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

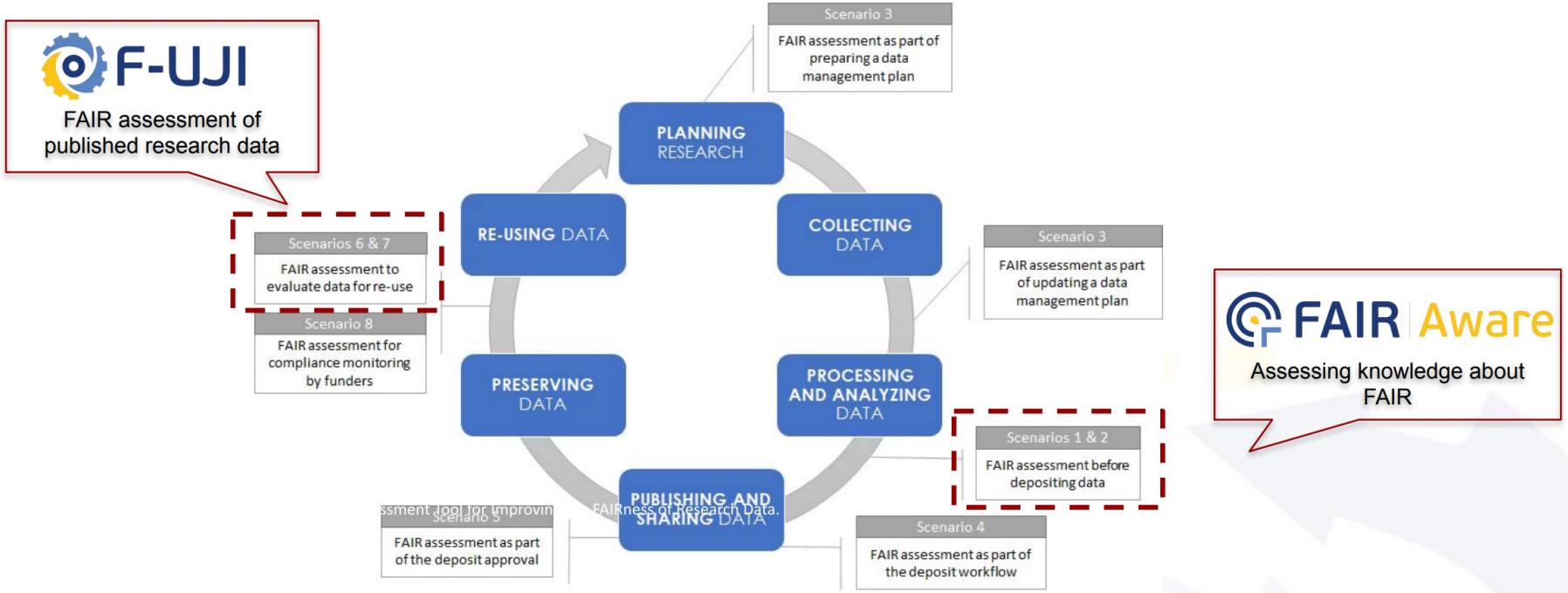
2021-06-02

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

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 inrfastructure vs. a dataset?

Infrastructure

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

Dataset / content

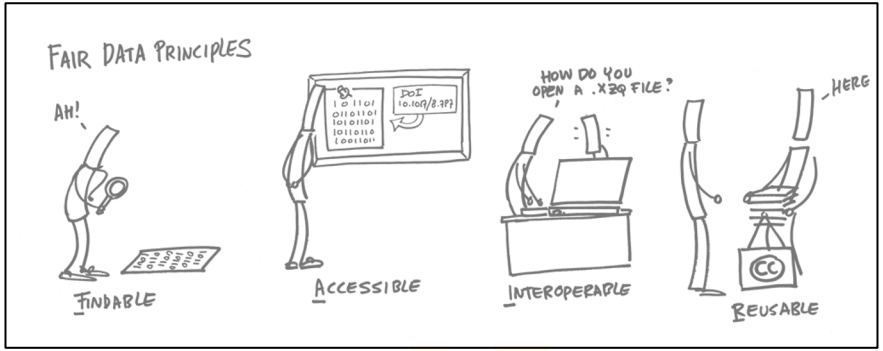
I – Interoperability

- Organisational
- Legal
- **Semantic**

R – Reusability (documentation)

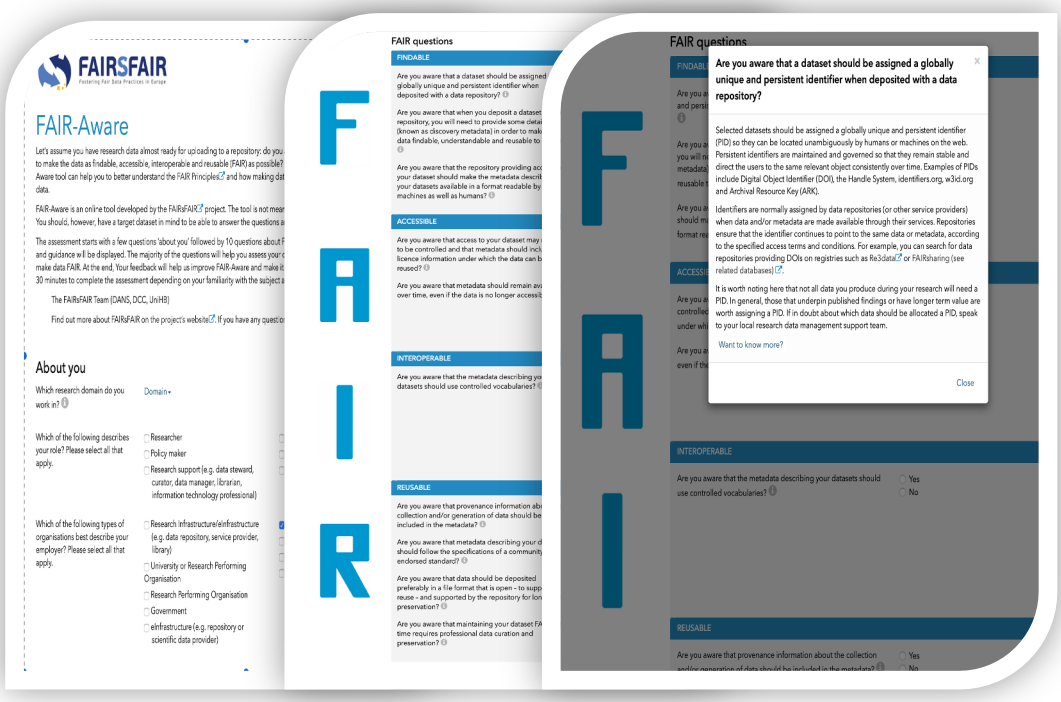
Making data FAIR as a data producer

- 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!



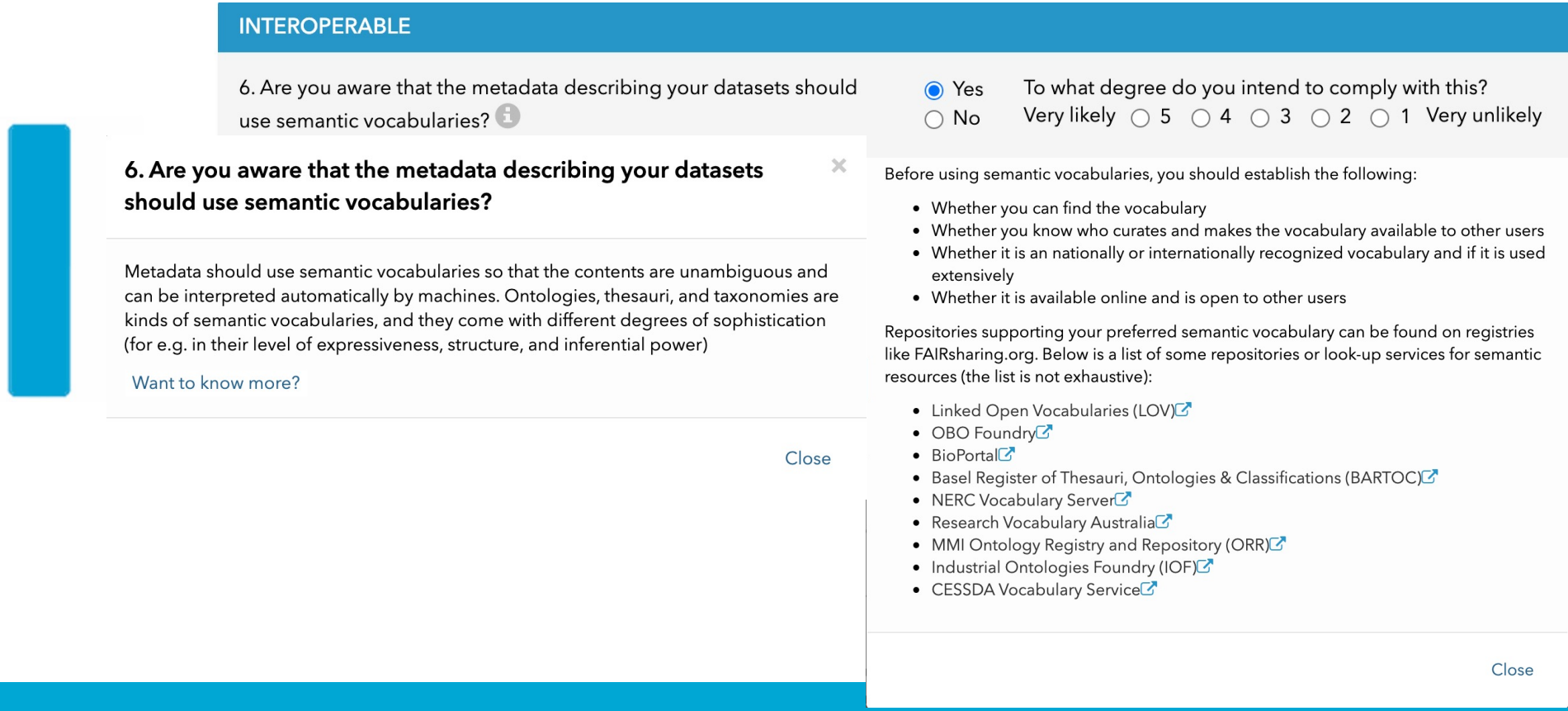
The FAIR-Aware tool

- 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



→ Check out at: **https://fairaware.dans.knaw.nl**

The FAIR-Aware tool



A closer look

Question 8 (Reusability):

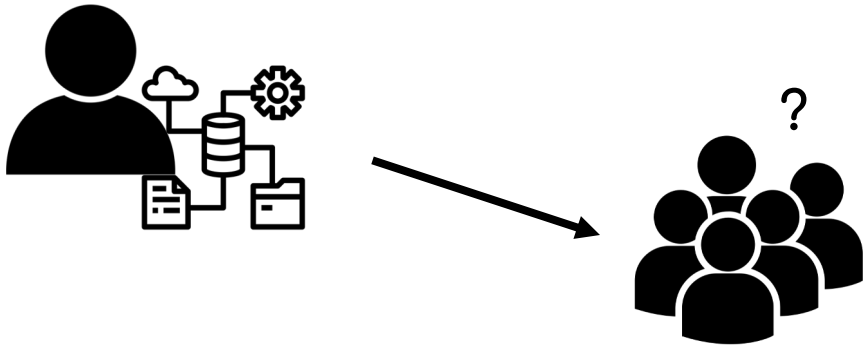
Are you aware that **metadata** describing your data should follow the specifications of a **community-endorsed standard**?

Images from: https://bit.ly/35lPlIK & https://bit.ly/2IvHGit Analogy from DDI: https://doi.org/10.5281/zenodo.4298302

FAIR-Aware guidance

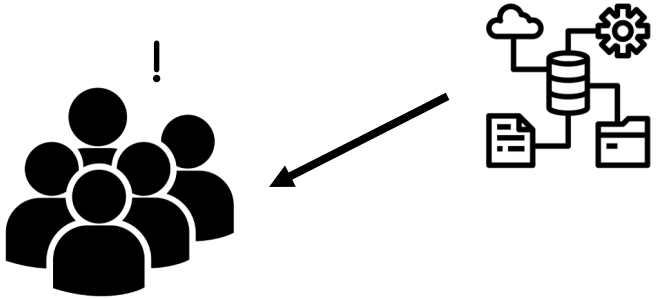
- Community-endorsed standards are important for communication and sharing within your domain
- Find a community-endorsed standard in a metadata registry: RDA | DCC
- Choose a repository that is specific to your domain or discipline on Re3data
- Is your domain lacking standards? (e.g., very limited,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

Life science example



Images from the Noun Project (Royyan Wijaya, ic2icon, Kevin, Alice Design)

Life science example

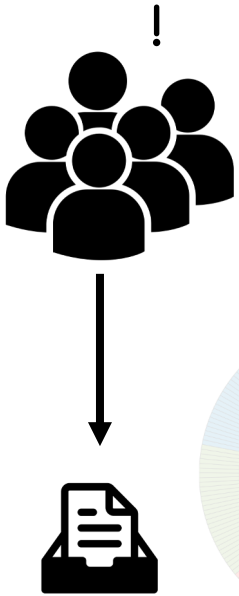


Metadata

RDA | Metadata Directory

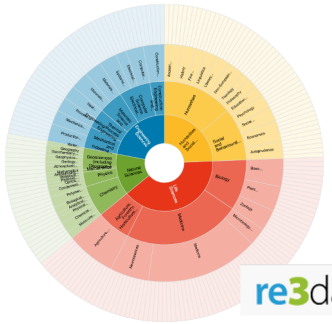
Images from the Noun Project (Royyan Wijaya, ic2icon, Kevin, Alice Design)

Life science example



Metadata

RDA | Metadata Directory

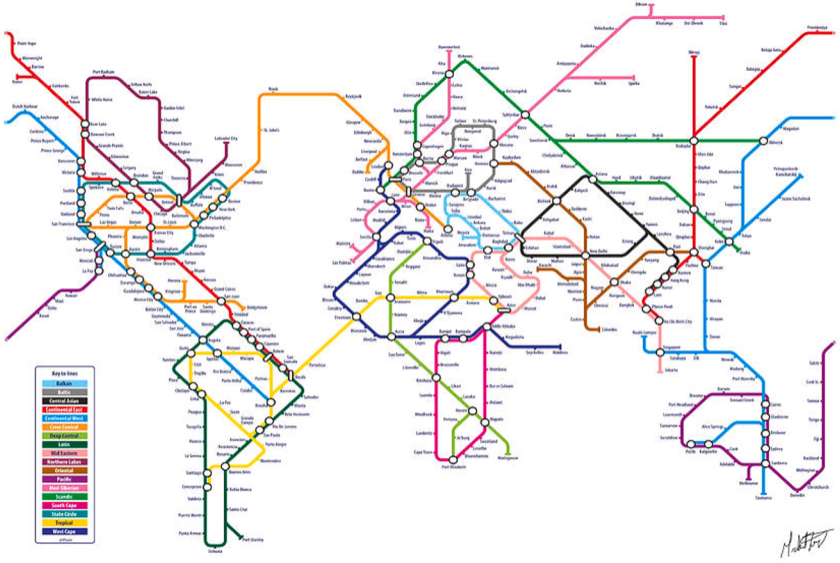


Images from the Noun Project (Royyan Wijaya, ic2icon, Kevin, Alice Design)

What can you do after today?

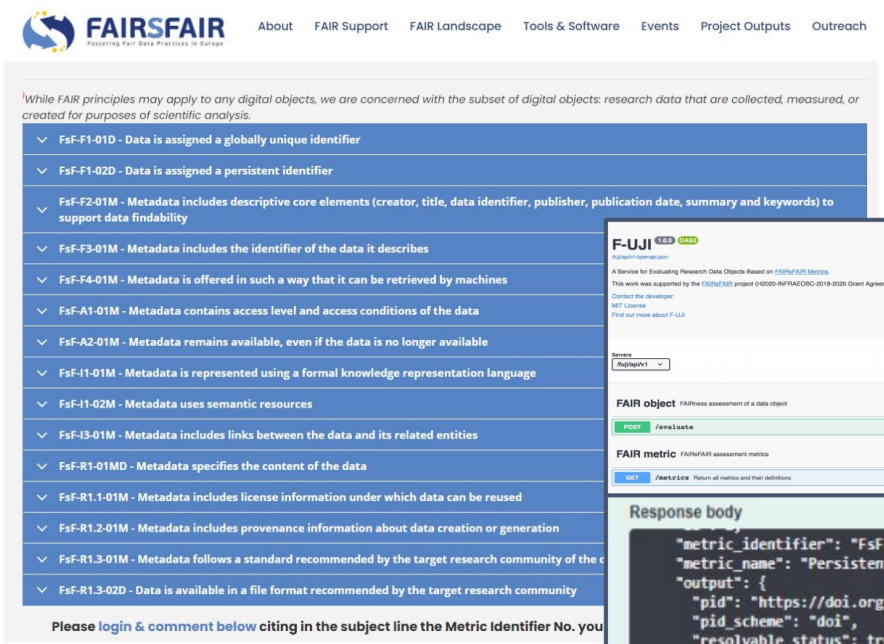
- Use the tool → https://fairaware.dans.knaw.nl/
- Share with others

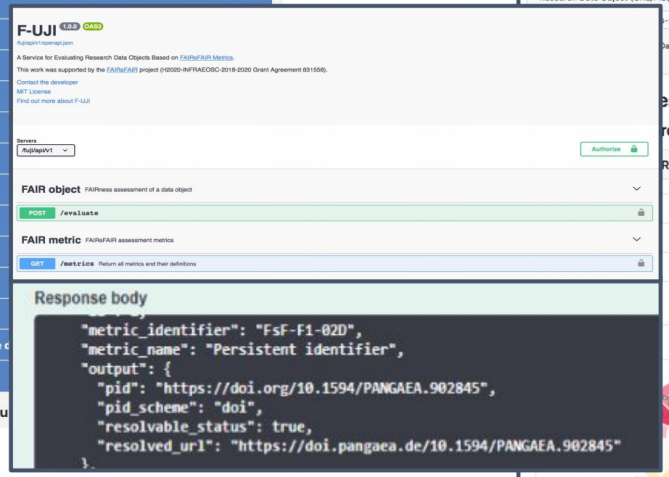
Contribute to a better data infrastructure for the life sciences!

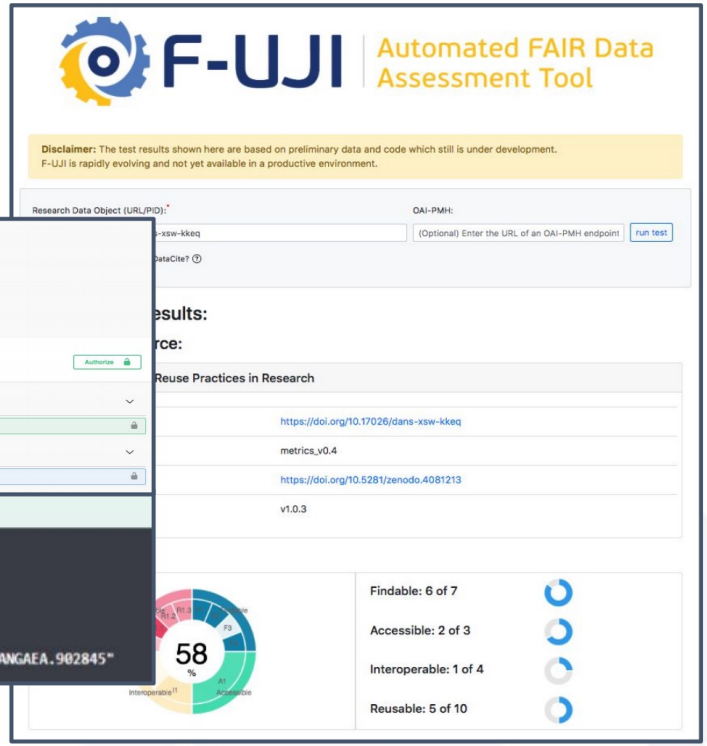


"World Metro Tube Map" - Michael Tompsett

FAIR assessment for repositories – F-UJI-tool







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/d
<link rel="schema.DCTERMS" href="http://purl.
<meta name="DC.title" content="Shell size var
<meta name="DC.creator" content="Huber, Rober
<meta name="DC.creator" content="Meggers, Hel
<meta name="DC.creator" content="Baumann, Kar
<meta name="DC.creator" content="Raymo, Maure
<meta name="DC.creator" content="Henrich, Rüd
<meta name="DC.publisher" content="PANGAEA" /
<meta name="DC.source" content="Supplement to
<meta name="DC.date" content="2000-09-24" sch
```

Typed links:

Accept:text/html

```
<link rel="describedby" href="https://doi.pangaea.
<link rel="describedby" href="https://doi.pangaea.
<link rel="describedby" href="https://doi.pangaea.
<link rel="item" href="https://doi.pangaea.de/10.1
<link rel="author" href="https://orcid.org/0000-00
```

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
      "logo": "https://cera-www.dkrz.de/WDCC/ui/cerasea
      "url": "https://cera-www.dkrz.de/WDCC/ui/cerasear
   },
  -"publisher": {
      "@type": "Organization",
      "name": "World Data Center for Climate (WDCC) at
```

From FAIR Principle to Test

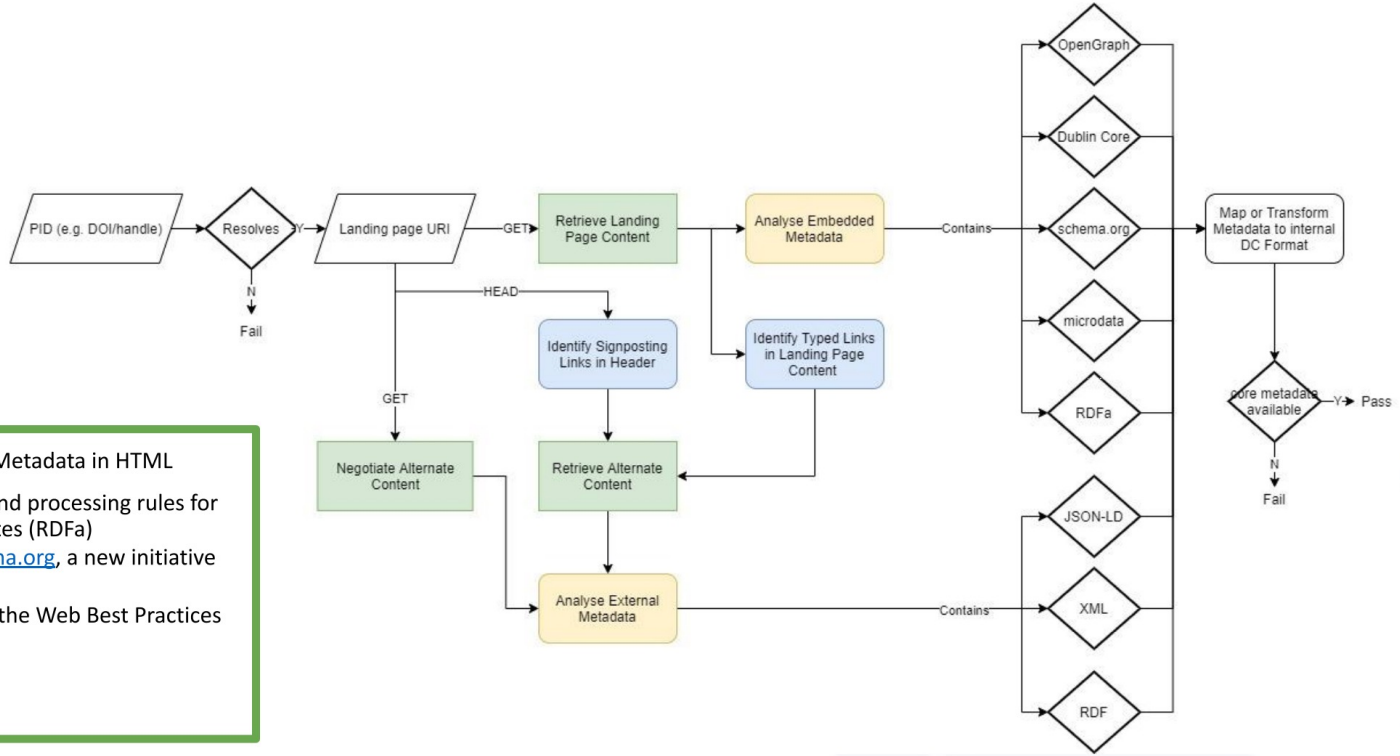
F3. metadata clearly and explicitly include the identifier of the data it describes

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
Datacite

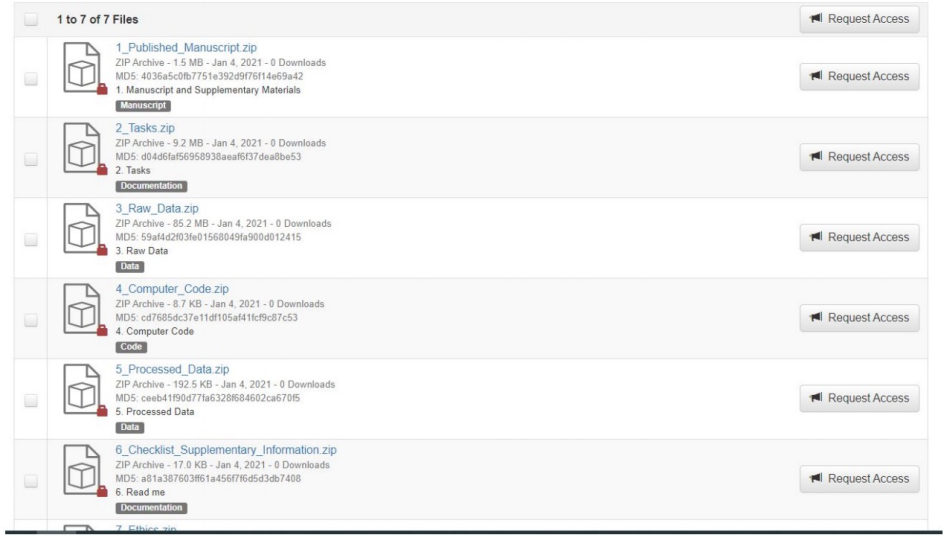
Domain agnostic metadata schemas

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