FAIRness and Metadata schemata

General Concepts

Francesca Martines, INAF – Osservatorio Astronomico di Palermo

Open Science: What, Why and How

What

Open science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination ***accessible*** to all levels of society, amateur or professional.

Open science is ***transparent*** and accessible knowledge that is shared and developed through collaborative networks

Open Science: What, Why and How

What

It encompasses practices such as publishing ***open research***, campaigning for ***open access***, encouraging scientists to practice ***open-notebook*** science (such as openly sharing data and code), ***broader dissemination*** and engagement in science and generally making it ***easier to publish, access and communicate scientific knowledge.***

Why

The rationale underlying Open Science is one simple principle, that **findings from publicly-funded research should be accessible to all**.

Open Science: What, Why and How

Why

OS allows the **reproduction** of the research findings, enables **transparency** in the research methodology, increases the researcher's societal **impact** and **saves money and time** both for researchers and research institutions.

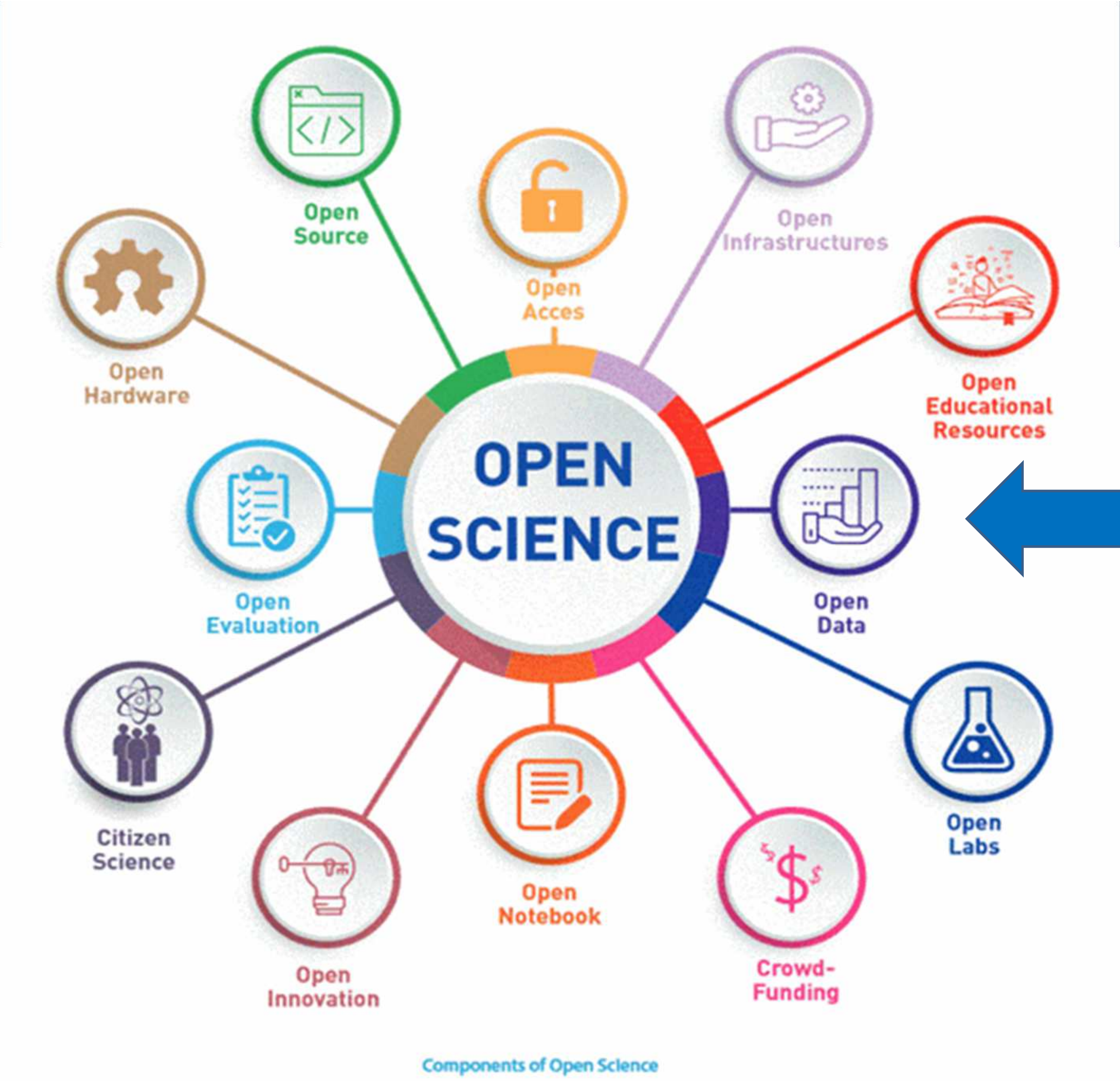
Pontika, Nancy; Knoth, Petr; Cancellieri, Matteo and Pearce, Samuel (2015). Fostering Open Science to Research using a Taxonomy and an eLearning Portal. In: iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business, 21-22 Oct 2015, Graz, Austria. DOI: https://doi.org/10.1145/2809563.2809571

How

We need a new perspective, new methodologies and tools

We have to rethink several concepts, including that of research evaluation

We must start from the components of Open Science


OPEN SCIENCE
Open Source
Open Acces
Open Infrastructures
Open Educational Resources
Open Data
Open Labs
Crowd-Funding
Open Notebook
Open Innovation
Citizen Science
Open Evaluation
Open Hardware
Components of Open Science

Open Data

Valid reasons not to participate in open science practices

Casper J. Albers*

Abstract

The past years have seen a sharp increase in the attention for open science practices. Such practices include pre-registration and registered reports, sharing of materials, open access publishing and attention to reproducibility of research. Despite the overwhelming amount of evidence highlighting the benefits of open science, some researchers remain reluctant. In this paper, I will outline valid reasons for researchers not to participate in open science practices.

Discussion

There are no valid reasons.

*Heymans Institute for Psychological Research, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. c.j.albers@rug.nl

FAIR Principles

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

Principles that must inspire practical implementations

FAIR Principles

Findable

Accessible

Interoperable

Reusable

https://www.go-fair.org/

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

https://www.go-fair.org/fair-principles/fairification-process/

Findable

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**

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

Globally unique and persistent identifiers remove ambiguity in the meaning of your published data by assigning a unique identifier to every element of metadata and every concept/measurement in your dataset.

Identifiers are essential to the human-machine interoperation that is key to the vision of **Open Science**. In addition, identifiers will help others to properly cite your work when reusing your data.

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

Examples of globally unique and persistent identifiers

- One particular person on planet earth has this globally unique and persistent identifier: **https://orcid.org/0000-0001-8888-635X**

- Here is an identifier that uniquely links to the results of a study estimating the FAIRness of different data repositories:

  doi:10.4121/uuid:5146dd06-98e4-426c-9ae5-dc8fa65c549f

- The number 163483 refers to the undergraduate student ID of **Mark Wilkinson**, the NCBI gi number for a bovine protease, and a part number for a Singer sewing machine. Hence, this is a poor example of F1!

F2. data are described with rich metadata (defined by R1 below) (I)

- In creating FAIR digital resources, **metadata** can (and should) be **generous and extensive**, including descriptive information about the context, quality and condition, or characteristics of the data.
- **Rich metadata allow a computer to automatically accomplish routine** and tedious sorting and prioritising tasks that currently demand a lot of attention from researchers.
- The rationale behind this principle is that **someone should be able to find data based on the information provided by their metadata**, even without the data's identifier.
- As such, compliance with F2 helps people to locate your data, and **increase re-use and citations.**

F2. data are described with rich metadata (defined by R1 below) (II)

This includes:

- **'intrinsic'** metadata (e.g., the data captured automatically by machines that generate data)
- **'contextual'** metadata (e.g., the protocol used, with both keywords and links to a formal protocol document)
- the **measurement devices** used
- the **units** of the captured data
- the **physical parameter space** of observed or simulated astronomical data sets

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

- The metadata and the dataset they describe are usually separate files.
- The **association** between a metadata file and the dataset should be made **explicit** by mentioning a dataset's globally unique and **persistent identifier** in the metadata.
- Many repositories will generate globally unique and persistent identifiers for deposited datasets that can be used for this purpose. The connection should be annotated in a formal manner.

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

- Identifiers and rich metadata descriptions alone **will not ensure 'findability'** on the internet.
- Perfectly good data resources may go unused simply because no one knows they exist.
- ***If the availability of a digital resource such as a dataset, service or repository is not known, then nobody (and no machine) can discover it.***

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

To be **Accessible:**

A1. (Meta)data are retrievable by their identifier using a **standardised communications protocol**

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

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

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

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

Principle A1 states that FAIR data retrieval should be mediated without specialised or proprietary tools or communication methods.

This principle focuses on **how data and metadata can be retrieved from their identifiers**.

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

To maximise data reuse, the protocol should be free (no-cost) and open (-sourced) and thus globally implementable to facilitate data retrieval. Anyone with a computer and an internet connection can access at least the metadata. Hence, this criterion will impact your choice of the repository where you will share your data.

Examples

- HTTP, FTP, SMTP, …
- A counter-example would be Skype, which is not universally-implementable because it is proprietary

A1.2 the protocol allows for an authentication and authorization procedure, where necessary (I)

The 'A' in FAIR does not necessarily mean 'open' or 'free'. Rather, it implies that **one should provide the exact conditions under which the data are accessible**.

Ideally, accessibility is specified in such a way that a machine can automatically understand the requirements, and then either automatically execute the requirements or alert the user to the requirements.

It often makes sense to request users to create a user account for a repository. This allows to authenticate the owner (or contributor) of each dataset, and to potentially set user-specific rights.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary (II)

Remember:

“As open as possible, as closed as necessary”

Open Science ≠ Open Access

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

There is a cost to maintaining an online presence for data resources. It happens that links become invalid and users waste time hunting for data that might no longer be there. Storing the metadata generally is much easier and cheaper.

Hence, principle A2 states that metadata should persist even when the data are no longer sustained. A2 is related to the registration and indexing issues described in F4 (*(meta)data are registered or indexed in a searchable resource*).

Metadata are valuable in and of themselves. Even if the original data are missing, tracking down people, institutions or publications associated with the original research can be extremely useful.

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

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

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

Data should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings. The main goal of this principle is to provide a “common understanding” of digital objects by means of a language for knowledge representation to be used to represent these objects.

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

The chosen language
- should have a formal specification
- the knowledge representation language specifications should be shared and accessible so others can read the specifications and learn the language
- to support interoperability, the language should be designed to be used in more than one scenario.

Examples

The RDF extensible knowledge representation model is a way to describe and structure datasets. You can refer to the Dublin Core Schema as an example.

I2. (meta)data use vocabularies that follow FAIR principles (I)

If we use vocabularies in our data or metadata, we should make sure that they are also FAIR in their own right so that others, humans or machines, can find, access, interoperate and reuse them.
The controlled vocabulary used to describe datasets needs to be documented and resolvable using globally unique and persistent identifiers.
This documentation needs to be easily findable and accessible by anyone who uses the dataset.

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

A qualified reference is **a cross-reference that explains its intent:** *X is regulator of Y* is a much more qualified reference than *X is associated with Y*, or *X see also Y*.

The goal is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data.

To be more concrete, you should specify if one dataset builds on another data set, if additional datasets are needed to complete the data, or if complementary information is stored in a different dataset. In particular, the scientific links between the datasets need to be described.

No need to say, all datasets need to be properly cited.

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

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

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

Principle focuses on the ability of a user (machine or human) to decide if the data is actually **useful** in a particular context. To make this decision, the **data publisher should provide** not just metadata that allows discovery, but also **metadata that richly describes the context under which the data was generated**.
Moreover, R1 states that the data publisher should not attempt to predict the data consumer's identity and needs, even including information that may seem irrelevant.

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

Some points to take into consideration (non-exhaustive list):

- Describe the scope of your data: for what purpose was it generated or collected?
- Mention any particularities or limitations about the data that other users should be aware of.
- Specify the date of generation/collection of the data, who prepared the data, the parameter settings, the name and version of the software used.
- Is it raw or processed data?
- Clearly specify and document the version of the archived and/or reused data.

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

Under ‘I’, we covered elements of *technical* interoperability. R1.1 is about ***legal*** **interoperability**. What usage rights do you attach to your data? This should be described clearly. Ambiguity could severely limit the reuse of your Clarity of licensing status will become more important with automated searches involving more licensing considerations. The conditions under which the data can be used should be clear to machines and humans.

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

For others to reuse your data, they should know where the data came from (i.e., clear story of origin/history, see R1), who to cite and/or how you wish to be acknowledged.

Include a description of the workflow that led to your data:

- Who generated or collected it?
- How has it been processed? Has it been published before?
- Does it contain data from someone else that you may have transformed or completed?
- Ideally, this workflow is described in a machine-readable format.

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

It is easier to reuse data sets if they are similar: same type of data, data organised in a standardised way, well-established and sustainable file formats, documentation (metadata) following a common template and using common vocabulary.

If community standards or best practices for data archiving and sharing exist, they should be followed. FAIR data should at least meet those standards.

Note that quality issues are not addressed by the FAIR principles. The data's reliability lies in the eye of the beholder and depends on the intended application.

PERSISTENT IDENTIFIER (I)

A **persistent identifier** (**PI** or **PID**) is a long-lasting reference to a document, file, web page, or other object.
Over centuries, writers and scholars developed standards for ***citation*** of paper-based documents so that readers could reliably and efficiently find a source that a writer mentioned in a footnote or bibliography.
An important aspect of persistent identifiers is that "persistence is purely a matter of service“: thhat means that persistent identifiers are only persistent to the degree that someone commits to resolving them for users.

PERSISTENT IDENTIFIER (II)

Examples

Open Researcher and Contributor ID - ORCID (people)

Research Organization Registry - ROR (Organizations)

Virtual International Authority File - VIAF (Publications)

International Standard Book Number - ISBN (Publications)

Handle System (Uniform Resource Identifiers)

Digital Object Identifier - DOI, with 200 million DOIs issued (Uniform Resource Identifiers)

Persistent Uniform Resource Locators (PURLs) (Uniform Resource Identifiers)

PERSISTENT IDENTIFIER: HANDLE AND DOI (I)

The **Handle** System provides a general-purpose global name service enabling secure name resolution over the Internet, to identify digital content independent of location.

The **DOI** system utilises the Handle System as one component in building an added value application, for the persistent, semantically interoperable, identification of intellectual property entities.

PERSISTENT IDENTIFIER: HANDLE AND DOI (II)

The Handle System technology provides **persistence** if used with appropriate social infrastructure. The International DOI Foundation (IDF) builds on the technical infrastructure of the Handle System a social structure guaranteeing persistence. Persistence is a function of organizations, not of technology; a persistent identifier system requires a persistent organization and defined processes.

Handle System protocols ensure interoperability for resolution purposes among a diverse set of implementations. At the application level, there is no requirement that **consistent** rules must be in place for multiple applications. The DOI system adds such a requirement.

PERSISTENT IDENTIFIER: HANDLE AND DOI (III)

Handles (including DOI names) will be resolved by the Handle System, but there is **no requirement** in the Handle System for declaring what is being identified, or **for ensuring semantic interoperability** across several identified resources.

The DOI system adds this facility, specifically designed for its area of applications, which is now being implemented and will be a feature of advanced applications of the DOI system.

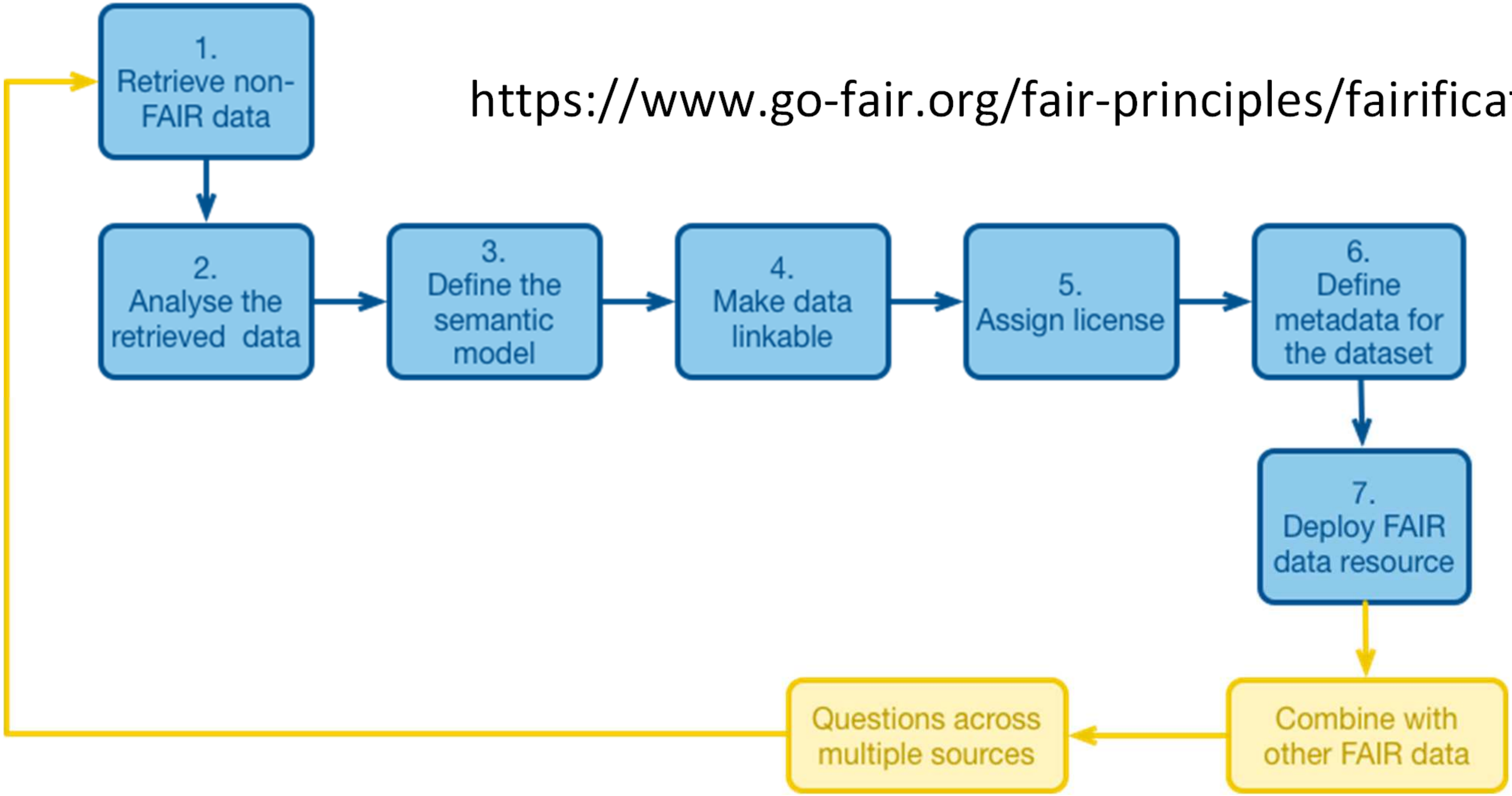
PERSISTENT IDENTIFIER: HANDLE AND DOI (IV)

The DOI system provides an optional tool to map an existing metadata scheme through a structured standard ontology, thereby ensuring semantic interoperability so that DOI names from different sources may be used as the key in building multi-component media objects or managing multiple assets.

The Handle system provides resolution of identifiers
The DOI system is an implementation on top of the Handle system that maintains a registry of metadata related to the identifiers

FAIRification (I)

https://www.go-fair.org/fair-principles/fairification-process/



FAIRification (II)

1. **Retrieve non-FAIR data:** gain access to the data to be FAIRified.
2. **Analyse the retrieved data:** inspect the content of the data: Which concepts are represented? What is the structure of the data? What are the relations between the data elements?

FAIRification (III)

3. Define the semantic model: define a 'semantic model' for the dataset, which describes the meaning of entities and relations in the dataset accurately, unambiguously, and in a computer-actionable way. Depending on the dataset, defining a proper semantic model may require a significant effort, even for experienced data modellers. A good semantic model should represent a consensus view in a particular domain, for a particular purpose. Therefore, it is good practice to search for existing models. Semantic models often contain multiple terms from existing ontologies and vocabularies. A vocabulary is a computer-readable file that captures terms, their URIs, and descriptions. An ontology can be roughly described as a vocabulary with hierarchies, meaningful relations among concepts, and their constraints. These conceptual models allow us to classify our data models and data items using the provided terms, concepts, and conceptual structures.

FAIRification (IV)

4.Make data linkable: The non-FAIR data can be transformed into linkable data by applying the semantic model defined in step 3. Currently, this is done using Semantic Web and Linked Data technologies. This step promotes interoperability and reuse, facilitating the integration of the data with other types of data and systems. However, the user should evaluate the feasibility of this step for the given data. It is a sensible thing to do for many types of data (e.g., structured data), but it may not be relevant for other types (e.g., the pixels or audio elements in images, audio data, and videos).

FAIRification (V)

5. Assign license: Although license information is part of the metadata, we have incorporated the **license assignment** as a separate step in the FAIRification process to highlight its importance. The absence of an explicit license may prevent others to reuse data, even if the data is intended to be open access.

6. Define metadata for the dataset: As explained by many of the **FAIR principles**, proper and rich metadata support all aspects of FAIR. (Read the GO FAIR **recommendation for metadata**.)

Deploy FAIR data resource: deploy or publish the FAIRified data, together with relevant metadata and a license, so that the metadata can be indexed by search engines and the data can be accessed, even if authentication and authorisation are required.

FAIRification (VI)

7. Deploy FAIR data resource: deploy or publish the FAIRified data, together with relevant metadata and a license, so that the metadata can be indexed by search engines and the data can be accessed, even if authentication and authorisation are required.

METADATA (I)

Metadata (or **metainformation**) is "data that provides information about other data", but not the content of the data itself, such as the text of a message or the image itself.

There are many distinct **types of metadata**, including:

Descriptive metadata – the descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.

Structural metadata – metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships, and other characteristics of digital materials.

METADATA (II)

Administrative metadata – the information to help manage a resource, like resource type, permissions, and when and how it was created.
Reference metadata – the information about the contents and quality of statistical data.
Statistical metadata– also called process data, may describe processes that collect, process, or produce statistical data.
Legal metadata – provides information about the creator, copyright holder, and public licensing, if provided.

METADATA (III)

Storage

Metadata can be stored either *internally*, in the same file or structure as the data (*embedded metadata*), or *externally*, in a separate file or field from the described data. A data repository typically stores the metadata *detached* from the data but can be designed to support embedded metadata approaches. Each option has advantages and disadvantages:

METADATA (IV)

Internal storage means metadata always travels as part of the data they describe; thus, metadata is always available with the data, and can be manipulated locally. This method creates redundancy (precluding normalization), and does not allow managing all of a system's metadata in one place. It arguably increases consistency, since the metadata is readily changed whenever the data is changed.

METADATA (V)

External storage allows collocating metadata for all the contents, for example in a database, for more efficient searching and management. Redundancy can be avoided by normalizing the metadata's organization. In this approach, metadata can be united with the content when information is transferred; or can be referenced (for example, as a web link) from the transferred content. On the downside, the division of the metadata from the data content, especially in standalone files that refer to their source metadata elsewhere, increases the opportunities for misalignments between the two, as changes to either may not be reflected in the other.

METADATA SCHEMATA (I)

A ***metadata standard*** or schema can be defined as:

a labeling, tagging or coding system used for recording cataloging information or structuring descriptive records. A metadata schema establishes and defines data elements and the rules governing the use of data elements to describe a resource.

METADATA SCHEMATA (II)

Structures

Metadata (metacontent) or, more correctly, the vocabularies used to assemble metadata (metacontent) statements, is typically structured according to a standardized concept using a well-defined metadata scheme, including **metadata standards** and **metadata models**. Tools such as *controlled vocabularies*, *taxonomies*, *thesauri*, *data dictionaries*, and *metadata registries* can be used to apply further standardization to the metadata. Structural metadata commonality is also of paramount importance in data model development and in database design.

METADATA SCHEMATA (III)

Syntax

Metadata syntax refers to the rules created to structure the fields or elements of metadata. A single metadata scheme may be expressed in a number of different markup or programming languages, each of which requires a different syntax. For example, Dublin Core may be expressed in plain text, HTML, XML, and RDF.

Granularity

The degree to which the data or metadata is structured is referred to as "granularity". "Granularity" refers to how much detail is provided. Metadata with a high granularity allows for deeper, more detailed, and more structured information and enables a greater level of technical manipulation

METADATA SCHEMATA (III)

Hierarchical, linear, and planar schemata

Metadata schemata can be **hierarchical** in nature where relationships exist between metadata elements and elements are nested so that parent-child relationships exist between the elements.

Metadata schemata can also be one-dimensional, or **linear**, where each element is completely discrete from other elements and classified according to one dimension only. An example of a linear metadata schema is the ***Dublin Core*** schema, which is one-dimensional.

Metadata schemata are often 2 dimensional, or **planar**, where each element is completely discrete from other elements but classified according to 2 orthogonal dimensions.

Examples in astronomy (I)

Flexible Image Transport System (**FITS**) is an open standard defining a digital file format useful for storage, transmission and processing of data: formatted as multi-dimensional arrays (for example a 2D image), or tables. FITS is the most commonly used digital file format in astronomy. The FITS standard was designed specifically for astronomical data, and includes provisions such as describing photometric and spatial calibration information, together with image origin metadata.

Astronomy Visualization Metadata (AVM) is a standard for tagging digital astronomical images stored in formats such as JPEG, GIF, PNG and TIFF. The AVM standard includes useful astronomical information about the subject of the image as well as the telescope used to take the image. This ensures that relevant information is transferred with the image when it is shared with others. AVM could be considered analogous to the FITS headers associated with raw astronomical data files.

Examples in astronomy (II)

Resource Metadata for the Virtual Observatory defines metadata terms and concepts necessary for discovery and use of astronomical data collections and services (based on DC)

Observation Data Model Core Components and its Implementation in the Table Access Protocol. It defines the core components of the Observation data model that are necessary to perform data discovery when querying data centers for astronomical observations of interest

DUBLIN CORE (I)

The **Dublin Core**, also known as the **Dublin Core Metadata Element Set** (**DCMES**), is a set of fifteen main metadata items for describing digital or physical resources.
The **Dublin Core Metadata Initiative** (**DCMI**) is responsible for formulating the Dublin Core; DCMI is a project of the Association for Information Science and Technology (ASIS&T), a non-profit organization.
Dublin Core has been formally standardized internationally as **ISO 15836** by the International Organization for Standardization (ISO) and as **IETF RFC 5013** by the Internet Engineering Task Force (IETF), as well as in the U.S. as **ANSI/NISO Z39.85** by the National Information Standards Organization (NISO).

DUBLIN CORE (II)

The core properties are part of a larger set of **DCMI Metadata Terms**. "Dublin Core" is also used as an adjective for Dublin Core metadata, a style of metadata that draws on multiple Resource Description Framework (RDF) vocabularies, packaged and constrained in Dublin Core application profiles.

The resources described using the Dublin Core may be digital resources (video, images, web pages, etc.) as well as physical resources such as books or works of art.

Dublin Core metadata may be used for multiple purposes, from simple resource description to combining metadata vocabularies of different metadata standards, to providing interoperability for metadata vocabularies in the linked data cloud and Semantic Web implementations.

DCMI Metadata Terms (I)

Contributor – "An entity responsible for making contributions to the resource".
Coverage – "The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant".
Creator – "An entity primarily responsible for making the resource".
Date – "A point or period of time associated with an event in the lifecycle of the resource".
Description – "An account of the resource".
Format – "The file format, physical medium, or dimensions of the resource".
Identifier – "An unambiguous reference to the resource within a given context".
Language – "A language of the resource".
Publisher – "An entity responsible for making the resource available".
Relation – "A related resource".
Rights – "Information about rights held in and over the resource".
Source – "A related resource from which the described resource is derived".
Subject – "The topic of the resource".
Title – "A name given to the resource".
Type – "The nature or genre of the resource".

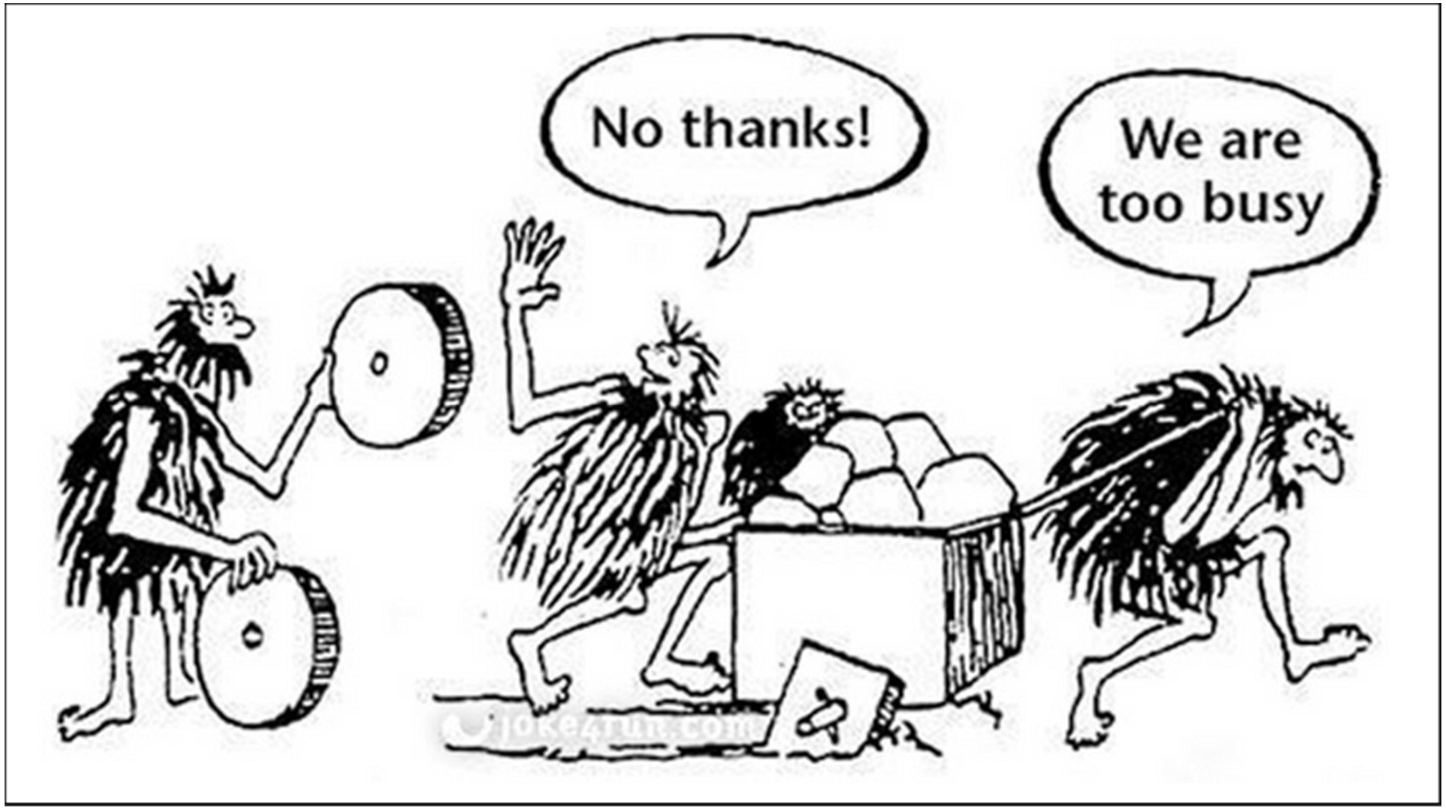
DCMI Metadata Terms (II)

abstract
accessRights
accrualMethod
accrualPeriodicity
accrualPolicy
alternative
audience
available
bibliographicCitation
conformsTo
created
dateAccepted
dateCopyrighted
dateSubmitted

educationLevel
extent
hasFormat
hasPart
hasVersion
instructionalMethod
isFormatOf
isPartOf
isReferencedBy
isReplacedBy
isRequiredBy
issued
isVersionOf
license

mediator
medium
modified
provenance
references
replaces
requires
rightsHolder
spatial
tableOfContents
temporal
valid

Open Science… and People



REFERENCES (1)

- GO FAIR Initiative https://www.go-fair.org/

    FAIR principles analysis https://www.go-fair.org/fair-principles/

    A framework guiding FAIRification https://www.go-fair.org/how-to-go-fair/

    FAIRification process (how to make data FAIR) https://www.go-fair.org/fair-principles/fairification-process/

- FAIR Data Points can be used to describe your data sets in a FAIR way, using standard metadata and make them available through simple WWW protocols. https://www.fairdatapoint.org/
- Metadata definition https://en.m.wikipedia.org/wiki/Metadata
- Dublin Core in brief https://en.wikipedia.org/wiki/Dublin_Core
- Dublin Core website https://www.dublincore.org/

REFERENCES (2)

- DCAT Data Catalog Vocabulary (DCAT) - Version 2
- Resource Metadata for the Virtual Observatory (IVOA), Dublin Core based https://www.ivoa.net/documents/RM/20070302/index.html
- General guidance, disciplinary metadata, Data Management Plan online and more by UK Digital Curation Center (DCC) https://dcc.ac.uk/guidance/standards/metadata
- https://dcc.ac.uk/guidance
- https://dcc.ac.uk/dmponline
- Research Data Alliance https://www.rd-alliance.org/
  - Metadata Standards Catalog
  - Guidelines for publishing structured data_V3.0_20210615.pdf

E (per ora) abbiamo finito, grazie!

francesca.martines@inaf.it