrieved by its identifier (satisfying A1-REQ-D). The CMR collection-level and granule-level metadata records are retrievable at the Earthdata Search

Portal<sup>35</sup> and through the CMR APIs using the data product DOI (satisfying A1-REQ-M). A CMR collection-level metadata record often contains information about and links to data access, visualization, and analysis services and tools that are applicable to that data collection. This allows machine users to automatically identify appropriate services and tools to access, visualize, and use the data.

Data product landing pages are currently established and maintained by different DAACs. The ESDS Program recently started a Web Unification effort to migrate all ESDS-funded public-facing web properties to a single web domain by December 2026. The Web Unification effort aims to improve discoverability and findability of NASA Earth science data, build efficient pathways for users to access data and information, improve website usability, eliminate redundancy, and maximize the effectiveness of communication efforts. Consistent and centrally managed data product landing pages across DAACs are a key component of this Web Unification effort. The new unified landing pages are driven by CMR metadata records and are expected to incorporate both the common and unique elements and features on landing pages currently hosted by different DAACs. They also offer more user-friendly features, such as responsive layout, efficient tabs and tables, and data visualization. The new unified landing pages will provide consistent ways for users to access data, documentation, and tools/services.

The ESDS Program is currently implementing an Earthdata Cloud Evolution<sup>36</sup>, a strategic vision to develop and operate multiple components of ESDIS in a commercial cloud environment. The migration of ESDIS data into the Earthdata Cloud benefits users by giving them new ways to access NASA's collection of Earth science datasets, improves the efficiency of data systems operations, increases user autonomy, maximizes flexibility, and offers shared services and controls. The Earthdata Cloud increases the accessibility of NASA Earth science data by providing users with direct and scalable access to large volumes of data across DAACs in the cloud, eliminating the need to download volumes of data. Users, should they want, still do retain the capability to download data. Despite the benefits that the Earthdata Cloud may offer, challenges remain and they need to be addressed to provide users with a smooth experience with Cloud-based data access and use. For example, multiple temporary short-lived Amazon Web Services (AWS) S3 credentials are needed when users need to access data from multiple DAACs in the Earthdata Cloud. The cost and unfamiliarity with the Cloud environment may hinder the ability of individual scientists to access data products hosted in the Cloud. In addition, the Earthdata Cloud is currently based on the AWS cloud platform, and improving data access across different cloud platforms will bring additional benefits to users.

---

<sup>35</sup> <https://search.earthdata.nasa.gov/search>

<sup>36</sup> <https://www.earthdata.nasa.gov/eosdis/cloud-evolution> (accessed: 2024-01-17)

The ESDIS Project has been actively adopting and developing standards-based enterprise tools and services to enhance accessibility to data, especially in and from the Earthdata Cloud. Such enterprise tools and services are rapidly evolving and being improved and expanded, but currently include:

- Earthdata Harmony services<sup>37</sup> allow users to seamlessly transform Earth observation data from different NASA DAACs into formats that are more analysis-ready by leveraging the CMR API and the NASA Earthdata access protocols.
- OPeNDAP<sup>38</sup> provides direct HTTPS-based access to subsetting of NASA data and metadata at the file/granule-level using a web browser, Panoply, Python or your favorite analysis tool.
- Earthdata GIS<sup>39</sup> provides a consistent user experience and enhances geospatial data access through platforms commonly used by the GIS community.

Examples of domain-specific data access services include the Sub-Orbital Order Tool (SOOT) for atmospheric airborne and field campaign data<sup>40</sup> and the Data Access Tool for ice and snow data.<sup>41</sup>

NASA ESDS generally leverages open and standardized protocols for data and metadata access (A1-REQ-IS). Hypertext Transfer Protocol Secure (HTTPS) is the most commonly used underlying protocol used across NASA ESDS for data and metadata access. All NASA Earthdata cloud data are accessible through HTTPS. The NASA Earthdata Cloud, which is hosted by AWS, provides access to data stored in S3 buckets, and the access is facilitated through HTTPS links. Higher-level data and metadata access protocols, including CMR, Harmony, OPeNDAP, and Earthdata GIS, offered by NASA ESDS build on top of HTTPS and follow open international/community standards, such as SpatioTemporal Asset Catalog (STAC) API and Open Geospatial Consortium (OGC) API (A1.1-REQ-IS). NASA ESDS data and metadata access services and applications support authentication and authorization through the free Earthdata Login system (A1.2-REQ-IS). These protocols are all standardized (satisfying A1-REQ-IS), open, free, universally implementable (satisfying A1.1-REQ-IS) and allow for an authentication and authorization procedure (satisfying A1.2-REQ-IS).

Practices for product versioning or whether to include multiple versions of data files in one collection tend to differ in particular communities. If a data product has multiple versions, the current best practice is to cross-reference the current version with previous version(s) in both collection-level metadata records and landing pages. The CMR UMM-C DOI field provides an

---

<sup>37</sup> <https://harmony.earthdata.nasa.gov/> (accessed: 2024-01-17)

<sup>38</sup> <https://opendap.earthdata.nasa.gov/> (accessed: 2024-01-17)

<sup>39</sup> <https://gis.earthdata.nasa.gov/portal/home/> (accessed: 2024-01-17)

<sup>40</sup> <https://asdc.larc.nasa.gov/soot/power-user> (accessed: 2024-05-17)

<sup>41</sup> <https://nsidc.org/data/data-access-tool> (accessed: 2024-05-17)

option (PreviousVersion) for including information on previous versions.<sup>42</sup> The landing page of the current version will indicate the existence of the previous version(s) and provide links to the landing page(s). At the landing page of the previous version(s), information will be provided about the latest version with the relevant DOI. This practice will enable the longevity of metadata even if the data are no longer available (meeting A2-REQ-M). For DAACs that do not yet implement a similar practice, we encourage them to do so. The Web Unification effort could help address this consistently. Recommendations A-1 and A-2 (below) are intended for: i) enhanced linkage between versions in CMR data product level metadata records and ii) enhanced linkage between versions in DOI registrations.

**RECOMMENDATION A-1 [Low/Moderate: CMR]**

Leverage the DataCite metadata schema to establish a common approach to associate DOIs of different versions of a data product and extend the Unified Metadata Model (UMM) to establish associations among data product metadata records.

**RECOMMENDATION A-2 [Medium/Easy: CMR/DAACs]**

The ESDIS CMR team and the DAACs work together to develop a common approach to cross-reference different versions of a data product and provide versioning information on data product landing pages.

### 5.3 Being Interoperable

The FAIR *Interoperability* principles strive to enable smooth integration and exchange of data and digital resources across diverse systems, tools, services, and research domains. This involves structuring data and metadata using standardized and widely accepted data models, formats, and vocabularies, while also including references to other pertinent data and metadata.

#### 5.3.1. Data interoperability - A grand challenge for global science

Data interoperability for cross-domains and disciplines is considered a grand challenge for global science (IOC 2020). Efforts are ongoing to improve the convergence of FAIR implementations (Schultes et al. 2020) with case studies of a wide range of science disciplines by the WorldFAIR project<sup>43</sup> to aim for enhanced interoperability across domains.

<sup>42</sup> <https://wiki.earthdata.nasa.gov/display/CMR/DOI> (accessed: 2024-04-25)

<sup>43</sup> <https://worldfair-project.eu/12-months-of-the-worldfair-project/> (accessed: 2024-02-22)

### *5.3.2. Key elements of data interoperability*

The term interoperability is generally understood from the perspective of interactions at the machine- or systems-level to imply an ability for data or tools from different sources to work together with minimal effort. Machine actionability is defined as a phrase "to indicate a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously acting, computational data explorer" (Wilkinson et al. 2016).

Key facets for enabling interoperability include semantic interoperability, which ensures shared meaning using ontologies and controlled vocabularies; syntactic interoperability, which relies on standardized structures and encoding rules; technical interoperability, which addresses compatibility of protocols and interfaces; and legal interoperability, which ensures that licensing enables lawful reuse and integration.

Effective data interoperability requires addressing semantic, syntactic, technical, and legal interoperability.<sup>44,45</sup>

### *5.3.3. Data interoperability in the context of FAIR*

In the context of FAIR, the Interoperability principles primarily focus on formal language for knowledge representation for both data and metadata, FAIR vocabularies, and references to other (meta)data.

#### *5.3.3.1. Formal knowledge representation language for data and metadata*

A formal knowledge representation language is a systematic, structured, and well-defined language used to represent and express knowledge in a clear, precise, and unambiguous manner to allow computers to process, understand, and use knowledge. Data file formats or metadata schemas are a couple of facets of information that such languages can represent. A formal data file format refers to a standardized and structured way of representing data that follows a predefined syntax and set of established rules; similarly, so is schema for metadata. They generally represent structured data and metadata in a way that can be easily interpreted and processed by machines, supporting syntactic interoperability.

The World Wide Web plays a critical role in facilitating data publishing, discovery, and sharing. The formal knowledge representation language, therefore, often refers to community standards and languages used for representing, organizing, and inferencing information on the web in a machine-actionable format. They enable the encoding of concepts, relationships, and metadata using standardized syntax and vocabulary (Semantic Representation) and support the principles

---

<sup>44</sup> <https://www.thehyve.nl/articles/plan-a-fair-journey> (accessed: 2023-06-01)

<sup>45</sup> <https://zenodo.org/records/162241> (accessed: 2024-04-16)

of Linked Data, which emphasize the use of URIs (Uniform Resource Identifiers) to create links between related resources on the web.

### 5.3.3.2. FAIR vocabulary

FAIR vocabulary refers to sustained, standardized, well-defined, and controlled vocabularies to ensure that data are described consistently with shared and unambiguous meanings and adhere to the FAIR principles. Examples of community-controlled vocabularies and ontologies are available in Peng et al. (2023; Table 4.7).

### 5.3.3.3. Qualified references to other (meta)data

A "qualified reference" is a reference to another relevant resource with the specific relationship clearly defined and machine interpretable (Jacobsen et al. 2020). Examples of qualified references include locations of input data sources and documentation, previous versions of data products and associated metadata records if multiple versions of data products exist, or additional service endpoints for obtaining data.

### 5.3.4. NASA ESDS enabling resources

Enabling interoperability is the most challenging area of being FAIR for NASA Earth science data products. Higher levels of interoperability would ideally facilitate seamless data and information sharing, integration, and exchanges among NASA tools and systems, across agencies that manage Earth science data products, and across disciplines. In NASA's context, the lack of interoperability between Agency data systems can make it increasingly difficult for both internal and external stakeholders to effectively discover, understand, integrate, access, and use valuable scientific and business data sources.<sup>46</sup>

Unfortunately, we are in the early stages of addressing this challenge. We will focus on ensuring and enhancing interoperability at the data product level within the Earth Science Division, as a solid step toward the goal of maximizing data interoperability.

NASA ESDS enabling resources are described below along with their compliances to category-specific requirements. The results are summarized in Table 5.4.

**Table 5.4.** A list of category-specific requirement IDs associated with the Interoperability principles (description in Table B1), NASA enabling resources and their compliance levels, and O'FAIR WG recommendation IDs along with their level of priority and effort required to implement.

FAIR-I REQ ID	NASA Enabling Resources	Compliance Level (Satisfying/Meeting/Supporting)	O'FAIR WG Recommendation ID
---------------	-------------------------	--	-----------------------------

<sup>46</sup> [https://www.nasa.gov/sites/default/files/atoms/files/nasa\\_data\\_strategy.pdf](https://www.nasa.gov/sites/default/files/atoms/files/nasa_data_strategy.pdf) (accessed: 2023-05-09)

			<i>(Priority/Effort)</i>
<b>I1-REQ-D</b>	netCDF; CF; ACCD; RDF; JSON-LD	Meeting; Supporting	I-1 ( <i>High/Moderate</i> )
<b>I1-REQ-M</b>	CMR UMM-C; GCMD	Meeting	I-4 ( <i>High/Easy</i> )
<b>I2-REQ-D</b>	CF Standard Names; ACVSNC	Meeting; Supporting	I-1 ( <i>High/Moderate</i> )
<b>I2-REQ-M</b>	GCMD; CF; ACDD	Supporting	I-3 ( <i>Low/Easy</i> ); I-5 ( <i>High/Moderate</i> )
<b>I3-REQ-D</b>	Linked data	Supporting	I-1 ( <i>High/Moderate</i> ); I-2 ( <i>Medium/Moderate</i> ); I-6 ( <i>High/Moderate</i> )
<b>I3-REQ-M</b>	UMM-C elements (AssociatedDOIs and AssociatedURLs)	Supporting	I-6 ( <i>High/Moderate</i> ); A-2 ( <i>Medium/Easy</i> )

#### 5.3.4.1. Data

Numerous machine-actionable data file formats have been approved for Earth science data products, including the network Common Data Form (netCDF)-4, the Hierarchical Data Format (HDF)-5, and Geographic Tagged Image File Format (GeoTIFF), Zarr, and Cloud Optimized GeoTIFF (COG).<sup>47</sup> We will use netCDF-4 as an example to demonstrate how the characteristics of the data file format are associated with the Interoperability requirements. NetCDF-4 is one of the recommended machine-readable data file formats by the ESDS Data Product Development Guide (DPDG) (Ramapriyan et al. 2024).

- Data file format: netCDF (formal, accessible, shared, and broadly applicable format for representing data in climate science and geoscience; meeting I1-REQ-D):
  - A common data model that defines the structure and organization of data within netCDF files;
  - Self-describing: they contain metadata that describes the structure and contents of the data within the file;

<sup>47</sup> <https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices#data-formats> (accessed: 2024-04-26)

- File metadata conventions: CF metadata conventions<sup>48</sup> and the Attribute Convention for Data Discovery (ACDD)<sup>49</sup>, a standardized way to describe data's dimensions, variables, and attributes (supporting I2-REQ-M);
  - CF compliance assumes use of CF Standard Name Table<sup>50</sup> (meeting I2-REQ-D), when available. For measurements common to field campaigns, CF standard names tend to have poor coverage.
- Platform-independent: making it easy to share and exchange data between different systems and software applications that support netCDF.

While netCDF facilitates the storage and organization of scientific data, it does not provide the formal semantics, ontological constructs, or inferencing capabilities.

Additional challenges to interoperability can come from the actual applications of the CF conventions and ACDD recommendations to different types of data which can be from either model or measurements. In these cases, additional attributes are introduced, as an extension of CF, to sustain data usability and provenance. As these practices are data source (e.g., instruments or measurements) dependent, it may be more difficult to make them interoperable.

#### 5.3.4.2. Metadata

- Metadata schema and standards:
  - CMR UMM-C (formal, accessible, shared, and ESDIS-wide with limited application in the Earth science community) (meeting I1-REQ-M).
  - UMM-C crosswalks to ISO 19115, ECHO-10, and DIF-10 metadata standards.
  - Other types of the metadata schemas in CMR: In addition to UMM-C, there are other UMM profiles that the CMR Metadata Team has created or is working on (Figure 5.1). We encourage using the schemas and terminology associated with the relevant types of metadata, when appropriate.

Similar to netCDF for data file format, while CMR UMM-C provides a managed metadata schema for NASA Earth science data products, it does not provide formal semantics, ontological constructs, or inferencing capabilities.

---

<sup>48</sup> <https://cfconventions.org/> (accessed: 2023-08-25)

<sup>49</sup> [https://wiki.esipfed.org/Attribute\\_Convention\\_for\\_Data\\_Discovery\\_1-3](https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3) (accessed: 2023-08-05)

<sup>50</sup> <http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html> (accessed: 2023-08-25)



Figure 5.1. Unified Metadata Model (UMM) for NASA's Earth Science Data and Information System (ESDIS) metadata. Based on a source diagram from: <https://www.earthdata.nasa.gov/unified-metadata-model-umm> (accessed: 2023-05-16).

The relevant documents describing conceptual models of the defined concepts shown in Figure 5.1 are available at: <https://wiki.earthdata.nasa.gov/display/CMR/UMM+Documents>

#### 5.3.4.3. Semantic representation

Community standards and languages used for representing, organizing, and inferring information on the web in a machine-actionable format include RDF (Resource Description Framework), RDFS (RDF Schema), OWL (Web Ontology Language), and SKOS (Simple Knowledge Organization System).

RDF and SKOS standards have been used to represent GCMD keywords.<sup>51</sup> Its level of FAIRness in semantic interoperability, however, can be improved by employing additional relevant ontologies (see the discussion on GCMD limitations in Subsection 5.3.4.5.)

In general, NASA ESDS enterprise tools and resources are lacking in adopting the standards and languages used for representing, organizing, and inferring data and information on the semantic web. We therefore recommend that actionable strategies be developed at the ESDIS level (Recommendation I-1).

#### 5.3.4.4. *Linked data*

##### i) Linked data for data product DOIs

While the data product DOI process is consistently defined at the ESDIS level, data product/DOI landing pages are created by individual DAACs. Historically they have primarily been focused on serving human end users. The technologies, standards (and their implementation) on these landing pages exhibited notable variability across different DAACs. A light analysis of the FAIRness assessment results by the FAIRsFAIR F-UJI tool<sup>52</sup> has revealed that data product DOIs resolve to corresponding landing pages via DataCite services, while landing pages which are hosted by DAACs do not always effectively or correctly provide DOIs for machine access. Steps can be taken to also bolster integration by tools and systems.

Fenner et al. (2019) have emphasized: "To verify that a persistent identifier resolves to a correct landing page, the persistent identifier must be embedded in the landing page in human-readable and machine-readable formats. The persistent identifier should be found somewhere on the landing page, but is ideally embedded in schema.org markup and/or using HTML meta tags."

The following examples are adapted from those provided by Fenner et al. (2019):

Example schema.org/JSON-LD:

```
<application type="application/ld+json"> {"@id": "https://doi.org/10.5067/GHGMR-4FJ04"}  
</application>
```

Example HTML meta tags:

```
<meta name="DC.identifier" content="https://doi.org/10.5067/GHGMR-4FJ04">
```

Here, JSON-LD, standing for JavaScript Object Notation for Linked Data, is a lightweight Linked Data format to easily read and write structured data on the web using open vocabularies like

---

<sup>51</sup> <https://gcmd.earthdata.nasa.gov/static/kms/> (accessed: 2024-03-29)

<sup>52</sup> <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool> (accessed: 2023-05-11)

schema.org. A quick introduction and start-up guide on JSON-LD can be found at: <https://moz.com/blog/json-ld-for-beginners>.

Fenner et al. (2018) have also stated that "Machine-readable metadata available from landing pages for datasets facilitate data citation by enabling easy integration with reference managers and other tools used in a data citation workflow. Embedding these metadata using the schema.org standard with the JSON-LD is emerging as the community standard."

We have observed that, currently, a human user may find data product DOIs, somewhere on the DAAC DOI landing pages, either in a Summary section or via the Citation tab. However, the information is not necessarily machine actionable. We therefore encourage all DAACs to accurately adopt community standards such as schema.org standard with JSON-LD when formulating their landing pages.

We conducted a preliminary analysis to examine how DOI information is incorporated into the HTML of landing pages, and the results revealed significant variations, ranging from not including any DOI information, partially or incorrectly implemented JSON-LD, to fully and correctly implemented the JSON-LD standards. It demonstrated that steps can be taken to improve/implement findability, accessibility, and interoperability for computational systems. Therefore, it may be beneficial to develop a consolidated landing page template for all DAACs. By doing so, it not only provides a unified user experience for both human and machine end users, it also ensures that web crawlers can effectively extract content from data product websites, leading to better indexing and improved visibility in search engine results.

We therefore recommend that the ESDIS Project systematically examine the matter for all DAACs and provide ESDIS-wide best practice recommendations to maximize the ability of NASA Earth science data products and information to be effectively accessed and integrated by machines, for example, under the on-going Web Unification effort (Recommendation I-1).

**RECOMMENDATION I-1 [*High/Challenge: ESDIS System Engineer/Web Unification Team*]**

Develop actionable strategies and guidelines to enable or enhance machine-actionability, utilizing community Linked Data standards and the semantic representation of NASA Earth science digital data and other associated digital resources.

NetCDF is a preferred machine-actionable file format for NASA Earth science data recommended in the DPDG (Ramapriyan and Leonard 2020, updated 2021; Ramapriyan et al. 2024). If data files are in netCDF, we recommend incorporating global attributes such as "id" and "naming\_authority" attributes, along with the metadata\_link attribute for landing page based on

the Attribute Convention for Data Discovery (ACDD), into all the individual data files (Recommendation I-2), as in the example below:

```
//global attributes:  
:id = "10.5067/GHGMR-4FJ04";  
:naming_authority = "org.doi";  
...  
:metadata_link = "https://doi.org/10.5067/GHGMR-4FJ04"  
...
```

This practice ensures that the data product DOI and its landing page are recorded in all file-level metadata. However, this workflow entails reserving a DOI early in the data product lifecycle in collaboration with the designated DAAC, so that even when sample files of the product are generated for testing, the DOI can be included in the file-level metadata (See Section 7 of the DPDG - Ramapriyan et al. 2024). When the data product is ready to be published, the reserved DOI would be formally registered, and an appropriate landing page generated. The same practice can be utilized to include other data product identifiers such as UUIDs or ARKs.

**RECOMMENDATION I-2 [*Medium/Moderate: Data Producers/DAACs*]**

Data product DOI (collection) should be included in individual data files (granules) in a machine-actionable fashion, if appropriate.

ii) Linked data for open Creative Common (CC) licenses

RDF, i.e., Resource Description Framework, is a data model and framework for describing resources on the web and their relationships in a standardized way. RDFa (RDF in Attributes) is an extension of HTML that allows embedding RDF data within HTML documents using attributes in the markup. Both RDF and RDFa are World Wide Web Consortium (W3C) standards.

The R1.1 principle states that (meta)data be released with a clear and accessible data usage license. They are detailed in Subsection 5.4.3.

Here is an example of linking an open CC-BY 4.0 license to the website:

```
<html>  
...  
<body>  
...  
<p>All content on this site is licensed under  
<a property="https://creativecommons.org/ns#license"
```

```
href="https://creativecommons.org/licenses/by/4.0/">  
a Creative Commons License</a>.</p>  
...  
</body>  
<html>
```

As noted in Subsection 5.4.3, the CC0 license should be used when appropriate to maximize data sharing. In such cases, the hybrid link in the example above can point to: <https://creativecommons.org/public-domain/cc0/>.

#### 5.3.4.5. *Vocabularies*

##### i) GCMD Keywords

GCMD keywords are broadly used by NASA DAACs and many other data centers, and the use of these keywords can be generally interpreted as meeting the I2 principle, albeit with some limitations - see the discussion below.

GCMD keywords contain science keywords, platforms, instruments, data centers, locations, projects, providers, services, resolution (Horizontal Resolution Ranges; Vertical Resolution Ranges, Temporal Resolution Ranges), Data Format, Measurement Name, etc.

GCMD keywords are a hierarchical set of controlled Earth Science vocabularies maintained in the Keyword Management System (KMS).<sup>53</sup>

- Each keyword category is assigned a category title and UUID;
- Each keyword is assigned a UUID;
- GCMD KMS API<sup>54</sup> generates UUIDs for all keywords.

The GCMD Keywords are actively maintained by the CMR Metadata Team, now ESDIS Metadata Stewardship Team, while the ESDIS Standards Coordination Office (ESCO) manages and coordinates the review of the keywords, according to the defined governance and review process (NASA ESDIS 2016). The community is welcome to propose updates and additions.

##### Limitations of GCMD

The GCMD keywords, notably the science keywords, are very useful and broadly adopted, but they do not adhere to some of the specific FAIR principles. The limitations have been explored in Parsons, Duerr, & Godøy (2022), and the pertinent aspects are outlined below.

<sup>53</sup> <https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords> (accessed: 2023-11-14)

<sup>54</sup> <https://wiki.earthdata.nasa.gov/display/ED/KMS+User%27s+Guide> (accessed: 2024-03-26)

The primary issue is with the identifiers for the science keywords. The science keywords are organized in a hierarchy (Category > Topic > Term > Variable > Detailed Variable) where the same Term or Variable may appear under multiple categories. For example, the Term "Sea Ice" appears under the Categories "Oceans" and "Cryosphere". The Term has the same definition regardless of what Category it comes under, but the Term has different URIs depending on what Category it is under. The keywords are available through a RESTful API and are downloadable in RDF, JSON, XML, and CSV (meeting the requirements of I1), but the multiple identifiers can confuse machines and automated workflows which expect one identifier for the same definition. This highlights a potential issue with hierarchical representation of terms rather than representing them through a fully linked graph or ontology. Basically, the science keywords are only nouns and include none of the predicates which would allow for greater understanding of context.

SWEET (Raskin and Pan 2005) begins to address these limitations, but it is not a direct match with the GCMD Science Keywords, and it is not nearly as comprehensive. Currently SWEET is a set of ~200 ontologies maintained by the Semantic Technologies Committee of the Earth Science Information Partners (ESIP)<sup>55</sup>, which welcomes participation from all the DAACs. Some of the GCMD Topics are described quite well by SWEET, are well interconnected with other widely adopted ontologies, and are fully FAIR; other GCMD Topics are incomplete or missing entirely from SWEET (Parsons, Duerr, & Godøy 2022)

In short, the GCMD Keywords are mostly FAIR. They are generally findable although it would be beneficial if they were registered in one or more vocabulary services or semantic repositories like the [ESIP Community Ontology Repository](#). They are accessible through common open protocols. They are mostly interoperable in that they are broadly used and represented in common knowledge representation languages although they lack some aspects of machine actionability as discussed above. They are reusable in that they are generally understood to be open although they have no formal machine-readable usage license, which we recommend GCMD include (Recommendation I-3).

**RECOMMENDATION I-3\* [Low/Easy: CMR]**

Release the GCMD keywords collection, and re-release the collection as it changes, with a formal CC0+Attribution Request usage license and link it to the GCMD landing page with the recommended citation. The recommended citation should include the license in a machine-actionable fashion.

\* As mentioned in Conclusion, a CC0 license has been assigned to the GCMD keywords collection.

<sup>55</sup> <https://github.com/ESIPFed/sweet> (accessed: 2024-03-26)

Compliance with the GCMD keywords is required by CMR.<sup>56</sup> Data producers and DAACs will go a long way to meeting the Interoperability requirements of FAIR by adopting the GCMD Keywords for their data (Recommendation I-4). Data can become even more interoperable if described using relevant ontologies such as SWEET.

**RECOMMENDATION I-4 [High/Easy: Data Producers/DAACs]**

Utilize GCMD Science Keywords (including their URIs) in metadata to describe their data products, when it is appropriate to do so.

ii) Other NASA vocabularies

Additional NASA vocabularies are listed in Table D1, including that for the Atmospheric Composition Variable Standard Name Convention (ACVSNC), which is an approved ESDS convention.<sup>57</sup> The ACVSNC adopts a decomposition approach with sets of controlled vocabularies to define measurands and associated attributes, which enhances discoverability and usability of field campaign data products. The decomposition approach used is primarily based on the physical state (e.g., gas, aerosol, cloud) of the measurands and the measurement characteristics. The ACVSNC vocabularies can be readily mapped to the I-ADOPT framework ontology, e.g., ACTRIS, which supports interoperability (Silverman et al. 2023).

The NASA vocabulary resources are available at various organizational levels, some of which are customized for specific domains or disciplines such as CASEI for the airborne and field campaign data. While they may not be as well managed as GCMD keywords, they still serve as valuable references.

iii) Recommendations for improving semantic representation of NASA vocabularies

Currently, most if not all of NASA vocabularies (Table D1) are not described for semantic web. We recommend NASA to collaborate with pertinent communities to develop actionable strategies to improve the semantic representation of NASA vocabularies, preferably starting with GCMD, working closely with the ESIP Semantic Committee along with the data producers and data users (Recommendation I-5).

---

<sup>56</sup> <https://wiki.earthdata.nasa.gov/pages/viewpage.action?spaceKey=CMR&title=CMR+Data+Partner+User+Guide> (accessed: 2024-04-15)

<sup>57</sup> <https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/acvsnc> (accessed: 2024-04-11)

**RECOMMENDATION I-5 [*High/Moderate: ESDIS System Engineer/CMR*]**

Explore community standards, develop actionable strategies to improve the semantic representation of NASA vocabularies, and support their implementation for GCMD science keywords, utilizing relevant ontologies like those in SWEET.

In the meantime, we propose the following actionable steps to the CMR team for GCMD: 1) registering keywords with a community ontology registry such as the ESIP Community Ontology Registry<sup>58</sup> (while some are already registered, others like science keywords are not); 2) Enforcing a single UUID for each keyword, regardless of its placement within a hierarchy; and 3) Augmenting RDF/JSON representations with predicates, such as the Term "is a subclass" of Topic.

#### 5.3.4.6. *Qualified references to other (meta)data*

Currently, there are no unified guidelines on what references are considered as qualified for NASA Earth science data products – most likely to be domain-dependent but such guidelines are needed. We recommend defining a set of core fields as qualified references for the Earth science data products along with best practices of embedding them in data product landing pages as linked data (Recommendation I-6).

In the meantime, we suggest (cross)referencing the following information for data products, using JSON-LD and html meta tags (supporting I3-REQ-D; R1.3-REQ-M) and UMM-C elements such as Associated DOIs and AssociatedURLs (supporting I3-REQ-M), including:

- Data product DOI;
- Data product usage license;
- Data product landing page URL;
- Collection-level metadata record, including that of previous versions of the data products if applicable;
- ATBD (Algorithm Theoretical Basis Document); One can use the APT tool to generate standardized ATBD<sup>59</sup>;
- Science paper – A peer-reviewed journal article to describe the science question(s) and interpret new research findings, often providing in-depth validation analysis to demonstrate scientific soundness the data product with data uncertainty estimates;
- Data paper – Usually a journal article to provide a detailed description of the data product to facilitate its (re)use, usually including a light validation analysis;
- User Guide;

<sup>58</sup> <https://esipfed.github.io/cor/> (accessed: 2024-04-11)

<sup>59</sup> <https://www.earthdata.nasa.gov/apt/> (accessed: 2023-10-04)

- Service endpoints for accessing data files.

How to link those references in a machine-actionable fashion currently varies across DAACs. See examples of linking data product DOIs in landing pages and netCDF data files in Subsection 5.3.4.4. Recommendation I-6 aims to improve the consistency in embedding them in data product landing pages as linked data. In addition, our recommendation A-2 helps provide machine-actionable versioning information.

**RECOMMENDATION I-6 [High/Moderate: ESDIS System Engineer/CMR]**

Develop an ESDIS-wide template for defining a set of core fields as qualified references and provide best practices of embedding them in data product landing pages as linked data and that of including them in CMR data product level metadata records.

#### 5.3.4.7. Application Programming Interfaces (APIs)

Well-documented APIs provide a clear set of functions, methods, and endpoints that other applications can use to access features and data.

NASA ESDS offers several documented APIs. Some examples of APIs are those for CMR, GCMD keywords, Harmony, and GIBS.<sup>60</sup>

#### 5.3.4.8. File-level resources

The Dataset Interoperability Working Group and the DPDG Working Group provide specific recommendations at the file-level for enabling or improving NASA Earth science data interoperability<sup>61</sup> that may be beneficial to review and adopt. Relevant examples include recommendations of utilizing CF for variable standard names, netCDF and GeoTIFF for preferred machine-readable data file formats. We encourage data producers to assess the existing templates and propose any necessary extensions to the CF conventions, to make their data interoperable.

## 5.4 Being Reusable

The FAIR *Reusability* principles promote the creation of data and digital resources that are comprehensively described so that they are readily understood, interpreted, and (re)used. This involves adhering to domain-relevant standards and providing comprehensive data documentation, including clear data usage license and detailed provenance.

<sup>60</sup> <https://www.earthdata.nasa.gov/engage/open-data-services-and-software/api> (accessed: 2024-05-29)

<sup>61</sup> <https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/dataset-interoperability-recommendations-for-earth-science> (accessed: 2023-07-24)

The NASA enabling resources are identified and evaluated – the results are described and summarized in Table 5.5.

**Table 5.5.** A list of category-specific requirement IDs associated with the Reusability principles (description in Table B1), NASA enabling resources and their compliance levels, and O'FAIR WG recommendation IDs along with their level of priority and effort required to implement.

<b>FAIR-R REQ ID</b>	<b>NASA Enabling Resources</b>	<b>Compliance Level</b> (Satisfying/Meeting/Supporting)	<b>O'FAIR WG Recommendation ID</b> (Priority/Effort)
<b>R1-REQ-D</b>	CMR UMM-C elements; GCMD keywords; netCDF global attributes	Supporting	R-1 (High/Moderate) R-3 (High/Easy)
<b>R1-REQ-M</b>	CMR UMM-C required elements; GCMD keywords	Meeting	R-2 (Medium/Moderate); I-3 (Low/Easy); R-4 (High/Easy)
<b>R1.1-REQ-D</b>	NASA Data Policy	Supporting	R-3 (High/Easy)
<b>R1.1-REQ-M</b>	Publication process of CMR UMM records; GCMD;	Supporting	I-3 (Low/Easy)
<b>R1.2-REQ-D</b>	Some of the UMM-C elements; netCDF global attributes	Supporting	R-5 (Medium/Challenging)
<b>R1.2-REQ-M</b>	A few of the UMM-C elements	Supporting	R-5 (Medium/Challenging); A-1 (Low/Moderate); A-2 (Medium/Easy)
<b>R1.3-REQ-D</b>	netCDF; HDF; GeoTIFF; DPDG;	Meeting	R-1 (High/Moderate)
<b>R1.3-REQ-M</b>	UMM (DIF 10; ECHO 10; ISO 19115); CF; ACDD; JSON-LD	Varying: mostly Meeting; ISO 19115 & JSON-LD - Satisfying)	

**5.4.1. Richly described with plurality of accurate and relevant attributes**

The degree of "richness" needed to implement the FAIR principles depends on specific domain and application. Determining a suitable degree of richness should be a collaborative effort between domain and data scientists, ensuring that the data are FAIR to both human and machine users, based on the context and intended application.

Reusability should ensure that NASA-generated data products are usable by a diverse user community. While it's important for users to consult with the science team during the mission period and later get support from the discipline-DAAC, sufficient and accurate information should be provided to allow users to gain a basic understanding of the data without needing to consult with the data producers.

From this point of view, the CMR metadata team has specified a rich set of metadata using the UMM with its various schemas (e.g., Collections, Granule, Services, Variables, Tools, and Visualizations). A list of metadata attributes derived from CF and ACDD has been included in the Appendices D and E of the DPDG document with a mapping to the UMM schemas (Ramapriyan et al. 2024). These attributes have been grouped into global and variable-level attributes. The latest version of the DPDG (Version 2.0) includes a mapping of each of the metadata attributes listed in Appendices D and E to the FAIR principles that the attribute supports (Armstrong et al. 2023; Ramapriyan et al. 2024). Section 6 of the DPDG v2.0 includes a list of metadata compliance checkers that can be used by data producers to assess that the metadata fields are all populated and are meaningful. The O'FAIR WG understands that these attributes may not be sufficient to describe some aspects of the variables and the measurements for research use. Some attributes needed for research use may be measurement-specific and need to be identified/developed through collaboration between data scientists and research scientists. A Preservation Content Specification (PCS) has been published identifying various categories of data, software and documentation that warrant preservation along with a given data product (Ramapriyan, Moses, & Smith 2022). It is essential for the teams associated with each mission or project to determine the specific artifacts to preserve within each category. This necessitates collaboration between data repositories and the mission, project, or research teams. Considering the long history of ESDIS and the comparatively recent publication of the aforementioned documents, it is likely that older data products may not have adhered to these requirements, specifications, or recommendations.

**RECOMMENDATION R-1 [*High/Moderate: Data Producers*]**

Data producers should follow the recommendations provided in the Data Product Development Guide (DPDG) document. Data producers should also raise questions and provide feedback when issues are encountered in application of the DPDG.

The quality of CMR collection-level and granule-level metadata in terms of accuracy and completeness of metadata elements as well as conformance to the defined metadata schemas can influence the FAIRness of datasets. For better compliance with the FAIR reusability principle, we recommend that the metadata associated with the data products and the associated documentation as well as software be made compliant with the above documents (Recommendation R-1). We also advocate a systematic and periodic effort, similar to that conducted by the IMPACT ARC team<sup>62</sup>, to evaluate metadata for various products offered by the DAACs in relation to CMR. An open-source tool developed by the IMPACT ARC team, known as pyQuARC<sup>63</sup>, can be used to automatically evaluate the consistency, completeness and robustness of CMR metadata records. pyQuARC identifies cases where supplementary contextual information may be incorporated. It also ensures consistency in information that is common to both data product-level and file-level metadata records. We therefore recommend data producers and DAACs to use this tool to evaluate their metadata quality (Recommendation R-2).

**RECOMMENDATION R-2 [*Medium/Moderate: Data Producers/DAACs*]**

Data producers and DAACs should utilize the pyQuARC API to evaluate the consistency, completeness and robustness of their data product-level metadata records and should seek solutions to the shortcomings identified during the process.

Equally important is soliciting feedback from the actual data users and producers about relevancy of metadata elements for their suitability. Generally speaking, it is beneficial to document issues encountered during application of these recommendations. This should be an iterative process to increase the data product FAIRness.

Several metadata compliance checkers are available and are referenced in Section 6 of the DPDG (Ramapriyan et al. 2024). Note, however, that compliance checkers typically check for presence or absence of the attributes, but not the accuracy of the attribute values. These checkers also do not check if there are attributes related to measurement specific information, e.g., sample processing, data processing assumptions, and measurement detection limits. These attributes are applicable only to certain measurements, while other measurements may need a different set of attributes. The DPDG is intended to be a "living document" that is openly available for comments and revisions, and additional attributes or checks can be included if deemed necessary. As to ensuring that the values are provided and are accurate, there is no substitute for manual effort and diligence by the data producers and archive managers.

---

<sup>62</sup> <https://www.earthdata.nasa.gov/data/projects/analysis-review-cmr-project> (accessed: 2023-09-21)

<sup>63</sup> <https://github.com/NASA-IMPACT/pyQuARC> (accessed: 2023-09-21)

Going forward, the WG suggests that the ESDIS Project should work to harmonize the metadata recommendations in Appendices D and E of the DPDG, the metadata requirements of CMR, and pyQuARC, in alignment with the definitions of Levels of Service<sup>64</sup> (NASA ESDS 2020).

#### *5.4.2. Rich Metadata for reusability*

Compliance with CMR UMM-C<sup>65</sup> and the metadata attributes called for in the DPDG (see below) as well as data producers' working with the designated DAACs to which their products are to be delivered to include any additional DAAC-specific metadata will ensure that the metadata are "rich" for reusability for the designated communities. (Crosswalks between the FAIR principles and UMM-C fields are captured in Table C1.)

- As mentioned before in the Findability principles, the required CMR UMM-C elements include Dataset Title, Short Name, Version, Abstract, DOI, Metadata Specification, Collection Progress, Processing Level, Data Center, Science Keywords, Temporal Extent, Spatial Extent, and Platform; These are common elements for data discovery. Combined with many additional elements (e.g., sampling, data processing), they are a good set of metadata for data reuse. (Meeting R1-REQ-M; Supporting R1-REQ-D).
- In the DPDG document (Ramapriyan et al. 2024), appendices D and E include tables listing metadata attributes (elements) that are required and/or recommended to constitute rich metadata from the points of view of interpretability, discovery, geolocation, provenance, etc. These attributes are also mapped to the various profiles (UMM-C, UMM-G, etc.) of the Unified Metadata Model.<sup>66</sup>
- Further, the attributes are obtained from a combination of CF Conventions, ACDD and the DAACs' metadata guidelines, which represent a subset of domain-relevant standards and best practices (meeting R1.3-REQ-M). Data products compliant with these recommendations are at a good starting point for enabling data reusability.
- Domain-relevant metadata are essential to support research use of data products, including description of data variables, source (e.g., measurement or model), assumptions and limitations. If the information is not all embedded in data files but available as additional related documentation files, these should be provided with proper linkages to the data products to which they apply as discussed in Section 4 of the DPDG.

#### *5.4.3. Data usage license*

A data usage license outlines the terms and conditions under which data can be accessed, used, shared, and redistributed.

---

<sup>64</sup> <https://www.earthdata.nasa.gov/engage/new-missions/level-of-service> (accessed: 2024-05-13)

<sup>65</sup> <https://wiki.earthdata.nasa.gov/display/CMR/UMM-C+Schema+Representation> (accessed: 2023-05-09)

<sup>66</sup> <https://wiki.earthdata.nasa.gov/display/CMR/UMM+Documents> (accessed: 2023-09-21)

License information is designed to be included in the Use Constraints element of UMM-C. Currently not a required element, but its inclusion is required to remain compliant with the FAIR-R1.1 principle. We therefore recommend that a machine-readable data usage license be clearly and explicitly included in a CMR UMM-C element that is mandatory (Recommendation R-3) (to satisfy R1.1-REQ-D).

NASA requires the use of open licenses for data products, and where no restrictions apply, recommends the use of Creative Commons Zero license (CC0)<sup>67</sup>, especially for NASA science mission data.<sup>68</sup> CC0 places data in the public domain by waiving all copyright and allows unrestricted reuse without requiring attribution. While attribution is not legally required under CC0, data creators can request it as a voluntary gesture aligned with Open Science norms. If attribution is strongly desired, CC0 + Attribution Request is the recommended approach.

Alternatively, the Creative Commons license Attribution 4.0 (CC-BY 4.0) license<sup>69</sup> retains copyright and requires users to provide appropriate credit to the original creator(s). CC-BY 4.0 allows broad reuse—including commercial and derivative works—but introduces greater complexity when applied to data products derived from multiple sources, particularly in ensuring proper attribution for each contributor. This complexity can be especially challenging for machine users, making CC0 generally the preferred license for scientific data.

Both CC0 and CC-BY 4.0 support attribution to original creators, but CC0 offers greater legal and technical interoperability, as outlined by the RDA-CODATA Legal Interoperability Interest Group (2016).

Additional details about these and other licenses are captured in the Software Package Data Exchange (SPDX) Specification and in general NASA guidance.<sup>70,71</sup> NASA does not currently provide a capability for providing machine-readable CC licenses. We therefore recommend utilizing those from the SPDX list (<https://spdx.org/licenses/>) (Recommendation R-3). Recommendations are also made to include the DOI as a mandatory item on every data product's landing page as well as in file level metadata (Recommendation R-4).

---

<sup>67</sup> <https://creativecommons.org/share-your-work/public-domain/cc0/> (accessed: 2023-07-03)

<sup>68</sup> [https://github.com/nasa/smd-open-science-guidelines/blob/main/request\\_for\\_comment/draft/rfc\\_002\\_how\\_to\\_make\\_data\\_fair.md](https://github.com/nasa/smd-open-science-guidelines/blob/main/request_for_comment/draft/rfc_002_how_to_make_data_fair.md) (accessed: 2024-05-31)

<sup>69</sup> <https://creativecommons.org/licenses/by/4.0/> (accessed: 2023-07-03)

<sup>70</sup> <https://spdx.org/licenses/CC-BY-4.0.html> (accessed: 2023-07-03)

<sup>71</sup> <https://science.data.nasa.gov/about/license> (accessed 2024-04-04)

**RECOMMENDATION R-3 [High/Easy: CMR]**

A machine-readable Creative Commons data usage license such as that from the Software Package Data Exchange (SPDX) Specification (<https://spdx.org/licenses/>) should be explicitly included in a mandatory CMR UMM-C element.

**RECOMMENDATION R-4 [High/Easy: DAAC]**

The data product DOI should be a mandatory, machine-readable, and linked-data element on its landing page.

In addition to a data usage license for data products, controlled vocabularies should also be released with an open data usage license. Specifically, we recommend a CC0+Attribution Request license for the GCMD keywords collection (Recommendation I-3).

At a minimum, unless otherwise prescribed, the data users are usually requested to cite the data they use in publications. Therefore, it is always beneficial to provide a recommended citation for a data product. The landing pages for most (if not all) datasets at the DAACs include suggested citations.

For ESDIS data, one can include the ESDIS Data Use License using CMR UMM-C in the "UseConstraints/LicenseURL" field<sup>72</sup>:

- Linkage: <https://www.earthdata.nasa.gov/engage/open-data-services-software-policies/data-use-guidance>
- Protocol: https
- Name: License URL
- Description: The Earth Science Data and Information System (ESDIS) data use policy for NASA data.
- MimeType: text/html
- ...
- Linkage: <https://spdx.org/licenses/CC0-1.0.html>
- Description: Machine-readable Creative Commons Zero v1.0 Universal data usage license

We also recommend displaying the data usage license on the landing page by embedded RDFa lite semantic annotations in HTML.<sup>73</sup> An example is included in Subsection 5.3.4.4.

<sup>72</sup> <https://wiki.earthdata.nasa.gov/display/CMR/Use+Constraints> (accessed: 2023-07-07)

<sup>73</sup> <https://www.w3.org/TR/rdfa-lite> (accessed: 2023-09-18)

#### 5.4.4. Provenance

(Meta)data provenance provides information about the origin, source, and history of (meta)data, and is crucial for understanding and assessing the reliability, quality, and trustworthiness of the data. The following related UMM-C elements (both required and not required) may directly or indirectly support (meta)data lineage/provenance but are not comprehensive.

- Version: This element identifies the current dataset version. Required. (Directly supports R1.2-REQ-D).
- Data Dates: This element is used to identify dates when the data or resource itself changed in some way. Type (Create/Update/Review/Delete) and Date. Not required. (Directly supports R1.2-REQ-D).
- 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. Not required. (Directly supports R1.2-REQ-M).
- Collection Progress: the production status of the dataset (Planned; Active/Complete; Deprecated; Not Applicable). Required. (Directly supports R1.2-REQ-D).
- Processing Level: the level at which the data in the collection are processed - aligning with the ESDIS data processing levels (0, 1A, 1B, 1C, 2, 2A, 2B, 3, 3A, 4).<sup>74</sup> Required. (Indirectly supports R1.2-REQ-D).

In research, data revision is somewhat frequent. Knowing what, why, and how the changes are made to the data is critical, including the differences in data processing procedures and/or ancillary data used in processing. We, therefore, recommend a more robust provenance for NASA Earth Science data products (Recommendation R-5).

Appendices D & E in the DPDG document list attributes pertaining to provenance with a subset of them defined as required. The "Related URL" field in both UMM-C and UMM-G can be used to link to resources that provide information such as processing and product history.<sup>75</sup> It might be helpful to have all data products to follow those attribute recommendations – Uniformity across DAACs is highly desirable.

Structured and machine-actionable ways of capturing information about data processing workflows, data derivation relationships, and versioning information can be achieved using PROV-O and schema.org predicates.<sup>76</sup> Implementation of our Recommendations A-1 and A-2 will help provide machine-actionable data product versioning information.

---

<sup>74</sup> <https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels> (accessed: 2023-09-18)

<sup>75</sup> <https://wiki.earthdata.nasa.gov/display/CMR/UMM+Documents> (access: 2024-04-03)

<sup>76</sup> <https://github.com/ESIPFed/science-on-schema.org/blob/master/guides/Dataset.md#provenance-relationships> (accessed: 2024-04-03)

Recognizing W3 PROV-O is quite general and complex, some previous efforts on implementing PROV-O for ESDS<sup>77</sup> have been carried out by the Provenance for Earth Science (PROV-ES) Working Group<sup>78</sup> (e.g., Ramapriyan et al. 2016). Practical guidance on implementing it for Earth science data products is still needed, especially on a set of metadata elements to be included in CMR UMM-C and/or UMM-G for a more robust data product provenance. We recommend developing specifications at the ESDIS level (Recommendation R-5).

**RECOMMENDATION R-5 [Medium/Challenging: ESDIS]**

Establish an ESDIS procedure/task force to develop technical and practical specifications for capturing a more robust provenance of NASA Earth science data products in CMR UMM-C and/or UMM-G.

#### 5.4.5. Community standards

CMR UMM-C is a NASA ESDS standard which can be crosswalked to Directory Interchange Format (DIF) and ISO 19115 (R1.3-REQ-M). However, UMM-C is shy of being a globally adopted community standard.

Standard data file formats (HDF, HDF-EOS, netCDF, etc.) have been used for most of the products archived by ESDIS, but given the flexibility offered by the formats in implementation, there has been some inconsistency across various data products and limited usability because of how data are structured. Additionally, the metadata attributes may be in free text which limits the interoperability and usability. Recommendations from the Dataset Interoperability Working Group (DIWG)<sup>79</sup> and the DPDG should be followed to improve interoperability as well as to comply with R1.3-REQ-D by being compliant with netCDF CF<sup>80</sup>, for example. Data producers, users, and the DAAC should proactively engage in discussions to define and agree upon the format and metadata template for the data products. Early coordination ensures the products meet both user needs and repository requirements, supporting long-term usability and FAIR compliance. In addition, we would like to reiterate the importance of vetting sample products with the expected user community representatives early in the product lifecycle (generating

---

<sup>77</sup>

<https://wiki.earthdata.nasa.gov/pages/viewpage.action?spaceKey=ESDSWG&title=A+Primer+for+Provenance+for+Earth+Science+Data+Systems> (accessed: 2024-05-23)

<sup>78</sup> <https://wiki.earthdata.nasa.gov/display/ESDSWG/Provenance+for+Earth+Science+%28PROV-ES%29+Working+Group> (Accessed: 2024-05-23)

<sup>79</sup> <https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/dataset-interoperability-recommendations-for-earth-science> (Accessed: 2024-05-23)

<sup>80</sup> <http://cfconventions.org/> (accessed: 2023-06-21)

sample products, having them checked out by independent users, responding to feedback and fixing any problems early before starting "full production" (Ramapriyan et al. 2024)

Specifically, the data file formats (e.g., netCDF; HDF, GeoTIFF) and file level metadata (e.g., DIF-10; ISO 19115) meet community standards (meeting I1-REQ-D; R1.3-REQ-D; R1.3-REQ-M).

#### *5.4.5.1. Extensions of CF Conventions for Measurements*

While the CF conventions and ACDD recommendations provide a solid set of guidelines to use HDF and netCDF formats for reporting scientific data products, it has also been realized that these guidelines can be insufficient to represent measurement-specific metadata in some cases. To overcome this deficiency, various communities or projects opted to develop templates as extensions to CF conventions. One prominent example is the Generic Earth Observation Metadata Standard (GEOMS) standards (<https://evdc.esa.int/documentation/geoms/>), which is actively maintained by a team of data scientists and research scientists. While helping to better scientifically describe the data products, these templates tend to have low consistency and therefore limited interoperability. This reflects that many metadata attributes are data source (or measurement) specific and the community/project efforts are independent from each other. One way to mitigate these challenges is to develop comprehensive metadata schema and standards. This effort will require synergistic collaboration between both data scientists and research scientists. A number of parallel ongoing efforts exist that are led by the World Meteorological Organization (WMO) and other organizations, which have achieved varying degrees of success

#### *5.4.5.2. Importance of variable-level metadata standards*

Variable-level metadata are a key element in making data usable, which should accurately describe the data, e.g., definitions, units, as well as limitations. Generic variable-level metadata needs to include contextual information. It is an actively pursued area with diverse approaches towards establishing some consistency across science disciplines. Appendix E of the DPDG lists a set of important variable-level attributes. We do, however, advocate the ESDIS Project to support and actively participate in community efforts such as those by OGC, WMO, RDA, and the Committee on Data of the International Science Council (CODATA). At the same time, we encourage data producers and DAACs to utilize any known standards and/or standard ways of capturing contextual information.

## **5.5 Roles and Responsibilities of Stakeholders**

To foster and improve the FAIRness of Earth science data products, a collaborative effort from diverse stakeholders is necessary. The specific roles and responsibilities may vary for stakeholders across different stages of a data product's lifecycle. While a comprehensive

delineation of roles and responsibilities is beyond the scope of this document, a brief description of what actions various stakeholders can take to support FAIR data is provided below.

We highly encourage data producers to work to understand the intent and familiarize themselves with NASA data product standards and best practices and the data publication process. NASA Earthdata Pub provides collective information on NASA policies and resources and serves as a good starting point.<sup>81</sup> A comprehensive set of responsibilities for data producers, related to but not yet directly focused on FAIR principles, are covered by the DPDG.

Recognizing that for those unfamiliar with its content, the DPDG document can present a steep learning curve, a cheat sheet has been developed to summarize key actions at each stage of the data product development lifecycle and directing users to the relevant sections of the DPDG, making the document easier to understand and navigate efficiently (Peng, Ramapriyan, & Armstrong 2025). Additionally, a cheat sheet was created for this document, offering an outline of the most critical information for easy reference (Peng et al. 2024b). It is equally important for the data producers to provide constructive feedback when issues are encountered during implementation of these standards and best practices, which will enable improving the existing standards and best practices.

DAACs play a critical role in ensuring that hosted data products are compliant with NASA established policies and procedures, including those for data preservation and sharing. In addition to ensuring data product quality, DAACs are responsible for assigning a DOI to and creating the DOI landing page for hosted data products.<sup>82,83</sup> Ways to access and cite data products are often described in the landing page. It should be the responsibility of the DAACs to document any issues encountered during the data ingest process. Further, DAACs should gather and document feedback from data producers and users about established policies and procedures to help improve compliance with the FAIR principles.

Service providers and tools such as the CMR metadata team and the Earthdata Search portal are the backbones of enabling the sharing of NASA Earth science data products. Earthdata Search is the gateway for search and discovery of NASA Earth science data, which is backed by CMR.<sup>84</sup> It is crucial for them to continue to address user needs and adopt community standards.

End users will have the final say regarding whether their data experiences meet their expectations and requirements, including data discovery, access, interoperability, and reuse.

---

<sup>81</sup> [https://pub.earthdata.nasa.gov/data\\_publication\\_guidelines](https://pub.earthdata.nasa.gov/data_publication_guidelines) (accessed: 2023-11-21)

<sup>82</sup> [https://github.com/chutneyboy/smd-open-science-guidelines/blob/main/guidance/guideline001\\_doi\\_registration.md](https://github.com/chutneyboy/smd-open-science-guidelines/blob/main/guidance/guideline001_doi_registration.md) (accessed 2024-04-04)

<sup>83</sup> <https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS> (accessed: 2024-05-31)

<sup>84</sup> <https://www.earthdata.nasa.gov/learn/earthdata-search> (accessed: 2024-02-12)

Therefore, NASA Earth science data end users play a critical role in the successful implementation of NASA ESDS open, free, and FAIR data policies. Over the years, several established metrics have been developed and collected by the ESDIS Metrics System (EMS) (NASA ESDIS 2024; Liu et al. 2022) for routine analysis at DAACs. Other metrics for ESDIS as well as each of the DAACs are also collected, such as the American Customer Satisfaction Index (ACSI) Reports (ACSI 2024). Revisiting those metrics would prove useful to ensure that they meet the requirements for implementing new FAIR data policies.

Since each DAAC may create their own survey questions, it can lead to difficulty in explaining the ACSI scores and therefore making improvement decisions at the ESDIS level. Cross-DAAC collaboration would be helpful for making the survey results more comparable.

## 6 DISCUSSION

Questions were brought up during the course of the O'FAIR WG, especially from the perspectives of data producers and users. We have compiled a selection of questions and corresponding answers that could be pertinent to others.

### 6.1 How Do We Deal with This for New Datasets Under Development?

SPD-41a data policy applies to digital scientific data from NASA SMD-funded scientific missions starting from the 2023 ROSES solicitations. Therefore, any new datasets under development should review this Guide document, follow identified FAIR-enabling practices and resources, and adopt recommendations as applicable.

### 6.2 How Do We “Retrofit” Existing Datasets to Make Them Compliant?

Compliance with the FAIR Principles is encouraged for NASA-funded legacy datasets. Although many factors go into prioritizing legacy datasets, identifying high-value datasets by users should be an important factor.

Before embarking on a journey towards improving the FAIRness of existing datasets, we encourage data producers or data managers to read this Guide document and the Inception Report (Peng et al. 2023) to become familiar with current NASA and community FAIR-enabling resources.

Several steps can be taken to improve the FAIRness of existing datasets, including:

- Work with the relevant DAAC to have a DOI assigned to the data product;
- Transform to a machine-actionable data format (e.g., netCDF, etc.) and using a suitable template for the intended data products (Note: templates may be dependent on the types of measurement or instrument);

- Make the metadata "richer". Using CMR UMM-C and UMM-G schemas help provide additional metadata to better describe the data variables and data source. For gridded data, a bare minimum entails providing enough information about axes, etc. to enable their handling in a viewer, e.g., Panoply for hdf/netcdf;
- Include an open data usage license such as CC0 or CC-BY 4.0 in the data product metadata and provide suggested citation.

### 6.3 What Are the Recommended Open Licenses for NASA-Hosted Data?

NASA scientific data should be released using an open data license. Unless there are other restrictions, scientific data from NASA-led missions should use CC0 with a machine-readable license in the data product metadata. CC0 is also preferable to maximize data reuse for NASA-funded scientific data if the copyright can be waived by data producers and associated organizations; otherwise, use a CC-BY 4.0 license. In either case (CC0 or CC-BY 4.0), a machine-readable recommended data citation should be provided for attributions - to allow for proper and accurate credit to the relevant entities.

Mutual agreement is required for data from external sources, which defines data use constraints such as data embargo period, data availability period, education/research use only, share alike (CC-BY-SA), non-commercial (CC-BY-NC), etc. in compliance with the "Data Rights" clause included in Cooperative Agreements and Contracts with data providers supported by NASA.<sup>85</sup>

### 6.4 What Does "Community Standards" Mean?

Developing and implementing "domain-relevant community standards" is paramount in enabling FAIR-compliance. What constitutes "community standards", however, can be arbitrary as "community" can be general and broad, such as the world geospatial science community, or narrow it to fit some specific needs, such as the ozone lidar community. This example demonstrates the complexity of defining a community, and consequently, in establishing community standards.

In addition, "user community" ties everything to products and services. Buy-in from the users guarantees that certain standards are actually used. In simple terms, the establishment of community standards should be driven by the needs of the relevant user communities with input from both domain and data scientists.

The challenge is how to best balance the efforts of efficiently and productively establishing domain-relevant community standards with the needs of the community and its users. It is important for standards-establishment entities to engage all stakeholders. Conversely, it is also

---

<sup>85</sup> <https://www.earthdata.nasa.gov/engage/open-data-services-software/data-rights-related-issues> (accessed: 2023-09-13)

essential for organizations and users to support and participate in these development activities, bringing in their unique perspectives.

## 6.5 Beyond FAIR

For NASA scientific data and information, in addition to compliance with the FAIR principles, there are other scientific and stewardship requirements. They include: ensuring scientific integrity, supporting open-source science, enabling reproducible science, promoting inclusiveness and equitability, long-term preservation, etc.

In order to ensure and maintain the scientific quality of NASA scientific data and to maximize their value and (re)usability, they need to be more than just compliant with the FAIR Principles. For example, in terms of reusability, releasing the scientific information with an open data usage license is a first step, but users often need additional contextual information to effectively understand and accurately use the data.

The following points are from the discussions among the O'FAIR WG members and listed here as takeaways:

- *Human-side*: While the FAIR Principles emphasize facilitating data sharing in machine-friendly environments, it is crucial to recognize that human users play a vital role in the data (re)use ecosystem. File formats widely used can vary significantly across different domains or among users with different levels of computing skills. To minimize obstacles to data utilization, offering data in multiple established machine-actionable formats or even a more human-friendly format (not requiring special tools or coding) can prove beneficial. Moreover, providing scripts in commonly-used programming languages such as Python or R—for reading, visualization, and analysis of the data—can greatly aid users, regardless of the original data format.
- *User-side*: User perspectives are essential in maximizing the value of NASA data.
- *Data producers-side*: Data producers are at the forefront of ensuring scientific quality of the data products. They can also take steps to support FAIR-compliance at the start of the data product lifecycle. Communication and training are important, specifically providing constructive feedback on issues encountered, understanding the intent, and contributing to standards/best practices development.
- *Data and information quality*: FAIR primarily focuses on data management and stewardship, while other quality aspects of data products such as data quality, providing information on uncertainty estimates and sources is critically important in reusing the data products (Ramapriyan et al. 2017) while the best practices for uncertainty reporting are still being improved.
- *Data preservation*: FAIR Principles do not explicitly address data preservation. To ensure long-term reusability, however, data, metadata and associated knowledge should be

gathered as completely as possible and curated and preserved. Which artifacts (documents, data, and software) should be preserved and when, are specified in the NASA Earth science data preservation content specification (Ramapriyan, Moses, & Smith 2022) and an ISO standard (ISO 19165-2 2020).

## 6.6 How to Address Dynamically Subsetted or Aggregated Time Series?

In this document, we have focused on defining Open and FAIR data practices to be employed during the creation of individual data products (collections) and associated data files (granules). However, most users work with a time series of data such as Synthetic Aperture Radar (SAR) images that cover a number of years of data, and at times even across collections. Furthermore, users might only be interested in a certain irregularly shaped area of interest or a subset rather than an entire collection or image. To make the aggregated data such as a time series findable and reusable, the user would benefit from a unique identifier for the entire time series. In addition, the user would need sufficient information to allow their study to be reproducible. What constitutes sufficient information for reproducibility for dynamically subsetted or aggregated data, and how to effectively cite all the associated artifacts, is beyond the scope of FAIR, and therefore this document. Examples of community efforts include RDA's recommendations on dynamic citations (Rauber et al. 2015), the RDA WG on complex citations (reliquary)<sup>86</sup>, and Buneman, Davison, & Frew (2016) on defining citable units.

Still, at a minimum, the time series should have a unique identifier to which the associated individual granules should be linked. The metadata should include enough information for all input granules and parameters used to generate the entire time series, which is especially critical to machine-learning training datasets.

## 6.7 How About Software?

This document primarily focuses on SMD-funded scientific data, excluding all research objects outside of the data and metadata categories such as algorithms, software, and other digital and physical research objects. Publishing the software as open source is key to meeting reproducibility requirements. Enabling FAIR software is a complex topic that requires domain knowledge in producing and managing research software. Discussion of these items are out of scope for this working group, so we will only touch on a few general recommendations based on leading community practices. Basically, software needs to be citable with a globally unique and resolvable persistent identifier, have a license appropriate to the software, be cited properly when used, be linked through metadata when cited, be permanently archived, and be searchable in the appropriate domain-specific archive (e.g., Barker et al. 2022; Stall et al. 2023). Good practices for software development and metadata curation, such as inclusion of contributors, can

---

<sup>86</sup> <https://www.rd-alliance.org/groups/complex-citations-working-group/outputs/> (accessed: 2024-05-20)

be found at <https://software-carpentry.org/>, <https://allcontributors.org>, and in Chue Hong et al. (2022).

## 6.8 How About Commercial Data Products?

As reiterated above, this document focuses on SMD-funded scientific data products. NASA also acquires data from commercial providers as a cost-effective way to complement Earth observations collected by NASA, other U.S. Government agencies, and international partners—for example, through the Commercial Satellite Data Acquisition (CSDA) Program.

Commercial data products are not included in this document, as they typically involve different data sharing requirements and workflows governed by end-user license agreements (EULAs) that include vendor-specific terms and conditions for scientific use. User licenses can vary, and such data products are not, by default, open and free.

## 7 CONCLUSION AND PATH FORWARD

Established in May 2022, the objectives of the O'FAIR WG are to explore the current landscape of FAIR practices and synthesize community FAIR practices to provide practical principle-by-principle guidance on enabling NASA Earth science data being FAIR, while adhering to being open and free.

The first deliverable of the WG, "An Overview of Community FAIR Practices – NASA O'FAIR WG Inception Report", was delivered to the ESDIS points of contact (POCs) in May 2023 and released to public in June 2023, along with a workbook that contains a collection of FAIR Principles interpretations and implementations as well as FAIRness assessment metrics and tools (Peng et al. 2023).

Building on that effort and continuous collection and synthesis of information, this Guide document is developed to provide practical guidance to stakeholders on being compliant with the FAIR Principles. The current NASA FAIR-enabling standards, practices, workflows, and tools/services for Earth science data products are highlighted. Their alignment with FAIR requirements is evaluated. Recommendations are provided when a gap is identified for data products, as well as for standards and practices, to be more FAIR. Recommendations are prioritized as *low*, *medium*, and *high*. Level of effort and/or resource requirement for a given recommendation to be implemented is denoted by *Easy*, *Moderate*, *Challenging*, or *Resources Intensive*. The responsible parties for implementing the recommendations are also identified.

A total of 16 recommendations resulted from this process, including 3 in the Findability dimension, 2 in Accessibility, 6 in Interoperability, and 5 in Reusability. It is worth noting that discussions and efforts are already underway to implement – or explore system and user

requirements for implementing – several of the O'FAIR WG recommendations. For example, recent updates to DOI and Associated DOI fields in CMR and the UMM metadata model reflect progress related to Recommendation A-1.<sup>87</sup> A CC0 license has been assigned to GCMD keywords, which is related to Recommendation I-3.<sup>88</sup>

Overall, NASA Earth science data managed by DAACs are findable and accessible by both human and machine end-users. Recommendations include assigning or registering globally unique persistent identifiers for missions/campaigns, CMR collection-level metadata records, and individual data producers and organizations. It is preferable to utilize established and sustained service providers for this purpose. Technologies for implementing those recommendations are mature and readily available with the required resources considered to be minimal.

In general, NASA Earth science data products support syntactic interoperability, with basic information on data provenance. Our recommendations aim to improve the semantic representation of data and metadata and linked data on the web with more robust data provenance by leveraging community standards. While the necessary technologies are becoming mature, it is necessary to develop strategies or templates at the ESDIS level. Therefore, moderate efforts and resources are required for the development and implementation.

The working group has observed that there are two competing forces at play when striving for FAIR compliance, particularly in enhancing data interoperability. Balancing domain-specific considerations with the need for convergence across diverse disciplines and domains presents a complex challenge.

We recognize that the existing controlled vocabularies such as GCMD keywords and CF standard names may not necessarily be able to meet every domain's needs, especially at the variable level. The variable-level vocabulary is recognized as a major challenge by many partner organizations, e.g., WMO, the European Space Agency (ESA), and the Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS). However, we encourage data producers to utilize them and other existing NASA and community standards, or follow a standardized way of capturing information as much as possible. It would be even better if they could submit a request to ESCO to accept that standardized way and publish it for others to use as well.

We also recognize that some needed domain-specific standards and best practices are still under improvement and development. We advocate for NASA's participation in developing them.

---

<sup>87</sup> <https://wiki.earthdata.nasa.gov/display/CMR/Associated+DOIs> (accessed: 2025-07-02)

<sup>88</sup> <https://www.earthdata.nasa.gov/data/tools/gcmd-keyword-viewer> (accessed: 2025-07-02)]

Aiming to harmonize FAIR implementations within individual domains by utilizing the FAIR Implementation Profiles (FIP) and improve interoperability via a cross-domain interoperability framework (CDIF)<sup>89</sup>, WorldFAIR project of CODATA is leading the effort in addressing this grand challenge (Hudson 2023). Our Guide document, while tailored to NASA Earth science data products, is helpful in harmonizing a FAIR implementation profile for Earth science data.

While the NASA DOI process is uniformly defined, it has come to light that significant variations exist in the execution workflows and the structure of landing pages across different DAACs. To ensure consistent user experiences across DAACs, adopting a standardized landing page layout could prove beneficial. We recommend DAACs to adopt community standards when formulating the landing page, such as Microdata and Schema Markup to provide structured data for product contents<sup>90</sup>, and the JSON-LD for Linked Data.<sup>91</sup> If deemed necessary, special attention should be given to enable automated FAIRness assessments and leverage insights gained from utilizing FAIRness evaluation tools such as F-UJI, to optimize data integration readiness by machines, while keeping in mind the limitations of such a tool for Earth science.

In summary, this Guide document describes enabling resources for stakeholders to ensure or improve compliance with the FAIR principles at the data product level. It offers actionable steps for stakeholders to readily share their data products in a computational environment.

Specifically, we encourage:

- NASA ESDS and ESDIS management to continue to advocate for quality-assured scientific data to be FAIR in support of open science;
- Data producers to read this Guide document to learn about the FAIR principles and NASA FAIR-enabling resources to ensure or improve their practices;
- DAACs to review this Guide document to bolster their procedures and capacity for computational systems to find, access, integrate, and reuse their data for both human and machine end users;
- Members of O'FAIR WG or its successor to periodically review and update the Guide document to keep up with the evolving NASA and community leading FAIR practices and incorporate feedback from data producers, DAACs, and end users.

Finally, we recommend that ESDS and/or ESDIS management initiate an effort to develop metrics to monitor the FAIR-compliance level for NASA Earth science data products. This should include

---

<sup>89</sup> <https://worldfair-project.eu/cross-domain-interoperability-framework> (accessed: 2023-09-07)

<sup>90</sup> <https://schema.org/docs/gs.html> (accessed: 2023-09-07)

<sup>91</sup> <https://json-ld.org/> (accessed: 2023-09-07)

gathering user feedback through ACSI surveys regarding how well users think ESDIS and DAACs are meeting the FAIR Principles.

## 8 ACKNOWLEDGEMENT

Support from NASA ESDSWG, OCSDO, ESDIS, DAACs, and affiliated NASA-funded projects has been essential for the O'FAIR WG to be established, for co-chairs to coordinate, and for members to participate in the working group activities.

Detailed information on CMR and Earthdata Search from the CMR team, especially Valerie Dixon, was very beneficial. We thank Lesley Wyborn for the information on the SPDX list. An earlier version of Figure 4.5 was reviewed by members of the NASA SMD Standards Working Group and the UAH/MSFC IMPACT project with valuable feedback.

We thank ESCO reviewers for their time, effort, and beneficial comments, which have improved the clarity of the document.

Finally, we express our gratitude to the members of the O'FAIR WG for their valuable contribution, active participation, and keen interest in this undertaking. Their names, affiliated organizations, roles, and/or subject areas are listed in the table below.

Name (First Last)	Affiliated Organization(s)	Role; Subject Area
Ge Peng	SSAI/NASA GSFC ESDIS, previously UA Huntsville/MSFC IMPACT	Chair; Scientific data stewardship
Robert R. Downs	SEDAC; Columbia University; ESCO	Co-Chair; Data archive
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	Geoinformatics, information quality
Mark A. Parsons	UA Huntsville/NASA Chief Science Data Office	Persistent Identifiers, SMD Standards and Guidelines, Data Stewardship
Manil Maskey	MSFC/ESDS HQ on AI/ML	POC
Keith Bryant	GES DISC	UI/UX/Usability
Zhong Liu	George Mason University/GES DISC	Data curator and user needs

Carl Mears	Remote Sensing Systems; NASA ACCESS	Data producer
Leigh Sinclair	GHRC.DAAC	
Stephanie Wingo	UA Huntsville/MSFC IMPACT; ADMG	Airborne
Rudiger Gens	ASF.DAAC	Metadata; Accessibility
Jeoffrey Stano	GHRC.DAAC	
Deborah Smith	UA Huntsville/IMPACT; ADMG	Airborne; Level of Services
Douglas Rao	NCICS/NCAI	Observer; AI/ML
Michele Thornton	ORNL.DAAC	Observer
Bhaskar Ramachandran	SSAI/LAADS.DAAC, GSFC	Data and information systems; Interoperability
Stephen (Steve) Olding	GSFC, ESDSWG	
Sara Lubkin	GSFC, ESDIS; ESCO	
Francis Lindsay	GSFC	ESDS/ESDIS POC
Siri Jodha S Khalsa	NSIDC.DAAC; ESCO	Observer; Standards
Shannon Leslie	NSIDC.DAAC	Observer
Karen Yuen	JPL/TOPS	Observer
Tammy Walker	ORNL DAAC	Observer; data management and user needs
Shawn R. Smith	COAPS/FSU; NASA ACCESS	In-situ marine data management; Interoperability; Vocabulary
Gao Chen	LaRC; ASDC	Atmospheric chemistry research, field campaign data management, and metadata standards
Ed Armstrong	PO.DAAC	Data management and stewardship
Katherine Saad	ESDS HQ	POC
Dave Meyer	GES DISC	Observer

Tyler Stevens	GCMD	CMR Metadata; POC
Rebecca Ringuette	ADNET Systems Inc.; Heliophysics Digital Resource Library, GSFC	Observer

## 9 REFERENCES

- ACSI, 2024: The American Customer Satisfaction Index (ACSI). Available online: <https://www.theacsi.org> (accessed: 2024-01-05)
- Albers, C., 2023: ESDS HQ Vision and Infusion. *2023 NASA ESDSWG Annual Meeting, Linthicum Heights, MD*. March 21-23, 2023. Slides are available at: <https://wiki.earthdata.nasa.gov/display/ESDSWG/2023+ESDSWG+Meeting+Agenda> (NASA ESDSWG wiki login credential required)
- Allen, C. and D.M.A. Mehler, 2019: Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*. <https://doi.org/10.1371/journal.pbio.3000246>
- Armstrong, E.M., H.K. Ramapriyan, P.J.T. Leonard, R.R. Downs, T.R. Walker, and G. Peng, 2023: Metadata Mapping to FAIR Principles. IN41B-0590. AGU 2023 Fall Meeting. <https://agu.confex.com/agu/fm23/meetingapp.cgi/Paper/1422385>
- Barker, M., Chue Hong, N.P., Katz, D.S., and others, 2022: Introducing the FAIR Principles for research software. *Sci Data* 9, <https://doi.org/10.1038/s41597-022-01710-x>
- Buneman, P., S. Davidson, and J. Frew, 2016: Why data citation is a computational problem. *Comm. ACM*, 59. <https://doi.org/10.1145/2893181>
- Chue Hong, N.P., D.S. Katz, M. Barker, A.-L. Lamprecht, C. Martinez, F.E. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P.A. Martinez, T. Honeyman, A. Struck, A. Lee, A. Loewe, B. van Werkhoven, C. Jones, D. Garijo, E. Plomp, F. Genova, and RDA FAIR4RS WG, 2022: FAIR Principles for Research Software (FAIR4RS Principles) (1.0). *Zenodo*. <https://doi.org/10.15497/RDA00068>
- Fenner, M., M. Crosas, G. Durand, S. Wimalaratne, and others, 2018: Listing of data repositories that embed schema.org metadata in dataset landing pages. *Zenodo*, <https://doi.org/10.5281/zenodo.1263942>
- Fenner, M., M. Crosas, J.S. Grethe, D. Kennedy, and others, 2019: A data citation roadmap for scholarly data repositories. *Sci Data* 6, <https://doi.org/10.1038/s41597-019-0031-8>
- Hudson, S., 2023: An Introduction to WorldFAIR. *Zenodo*. <https://doi.org/10.5281/zenodo.7753387>
- IOC (International Science Council), 2020: Making data work for cross-domain grand challenges. <https://codata.org/wp-content/uploads/2020/11/Decadal-Programme-Two-Sider-1.pdf>

ISO 19165-2, 2020: Geographic information - Preservation of digital data and metadata - Part 2: Content specifications for Earth observation data and derived digital products, <https://www.iso.org/standard/73810.html>

Jacobsen, A, R.M. de Azevedo, N. Juty, D. Batista, S. Coles, and others. (2020) 'FAIR principles: Interpretations and implementation considerations'. *Data Intelligence*, 2:10-29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024)

Liu, Z., C.-L. Shie, A.J. Rittrivi, G.-D. Lei, and others, 2022: Developing Metrics for NASA Earth Science Interdisciplinary Data Products and Services. *Data Science Journal*, 21(1), <https://doi.org/10.5334/dsj-2022-005>

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

NASA ESDIS, 2016: Global Change Master Directory (GCMD) Keyword Governance and Community Guide Document. Available at: [https://www.earthdata.nasa.gov/s3fs-public/imported/KeywordsCommunityGuide\\_Baseline\\_v1\\_SIGNED\\_FINAL.pdf](https://www.earthdata.nasa.gov/s3fs-public/imported/KeywordsCommunityGuide_Baseline_v1_SIGNED_FINAL.pdf)

NASA ESDIS, 2024: ESIDS Metrics System (EMS). Available online: <https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/ems> (accessed: 2024-01-05)

NASA ESDIS CMO (Configuration Management Office), 2023: Metadata requirements base reference. Appendix B. Metadata Requirements Base Reference for Unified Metadata Model - Collection (UMM-C). 423-RQMT-003-B, Rev C. Release date: February 16, 2023. Available at: <https://wiki.earthdata.nasa.gov/display/CMR/UMM+Documents>

NASA ESDS, 2020: Level of Service Model. Version: v1.0-20200220. Available at: [https://www.earthdata.nasa.gov/s3fs-public/imported/LoSModelDescription\\_V102202020.pdf](https://www.earthdata.nasa.gov/s3fs-public/imported/LoSModelDescription_V102202020.pdf)

NASA SMD, 2019: SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024. Available at: [https://science.nasa.gov/wp-content/uploads/2023/05/SDMWGStrategy\\_Final-1.pdf](https://science.nasa.gov/wp-content/uploads/2023/05/SDMWGStrategy_Final-1.pdf)

NASA SMD, 2022: Scientific Information Policy for the Science Mission Directorate. Document ID: 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, 2023: Open-Source Science Guidance. Updated 2025. Available at: <https://assets.science.nasa.gov/content/dam/science/cds/open-science/SMD%20Open-Source%20Science%20Guidance%20v2.2%2020250205.pdf>

Nosek, B., 2021: The Future of Open Science: The Need to Change Culture to Change Science. Version: April 26, 2021. Available at: <https://www.apa.org/members/content/culture-science-slides.pdf>

OSTP (Office of Science and Technology Policy), 2022: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. Version: 5 August 2022. Available at: <https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

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

Parsons, M. A., R. Duerr, and Ø. Godøy, 2022: The evolution of a geoscience standard: An instructive tale of science keyword development and adoption. *Geoscience Frontiers*, 101400. <https://doi.org/10.1016/j.gsf.2022.101400>

Peng, G., 2022: What can scientists do to support open science? Town Hall: TH43G – Open Science Realities: How Open is Open? 15 December 2022, AGU Fall Meeting, Chicago, IL.

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

Peng, G., R.R. Downs, H.K. Ramapriyan, M.A. Parsons, D.F. Moroni, Z. Liu, S.J.S. Khalsa, C. Mears, Y. Wei, B. Ramachandran, S. Smith, and NASA O'FAIR WG, 2023: An Overview of Community FAIR Practices – NASA O'FAIR WG Inception Report. Document ID: NASA-OFAIR-ESDSWG-DOC-0001. CC-BY 4.0 International. Version: v01r00-20230508, <https://doi.org/10.5067/DOC/ESCO/ESDSWG-0001V1>

Peng, G., G. Berg-Cross, M. Wu, R.R. Downs, S.R. Shrestha, L. Wyborn, N. Ritchey, H.K. Ramapriyan, S.J. Clark, J. Wood, Z. Liu, and A. Marouane, 2024a: Harmonizing Quality Measures of FAIRness Assessment Towards Machine-Actionable Quality Information. *Inter. J. Digital Earth*, 17(1). <https://doi.org/10.1080/17538947.2024.2390431>

Peng, G., R.R. Downs, H.K. Ramapriyan, and O'FAIR WG., 2024b: Open, Free & FAIR Data Guide Cheat Sheet for NASA-funded Earth Science Data Products (v01r00-20240926). *Zenodo*. <https://doi.org/10.5281/zenodo.11520035>

Peng, G., H.K. Ramapriyan, and E.M. Armstrong, 2025: DPDG Document Cheat Sheet: A Quick Guide for NASA-funded Earth Science Data Producers. *NASA Earthdata*. <https://doi.org/10.5067/DOC/ESCO/ESDSWG-0003V1>

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 version 1. NASA Earth Science Data and Information System Standards Office, Updated: 21 October 2021. <https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1>

Ramapriyan, H.K., P.J.T. Leonard, E.M. Armstrong, S.J.S. Khalsa, D.K. Smith, L.F. Iredell, D.M. Wright, G.J. Huffman, and T.R. Walker, 2024: Data Product Development Guide (DPDG) for Data Producers version 2.0. NASA Earth Science Data and Information System Standards Coordination Office. <https://doi.org/10.5067/DOC/ESCO/RFC-041VERSION2>

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. <https://doi.org/10.5067/DOC/ESO/423-SPEC-001>

Ramapriyan, H., G. Peng, D. Moroni, and C.-L. Shie, 2017: Ensuring and Improving Information Quality for Earth Science Data and Products. *D.-Lib Magazine*, **23**, <https://doi.org/10.1045/july2017-ramapriyan>

Ramapriyan, H., J. Goldstein, H. Hua, R. Wolfe, and the PROV-ES Working Group, 2016: Tracking and Establishing Provenance of Earth Science Datasets: A NASA-based example. Mattoso, M and Glavic, B eds. *Provenance and Annotation of Data and Processes* 6th International Provenance and Annotation Workshop, IPAW 2016 McLean, Proceedings. June 7–8, 2016, VA, USA. <https://doi.org/10.1007/978-3-319-40593-3>

Raskin, R.G., and M.J. Pan, 2005: Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Computers & Geosciences Application of XML in the Geosciences*, 31(9), 1119-1125. <https://doi.org/10.1016/j.cageo.2004.12.004>

Rauber, A., A. Asmi, D. van Uytvanck, and S. Proell, 2015: Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC). *Zenodo*. <https://doi.org/10.15497/RDA00016>

RDA-CODATA Legal Interoperability Interest Group, 2016: Legal Interoperability of Research Data: Principles and Implementation Guidelines. *Zenodo*. <https://doi.org/10.5281/zenodo.162241>

Schultes, E., B. Magagna, K.M. Hettne, R. Pergl, M. Suchánek, and T. Kuhn, 2020: Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In G. Grossmann & S. Ram (Eds.), *Advances in Conceptual Modeling* (Vol. 12584, pp. 138–147). *Springer International Publishing*. [https://doi.org/10.1007/978-3-030-65847-2\\_13](https://doi.org/10.1007/978-3-030-65847-2_13)

Silverman, M.L., M. Fiebig, M. Shook, B. Huffer, M. Buzanowicz, S. Leavor, J. Kusterer, and G. Chen, 2023: Applying community standards for domain-relevant metadata to enhance data

product FAIRness. Poster. *2023 AGU Fall Meeting*, 14 December 2023. IN41B-0578. Abstract available at: <https://agu.confex.com/agu/fm23/meetingapp.cgi/Paper/1395474>

Stall, S., G. Bilder, M. Cannon, *and others*, 2023: Journal Production Guidance for Software and Data Citations. *Sci Data* 10, <https://doi.org/10.1038/s41597-023-02491-7>

U.S. Law 115-254, 2018; Geospatial Data Act of 2018. Subtitle F, Sections 751 - 759. Available at: <https://www.fgdc.gov/gda/geospatial-data-act-of-2018.pdf>

U.S. Law 115-435, 2019: Title II: OPEN Government Data Act. Version: 14 January 2019. Available at: <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>

Wanchoo, L., D. Turcios, Y. Hao, and H. Weir, 2024: EOSDIS FY2024 Annual Metrics Report. Accessed: March 13, 2025. 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. *Sci. 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>

Wilkinson, M.D., S. Sansone, E. Méndez, and others, 2022: Community-driven governance of FAIRness assessment: an open issue, an open discussion. *Open Research Europe*. <https://doi.org/10.12688/openreseurope.15364.1>

## APPENDIX A. KEY TERMS AND DEFINITIONS

This appendix defines the key terms that are utilized in this Guide document with source policy documents. Notes provide additional information or nuances of individual terms.

*Table A1. Key terms and definitions*

Term	Definition	Source/Notes
Machine-readable format	<p>The term "machine-readable"* means data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.</p> <p>* Machine-actionable data formats go beyond machine-readable with embedded instructions or metadata that enable automated actions based on the data.</p>	<p>U.S. Law 115-435 (2019).</p> <p>Machine-readable formats are structured, including CSV, XML, JSON.</p> <p>Machine-actionable formats include HDF5, netCDF-4, JSON-LD, RDF.</p>
Metadata	<p>The term "metadata" means information about the data; it can refer to structural information on how the components of data are organized or descriptive information about the content and context of the data such as titles, abstracts, keywords, and other descriptors that help identify and understand the data.</p>	<p>U.S. Law 115-435 (2019).</p> <p>There are many different ways to categorize metadata based on different purposes. In addition to the structural and descriptive metadata, other types of metadata may include:</p> <ul style="list-style-type: none"> <li>• Administrative Metadata - providing information about the creation, management, and ownership of the data, such as the creator, creation date, access permissions, and version.</li> <li>• Technical Metadata - describing the technical characteristics of the data such as file formats, resolution, compression, and other technical specifications relevant to the data.</li> <li>• Rights Metadata - specifying about the intellectual property rights and usage restrictions associated with the data, helping users understand how they can use and share the data legally.</li> </ul>

Open license	<p>The term "open license" means a legal guarantee that a data asset is made available - (A) at no cost to the public; and (B) with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting such asset.</p>	<p>U.S. Law 115-435 (2019)</p> <p>An open license allows you to maintain your copyright while enabling the public to use and remix your creative work.</p> <p>Creative Commons (CC) is a global body that provides open-copyright licenses in a standardized way.</p> <p>CC0 explicitly waives copyright and related rights, allowing for unrestricted use, modification, and distribution of the work.</p>
Open science	<p>The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity.</p>	<p>OSTP and NSTC (2023-01-11)  <a href="https://open.science.gov/">https://open.science.gov/</a></p>

## APPENDIX B. CATEGORY-SPECIFIC FAIR REQUIREMENTS

**Table B1.** Category-specific requirements mapped from the FAIR Principles and associated FAIR category. Descriptions of individual FAIR principles are in blue. The naming convention for requirement identifiers (REQ IDs) is: {FAIR ID}-REQ-{Category ID}, where, {FAIR ID}=[F1, F2, F3, F4; A1, A1.2, A1.2, A2; I1, I2, I3; R1, R1.1, R1.2, R1.3], denoting individual FAIR principles, respectively; {Category ID}=[D; M; IS], denoting the Data, Metadata, and Infrastructure category, respectively. Source: Peng (2023).

FAIR ID	REQ ID	Description	D	M	IS
<b>F1</b>		<b>(meta)data are assigned a globally unique and eternally persistent identifier</b>	X	X	
	<b>F1-REQ-D</b>	data are assigned a globally unique and eternally persistent identifier (PID-D)	X		
	<b>F1-REQ-M</b>	metadata are assigned a globally unique and eternally persistent identifier		X	
<b>F2</b>		<b>data are described with rich metadata (defined by R1 below)</b>	X	X	
	<b>F2-REQ-D</b>	data are described with rich metadata	X		
	<b>F2-REQ-M</b>	rich metadata are defined by R1 below		X	
<b>F3</b>		<b>metadata clearly and explicitly include the identifier of the data it describes</b>	X	X	
	<b>F3-REQ-M</b>	metadata clearly and explicitly include the PID-D		X	
<b>F4</b>		<b>(meta)data are registered or indexed in a searchable resource</b>	X	X	X
	<b>F4-REQ-D</b>	data are registered or indexed	X		
	<b>F4-REQ-M</b>	metadata are registered or indexed		X	
	<b>F4-REQ-IS</b>	the registering or indexing resource is searchable			X
<b>A1</b>		<b>(meta)data are retrievable by their identifier using a standardised communications protocol</b>	X	X	X
	<b>A1-REQ-D</b>	data are retrievable by their identifier	X		
	<b>A1-REQ-M</b>	metadata are retrievable by their identifier		X	

	<b>A1-REQ-IS</b>	the retrieving communications protocol is standardized			X
<b>A1.1</b>		<b>the protocol is open, free, and universally implementable</b>			X
	<b>A1.1-REQ-IS</b>	the protocol is open, free, and universally implementable			X
<b>A1.2</b>		<b>the protocol allows for an authentication and authorization procedure, where necessary</b>			X
	<b>A1.2-REQ-IS</b>	the protocol allows for an authentication and authorization procedure, where necessary			X
<b>A2</b>		<b>metadata are accessible, even when the data are no longer available</b>		X	
	<b>A2-REQ-M</b>	metadata are accessible permanently		X	
<b>I1</b>		<b>(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation</b>	X	X	
	<b>I1-REQ-D</b>	data use a formal, accessible, shared, and broadly applicable language for knowledge representation	X		
	<b>I1-REQ-M</b>	metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation		X	
<b>I2</b>		<b>(meta)data use vocabularies that follow FAIR principles</b>	X	X	
	<b>I2-REQ-D</b>	data use vocabularies that follow FAIR principles	X		
	<b>I2-REQ-M</b>	metadata use vocabularies that follow FAIR principles		X	
<b>I3</b>		<b>(meta)data include qualified references to other (meta)data</b>	X	X	
	<b>I3-REQ-D</b>	data include qualified references to other (meta)data	X		
	<b>I3-REQ-M</b>	metadata include qualified references to other (meta)data		X	
<b>R1</b>		<b>(meta)data are richly described with a plurality of accurate and relevant attributes</b>	X	X	
	<b>R1-REQ-D</b>	data are richly described with a plurality of accurate and relevant attributes	X		
	<b>R1-REQ-M</b>	metadata are richly described with a plurality of accurate and relevant attributes		X	

<b>R1.1</b>		<b>(meta)data are released with a clear and accessible data usage license</b>	X	X	
	<b>R1.1-REQ-D</b>	data are released with a clear and accessible data usage license	X		
	<b>R1.1-REQ-M</b>	metadata are released with a clear and accessible data usage license		X	
<b>R1.2</b>		<b>(meta)data are associated with detailed provenance</b>	X	X	
	<b>R1.2-REQ-D</b>	data are associated with detailed provenance	X		
	<b>R1.2-REQ-M</b>	metadata are associated with detailed provenance		X	
<b>R1.3</b>		<b>(meta)data meet domain-relevant community standards</b>	X	X	
	<b>R1.3-REQ-D</b>	data meet domain-relevant community standards	X		
	<b>R1.3-REQ-M</b>	metadata meet domain-relevant community standards		X	

## APPENDIX C. CROSSWALKS BETWEEN FAIR AND CMR UMM-C FIELDS

*Table C1. Crosswalks between the FAIR principles and CMR UMM-C fields*

FAIR ID	Description	CMR UMM-C Field(s)
F1	(meta)data are assigned a globally unique and eternally persistent identifier	DOI/[DOI: "DOI Value"; Authority: " <a href="https://doi.org/">https://doi.org/</a> "]* * Only with a valid DOI authority.
F2	data are described with rich metadata (defined by R1 below)	EntryTitle, CollectionDataType, CollectionCitations/Creator, Abstract, DOI/DOI, Version, ContactPerson, AncillaryKeywords, LocationKeywords, ScienceKeywords, TemporalKeywords, CollectionCitations/Publisher, DataDates, MetadataDates, UseConstraints, AssociatedDOIs/DOI, PublicationReferences, DataCenters, SpatialExtent, TemporalExtents, DataFormat, ISOTopicCategory, AccessConstraints, Quality, Purpose, ArchiveAndDistributionInformation, Characteristics, ProcessingLevel, RelatedURLs, Project, Platform, Instrument
F3	metadata clearly and explicitly include the identifier of the data it describes	DOI/DOI (required field)
F4	(meta)data are registered or indexed in a searchable resource	RelatedUrls/URLSubtype=Earthdata Search
A1	(meta)data are retrievable by their identifier using a standardised communications protocol	DOI; RelatedUrls/URLType=GET DATA; RelatedUrls/URLType=DATA SET LANDING PAGE; RelatedUrls/URLType=USE SERVICE API; RelatedUrls/URLSubtype=Earthdata Search, ShortName + Version, EntryTitle
A1.1	the protocol is open, free, and universally implementable	Data: UseConstraints Metadata: CMR API or Earthdata Search (Registered with CMR and searchable via Earthdata Search Portal) (RelatedUrls/URLSubtype=Earthdata Search)
A1.2	the protocol allows for an authentication and authorization procedure, where necessary	Earthdata Login* DirectDistribution (AWS)  * This item pertains to infrastructure, which is why it may not have been included in the metadata.
A2	metadata are accessible, even when the data are no longer available	MetadataDates MetadataDates/DateType=DELETE (if this is not present, assume that metadata is still available)

I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	Data Format; MetadataSpecification
I2	(meta)data use vocabularies that follow FAIR principles	GCMD keywords vocabulary and controlled vocabulary lists documented in the Earthdata Wiki
I3	(meta)data include qualified references to other (meta)data	MetadataAssociation (Type = INPUT), AssociatedDOIs; RelatedUrls (Type = ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD) or PUBLICATION); PublicationReferences
R1	(meta)data are richly described with a plurality of accurate and relevant attributes	AdditionalAttributes; EntryTitle, CollectionDataType, CollectionCitations/Creator, Abstract, DOI/DOI, Version, ContactPerson, AncillaryKeywords, LocationKeywords, ScienceKeywords, TemporalKeywords, CollectionCitations/Publisher, DataDates, MetadataDates, UseConstraints, AssociatedDOIs/DOI, PublicationReferences, DataCenters, SpatialExtent, TemporalExtents, DataFormat, ISOTopicCategory, AccessConstraints, Quality, Purpose, ArchiveAndDistributionInformation, Characteristics, ProcessingLevel, RelatedURLs, Project, Platform, Instrument
R1.1	(meta)data are released with a clear and accessible data usage licence	UseConstraints/LicenseURL example: "Linkage: <a href="https://earthdata.nasa.gov/earth-observation-data/data-use-policy">https://earthdata.nasa.gov/earth-observation-data/data-use-policy</a> Protocol: https Name: License URL Description: The Earth Science Data and Information System (ESDIS) data use policy for NASA data. MimeType: text/html"
R1.2	(meta)data are associated with detailed provenance	MetadataDates, DataDates; CollectionCitations/Creator; ContactPersons/Role=Metadata Author; Version; CollectionCitations/Publisher; Data Centers
R1.3	(meta)data meet domain-relevant community standards	MetadataSpecification; Data Format

## APPENDIX D. NASA EARTH SCIENCE GLOSSARIES AND VOCABULARIES

*Table D1. NASA Earth science glossaries and vocabularies*

Vocabulary Title	Related URL(s)
ADMG Airborne and Field Data Inventory Definitions	<a href="https://www.earthdata.nasa.gov/esds/impact/admg/admg-definitions">https://www.earthdata.nasa.gov/esds/impact/admg/admg-definitions</a>
Atmospheric Composition Variable Standard Name Convention (ACVSNC)	<a href="https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/acvsnc">https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/acvsnc</a> <a href="https://asdc.larc.nasa.gov/glossary">https://asdc.larc.nasa.gov/glossary</a>
CASEI long and short names of Campaigns, Platforms, and Instruments	<a href="https://impact.earthdata.nasa.gov/casei/">https://impact.earthdata.nasa.gov/casei/</a> * * Under "Explore"
Earth Observatory Glossary	<a href="https://earthobservatory.nasa.gov/glossary/all">https://earthobservatory.nasa.gov/glossary/all</a>
Earth Satellite Missions	<a href="https://eosps.nasa.gov/content/all-missions">https://eosps.nasa.gov/content/all-missions</a>
Earth Satellite Remote Sensors	<a href="https://www.earthdata.nasa.gov/sensors">https://www.earthdata.nasa.gov/sensors</a>
ESDIS Acronym and Symbol List	<a href="https://www.earthdata.nasa.gov/learn/acronym-list">https://www.earthdata.nasa.gov/learn/acronym-list</a>
ESDIS Data Product processing Levels	<a href="https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels">https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels</a>
ESDIS Glossary	<a href="https://www.earthdata.nasa.gov/learn/glossary">https://www.earthdata.nasa.gov/learn/glossary</a>
GCMD Keywords	<a href="https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords">https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords</a>
	<a href="https://gcmd.earthdata.nasa.gov/KeywordViewer">https://gcmd.earthdata.nasa.gov/KeywordViewer</a>
Native Projection Names in CMR	<a href="https://wiki.earthdata.nasa.gov/display/CMR/Native+Projection+Names">https://wiki.earthdata.nasa.gov/display/CMR/Native+Projection+Names</a>
SMD Glossary of Science Terms	<a href="https://science.nasa.gov/glossary">https://science.nasa.gov/glossary</a>
Short and Long Names of Data Centers Along with Their UUIDs in CMR	<a href="https://wiki.earthdata.nasa.gov/display/CMR/Data+Center">https://wiki.earthdata.nasa.gov/display/CMR/Data+Center</a>

## LIST OF FIGURES

**Figure 4.1.** A schematic diagram depicts the complexity of the FAIR principles in terms of multi-dimensions, multi-layers, and multi-facets.

**Figure 4.2.** A schematic diagram shows varying degrees of FAIRness.

**Figure 4.3.** Continuum of Open-Source Science.

**Figure 4.4.** An illustrative depiction of the actions that scientists and data producers can take to support open-source science, including pertinent information to be documented and disseminated for data users across the project lifecycle. A method paper may be published during the planning phase.

**Figure 4.5.** Crosswalk from the data requirements that are defined in the SPD-41a to the FAIR principles.

**Figure 5.1.** Unified Metadata Model (UMM) for NASA's Earth Science Data and Information System (ESDIS) metadata.

## LIST OF TABLES

**Table 5.1.** Characteristics of a Digital Object Identifier (DOI), Universally Unique Identifier (UUID), and local identifier (ID), as well as their compliance level to the F1 principle (i.e., (meta)data are assigned a globally unique and eternally persistent identifier)

**Table 5.2.** A list of category-specific requirement IDs associated with the Findability principles, NASA enabling resources and their compliance levels, and O'FAIR working group (WG) recommendation identifiers (IDs) along with their level of priority and effort required to implement.

**Table 5.3.** A list of category-specific requirement IDs associated with the Accessibility principles, NASA enabling resources and their compliance levels, and O'FAIR WG recommendations along with their level of priority and effort required to implement.

**Table 5.4.** A list of category-specific requirement IDs associated with the Interoperability principles, NASA enabling resources and their compliance levels, and O'FAIR WG recommendations along with their level of priority and effort required to implement.

**Table 5.5.** A list of category-specific requirement IDs associated with the Reusability principles, NASA enabling resources and their compliance levels, and O'FAIR WG recommendations along with their level of priority and effort required to implement.

**Table A1.** Key terms and definitions

**Table B1.** Category-specific requirements mapped from the FAIR Principles

**Table C1.** Crosswalks between the FAIR principles and CMR UMM-C fields

**Table D1.** NASA Earth science glossaries and vocabularies

## AUTHOR LIST

### Ge Peng

Science Systems and Applications, Inc. (SSAI)/Earth Science Data and Information System (ESDIS) Project of NASA's Goddard Space Flight Center (GSFC), Lanham, MD, USA.

Previously, 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

SSAI/NASA FSGC ESDIS, Lanham, MD, USA

ORCID: 0000-0002-8425-8943

### Yaxing Wei

ORNL.DAAC, Oak Ridge National Laboratory, Oak Ridge, TN, USA

ORCID: 0000-0001-6924-0078

### Bhaskar Ramachandran

LAADS DAAC, GSFC/SSAI, Greenbelt, MD, USA

ORCID: 0000-0002-1684-8888

### Mark A. Parsons

UAH/NASA Office of Chief Science Data Officer (OCSDC), Huntsville, AL, USA

ORCID: 0000-0002-7723-0950

### Zhong Liu

George Mason University and NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA

ORCID: 0000-0001-8150-7556

## CONTRIBUTING MEMBERS OF THE O'FAIR WORKING GROUP

### Gao Chen

NASA Langley Research Center (LaRC), Hampton, VA, USA  
ORCID: 0009-0001-1699-7824

### Edward M. Armstrong

NASA Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, USA  
ORCID: 0000-0002-5595-9353

### Rudiger Gens

Alaska Satellite Facility (ASF), University of Alaska, Fairbanks, AK, USA  
ORCID: 0000-0002-1962-9787

### David F. Moroni

NASA JPL, California Institute of Technology, Pasadena, CA, USA  
ORCID: 0000-0003-2994-557X

### Carl Mears

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

### Jenny Wood

ESSC/MSFC IMPACT, UAH, Huntsville, AL, USA  
ORCID: 0009-0007-9932-4716

### Tyler B. Stevens

KBR/NASA GSFC, Greenbelt, MD, USA  
ORCID: 0009-0007-5564-5403

### Rebecca Ringuette

ADNET Systems, Inc./NASA GSFC, Greenbelt, MD, USA  
ORCID: 0000-0003-0875-2023

### Stephen Olding

SSAI/NASA GSFC, Greenbelt, MD, USA  
ORCID: 0000-0002-5276-5545

### Shawn R. Smith

Center for Ocean-Atmospheric Prediction Studies (COAPS), Florida State University, Tallahassee, FL, USA  
ORCID: 0000-0003-1392-3077