



FAIRsFAIR

Fostering Fair Data Practices in Europe



FAIRsFAIR Tools for Assessing FAIR

Hanna Koivula (CSC – IT Center for Science)

FAIRsFAIR WP2 Semantics, Interoperability, and Services

2021-06-02

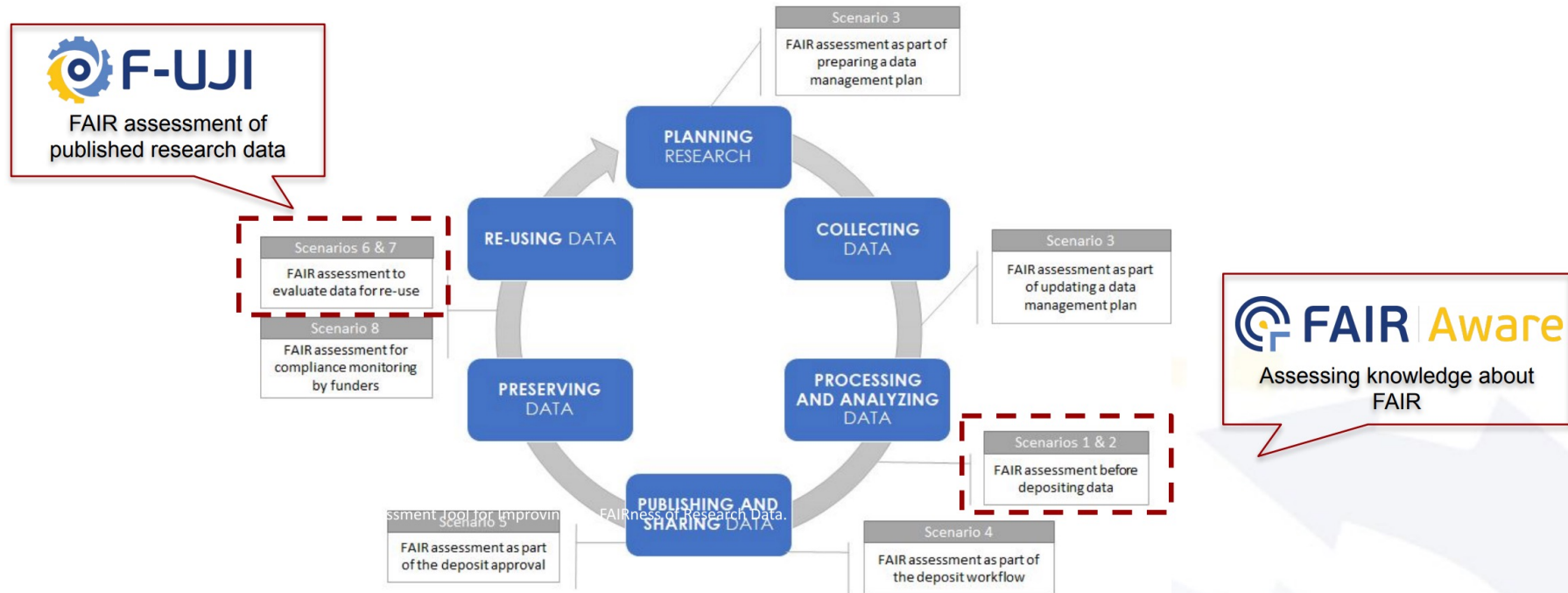


FAIRsFAIR Adopts the RDA FAIR Data Maturity Model Specification and Guidelines

FAIRsFAIR used the RDA [FAIR Data Maturity Model Specification and Guidelines Recommendation](#) of the [FAIR Data Maturity Model Working Group](#) as a basis to develop this set of minimum metrics for assessing the FAIRness of research data objects and tools to address researchers and data repositories.



FAIR and research data life cycle



Research data lifecycle; figure adapted from (Mosconi et al., 2019) and scenarios of FAIR assessment of datasets therein.

FAIR maturity of an infrastructure vs. a dataset?

Infrastructure

- F - Findability
- A – Accessibility
- I – Interoperability (Technical)
- R - Reusability (Technical)

Dataset / content

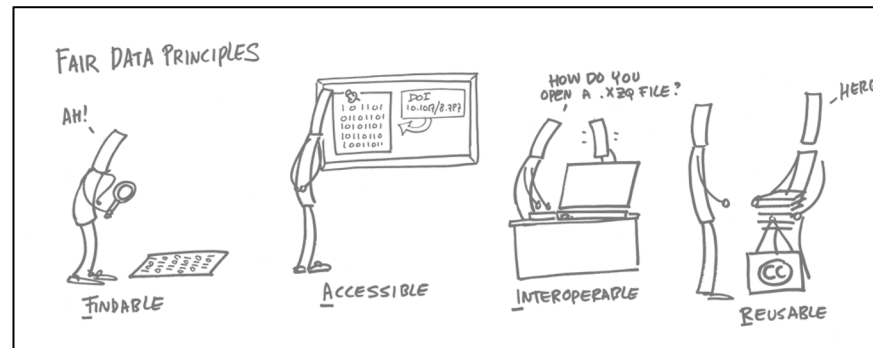
I – Interoperability

- Organisational
- Legal
- **Semantic**

R – Reusability (documentation)

Making data FAIR as a data producer

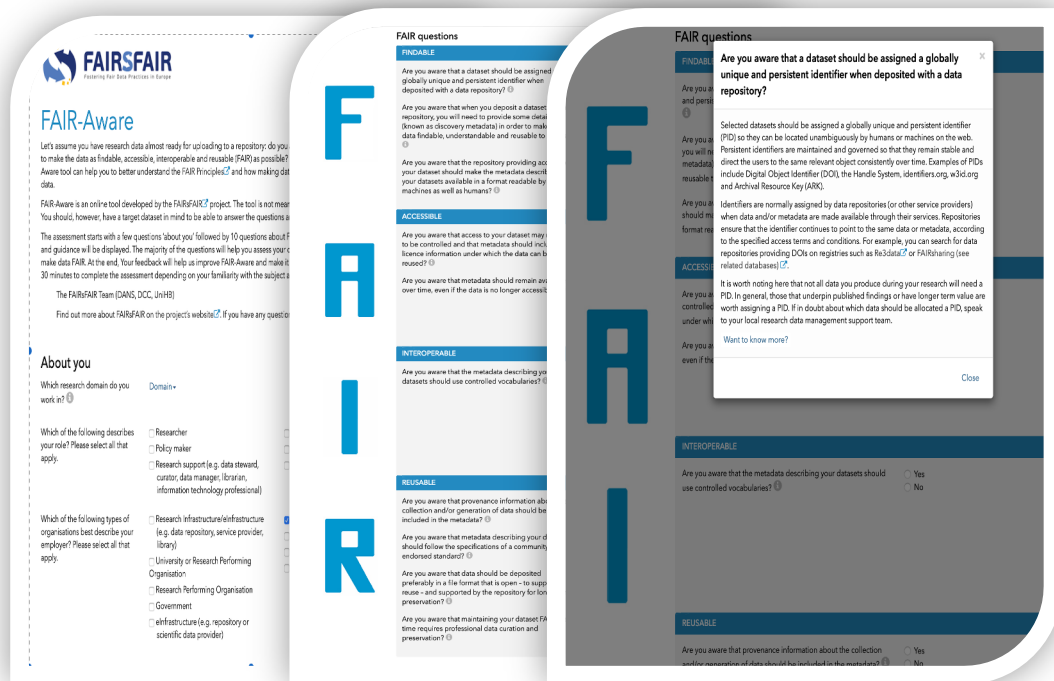
- Collaboration you ↔ repository
- Takes work!
- It's not an all-or-nothing principle
 - Any step towards FAIR reaps benefits
- It's not an one-size-fits-all solution
 - Discipline-specific **standards**
 - You might need **tailored support**
- So, what can **you** do?
 - Step 1: Educate yourself!



The FAIR-Aware tool




- For researchers and data stewards
- Online tool to raise awareness and educate on the FAIR data principles
- 10 simple questions with practical tips to improve data FAIRness before deposit



The FAIR-Aware tool

INTEROPERABLE

6. Are you aware that the metadata describing your datasets should use semantic vocabularies? 

6. Are you aware that the metadata describing your datasets should use semantic vocabularies?

Metadata should use semantic vocabularies so that the contents are unambiguous and can be interpreted automatically by machines. Ontologies, thesauri, and taxonomies are kinds of semantic vocabularies, and they come with different degrees of sophistication (for e.g. in their level of expressiveness, structure, and inferential power)

[Want to know more?](#)

Close

☒ Yes To what degree do you intend to comply with this?
☐ No Very likely ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 Very unlikely

Before using semantic vocabularies, you should establish the following:

- Whether you can find the vocabulary
- Whether you know who curates and makes the vocabulary available to other users
- Whether it is an nationally or internationally recognized vocabulary and if it is used extensively
- Whether it is available online and is open to other users

Repositories supporting your preferred semantic vocabulary can be found on registries like FAIRsharing.org. Below is a list of some repositories or look-up services for semantic resources (the list is not exhaustive):

- [Linked Open Vocabularies \(LOV\)](#)
- [OBO Foundry](#)
- [BioPortal](#)
- [Basel Register of Thesauri, Ontologies & Classifications \(BARTOC\)](#)
- [NERC Vocabulary Server](#)
- [Research Vocabulary Australia](#)
- [MMI Ontology Registry and Repository \(ORR\)](#)
- [Industrial Ontologies Foundry \(IOF\)](#)
- [CESSDA Vocabulary Service](#)

Close

A closer look

Question 8 (Reusability):

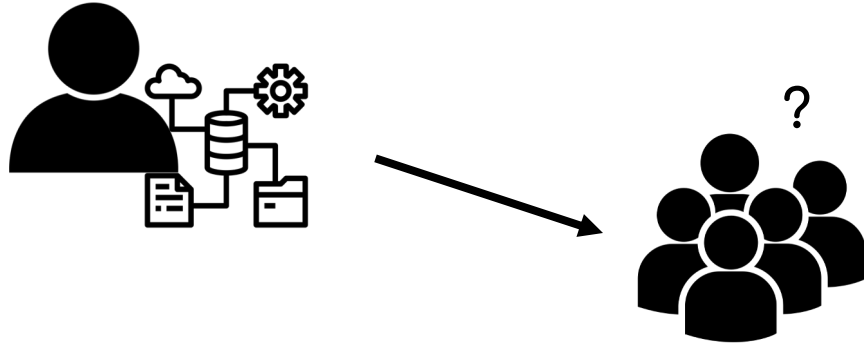
Are you aware that **metadata** describing your data should follow the specifications of a **community-endorsed standard**?



FAIR-Aware guidance

- Community-endorsed standards are important for communication and sharing within your domain
- Find a community-endorsed standard in a metadata registry: [RDA](#) | [DCC](#)
- Choose a repository that is specific to your domain or discipline on [Re3data](#)
- Is your domain lacking standards? (e.g., very limited, still in development)
 - Contact a research data management expert to identify possible solutions

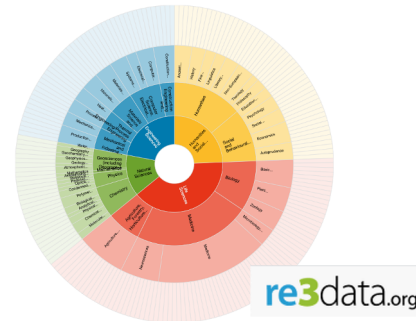
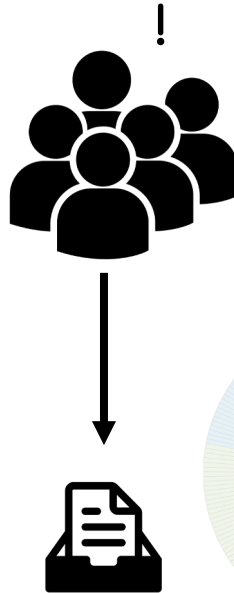
Life science example



Life science example



Life science example



Metadata

RDA | Metadata Directory

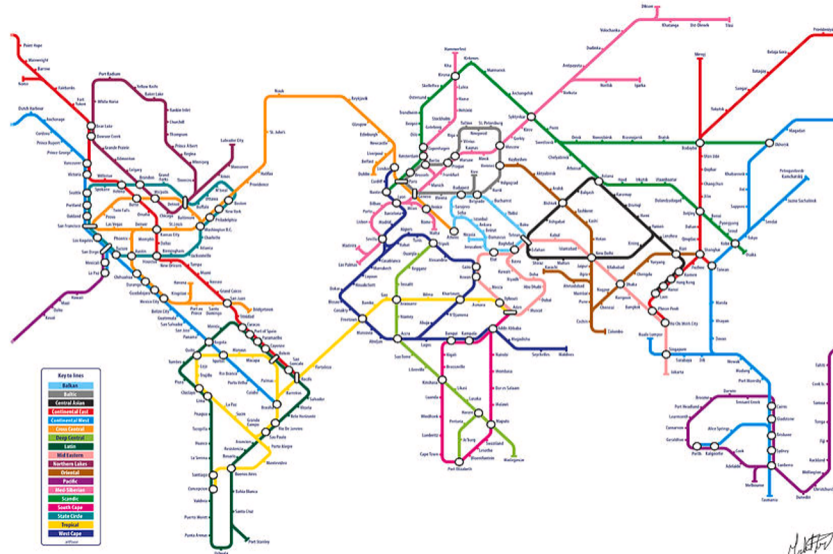


What can you do after today?

- Use the tool
- Share with others

→ <https://fairaware.dans.knaw.nl/>

Contribute to a better data infrastructure for the life sciences!



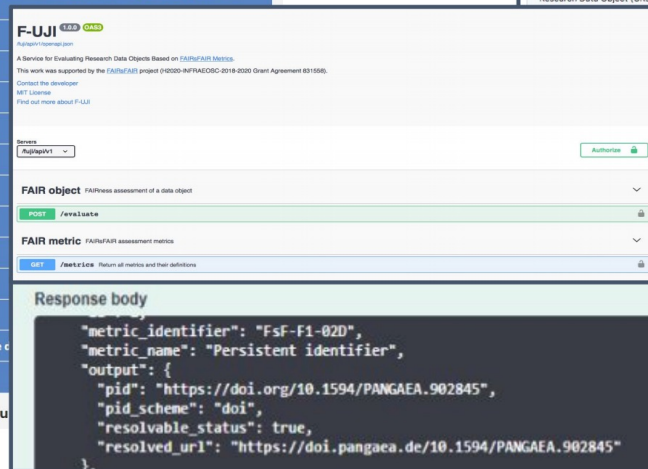
"World Metro Tube Map"
- Michael Tompsett

FAIR assessment for repositories – F-UJI-tool

¹While FAIR principles may apply to any digital objects, we are concerned with the subset of digital objects: research data that are collected, measured, or created for purposes of scientific analysis.

- ✓ FsF-F1-01D - Data is assigned a globally unique identifier
- ✓ FsF-F1-02D - Data is assigned a persistent identifier
- ✓ FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability
- ✓ FsF-F3-01M - Metadata includes the identifier of the data it describes
- ✓ FsF-F4-01M - Metadata is offered in such a way that it can be retrieved by machines
- ✓ FsF-A1-01M - Metadata contains access level and access conditions of the data
- ✓ FsF-A2-01M - Metadata remains available, even if the data is no longer available
- ✓ FsF-I1-01M - Metadata is represented using a formal knowledge representation language
- ✓ FsF-I1-02M - Metadata uses semantic resources
- ✓ FsF-I3-01M - Metadata includes links between the data and its related entities
- ✓ FsF-R1-01MD - Metadata specifies the content of the data
- ✓ FsF-R1.1-01M - Metadata includes license information under which data can be reused
- ✓ FsF-R1.2-01M - Metadata includes provenance information about data creation or generation
- ✓ FsF-R1.3-01M - Metadata follows a standard recommended by the target research community of the data
- ✓ FsF-R1.3-02D - Data is available in a file format recommended by the target research community

Please login & comment below citing in the subject line the Metric Identifier No. you




F-UJI

Automated FAIR Data Assessment Tool

Disclaimer: The test results shown here are based on preliminary data and code which still is under development. F-UJI is rapidly evolving and not yet available in a productive environment.

Research Data Object (URL/PID): OAI-PMH:

Results:

Force:

Reuse Practices in Research

<https://doi.org/10.17026/dans-xsw-kkeq>
 metrics_v0.4
<https://doi.org/10.5281/zenodo.4081213>
 v1.0.3



Findable: 6 of 7

Accessible: 2 of 3

Interoperable: 1 of 4

Reusable: 5 of 10

<https://github.com/pangaea-data-publisher/fuji>
<https://www.f-uji.net>

Metadata for machines and for humans

Embedded:

Accept:text/html

```
<!--BEGIN: Dublin Core description-->
<link rel="schema.DC" href="http://purl.org/dc/terms/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" content="Shell size variation in the North Sea" />
<meta name="DC.creator" content="Huber, Robert" />
<meta name="DC.creator" content="Meggers, Helmut" />
<meta name="DC.creator" content="Baumann, Karl" />
<meta name="DC.creator" content="Raymond, Maureen" />
<meta name="DC.creator" content="Henrich, Rüdiger" />
<meta name="DC.publisher" content="PANGAEA" />
<meta name="DC.source" content="Supplement to PANGAEA 2000-09-24" />
<meta name="DC.date" content="2000-09-24" />
```

Typed links:

Accept:text/html

```
<link rel="describedby" href="https://doi.pangaea.de/10.1593/PANGAEA2000-09-24" />
<link rel="describedby" href="https://doi.pangaea.de/10.1593/PANGAEA2000-09-24" />
<link rel="describedby" href="https://doi.pangaea.de/10.1593/PANGAEA2000-09-24" />
<link rel="item" href="https://doi.pangaea.de/10.1593/PANGAEA2000-09-24" />
<link rel="author" href="https://orcid.org/0000-0001-9146-0001" />
```

Content negotiation:

Accept:application/ld+json

```
{
  "@context": [Array[1]
    0: "http://schema.org"
  ],
  "@type": "Dataset",
  "provider": {
    "@type": "Organization",
    "name": "World Data Center for Climate (WDCC) at CERA",
    "logo": "https://cera-www.dkrz.de/WDCC/ui/cerasearch/logo.png",
    "url": "https://cera-www.dkrz.de/WDCC/ui/cerasearch"
  },
  "publisher": {
    "@type": "Organization",
    "name": "World Data Center for Climate (WDCC) at CERA"
  }
}
```

From FAIR Principle to Test

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

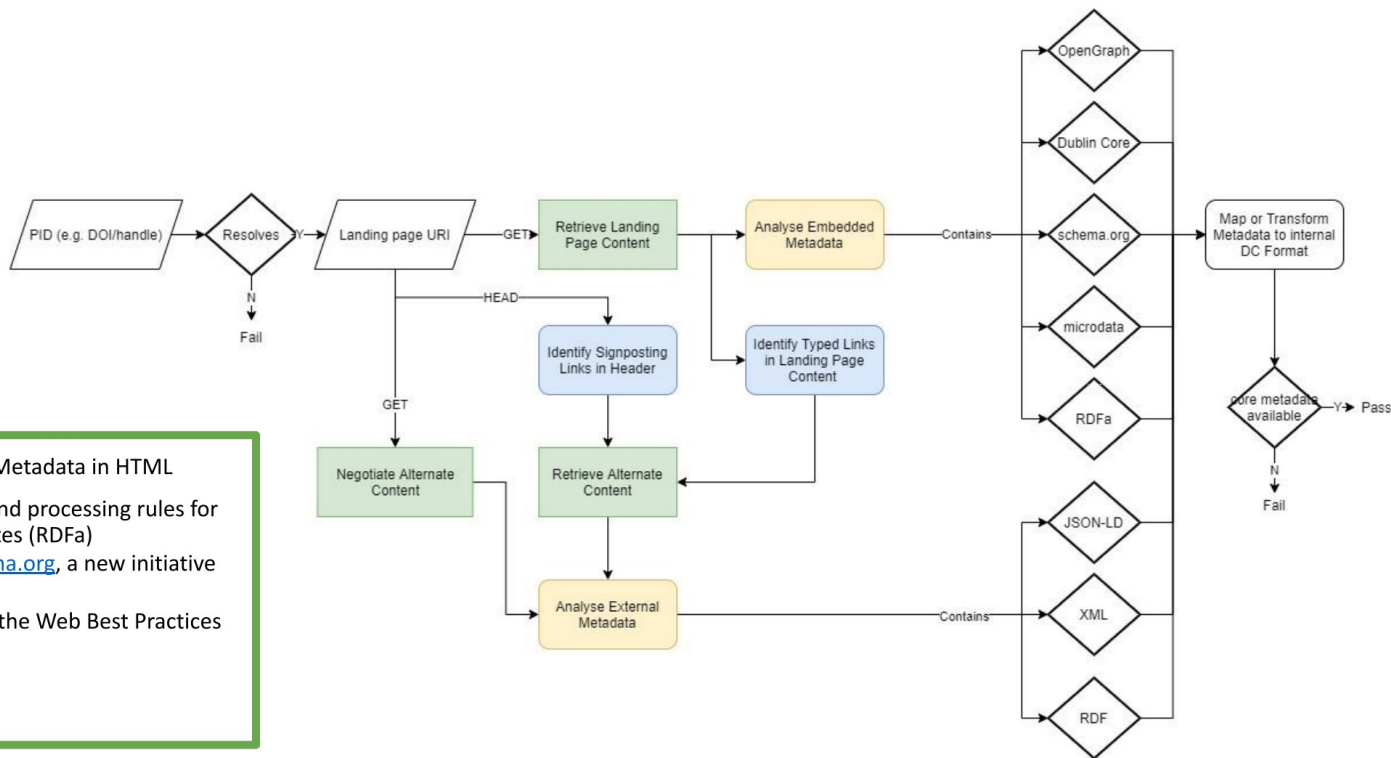
F2. data are described with rich metadata

- FsF-F3-01M - **Metadata** includes the identifier of the data it describes
- RDA-F3-01M **Metadata** includes the identifier for the data

Level:	Message:
INFO	Found data links in Schema.org metadata : [{"url": "https://doi.pangaea.de/10.1594/PANGAEA.893034?format=zip", "type": "application/zip"}]
INFO	Found data links in response header (signposting) : 1
INFO	Found data links in HTML head (link rel=item) : 1
INFO	Object identifier specified https://doi.org/10.1594/PANGAEA.893034
SUCCESS	Number of object content identifier found - 1



F-UJI: Metadata discovery



RFC2731: Encoding Dublin Core Metadata in HTML

W3C recommendation: Syntax and processing rules for embedding RDF through attributes (RDFa)

...Today we're announcing schema.org, a new initiative from Google, Bing and Yahoo!...

W3C recommendation: Data on the Web Best Practices

RFC5988: web linking

Signposting.org

Datcite

Domain agnostic metadata schemas

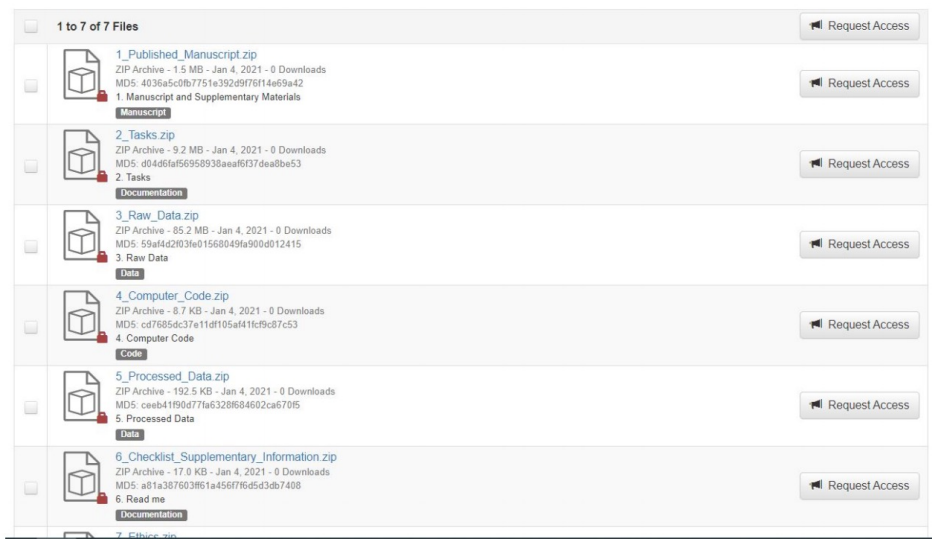
Widely used:

- Dublin Core (Metadata standard and vocabulary)
- Schema.org (Generic metadata vocabulary and schemas)
- DCAT (Generic data-cataloging metadata and vocabulary)
- (DataCite)



Some recommendations

- Avoid storing multiple unrelated data objects within one dataset
- Avoid storing additional metadata as part of a data set (e.g. pdf)
- Indicate access levels rather than hiding links for protected files



```
{"@type":"DataDownload","name":"7_Ethics.zip",  
"fileFormat":"application/zip","contentSize":3  
83347,"description":"7. Ethics Protocol"}}
```





<https://github.com/pangaea-data-publisher/fuji>

<https://www.f-uji.net>

