OPEN, FREE & FAIR DATA GUIDE CHEAT SHEET

FOR NASA-FUNDED EARTH SCIENCE DATA PRODUCTS

- Keep a holistic view with FAIR core concepts.
- Follow NASA established processes and practices.
- Utilize NASA enterprise systems and tools.
- Leverage community standards and practices.

WHAT YOU SHOULD KNOW

Making data FAIR is a multi-dimensional, multi-layered, and multi-faceted problem (Figure 1) that requires multi-domain information and knowledge.

- **FAIR Principles consist of 15 FAIR sub-principles in 4 quality dimensions (Findability, Accessibility, Interoperability, and Reusability), covering three categories (Data, Metadata, and Infrastructure) (Figure 2).**
- **The FAIR principles can be mapped to 28 category-specific requirements (Figure 3) and 21 unique category-specific unique core concepts (Figure 2).**

FAIR ID / CORE CONCEPT

F1 IDENTIFIER	F2 RICH METADATA	F3 LINKED PID-D	F4 RESOURCE	A1, A1.1 PROTOCOL	A1.2 PROCEDURE	A2 PERMANENT METADATA
Obtain and use unique, persistent, and resolvable identifiers: ▪ Data product: DOI from DataCite (via a DAAC) or Zenodo. ▪ Metadata records: UUID or ARKs. Additionally, use these IDs if appropriate: ▪ Person (Data producer and (meta)data curator): ORCID. ▪ Organization: ROR ID. ▪ Data center or Repository: re3data ID. ▪ Instrument/Platform: PIDInst ▪ Other artifacts: RRID	▪ Use CMR UMM-C schema. ▪ Ensure inclusion of required UMM-C metadata elements such as: DOI, Title, Abstract, Creator and/or Publisher [Data Centers], Version, Platform, Spatial Extent, Temporal Extents, Processing Level, Science Keywords, and Metadata Specification.	In both human- and machine-readable fashion: ▪ Include the data product DOI in collection-level metadata record, using the CMR UMM-C field: DOI. ▪ Include the DOI in file-level metadata.* ▪ Link the DOI to the landing page.* ▪ Include citation with DOI on the landing page. * See the side panel on the right for examples.	Integrate with NASA's Earth data publication process: ▪ Register data product DOIs and UMM-C/G metadata records in CMR ▪ Integrate to Earthdata Search Portal Ensure integration to additional searchable resources, including: ▪ DataCite, if it minted DOI ▪ Google Search Engine	▪ Register for an Earthdata account. ▪ Use Earthdata Search Portal for retrieving metadata records. ▪ Download data with Earthdata Login.	▪ Use Earthdata (Cloud/GIS) login for data access. ▪ Use Earthdata Search Portal or CMR API for metadata access.	▪ Maintain DOI landing pages for all publicly-released versions of data products. ▪ Capture version history in the latest data product version's landing page. ▪ Cross-reference versions from respective landing pages.

FAIR ID / CORE CONCEPT – CONT.

I1 LANGUAGE	I2 VOCABULARY	I3 REFERENCE	R1 ATTRIBUTE	R1.1 LICENSE	R1.2 PROVENANCE	R1.3 STANDARD
▪ Utilize ESCO approved human- and machine-readable file formats like: ○ netCDF file format with CF and ACCD, described by CDL or NcML ○ XML ▪ Utilize CMR UMM-C/G metadata profiles or supported profiles (ECHO-10; DIF-10; ISO 19915). ▪ Use common vocabularies like those in I2. ▪ Use RDF, JSON-LD and OWL. ▪ Follow RDF Schema and SKOS standards.	Use community or NASA ESCO approved vocabularies*, preferable with both human-readable terms and machine-readable URIs: ▪ Keywords or terms (Sciences, Platforms, Instruments, Projects, etc.): GCMD; CASEI ▪ Variable names: CF; ACVSNC Use project-specific common glossaries* if appropriate. * Work with ESCO for approval, updates, and additions.	Utilize JSON-LD, html meta tags, and UMM-C Associated DOIs and AssociatedURLs elements to include/cross-reference documents like: ▪ User guide ▪ ATBD ▪ Science paper ▪ Data paper ▪ Service endpoints for accessing data files ▪ DOIs of source data and previous versions	When publicly releasing data products and associated metadata records: ▪ Include an open license, following practices in R1.1. ▪ Include data provenance, following practices in R1.2. Additionally, ▪ include those in I3 if appropriate. ▪ Review and include relevant attributes from Appendices D and E in DPDG (Ramapriyan et al. 2024). * See the side panel on the right for examples.	▪ Use CC0 1.0 + Attribution for mission data products. ▪ Use CC0 1.0 + Attribution or CC-BY 4.0 for NASA-funded scientific data products. ▪ Make it both human and machine-readable* and include recommended citation in metadata and on DOI/data product landing pages.	▪ Use UMM-C metadata fields (Version, Data Dates, Metadata Dates, DOI/PreviousVersion, RelatedURL, Data Center, Collection Citation/Creator). ▪ Use global attributes in netCDF data files for source data, algorithm used, and data producer and/or publisher.	▪ Follow a standard for enabling data identifiers globally unique, persistent, and resolvable (e.g., DOI, Handler, URI). ▪ Use a standard-based communication protocol for (meta)data retrieval from their identifiers that allow for security procedures (e.g., HTTPS, SFTP, REST). ▪ Utilize standards for knowledge representation language for (meta)data (e.g., RDF, OWL, XML, JSON-LD; DCAT; PROV-O; ISO 19115). ▪ Use Linked Data standards for linking data product DOI, license, provenance, and other information and resources in a machine-actionable format.

RECOMMENDED PRACTICES FOR LINKED DATA

Linked Data standards should be utilized for machine-actionability to ensure enhanced search engine visibility and semantic understanding. Examples of linked data product DOIs and usage license in collection- and file-level metadata, and in landing page markups are provided below.

Examples of Including Data Product DOI

❖ **In CMR UMM-C:**
DOI/DOI:" 10.5067/GHGMR-4FJ04"
DOI/Authority: "https://doi.org/"

❖ **In NetCDF data files:**
```
//global attributes:
:id = "10.5067/GHGMR-4FJ04";
:naming_authority = "org.doi";
:metadata_link = "https://doi.org/10.5067/GHGMR-4FJ04"
```

❖ **In Landing page HTML:**
- schema.org/JSON-LD:
  <application type="application/ld+json"> {"@id": "https://doi.org/10.5067/GHGMR-4FJ04"} </application>
- **HTML meta tags:**
  <meta name="DC.identifier" content="https://doi.org/10.5067/GHGMR-4FJ04"

Examples of Including Data Usage License in HTML

❖ **CC0 1.0 + Attribution License**

```
<!-- Machine-readable link -->
<link rel="license" href="https://creativecommons.org/publicdomain/zero/1.0/">
<!-- Human-readable content -->
<p>This content is dedicated to the public domain under the <a rel="license"
    href="https://creativecommons.org/publicdomain/zero/1.0/">Creative
    Commons CC0 1.0 Universal Public Domain Dedication</a>.</p>
<p>You can copy, modify, distribute, and perform the work, even for commercial
    purposes, all without asking permission.</p>
<p>For more information about the CC0 1.0 license, visit <a
    href="https://creativecommons.org/publicdomain/zero/1.0/">here</a>.</p>
<p>If you use this work, please consider citing it as follows:</p>
<p>Author(s). (Year). <em>Title of the Work</em>. Retrieved from <a
    href="URL">URL</a>.</p>
```

❖ **CC-BY 4.0 License**

```
<link rel="license" href="https://creativecommons.org/licenses/by/4.0/">
<p>This content is licensed under a <a rel="license"
    href="https://creativecommons.org/licenses/by/4.0/">Creative Commons
    Attribution 4.0 International License</a>.</p>
<p>You are free to:</p> <ul>
    <li>Share — copy and redistribute the material in any medium or format</li>
    <li>Adapt — remix, transform, and build upon the material</li> </ul>
<p>Under the following terms:</p> <ul> <li><strong>Attribution</strong> — You
    must give appropriate credit, provide a link to the license, and indicate if
    changes were made. You may do so in any reasonable manner, but not in any
    way that suggests the licensor endorses you or your use.</li> </ul>
```

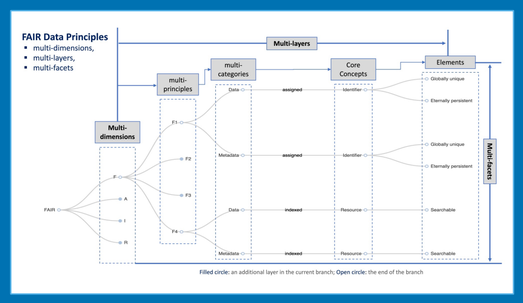

Figure 1. A schematic diagram depicts the complexity of the FAIR principles in terms of multi-dimensions, multi-layers, and multi-facets. Source: Peng et al. (2024a). License: CC-BY 4.0

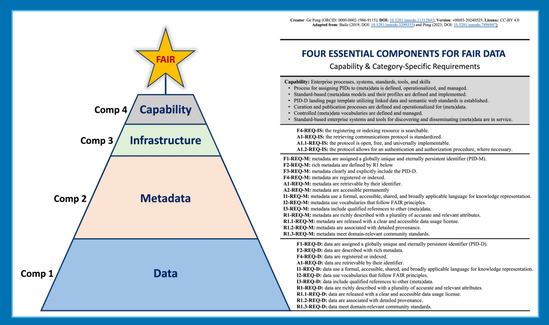

Figure 3. Capability & category-specific FAIR requirements. Source: Peng (2024). License: CC-BY 4.0

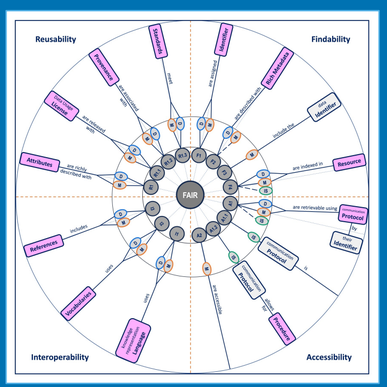

Figure 2. FAIR dimensions and sub-principles IDs (filled circles); categories (filled ovals) for data (D), metadata (M), and infrastructure (IS); and core concepts (round triangles). The pink-filled core concepts are FAIRness quality measures. Source: Peng et al. (2024a). License: CC-BY 4.0

DEVELOPED BY

NASA ESDS O'FAIR Working Group

Recommended Citation:

Peng. G., R.R. Downs, H. K. Ramapriyan, and NASA ESDS O'FAIR WG, 2024: Open, Free & FAIR Data Guide Cheat Sheet for NASA-funded Earth Science Data Products. Version: v01r00-20240926. DOI: https://doi.org/10.5281/zenodo.11520035. License: CC0 1.0 + Attribution.

Acknowledgement:

The work for this cheat sheet and the O'FAIR data guide document (i.e., Peng et al. 2024b) was carried out across multiple United States government-funded institutions under contracts with NASA (see details in Peng et al. 2024b). Government sponsorship is acknowledged.

ACRONYMS

ACDD	Attribute Convention for Data Discovery
ACVSNC	Atmospheric Composition Variable Standard Name Convention
API	Application Programming Interface
ARK	Archival Resource Key
ATBD	Algorithm Theoretical Basis Document
CASEI	Catalog of Archived Suborbital Earth Science Investigations
CC	Creative Commons
CDL	Common Data Language
CF	Climate and Forecast
CMR	Common Metadata Repository
DAAC	Distributed Active Archive Center
DCAT	Data Catalog Vocabulary
DIF	**Directory Interchange Format**
DOI	Digital Object Identifier
DPDG	Data Product Development Guide
ECHO	Earth Observing System Clearinghouse
ESCO	ESDIS Standards Coordination Office
ESDIS	Earth Science Data and Information System
ESDS	Earth Science Data Systems
GIS	Geographic Information System
JSON-LD	JavaScript Object Notation for Linked Data
HTTPS	Hypertext Transfer Protocol Secure
ISO	International Organization for Standardization
NASA	National Aeronautics and Space Administration
NcML	NetCDF Markup Language
netCDF	network Common Data Form
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PIDInst	Persistent Identification of Instruments
PROV-O	PROV Ontology
RDF	Resource Description Framework
re3data	Registry of Research Data Repositories
REST	REpresentational State Transfer
ROR	Research Organizations Registry
RRID	Research Resource Identifiers
SFTP	Secure File Transfer Protocol
SKOS	**Simple Knowledge Organization System**
UMM-C	Unified Metadata Model–Collection-level
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally Unique Identifiers
XML	Extensible Markup Language

REFERENCES

Peng. G., 2024: Four essential components for FAIR data. *Zenodo*. https://doi.org/10.5281/zenodo.11315841

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*. https://doi.org/10.1080/17538947.2024.2390431

Peng, G., R.R. Downs, H. K. Ramapriyan, Y. Wei, B. Ramachandran, M. Parsons, Z. Liu, and NASA O'FAIR WG, 2024b: A Practical Guide for Open, Free & FAIR NASA Earth Science Data Products. https://doi.org/10.5067/DOC/ESCO/ESDSWG-0002V1

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