INSIGHT CHIASMA >PINK

ChIPs Workshop on FAIRification

14th January 2025 | 14:00 – 18:00 (CET)

Welcome & Introduction to the FAIR data principles

Iseult Lynch, University of Birmingham

UNIVERSITY OF BIRMINGHAM

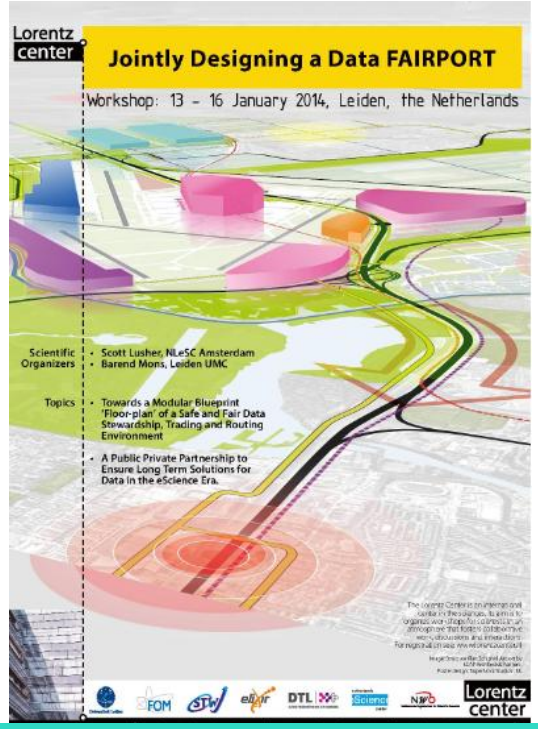
Introduction to the FAIR data principles

Interpretation and implementation considerations

Acknowledgements:
Slides adapted from materials developed by Dr. Panče Panov, Jožef Stefan Institute, and Dr. Tassos Papadiamantis, UoB

What is FAIR?

- **F**indable, **A**ccessible, **I**nteroperable, **R**eusable data principles
- 4 foundational principles explicitly described by 15 FAIR guiding principles
- Intended as a guide to enable digital resources to become more FAIR for machines and also for humans.
- Emerged from a multi-stakeholder vision of an infrastructure supporting machine-actionable data reuse, i.e., reuse of data that can be processed by computers



https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html

www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Overview of the FAIR data principles

Box 2 | The FAIR Guiding Principles

https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Machine Actionability

- The four foundational principles describe the core objectives of the principles that, if achieved, should enable machines to make optimal use of data resources.
- This is achieved, technically, by making every digital resource FAIR via some technical implementation choice

Machine Actionability

- **Findability:** Digital resources should be easy to find for both humans and computers
- **Accessibility**: Protocols for retrieving digital resources should be made explicit, for both humans and machines
- **Interoperability:** When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity.
- **Reusability:** Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource should be reused; if a digital resource can be reused, and under what conditions; and who to credit if it is reused.

(Meta)data

- The concepts of “data” and “metadata” occur throughout the 15 FAIR guiding principles.
  - Data is used to refer to all digital resources (not just data in the restricted sense, but also, for example, software tools).
- Metadata is any description of a resource that can serve the purpose of enabling findability and/or reusability and/or interpretation and/or assessment of that resource. (e.g., **in PINK = data documentation**)
- FAIR treats every data/metadata pair in-isolation;
  - metadata is the descriptor, and data is the thing being described, unambiguously, within the context of that pair
- Metadata must also be a FAIR digital resource in its own right

Machine actionable metadata

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Machine-actionable metadata

Technical infrastructure

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Machine-actionable metadata

Technical infrastructure (accepted generic services)

Social decisions

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Machine-actionable metadata

Technical infrastructure (accepted generic services)

Social decisions (domain specific agreements)

Implementation of the FAIR principles

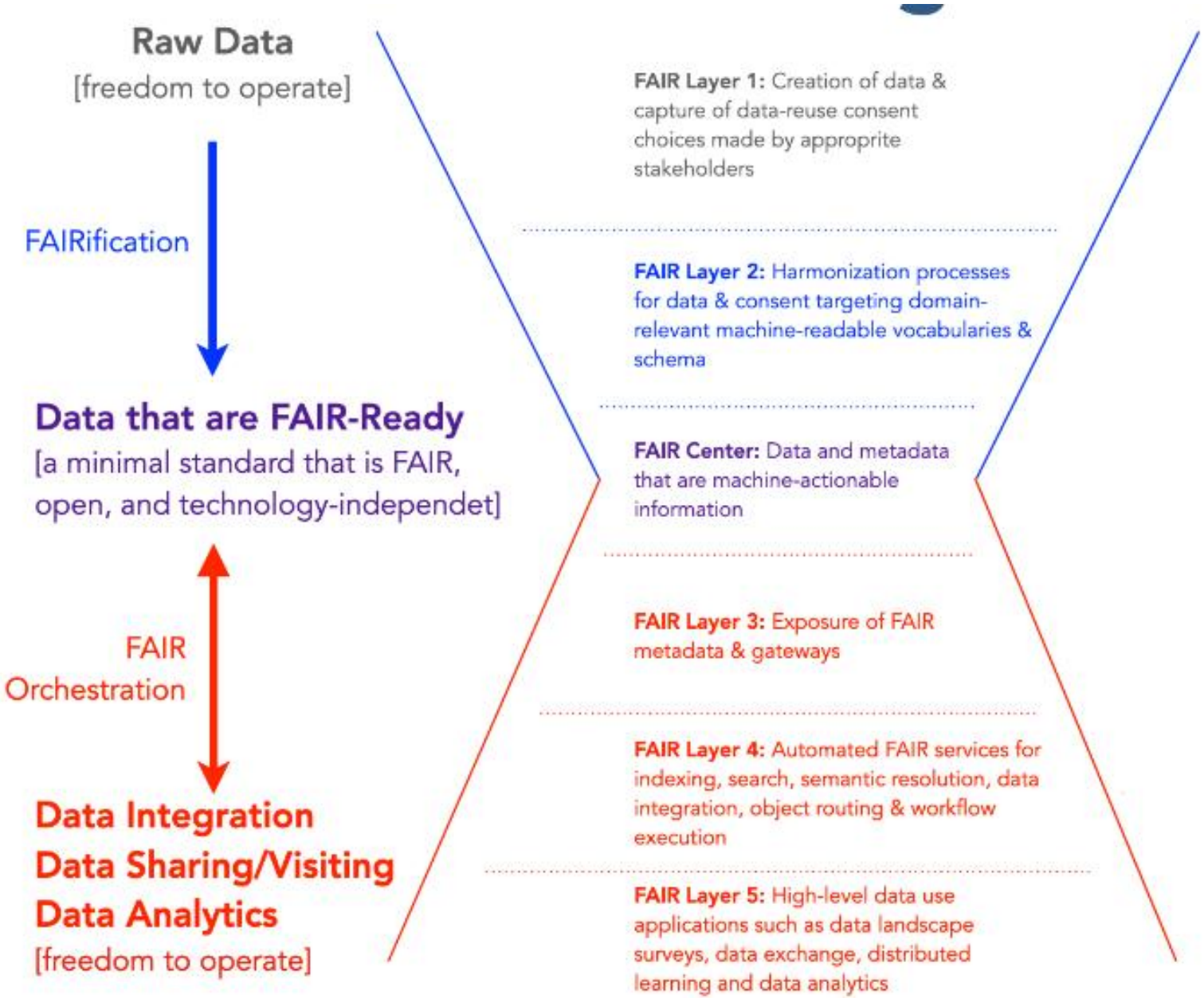
- Several alternative routes towards the implementation of the FAIR principles, some specialized for different types of digital resources
- Communities have already published documents that can guide implementation choices
  - Existing technologies should be used where possible
  - The FAIRfication process can typically be broken down into steps, allowing the different facets of FAIRness to be prioritized depending on the resource under consideration and the cost-benefit
  - Different types of stakeholders adopt complementary roles with respect to implementing FAIR principles
    - a domain expert, an information scientist, a system engineer, a data archivist, a data analysis expert)
  - The implementation decisions for certain kinds of stakeholders can be shared and reused across domains or communities.

FAIR Hourglass

FAIRification: the process by which data captured using localized or domain-specific practices are transformed into formats that follow open standards for interoperability.

FAIR Orchestration: putting FAIR-ready data into action by exposing them to software applications and services that can perform operations on them:

- Indexing;
- ontology-based access control of restricted data;
- disambiguation of semantic content;
- merging appropriate datasets and eventuality running analyses leading to new insights



Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

FAIR enabling resources (FERs)

FAIR principle	FER Type	Definition
F1	Identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	Metadata schema	A **specification** that specifies the structured representation of metadata describing attributes of data or other digital objects in terms of semantics, syntax and optionality.
F3	Metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	Registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	Communication protocol	A **specification** of how messages are structured and exchanged.
A1.2	Authentication & authorisation service	A **service** that mediates access to digital objects according to specifed conditions.
A2	Metadata preservation policy	A data **policy** that describes the conditions under which metadata should be provided in the future.
I1	Knowledge representation language	A language **specification** that enables knowledge to be processed by machines.
I2	Structured vocabularies	A **specification** for a controlled list of uniquely identified and unambigous concepts with their definitions represented using web standards.
I3	Semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects according to the Linked Data principles. This can include semantic data models and ontologies.
R1.1	Data usage license	A data **policy** that specifies legal restrictions on the reuse of the data.
R1.2	Provenance model	A **specification** that specifies metadata describing the origin and lineage of data or other digital objects.
R1.3	The FAIR implementation Profile	A FAIR Implementation Profile (FIP) is a list of declared technology choices intended to implement each of the FAIR Guiding Principles, made as a collective decision by the members of a particular community of practice.

The FER type version 2.0.0 was created in December 2023 by Barbara Magagna and Erik Schultes; GO FAIR Foundation

What FAIR is not

- FAIR is not a standard
- FAIR is not only semantic web or linked open data (LOD)
- FAIR is not equal to "Open" or "Free"
  - Data are often Open but not FAIR
  - Some data can never be Open, yet they can be FAIR
- FAIR is not explicit about data quality, trustworthiness, responsibility, ethics, etc.

Interpretation of the FAIR principles

In the interests of time, for this session I will focus on the domain-specific principles, but the slide deck contains all sub-principles, and will be shared.

Findable – The "F" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

F1: (meta)data are assigned a globally unique and persistent identifier

Interpretation

- **Essential element:** Fundamental to all FAIR principles.
- **Globally unique:** Identifier must unambiguously refer to one resource universally, not just locally.
- **Persistence:** Identifier is never reused and continues to identify the same resource over time, even if the resource changes or ceases to exist.
- **Practical implementation:** Typically involves third-party generation of identifiers with guaranteed longevity and independence from the originating project or organization.

F1: (meta)data are assigned a globally unique and persistent identifier

Implementation consideration

- **Challenges:** Ensuring identifier longevity beyond the lifespan of the originating project or community.
- **Reliance on third parties:** Requires third-party organizations to maintain and guarantee the longevity of identifiers.
- **Community choices:** Each community should select appropriate identifier registration services ensuring global uniqueness, persistence, and resolvability.
- **Example:** Digital Object Identifier (DOI) (https://www.doi.org/) is a widely used example that guarantees global uniqueness and persistence, and provides additional services like directing calls to the source data's current location.
- **Additional support:** Persistent identifiers like DOIs can also support metadata availability beyond the data’s lifespan, ensuring continued metadata accessibility.

F1. (Meta)data are assigned a globally unique and persistent identifier

FER: Identifier service

- **Definition:** Provides unique, persistent identifiers for digital objects, ensuring they can always be found.
- **Purpose:** Ensures that data can be reliably located and accessed over time, regardless of changes in its location or ownership.
- **Examples**: Digital Object Identifier (DOI) for publications, ORCID for researchers, Handle System for various digital objects.

Benefits

- **Long-term Access:** Ensures data remains accessible even if the location changes.
- **Citation and Tracking:** Supports proper citation of data in academic work and tracking of its use.
- **Data Integrity:** Helps in maintaining the integrity and authenticity of data.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

F2: Data are described with rich metadata

Interpretation

- **Discovery of resources:** Facilitates the discovery of digital resources through search or filtering.
- **Rich metadata:** Digital resources must be described with detailed descriptors of their content.
- **Metadata richness:** The more detailed the metadata, the easier it is to find resources through refined searches.
- **Importance:** A resource that is not well-described with metadata cannot be accurately discovered.
- **Support for search engines:** Encourages data providers to consider various search facets and provide both generic and domain-specific descriptors to enable global and local search engines to locate the resource.

F2: Data are described with rich metadata

Implementation consideration

- **Domain-specific challenges:** Each community must define metadata descriptors necessary for optimizing findability.
- **Defining richness:** Determine the minimal richness of metadata to serve its intended purpose while aligning with other FAIR principles.
- **Machine-actionable templates:** Create templates to capture uniform and harmonized metadata across similar data resources within the community.
- **Metadata curation:** Ensure that metadata is continuously updated and curated to maintain its accuracy and usefulness.
- **Examples:** Metadata schemata like the Data Documentation Initiative (DDI) (https://ddialliance.org/), HCLS Dataset Descriptors (https://www.w3.org/TR/hcls-dataset/), and domain-specific minimal information models available on FAIRsharing (https://fairsharing.org/).

F2. Data are described with rich metadata

FER: Metadata schema

- **Definition:** A structured framework that defines how metadata should be recorded and organized.
- **Purpose:** Facilitates the discovery, understanding, and use of data by ensuring consistent and meaningful descriptions.
- **Examples:** Dublin Core for general metadata, DataCite for research data, Schema.org for web data.

Benefits

- **Enhanced Searchability:** Makes data easier to find through detailed and standardized descriptions.
- **Consistency:** Ensures that similar data is described in the same way across different datasets and platforms.
- **Data Sharing and Reuse:** Promotes effective sharing and reuse of data by providing clear, standardized information.

Metadata Schema 4.4

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

CEDAR
https://metadatacenter.org/

PROV-DM: The PROV Data Model
W3C Recommendation 30 April 2013

CEDAR

DataFAIRy Bioassay Annotation

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
[illegible]3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Accessible– The "A" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

A2: Metadata are accessible, even when the data are no longer available

Interpretation

- **Focus on longevity:** Emphasizes the importance of keeping relevant digital resources available for the future.
- **Inaccessibility scenarios:** Data may become inaccessible by design (e.g., financial or legal constraints) or by accident.
- **Importance of metadata:** Ensures that high-quality metadata remains accessible to describe the resources, allowing users to understand their nature and provenance even if the data itself is unavailable.
- **Dependence on principle F3**: Relies on the metadata containing the identifier of the data (as per F3), allowing the historical metadata record to be discovered if the data is no longer available.
- **Data citation**: This aspect of accessibility is further supported by the Joint Declaration of Data Citation Principles.

A2: Metadata are accessible, even when the data are no longer available

Implementation Considerations

- **Persistence policy:**
  - Communities need to define a persistence policy for metadata describing data that may not always be available.
  - Create machine-actionable templates for persistence policy documents for metadata.
  - Develop a machine-actionable scheme to reference the metadata persistence policy.
- **Digital curation principles:**
  - Early attempts to address this principle align with digital curation practices, including FAIR-compliant Data Management Plans (DMPs).
  - Continuous efforts are required to improve long-term stewardship of reusable digital resources.

A2. Metadata are accessible, even when the data are no longer available

FER: Metadata Preservation Policy

- **Definition**: Guidelines and practices for maintaining the usability and accessibility of metadata over time.
- **Purpose**: Ensures that metadata remains intact, understandable, and accessible for the long term.
- **Examples**: Policies from institutions like Data Archiving and Networked Services (DANS), National Digital Stewardship Alliance (NDSA) guidelines.

Benefits

- **Long-term Usability:** Ensures metadata can be used and understood in the future.
- **Data Curation:** Supports ongoing maintenance and improvement of metadata.
- **Archival Standards:** Helps comply with standards for digital preservation and archiving.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms gua global uniqueness, (2) policy document that guarantees persist resolution of the identifier to machine-actionable metadata de object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attribut other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-d machine-actionable link between metadata and the data they
F4	registry	A **service** that indexes metadata and data and provides search index.
A1.1	communication protocol	A **specification** that defines how messages are structured and
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to s conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metada provisioned in the future (maybe part of a data management p
I1	knowledge representation language	A language **specification** whereby knowledge can be made pro machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concept definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entitie data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digita be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the orig of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Interoperable– The "I" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

I2: (meta)data use vocabularies that follow FAIR principles

Interpretation

- **Vocabulary definition:** Methods that unambiguously represent concepts within a given domain.
- **Importance of structured terms:** Use of shared, formally structured vocabularies, including flat vocabularies, hierarchical thesauri, data models, and ontologies, is essential.
- **FAIR vocabularies:** Vocabularies must be findable, accessible, interoperable, and reusable to ensure machines and users understand the terms used in metadata.
- **Label ambiguity:** Simple labels (e.g., "temperature") are insufficient for machine understanding without context.
- **Detection of false agreements/disagreements:** Vocabulary terms must be distinguishable by machines to avoid misinterpretation.

I2: (meta)data use vocabularies that follow FAIR principles

Implementation Considerations

- **Ensuring FAIR terminology systems**:
  - Communities need to ensure that units of measure, classifications, and relationship definitions are FAIR.
  - Avoid proprietary thesauri that are not universally accessible.
- **Accessibility issues:**
  - Proprietary systems may restrict access, making data less useful even if technically accessible.
- **Examples:**
  - Ontologies in the "Web Ontology Language" (OWL) shared via accessible registries like BioPortal for life science ontologies (https://bioportal.bioontology.org/).
  - BioPortal provides a machine-accessible search interface and complies with the Findability requirements of FAIR.

I2. (Meta)data use vocabularies that follow FAIR principles

FER: Structured vocabulary

- **Definition**: Standardized sets of terms used to describe data elements consistently.
- **Purpose**: Ensures clarity and consistency in data descriptions, facilitating interoperability.
- **Examples**: MeSH (Medical Subject Headings) for biomedical terms, SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms), AGROVOC for agricultural data.

Benefits

- **Standardization:** Ensures data is described consistently across different datasets and platforms.
- **Interoperability:** Facilitates integration and use of data from different sources.
- **Enhanced Discovery:** Makes it easier to find data using standardized terms.

BioPortal
- ZonMw Generic Terms
- ZonMw COVID-19 Vocabulary

ONTO COMMONS

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

I3: (meta)data include qualified references to other (meta)data

Interpretation

- **Interconnected data:** Emphasizes that data and metadata should not exist in isolation; they should be interconnected to create a meaningful network of data and services.
- **Qualified references:** References to other resources, specifying the nature of the relationship (e.g., linking to prior or next versions).
- **Examples of relationships:**
  - Links between different versions of a metadata file (e.g., “prior version” or “next version”).
  - Linking datasets to related resources (e.g., city data to Wikidata, geographical, and geospatial data).
- **Intrinsic and secondary metadata:** Different metadata files can describe the same digital resource from various perspectives (e.g., intrinsic metadata, provenance metadata, secondary metadata).
- **Distinguishing metadata and resources:** It is good practice to clearly distinguish between metadata (files/containers) and the resources they describe.

I3: (meta)data include qualified references to other (meta)data

Implementation considerations

- **Selection of relationships:** Choose appropriate relationships from existing vocabularies or create new ones following FAIR principles.
- **Use of upper ontologies:**
  - Upper ontologies, like the Basic Formal Ontology (BFO), provide predefined relationships that can be used directly or as a basis for new, more specific relationships.
  - Leveraging upper ontologies helps ensure that new relationships are interoperable and comprehensible by agents familiar with higher-level concepts.
- **Enhancing Interoperability:** Creating well-defined relationships enhances the ability of machines to interpret the connections between different data resources accurately.

I3. (Meta)data include qualified references to other (meta)data

FER: Semantic model

- **Definition**: Frameworks that define the relationships between data elements
- **Purpose**: Provides a structured representation of data semantics, enhancing data interoperability and integration.
- **Examples**: Schema.org for web data, SKOS (Simple Knowledge Organization System) for linking concepts.

Benefits

- **Data Interoperability**: Improves the ability to use data across different systems and applications.
- **Integration**: Supports the combination of data from different sources.
- **Advanced Queries**: Enables complex querying and analysis by providing detailed semantic relationships.

DublinCore

DDI

fairdata/fairdatapoint ☆

By fairdata • Updated a month ago

Serve FAIR Metadata from this docker image

Data Catalog Vocabulary (DCAT) - Version 3

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteein global uniqueness, (2) policy document that guarantees persistent and resolution of the identifier to machine-actionable metadata describing object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of dat other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-direction machine-actionable link between metadata and the data they describe
F4	registry	A **service** that indexes metadata and data and provides search over tha index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchang
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are t provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with th definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities descrit data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and li of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Reusable– The "R" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Interpretation

- **Distinction from F2:** While F2 focuses on enabling effective attribute-based search and query (findability), R1 is about assessing the appropriateness of resources for reuse.
- **Relevance assessment:** Helps determine if discovered resources are suitable for specific tasks.
  - **Example:** Not all gene expression data are relevant for studies on heat stress effects.
- **Detailed metadata:** Beyond high-level facets, detailed metadata provides operational instructions for reuse.
- **Generosity in metadata:** Authors should provide extensive metadata to support various use cases and user needs.
- **Plurality of attributes:** Emphasizes the inclusion of diverse and detailed metadata attributes to ensure comprehensive resource descriptions.
- **Support for diverse use-cases:** Detailed metadata aids in determining how to include and process resources in analyses.

R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Implementation Considerations

- **Consider high-level and detailed facets:** Providers should consider both high-level and detailed metadata facets to support search and operational reuse.
- **Support wide variety of factors:** Consider the various factors that might influence the suitability of a resource for different tasks.
- **Critical attributes:** Sub-principles R1.1, R1.2, and R1.3 define essential types of attributes that contribute to comprehensive metadata descriptions.

R1.1: (meta)data are released with a clear and accessible data usage license

Interpretation

- **Mandatory licensing:** digital resources and their metadata must include a license describing usage conditions, even if unconditional.
- **Legal clarity:** resources cannot be legally used without a clear, accessible license; an absent license creates legal uncertainty and deters reuse.
- **Differentiating licenses:** licenses may differ between data resources and their metadata, impacting findability.
- **Public domain statement:** a clear equivalent such as terms of use or a computer protocol may be necessary.
- **Avoid restrictive licenses:** combining resources with restrictive licenses can have adverse effects and preclude their use.
- **Open licenses:** to facilitate reuse, choose licenses that are as open as possible.

R1.1: (meta)data are released with a clear and accessible data usage license

Implementation Considerations

- **Clarity in licensing:** distinguish between licenses for data and metadata to avoid ambiguity.
- **Community choice:** communities must choose suitable usage licenses for their digital resources and metadata, considering broader reuse.
- **CC0 license:** recommended for data to maximize reuse potential.
- **Qualified link:** ensure metadata records include a qualified link to the chosen license.

R1.1. (Meta)data are released with a clear and accessible data usage license

FER: Usage license

- **Definition:** Legal frameworks governing data use and sharing
- **Purpose:** Clarifies rights and responsibilities regarding data use, ensuring legal and ethical compliance.
- **Examples:** Creative Commons licenses for flexible data sharing, Open Data Commons licenses for open data, MIT License for software.

Benefits

- **Legal Compliance**: Ensures data use complies with legal requirements.
- **Promotes Sharing:** Encourages data sharing by clearly defining usage terms.
- **Protects IP:** Safeguards the intellectual property rights of data creators.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3 resolution of the identifier to machine-actionable metadata describing th object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible b machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with the definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describin data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object ca be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and line of data or other digital objects.
R1.3	FIP as a whole	

R1.2: (meta)data are associated with detailed provenance

Interpretation

- **Detailed provenance:** Includes how and why the resource was generated, by whom, under what conditions, using what starting data or source resource, funding, ownership, credit, and any post-generation data processing.
- **Assessment aid:** Provenance information helps people and machines determine if a resource meets their criteria for reuse and what data manipulation might be necessary.

R1.2: (meta)data are associated with detailed provenance

Implementation considerations

- **Metadata choices:** Communities need to choose metadata descriptions that optimize provenance to support both machine and human reusability.
- **Focus on reuse:** Implementation considerations are similar to those for principle F2 but focus more on appropriateness for reuse rather than findability.
- **Provenance templates:** Use community-specific templates like the PROV-Template approach to structure provenance information.
- **Reducing burden:** Templates help community members avoid the complexity of the PROV ontology (https://www.w3.org/TR/prov-o/) by providing predefined structures.
- **Early tools:** Tools like CEDAR (https://metadatacenter.org/), and the Data Stewardship Wizard (https://ds-wizard.org/) are being developed to simplify the creation of FAIR metadata.

R1.2. (Meta)data are associated with detailed provenance

FER: Provenance model

- **Definition:** A specification that defines metadata fields describing the origin and lineage of data or other digital objects.
- **Purpose:** Ensures transparency, traceability, and accountability in data handling by documenting its creation, modification, and ownership history.
- **Examples:** PROV-DM (Provenance Data Model), W3C PROV standards.

Benefits

- **Data Trustworthiness**: Enhances the reliability and credibility of data by providing detailed information about its origin and changes over time.
- **Reproducibility:** Supports the replication of research by documenting the complete history of data, including how it was generated and processed.
- **Data Integrity:** Ensures that data has not been tampered with by maintaining a comprehensive record of its lineage and transformations.

Metadata Schema 4.4

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

CEDAR
https://metadatacenter.org/

PROV-DM: The PROV Data Model
W3C Recommendation 30 April 2013

CEDAR

DataFAIRy Bioassay Annotation

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

R1.3: (meta)data meet domain-relevant community standards

Interpretation

- **Follow standards:** Where community standards or best practices for data archiving and sharing exist, they should be followed.
- **Minimal information standards:** Defined by disciplinary communities to describe the minimal set of metadata items required for data quality assessment and reproducibility.
- **Interdisciplinary reusability:** Generally requires richer metadata beyond minimal standards.
- **Resource:** Consult FAIRsharing (https://fairsharing.org/) for a list of such standards.

R1.3: (meta)data meet domain-relevant community standards

Implementation considerations

- **Practice selection:** Communities must choose appropriate practices for data and metadata, considering inter-domain interoperability.
- **Metadata elements:** Communities should decide which elements in their "boutique" standards should also be represented using more global standards (principles F2 and R1.2), even if it results in metadata duplication.
- **Example standards:** MIAME standard for minimal information and various DCAT profiles for metadata.

R1.3.(Meta)data meet domain-relevant community standards

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

Take home key points – How to make your data FAIR?

- **Findability:** Assign globally unique and persistent identifiers, describe data with rich metadata, ensure metadata links clearly to data, and register data in searchable resources.
- **Accessibility:** Use standardized, open, and free protocols for data retrieval, allow for necessary authentication and authorization, and ensure metadata remains accessible even if data isn't.
- **Interoperability:** Use formal, shared languages for knowledge representation, employ FAIR-compliant vocabularies, and include qualified references to interlink data.
- **Reusability:** Provide rich descriptions with clear usage licenses, detailed provenance, and ensure adherence to community standards.

ChIPs Workshop on FAIRification

AGENDA

- 14:00 – 14:30 | **Greetings and Introduction on FAIR principles and their current state**
  Iseult Lynch (University of Birmingham)
- 14:30 – 15:30 | **Session on Data**
  > Data curation and FAIRification for knowledge graph development
  Angela Serra (Tampere University)

  > Semantic data documentation for FAIR data
  Jesper Friis (Sintef)
- 15:30 – 16:00 | Coffee Break
- 16:00 – 17:00 | **Session on Models**
  > Harmonizing model-characterization templates across different computational and experimental fields: Towards improved readability of regulatory model validity
  Martin Paparella (Medizinische Universitat Innsbruck)

  > FAIRifying Computational Models: Advancing Transparency and Reusability in the SSbD framework
  ***Nikiforou Fotini** (Aristotle University of Thessaloniki)*
- 17:00 – 18:00 | **Open Session**
  In Vitro Model Complexity: A Challenge for FAIRification?
  Moderator: *Steffi Friedrichs (Acumenist)*

Interpretation of the FAIR principles

The unabridged set of slides

Findable – The "F" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

F1: (meta)data are assigned a globally unique and persistent identifier

Interpretation

- **Essential element:** Fundamental to all FAIR principles.
- **Globally unique:** Identifier must unambiguously refer to one resource universally, not just locally.
- **Persistence:** Identifier is never reused and continues to identify the same resource over time, even if the resource changes or ceases to exist.
- **Practical implementation:** Typically involves third-party generation of identifiers with guaranteed longevity and independence from the originating project or organization.

F1: (meta)data are assigned a globally unique and persistent identifier

Implementation consideration

- **Challenges:** Ensuring identifier longevity beyond the lifespan of the originating project or community.
- **Reliance on third parties:** Requires third-party organizations to maintain and guarantee the longevity of identifiers.
- **Community choices:** Each community should select appropriate identifier registration services ensuring global uniqueness, persistence, and resolvability.
- **Example:** Digital Object Identifier (DOI) (https://www.doi.org/) is a widely used example that guarantees global uniqueness and persistence, and provides additional services like directing calls to the source data's current location.
- **Additional support:** Persistent identifiers like DOIs can also support metadata availability beyond the data’s lifespan, ensuring continued metadata accessibility.

F1. (Meta)data are assigned a globally unique and persistent identifier

FER: Identifier service

- **Definition:** Provides unique, persistent identifiers for digital objects, ensuring they can always be found.
- **Purpose:** Ensures that data can be reliably located and accessed over time, regardless of changes in its location or ownership.
- **Examples**: Digital Object Identifier (DOI) for publications, ORCID for researchers, Handle System for various digital objects.

Benefits

- **Long-term Access:** Ensures data remains accessible even if the location changes.
- **Citation and Tracking:** Supports proper citation of data in academic work and tracking of its use.
- **Data Integrity:** Helps in maintaining the integrity and authenticity of data.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

F2: Data are described with rich metadata

Interpretation

- **Discovery of resources:** Facilitates the discovery of digital resources through search or filtering.
- **Rich metadata:** Digital resources must be described with detailed descriptors of their content.
- **Metadata richness:** The more detailed the metadata, the easier it is to find resources through refined searches.
- **Importance:** A resource that is not well-described with metadata cannot be accurately discovered.
- **Support for search engines:** Encourages data providers to consider various search facets and provide both generic and domain-specific descriptors to enable global and local search engines to locate the resource.

F2: Data are described with rich metadata

Implementation consideration

- **Domain-specific challenges:** Each community must define metadata descriptors necessary for optimizing findability.
- **Defining richness:** Determine the minimal richness of metadata to serve its intended purpose while aligning with other FAIR principles.
- **Machine-actionable templates:** Create templates to capture uniform and harmonized metadata across similar data resources within the community.
- **Metadata curation:** Ensure that metadata is continuously updated and curated to maintain its accuracy and usefulness.
- **Examples:** Metadata schemata like the Data Documentation Initiative (DDI) (https://ddialliance.org/), HCLS Dataset Descriptors (https://www.w3.org/TR/hcls-dataset/), and domain-specific minimal information models available on FAIRsharing (https://fairsharing.org/).

F2. Data are described with rich metadata

FER: Metadata schema

- **Definition:** A structured framework that defines how metadata should be recorded and organized.
- **Purpose:** Facilitates the discovery, understanding, and use of data by ensuring consistent and meaningful descriptions.
- **Examples:** Dublin Core for general metadata, DataCite for research data, Schema.org for web data.

Benefits

- **Enhanced Searchability:** Makes data easier to find through detailed and standardized descriptions.
- **Consistency:** Ensures that similar data is described in the same way across different datasets and platforms.
- **Data Sharing and Reuse:** Promotes effective sharing and reuse of data by providing clear, standardized information.

Metadata Schema 4.4

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

CEDAR
https://metadatacenter.org/

PROV-DM: The PROV Data Model
W3C Recommendation 30 April 2013

CEDAR

DataFAIRy Bioassay Annotation

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
[illegible]3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

F3: metadata clearly and explicitly include the identifier of the data it describes

Implementation Considerations

- **Community challenges:** Selecting a machine-actionable metadata model that links resources and metadata explicitly.
- **Example technology:** FAIR Data Point (https://www.fairdatapoint.org/), based on the Data Catalogue model (DCAT) (https://www.w3.org/TR/vocab-dcat-3/), which provides unique identifiers for multiple layers of metadata and a predictable, searchable path through these layers to the data object.
- **Metadata models:** Communities must choose models that ensure a clear and searchable connection between data and its metadata.

F3: metadata clearly and explicitly include the identifier of the data it describes

Interpretation

- **Explicit identification:** Descriptions of digital resources must contain the identifier of the resource being described.
- **Unambiguous description:** Ensure that identifiers are included explicitly and unambiguously in metadata.
- **Independent storage:** Crucial when resources and metadata are stored independently but are persistently linked.
- **Twofold purpose:**
  - **Basic clarity:** Metadata should clearly indicate what it is describing.
  - **Technical necessity:** Allows for discovery of metadata via the resource's identifier, especially when digital objects have structures that disallow additional fields.
- **Search optimization:** Using the identifier as a search term to discover metadata records effectively.

F3. Metadata clearly and explicitly include the identifier of the data they describe

FER: Metadata-data linking schema

- **Definition:** Structures that link metadata directly to the corresponding data it describes.
- **Purpose:** Ensures accurate and efficient retrieval of data through its metadata.
- **Examples:** JSON-LD (JavaScript Object Notation for Linked Data), RDF (Resource Description Framework).

Benefits

- **Improved Discoverability:** Enhances the ability to find data based on its metadata.
- **Accuracy:** Ensures that metadata accurately represents the data it describes.
- **Supports Complex Queries:** Enables more advanced and precise searches by linking detailed metadata with data.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

F4: (meta)data are registered or indexed in a searchable resource

Interpretation

- **Searchable resource:** Digital resources must be registered or indexed in a searchable resource.
- **Infrastructure:** Searchable resources provide the necessary infrastructure to discover metadata records.
- **Discovery mechanism:** Utilizes attributes in the metadata (F2) or the identifier of the data object (F3) for discovery.
- **Community choices:**
  - Each community should choose and publicly declare which search engine they will use, whether general or field-specific.
  - Ensure metadata is provided in a standard indexed by the chosen search engine.
  - Provide a machine-readable interface definition to enable automated searches.

F4: (meta)data are registered or indexed in a searchable resource

- **Challenges:**
  - **No single search source:** There is currently no single search engine that indexes all metadata fields across all domains.
  - **Lack of uniform search execution:** Different search tools require tool-specific software, complicating the search process.
  - **Automated search limitations:** Many search engines forbid automated searches, limiting their use in FAIR-enabled software.
  - **Emerging initiatives:** Various initiatives aim to provide well-defined, machine-accessible search interfaces over indexed metadata, though none cover all metadata properties or domains.
- **Example:** Google Dataset Search (https://datasetsearch.research.google.com/) supports manual exploration but has limitations, such as indexing only certain metadata types and forbidding automated searches under its Terms of Service.

F4. (Meta)data are registered or indexed in a searchable resource

FER: Registry

- **Definition**: A service that indexes metadata and data, providing search capabilities over the indexed information.
- **Purpose:** Facilitates the discovery, management, and sharing of data by providing an efficient and centralized search interface.
- **Examples:** Zenodo for research outputs, Dryad for data publications, Figshare for general data sharing.

Benefits:

- **Efficient Data Discovery**: Enhances the ability to find relevant data through powerful search functionalities.
- **Centralized Management:** Provides a single point of access for managing and searching through large volumes of data.
- **Supports Compliance:** Helps researchers meet data management and sharing requirements by providing a structured and reliable service for data indexing and search.

Google

Microsoft Bing

health RI Portal

SeaDataNet

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Accessible– The "A" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

A1: (meta)data are retrievable by their identifier using a standardized communications protocol

Interpretation

- **Purpose of identification:** Ensures the ability to retrieve the digital resource record using a defined mechanism.
- **Retrievability as accessibility:** Focuses on providing the capability to access the resource without additional barriers.
- **Mechanized access:** Requires that the identifier (F1) follows a globally accepted schema tied to a standardized, high-level communication protocol.
- **Standardized protocol importance:** Provides a predictable way for agents to access resources, regardless of access restrictions.
- **Example protocol:** Hypertext Transfer Protocol (HTTP) is a standardized protocol, but FAIR also allows non-mechanized protocols for sensitive data, as long as they are explicit and clearly defined.

A1: (meta)data are retrievable by their identifier using a standardized communications protocol

Implementation Considerations

- **Challenges:**
  - Ensuring that the standardized protocol is accessible to all potential agents, including those behind firewalls.
  - Making sure that the chosen protocol is globally accepted and well-documented.
- **Example:**
  - HTTP is commonly used for mechanized access, supporting predictability and standardization.
  - Verbal requests or other non-mechanized protocols can be used for sensitive data, provided the access method is clearly defined.

A1.1: The protocol is open, free, and universally implementable

Interpretation

- **Open protocol:** The mechanism for accessing digital resources should not create bottlenecks.
- **Accessibility:** Describes the access process, not restrictions on resource use.
- **Web protocols:** Protocols like HTTP are models of open, free, and universally implementable protocols.
- **Cost reduction:** Well-defined and open protocols lower the cost of accessing digital resources.
- **Equitable access:** Free protocols ensure access for those without monetary means.
- **Universal implementation:** Protocols must be available globally, not restricted by region or community, covering both “gratis” and “libre” meanings of "free".

A1.1: The protocol is open, free, and universally implementable

Implementation considerations

- **Challenges:**
  - Fully document protocols that are not open or free, and make these details part of the machine-readable metadata.
- **Community choices:**
  - Select standardized communication protocols that are open, free, and universally implementable.
- **Example:**
  - HTTP is the most common example, underlying the majority of web traffic.
  - HTTP supports requesting metadata in preferred formats and is widely supported by various software and programming languages.

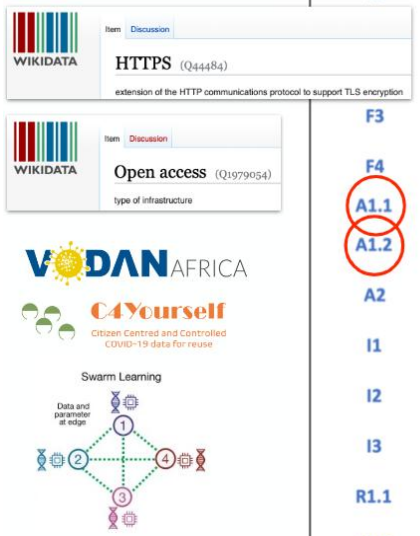
A1.1 The protocol is open, free, and universally implementable

FER: Communication protocol

- **Definition:** Standards and rules that govern how data is exchanged between systems.
- **Purpose:** Ensures smooth, secure, and standardized data transfer across different platforms and systems.
- **Examples:** HTTP (HyperText Transfer Protocol), FTP (File Transfer Protocol), OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

Benefits

- **Interoperability:** Ensures different systems can work together and exchange data seamlessly.
- **Security:** Protects data during transfer, ensuring it reaches its destination without being compromised.
- **Real-time Exchange:** Supports timely data sharing and updates between systems.



FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

A1.2: The protocol allows for an authentication and authorization procedure, where necessary

Interpretation

- **FAIR vs. Open**: FAIR does not mean all resources are openly accessible; some have access restrictions due to ethical, legal, or contractual constraints.
- **Authentication:** Ensures the requester is who they claim to be.
- **Authorization:** Ensures the requester's credentials and profile match the access conditions of the resource.
- **Permitted use:** Verifies that the intended use matches permitted use cases (e.g., non-commercial purposes).
- **Technical implementation:** Requires an additional authentication and authorization procedure if not defined by the access protocol (see A1.1).
- **Requesters:** Can be humans or machine agents (proxies for humans or organizations).
- **Requirement:** FAIR resources must provide such a protocol for access control.
- **AAI:** An Internet of FAIR Data and Services relies on Authentication and Authorization Infrastructure (AAI).

A1.2: The protocol allows for an authentication and authorization procedure, where necessary

Interpretation considerations

- **Community Choices:**
  - Select protocols to control access to meta(data) that are as generic as possible but also domain-specific where necessary.
  - Harmonize AAI approaches
- **Example Protocols:**
  - HTTP protocol is a common example for implementing access control.
  - Life science AAI protocol is another example.

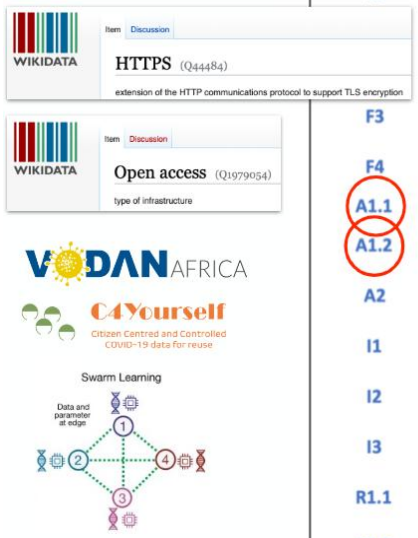
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

FER: Authorization and Authentication Service

- **Definition:** Systems that manage who can access data and verify their identity.
- **Purpose:** Ensures that only authorized users can access sensitive or restricted data, protecting it from unauthorized use.
- **Examples:** OAuth for secure access delegation, LDAP (Lightweight Directory Access Protocol) for directory services.

Benefits

- **Data Security:** Protects sensitive data from unauthorized access.
- **Access Control:** Manages user permissions and access levels effectively.
- **Compliance:** Ensures adherence to data protection regulations and policies.



FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

A2: Metadata are accessible, even when the data are no longer available

Interpretation

- **Focus on longevity:** Emphasizes the importance of keeping relevant digital resources available for the future.
- **Inaccessibility scenarios:** Data may become inaccessible by design (e.g., financial or legal constraints) or by accident.
- **Importance of metadata:** Ensures that high-quality metadata remains accessible to describe the resources, allowing users to understand their nature and provenance even if the data itself is unavailable.
- **Dependence on principle F3**: Relies on the metadata containing the identifier of the data (as per F3), allowing the historical metadata record to be discovered if the data is no longer available.
- **Data citation**: This aspect of accessibility is further supported by the Joint Declaration of Data Citation Principles.

A2: Metadata are accessible, even when the data are no longer available

Implementation Considerations

- **Persistence policy:**
  - Communities need to define a persistence policy for metadata describing data that may not always be available.
  - Create machine-actionable templates for persistence policy documents for metadata.
  - Develop a machine-actionable scheme to reference the metadata persistence policy.
- **Digital curation principles:**
  - Early attempts to address this principle align with digital curation practices, including FAIR-compliant Data Management Plans (DMPs).
  - Continuous efforts are required to improve long-term stewardship of reusable digital resources.

A2. Metadata are accessible, even when the data are no longer available

FER: Metadata Preservation Policy

- **Definition**: Guidelines and practices for maintaining the usability and accessibility of metadata over time.
- **Purpose**: Ensures that metadata remains intact, understandable, and accessible for the long term.
- **Examples**: Policies from institutions like Data Archiving and Networked Services (DANS), National Digital Stewardship Alliance (NDSA) guidelines.

Benefits

- **Long-term Usability:** Ensures metadata can be used and understood in the future.
- **Data Curation:** Supports ongoing maintenance and improvement of metadata.
- **Archival Standards:** Helps comply with standards for digital preservation and archiving.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms gua global uniqueness, (2) policy document that guarantees persist resolution of the identifier to machine-actionable metadata de object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attribut other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-d machine-actionable link between metadata and the data they
F4	registry	A **service** that indexes metadata and data and provides search index.
A1.1	communication protocol	A **specification** that defines how messages are structured and
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to s conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metada provisioned in the future (maybe part of a data management p
I1	knowledge representation language	A language **specification** whereby knowledge can be made pro machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concept definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entitie data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digita be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the orig of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Interoperable– The "I" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Interpretation

- **Challenges for consumers:** Difficulty in understanding and combining digital resources due to ambiguous or non-machine-interpretable content descriptors.
- **Community-defined formats:** Work well within narrow scopes but fail with broader, more diverse data sets.
- **Interoperation and integration:** Expensive and often impossible without clear, unambiguous descriptors.
- **Goal of FAIR:** Enable machines to easily use digital resources through a common language for knowledge representation.
- **Contextual descriptions:** Essential for distinguishing between similarly named data fields (e.g., "temperature" in weather vs. body temperature data).
- **Global understanding:** Requires a globally understood language for machines, defining entities and relationships within the data.
- **Prerequisite for FAIR data and services:** Achieving a common understanding of digital resources by machines is essential.

I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Implementation considerations

- **Choosing technologies:** Communities must select technologies for knowledge representation or manage multiple representations.
- **Consistency across resources:** Ensure consistent interpretation of data items across resources by all agents (human and machine).
- **Generic reuse:** FAIR aims for data to be reused by generic agents, not just community-specific ones.
- **Widely used formats:** Knowledge should be available in widely accepted formats, potentially duplicating community-specific formats.
- **Example technology:** The Resource Description Framework (RDF) is widely accepted for representing knowledge on the web in a machine-accessible format.
- **Translators:** Communities using specific formats should provide translators to more widely used formats like RDF.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

FER: Knowledge Representation Language

- **Definition**: Formal languages used to represent information about data in a structured, machine-readable format.
- **Purpose**: Facilitates understanding, sharing, and reasoning about data by machines.
- **Examples**: RDF (Resource Description Framework), OWL (Web Ontology Language).

Benefits

- **Semantic Understanding:** Enhances the ability of machines to understand and process data semantically.
- **Data Integration:** Supports integration of data from different sources by providing a common representation.
- **Machine Readability:** Enables automated reasoning and processing of data.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
.2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

I2: (meta)data use vocabularies that follow FAIR principles

Interpretation

- **Vocabulary definition:** Methods that unambiguously represent concepts within a given domain.
- **Importance of structured terms:** Use of shared, formally structured vocabularies, including flat vocabularies, hierarchical thesauri, data models, and ontologies, is essential.
- **FAIR vocabularies:** Vocabularies must be findable, accessible, interoperable, and reusable to ensure machines and users understand the terms used in metadata.
- **Label ambiguity:** Simple labels (e.g., "temperature") are insufficient for machine understanding without context.
- **Detection of false agreements/disagreements:** Vocabulary terms must be distinguishable by machines to avoid misinterpretation.

I2: (meta)data use vocabularies that follow FAIR principles

Implementation Considerations

- **Ensuring FAIR terminology systems**:
  - Communities need to ensure that units of measure, classifications, and relationship definitions are FAIR.
  - Avoid proprietary thesauri that are not universally accessible.
- **Accessibility issues:**
  - Proprietary systems may restrict access, making data less useful even if technically accessible.
- **Examples:**
  - Ontologies in the "Web Ontology Language" (OWL) shared via accessible registries like BioPortal for life science ontologies (https://bioportal.bioontology.org/).
  - BioPortal provides a machine-accessible search interface and complies with the Findability requirements of FAIR.

I2. (Meta)data use vocabularies that follow FAIR principles

FER: Structured vocabulary

- **Definition**: Standardized sets of terms used to describe data elements consistently.
- **Purpose**: Ensures clarity and consistency in data descriptions, facilitating interoperability.
- **Examples**: MeSH (Medical Subject Headings) for biomedical terms, SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms), AGROVOC for agricultural data.

Benefits

- **Standardization:** Ensures data is described consistently across different datasets and platforms.
- **Interoperability:** Facilitates integration and use of data from different sources.
- **Enhanced Discovery:** Makes it easier to find data using standardized terms.

BioPortal
- ZonMw Generic Terms
- ZonMw COVID-19 Vocabulary

ONTO COMMONS

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

I3: (meta)data include qualified references to other (meta)data

Interpretation

- **Interconnected data:** Emphasizes that data and metadata should not exist in isolation; they should be interconnected to create a meaningful network of data and services.
- **Qualified references:** References to other resources, specifying the nature of the relationship (e.g., linking to prior or next versions).
- **Examples of relationships:**
  - Links between different versions of a metadata file (e.g., “prior version” or “next version”).
  - Linking datasets to related resources (e.g., city data to Wikidata, geographical, and geospatial data).
- **Intrinsic and secondary metadata:** Different metadata files can describe the same digital resource from various perspectives (e.g., intrinsic metadata, provenance metadata, secondary metadata).
- **Distinguishing metadata and resources:** It is good practice to clearly distinguish between metadata (files/containers) and the resources they describe.

I3: (meta)data include qualified references to other (meta)data

Implementation considerations

- **Selection of relationships:** Choose appropriate relationships from existing vocabularies or create new ones following FAIR principles.
- **Use of upper ontologies:**
  - Upper ontologies, like the Basic Formal Ontology (BFO), provide predefined relationships that can be used directly or as a basis for new, more specific relationships.
  - Leveraging upper ontologies helps ensure that new relationships are interoperable and comprehensible by agents familiar with higher-level concepts.
- **Enhancing Interoperability:** Creating well-defined relationships enhances the ability of machines to interpret the connections between different data resources accurately.

I3. (Meta)data include qualified references to other (meta)data

FER: Semantic model

- **Definition**: Frameworks that define the relationships between data elements
- **Purpose**: Provides a structured representation of data semantics, enhancing data interoperability and integration.
- **Examples**: Schema.org for web data, SKOS (Simple Knowledge Organization System) for linking concepts.

Benefits

- **Data Interoperability**: Improves the ability to use data across different systems and applications.
- **Integration**: Supports the combination of data from different sources.
- **Advanced Queries**: Enables complex querying and analysis by providing detailed semantic relationships.

DublinCore

DDI

fairdata/fairdatapoint ☆

Data Catalog Vocabulary (DCAT) - Version 3

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

Interpretation of the FAIR principles

Reusable– The "R" principles

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29. https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation

R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Interpretation

- **Distinction from F2:** While F2 focuses on enabling effective attribute-based search and query (findability), R1 is about assessing the appropriateness of resources for reuse.
- **Relevance assessment:** Helps determine if discovered resources are suitable for specific tasks.
  - **Example:** Not all gene expression data are relevant for studies on heat stress effects.
- **Detailed metadata:** Beyond high-level facets, detailed metadata provides operational instructions for reuse.
- **Generosity in metadata:** Authors should provide extensive metadata to support various use cases and user needs.
- **Plurality of attributes:** Emphasizes the inclusion of diverse and detailed metadata attributes to ensure comprehensive resource descriptions.
- **Support for diverse use-cases:** Detailed metadata aids in determining how to include and process resources in analyses.

R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Implementation Considerations

- **Consider high-level and detailed facets:** Providers should consider both high-level and detailed metadata facets to support search and operational reuse.
- **Support wide variety of factors:** Consider the various factors that might influence the suitability of a resource for different tasks.
- **Critical attributes:** Sub-principles R1.1, R1.2, and R1.3 define essential types of attributes that contribute to comprehensive metadata descriptions.

R1.1: (meta)data are released with a clear and accessible data usage license

Interpretation

- **Mandatory licensing:** digital resources and their metadata must include a license describing usage conditions, even if unconditional.
- **Legal clarity:** resources cannot be legally used without a clear, accessible license; an absent license creates legal uncertainty and deters reuse.
- **Differentiating licenses:** licenses may differ between data resources and their metadata, impacting findability.
- **Public domain statement:** a clear equivalent such as terms of use or a computer protocol may be necessary.
- **Avoid restrictive licenses:** combining resources with restrictive licenses can have adverse effects and preclude their use.
- **Open licenses:** to facilitate reuse, choose licenses that are as open as possible.

R1.1: (meta)data are released with a clear and accessible data usage license

Implementation Considerations

- **Clarity in licensing:** distinguish between licenses for data and metadata to avoid ambiguity.
- **Community choice:** communities must choose suitable usage licenses for their digital resources and metadata, considering broader reuse.
- **CC0 license:** recommended for data to maximize reuse potential.
- **Qualified link:** ensure metadata records include a qualified link to the chosen license.

R1.1. (Meta)data are released with a clear and accessible data usage license

FER: Usage licence

- **Definition:** Legal frameworks governing data use and sharing
- **Purpose:** Clarifies rights and responsibilities regarding data use, ensuring legal and ethical compliance.
- **Examples:** Creative Commons licenses for flexible data sharing, Open Data Commons licenses for open data, MIT License for software.

Benefits

- **Legal Compliance**: Ensures data use complies with legal requirements.
- **Promotes Sharing:** Encourages data sharing by clearly defining usage terms.
- **Protects IP:** Safeguards the intellectual property rights of data creators.

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

R1.2: (meta)data are associated with detailed provenance

Interpretation

- **Detailed provenance:** Includes how and why the resource was generated, by whom, under what conditions, using what starting data or source resource, funding, ownership, credit, and any post-generation data processing.
- **Assessment aid:** Provenance information helps people and machines determine if a resource meets their criteria for reuse and what data manipulation might be necessary.

R1.2: (meta)data are associated with detailed provenance

Implementation considerations

- **Metadata choices:** Communities need to choose metadata descriptions that optimize provenance to support both machine and human reusability.
- **Focus on reuse:** Implementation considerations are similar to those for principle F2 but focus more on appropriateness for reuse rather than findability.
- **Provenance templates:** Use community-specific templates like the PROV-Template approach to structure provenance information.
- **Reducing burden:** Templates help community members avoid the complexity of the PROV ontology (https://www.w3.org/TR/prov-o/) by providing predefined structures.
- **Early tools:** Tools like CEDAR (https://metadatacenter.org/), and the Data Stewardship Wizard (https://ds-wizard.org/) are being developed to simplify the creation of FAIR metadata.

R1.2. (Meta)data are associated with detailed provenance

FER: Provenance model

- **Definition:** A specification that defines metadata fields describing the origin and lineage of data or other digital objects.
- **Purpose:** Ensures transparency, traceability, and accountability in data handling by documenting its creation, modification, and ownership history.
- **Examples:** PROV-DM (Provenance Data Model), W3C PROV standards.

Benefits

- **Data Trustworthiness**: Enhances the reliability and credibility of data by providing detailed information about its origin and changes over time.
- **Reproducibility:** Supports the replication of research by documenting the complete history of data, including how it was generated and processed.
- **Data Integrity:** Ensures that data has not been tampered with by maintaining a comprehensive record of its lineage and transformations.

Metadata Schema 4.4

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

CEDAR
https://metadatacenter.org/

PROV-DM: The PROV Data Model
W3C Recommendation 30 April 2013

CEDAR

DataFAIRy Bioassay Annotation

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	

R1.3: (meta)data meet domain-relevant community standards

Interpretation

- **Follow standards:** Where community standards or best practices for data archiving and sharing exist, they should be followed.
- **Minimal information standards:** Defined by disciplinary communities to describe the minimal set of metadata items required for data quality assessment and reproducibility.
- **Interdisciplinary reusability:** Generally requires richer metadata beyond minimal standards.
- **Resource:** Consult FAIRsharing (https://fairsharing.org/) for a list of such standards.

R1.3: (meta)data meet domain-relevant community standards

Implementation considerations

- **Practice selection:** Communities must choose appropriate practices for data and metadata, considering inter-domain interoperability.
- **Metadata elements:** Communities should decide which elements in their "boutique" standards should also be represented using more global standards (principles F2 and R1.2), even if it results in metadata duplication.
- **Example standards:** MIAME standard for minimal information and various DCAT profiles for metadata.

R1.3.(Meta)data meet domain-relevant community standards

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	