



FAIRSFAR
Fostering Fair Data Practices in Europe

Some recommendations for the practical, machine-friendly implementation of the FAIR F3 principle

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

EOSC-Nordic FAIRification webinar, FAIR step 2
Feb. 3, 2021



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

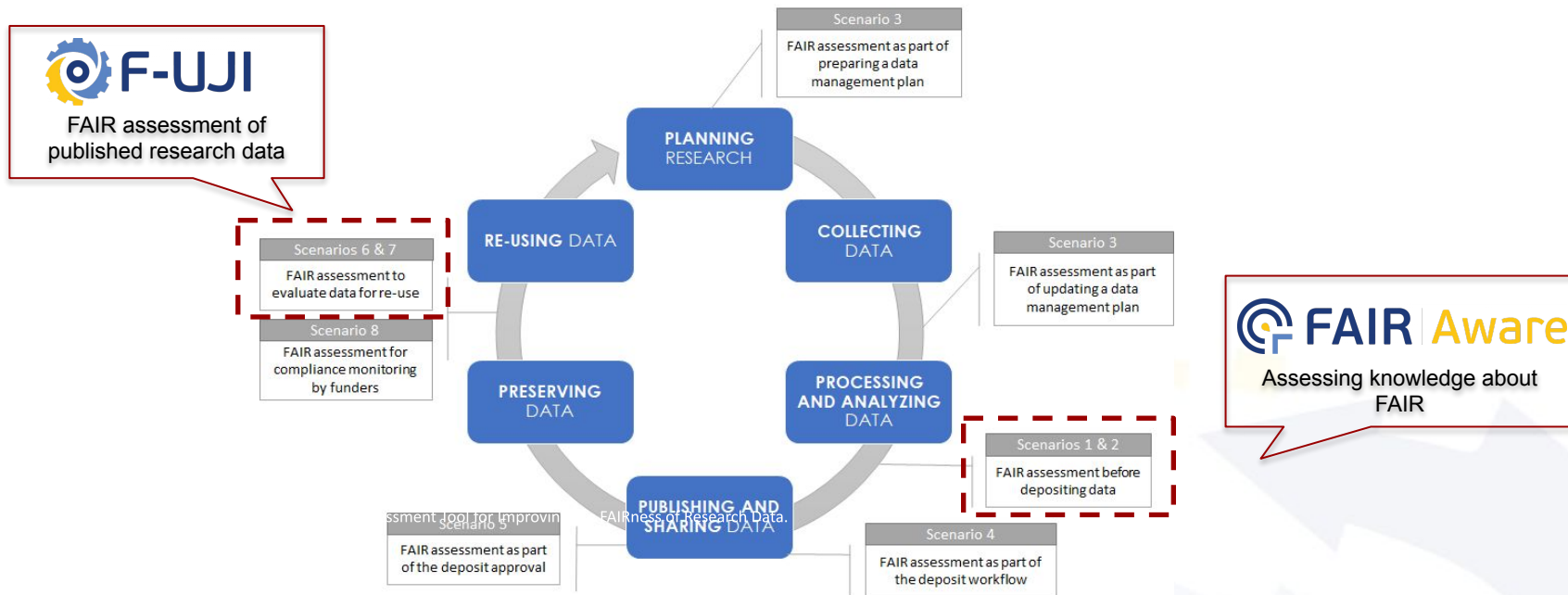


European Commission Expert Group on FAIR Data. 2018. 'Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.'

<https://doi.org/10.2777/1524>

Assessment Scenarios

For more information, see D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories,
<https://doi.org/10.5281/zenodo.3678715>




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

F-UJI -FAIRsFAIR FAIR metric assessment 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 FAIR metric assessment tool

A Service for Evaluating Research Data Objects Based on FAIRsFAIR Metrics.
 This work was supported by the FAIRsFAIR project (101000-NPFAEORC-2018-2020 Grant Agreement 831558).
 Contact the developer
 Get License
 Find out more about F-UJI

Search

FAIR object FAIRsFAIR assessment of a data object

POST /eval/instance

FAIR metric FAIRsFAIR assessment metrics

GET /metrics Return all metrics and their definitions

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/PIID):¹

OAI-PMH:

(Optional) Enter the URL of an OAI-PMH endpoint

DataCite?

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>

From FAIR principle to test

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

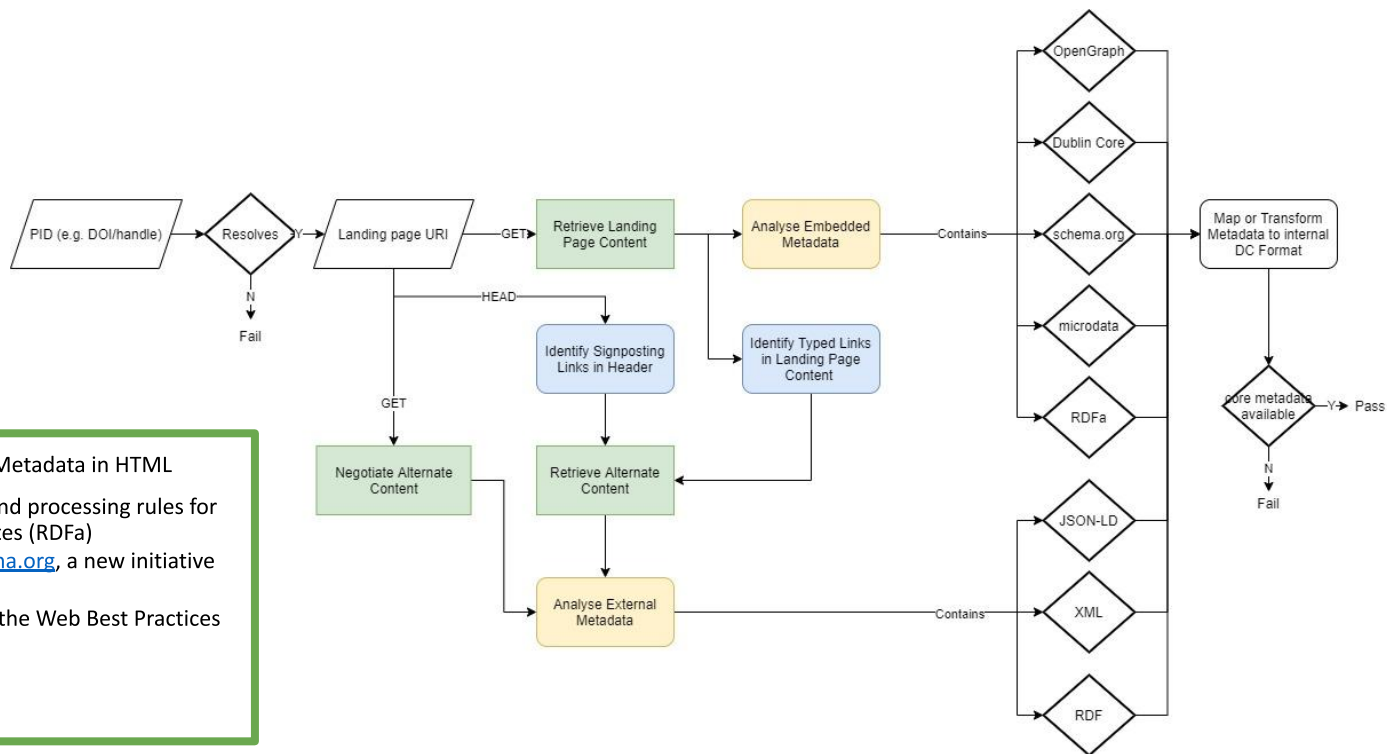
F2. data are described with rich metadata

- 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 (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

Common ways to include data links

- Typed links, signposting links
- Schema.org
- DataCite
- DCAT (RDF)
- Domain specific formats



HTML embedded - Typed links, signposting

RFC5988: web linking; signposting.org

Links to other representations of metadata or items (HTML):

```
<html><head>  
    <link rel="[relation type]" src="[URI]" type="[mime type]">  
</head></html>
```

Relation types:

- **item**: link to data download (quite frequent)
- **alternate**: link to metadata (DCAT, CKAN)
- **(meta**: link to metadata (rare, DDI))
- **describedby**: signposting



Typed links, signposting

signposting.org (HTTP Header):

HTTP/1.1 200 OK

Content-length: 8424

Content-type: text/html; charset=UTF-8

Link: <https://doi.pangaea.de/10.1594/PANGAEA.867908?format=citation_bibtex>

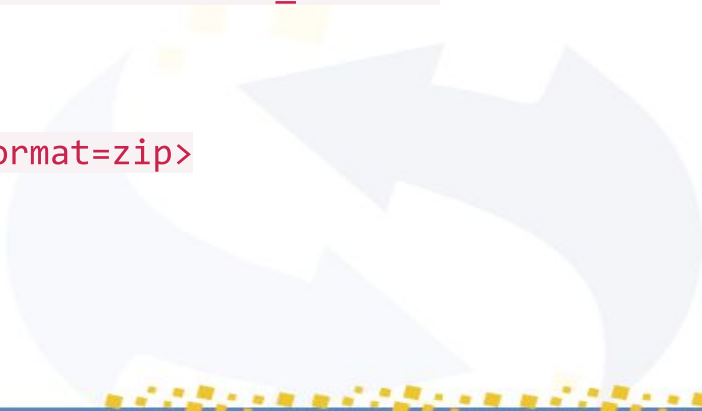
; rel="describedby"

; type="application/x-bibtex",

<<https://doi.pangaea.de/10.1594/PANGAEA.867908?format=zip>>

; rel="item"

; type="application/zip",



Schema.org metadata

Schema.org/Dataset

Usage:

- Embedded in HTML
- Content negotiation
- distribution property:

```
"distribution": [  
  {  
    "@type": "DataDownload",  
    "encodingFormat": "text/html",  
    "contentUrl": "https://doi.pangaea.de/10.1594/PANGAEA.875642?format=html"  
  }  
]
```

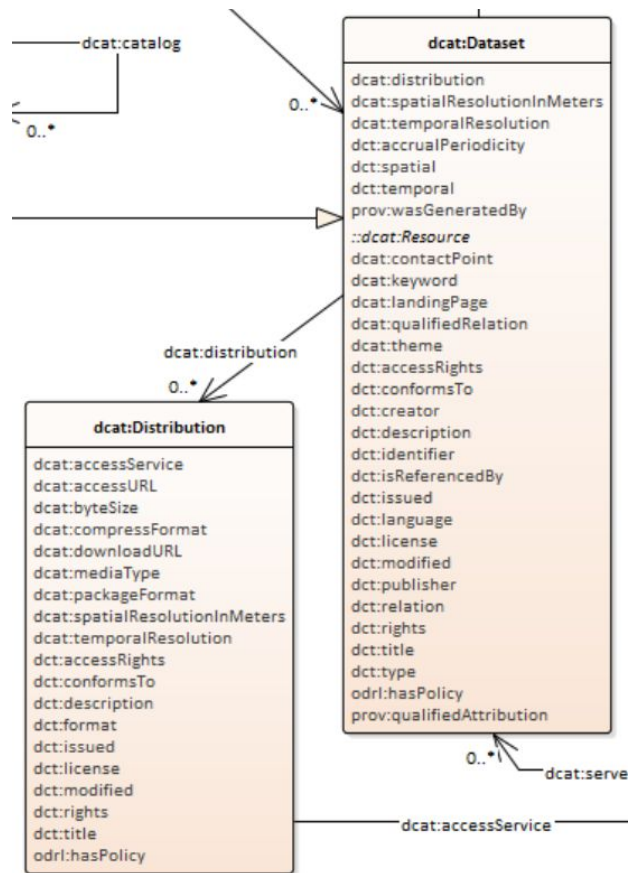
RDF metadata (DCAT)

Usage:

- Embedded, RDFa
- Content negotiation
- Typed or signposting links
- dcat:Distribution

```

:dataset-001-csv
  a dcat:Distribution ;
  dcat:downloadURL <http://www.example.org/files/001.csv> ;
  dct:title "CSV distribution of imaginary dataset 001"@en
;
  dcat:mediaType
<http://www.iana.org/assignments/media-types/text/csv> ;
  dcat:byteSize "5120"^^xsd:decimal ;
  
```



DataCite

Not yet in the official schema, but...

Also has a property **contentUrl**

```
{  
  data:  
  {  
    id: "10.1594/yourDOI",  
    type: "dois",  
    attributes:  
    {  
      contentURL: "someDataLink",  
    }  
  }  
}
```



Content negotiation - Domain specific XML

Usage:

- Content negotiation
- Typed links

Examples:

- ISO19139
- EML
- DDI
- Etc, etc etc..

Not yet (fully) implemented in F-UJI



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

1 to 7 of 7 Files		Request Access
<input type="checkbox"/>	<div>1_Published_Manuscript.zip ZIP Archive - 1.5 MB - Jan 4, 2021 - 0 Downloads MD5: 4036a5c0b7751e39249f76f14e69a42 1. Manuscript and Supplementary Materials Manuscript</div>	Request Access
<input type="checkbox"/>	<div>2_Tasks.zip ZIP Archive - 9.2 MB - Jan 4, 2021 - 0 Downloads MD5: d04d5fa56958938aaaf637dea8be53 2. Tasks Documentation</div>	Request Access
<input type="checkbox"/>	<div>3_Raw_Data.zip ZIP Archive - 85.2 MB - Jan 4, 2021 - 0 Downloads MD5: 59af4d2f03fe01568049fa900d012415 3. Raw Data Data</div>	Request Access
<input type="checkbox"/>	<div>4_Computer_Code.zip ZIP Archive - 8.7 KB - Jan 4, 2021 - 0 Downloads MD5: cd7685dc37e11df105af41fc9c87c53 4. Computer Code Code</div>	Request Access
<input type="checkbox"/>	<div>5_Processed_Data.zip ZIP Archive - 192.5 KB - Jan 4, 2021 - 0 Downloads MD5: ceeb41f90d77fa6328f68460ca67015 5. Processed Data Data</div>	Request Access
<input type="checkbox"/>	<div>6_Checklist_Supplementary_Information.zip ZIP Archive - 17.0 KB - Jan 4, 2021 - 0 Downloads MD5: a81a387603f61a45687f6d5d3db7408 6. Read me Documentation</div>	Request Access
<input type="checkbox"/>	<div>7_Ethics.zip</div>	

```
{"@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>