



FAIRSFair

Fostering Fair Data Practices in Europe

Domain agnostic metadata schemas (and how machines (FAIR assessment tools) can find them)



Robert Huber
(MARUM, Universität Bremen)
rhuber@uni-bremen.de

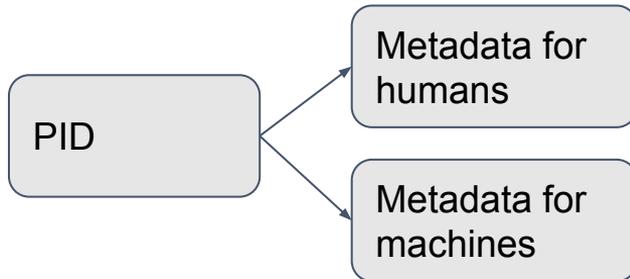
EOSC-Nordic FAIRification webinar, FAIR step 3
April 29, 2021



FAIR principles to be considered for domain agnostic metadata provision

FAIR is for machine & humans
And for data & metadata

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata



“The emphasis placed on FAIRness being applied to both human-driven and machine-driven activities, is a specific focus of the FAIR Guiding Principles that distinguishes them from many peer initiatives (discussed in the subsequent section).” (Wilkinson et al, 2016)

Metadata for machines and humans

PID resolves to a human readable landing page

- Embedding metadata in HTML page (landing page)
- Retrieving machine metadata using content negotiation
- Providing machine readable links in HTML (signposting, typed links)



Metadata for machines and 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 variations in the
<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 4242101" />
<meta name="DC.date" content="2000-09-24" />
```

Typed links:

Accept:text/html

```
<link rel="describedby" href="https://doi.pangaea.de/10.1017/S1522172400000000" />
<link rel="describedby" href="https://doi.pangaea.de/10.1017/S1522172400000000" />
<link rel="item" href="https://doi.pangaea.de/10.1017/S1522172400000000" />
<link rel="author" href="https://orcid.org/0000-0001-9146-3040" />
```

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

FAIRsFAIR - FAIR Data Assessments

- FAIR assessment implementation comprises the development of two main components – **assessment metrics** and **tool**.

Priority Recommendations

Rec. 8: Facilitate automated processing

Rec. 12: Develop metrics for FAIR Digital Objects

Supporting Recommendations

Rec. 25: Implement FAIR metrics to monitor uptake



European Commission Expert Group on FAIR Data. 2018. 'Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.'
<https://doi.org/10.2777/1524>

F-UJI -FAIRsFAIR FAIR metric assessment tool

While 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 FAIRsFAIR

A Service for Evaluating Research Data Objects Based on FAIRsFAIR Metrics.
 This work was supported by the FAIRsFAIR project (101020-NF14E0SC-2018-2020 Grant Agreement 631558).

Contact the developer
 GET License
 Find out more about F-UJI

Server:

FAIR object FAIRsFAIR assessment of a data object

FAIR metric FAIRsFAIR assessment metrics

Response body

```

{
  "metric_identifier": "FsF-F1-02D",
  "metric_name": "Persistent identifier",
  "output": {
    "pid": "https://doi.org/10.1594/PANGAEA.902845",
    "pid_scheme": "doi",
    "resolvable_status": true,
    "resolved_url": "https://doi.pangaea.de/10.1594/PANGAEA.902845"
  }
}

```



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):* DAI-PMH:

dataCite?

Results:

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



58%

Interoperable¹

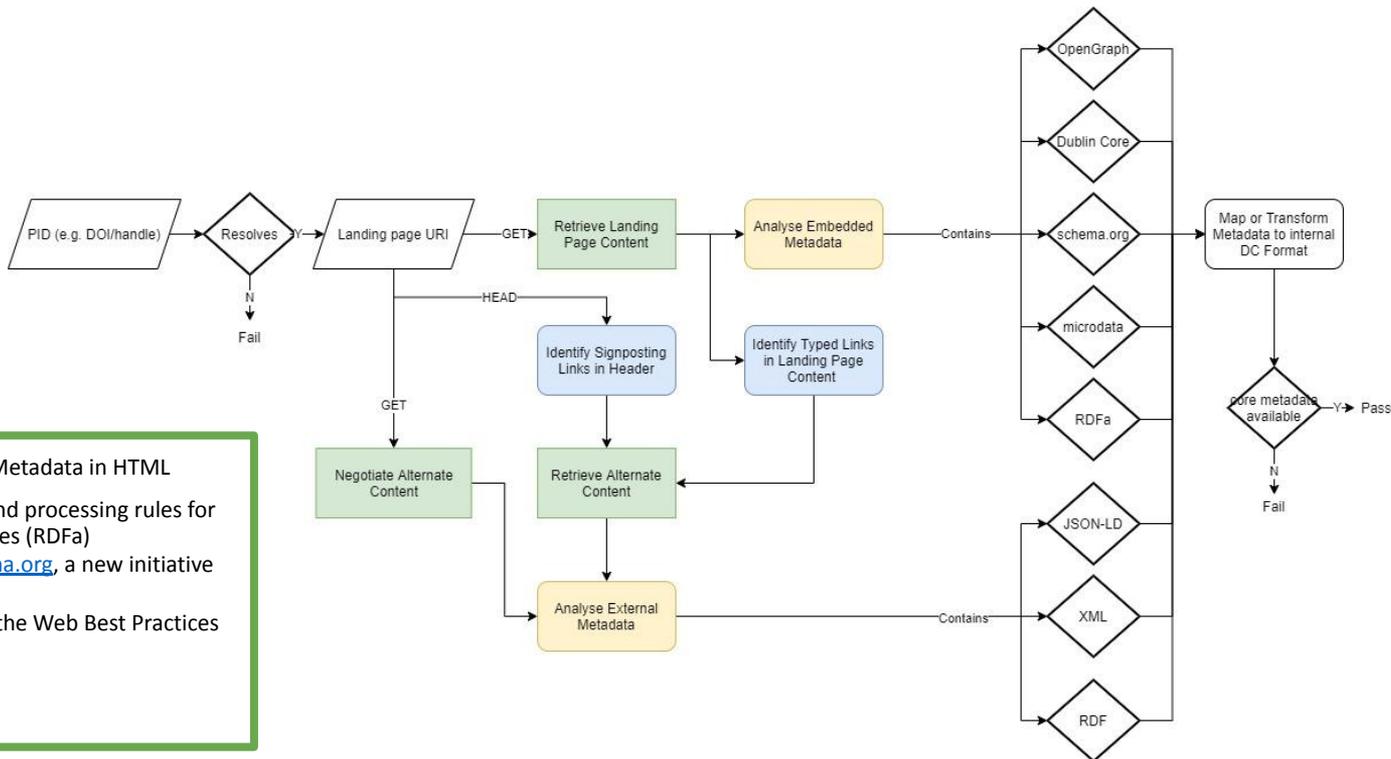
Findable: 6 of 7 

Accessible: 2 of 3 

Interoperable: 1 of 4 

Reusable: 5 of 10 

F-UJI: metadata discovery (F2)



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
- DCAT
- Schema.org
- (DataCite schema)



Dublin Core



- Metadata standard and vocabulary
- Can be expressed in:
 - XML (schema ...)
 - RDF
 - Microdata, RDFa
 - XHTML (`<meta name="DC.element" content="Value" />`)
- Protocol / Interface:
 - OAI-PMH
 - SPARQL
- Pro:
 - embedding in landing page
- Con:
 - Very generic, hard to link data

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:title>Homage to Catalonia,</dc:title>
    <dc:creator>Orwell, George,1903-1950.</dc:creator>
    <dc:type>text</dc:type>
    <dc:publisher>London, Secker and Warburg</dc:publisher>
    <dc:date>[1938]</dc:date>
    <dc:language>eng</dc:language>
  </rdf:Description>
</rdf:RDF>
```

ENDORSED BY:



Schema.org

schema.org

- Generic metadata vocabulary and schemas (e.g. Dataset)
- Can be expressed in:
 - RDF
 - JSON-LD
- Protocol / Interface:
 - SPARQL
 - Google Dataset Search ;)
- Pro:
 - Suitable for embedding in landing page (human & machine)
 - SEO included
- Con:
 - Difficult to validate

```
["@context": "http://schema.org/",
"@id": "https://doi.org/10.1594/PANGAEA.907686",
"@type": "Dataset",
"identifier": "https://doi.org/10.1594/PANGAEA.907686",
"url": "https://doi.pangaea.de/10.1594/PANGAEA.907686",
"creator": [
  {
    "@id": "https://orcid.org/0000-0003-3000-0020",
    "@type": "Person",
    "name": "Robert Huber",
    "familyName": "Huber",
    "givenName": "Robert",
    "identifier": "https://orcid.org/0000-0003-3000-0020",
    "email": "rhuber@uni-bremen.de"
  },
  {
    "@type": "Person",
    "name": "Robert Darga",
    "familyName": "Darga",
    "givenName": "Robert"
  },
  {
    "@type": "Person",
    "name": "Hans Lauterbach",
    "familyName": "Lauterbach",
    "givenName": "Hans"
  }
],
"name": "Roundness of grains from sediments from the surroundings of Lake Tüttensee (Bavaria, Germany)",
"publisher": {
  "@type": "Organization",
  "name": "PANGAEA",
  "disambiguatingDescription": "Data Publisher for Earth & Environmental Science",
```

ENDORSED BY:



DCAT (v2)

- Generic data cataloging metadata vocabulary
- Can be expressed in:
 - RDF
 - JSON-LD
- Protocol / Interface:
 - SPARQL
- Pro:
 - Designed for data catalogs
 - Uses Dublin Core etc.
 - Extensible
 - Widely used incl. Open Data (gov)
 - Accepted by Google Dataset Search
- Con:
 - Difficult to validate

DCAT

ENDORSED BY:





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

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