Assessing the FAIRness of your content

Robert Huber, PANGAEA - University Bremen
Task 4.5 – FAIR Data Assessments: Pilots

• 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

https://doi.org/10.2777/1524
Research data lifecycle; figure adapted from (Mosconi et al., 2019) and scenarios of FAIR assessment of datasets therein.

For more information, see D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories, https://doi.org/10.5281/zenodo.3678715
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-F1-01M - Metadata is represented using a formal knowledge representation language
- FsF-F1-02M - Metadata uses semantic resources
- FsF-F3-01M - 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 are referring to - e.g., "FsF-R1.3-01M"

Object Assessment Metrics V0.4

We would love to hear your feedback!

FAIR Principle

F1: (Meta) data are assigned globally unique and persistent identifiers

Metric

Data is assigned a persistent identifier.

Practical Test

Identifier is based on persistent identification scheme.

Identifier is resolved.

....

....

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

https://www.f-uji.net
High Level Flow (Data Gathering)

Identifier (e.g., URL, PID), OAI-PMH endpoint (optional)

Parse request

Extract metadata from the data page based on the identifier provided.

Is OAI-PMH endpoint provided?

Content-negotiation

Extract metadata standards via the endpoint

Metadata at the repository-level

Is a persistent identifier?

Collate metadata of the object

Extract repository metadata (api, metadata standards) through re3data

Retrieve metadata from PID provider (datacite)

Collate metadata of the object

Metadata at the object-level
F-UJI – An Automated FAIR Data Assessment Tool

FAIRSFAIR

F-UJI 1.0.0 OASIS

A Service for Evaluating Research Data Objects Based on FAIRSFAIR Metrics.

This work was supported by the FAIRSFAIR project (F2020-INFRAEDU-2018-2020 Grant Agreement 831536).

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

FAIR object
FAIRness assessment of a data object

FAIR metric
FAIRSFAIR assessment metrics

POST /evaluate

GET /metrics
Return all metrics and their definitions
F-UJI – An Automated FAIR Data Assessment Tool

https://www.f-uji.net
## Testing & Consultation – the pilots

<table>
<thead>
<tr>
<th>Repository</th>
<th>Certification</th>
<th>Subject Areas</th>
<th>Datasets Evaluated (as of 25.09.2020)</th>
<th>FAIR Data Improvement</th>
<th>Repository Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>PANGAEA</td>
<td>CoreTrustSeal, WDS Regular Member</td>
<td>Earth and Environmental Science</td>
<td>500 Completed</td>
<td></td>
<td>Uwe Schindler, Michael Diepenbroek</td>
</tr>
<tr>
<td>Phaidra-Italy</td>
<td>CoreTrustSeal</td>
<td></td>
<td>500 Completed</td>
<td></td>
<td>Yuri Carrer, Cristiana Bettella, GianLuca Drago, Giulio Turetta</td>
</tr>
<tr>
<td>CSIRO Data Portal</td>
<td>CoreTrustSeal</td>
<td>Multiple disciplines</td>
<td>500 Completed</td>
<td></td>
<td>Mikaela Lawrence, Dominic Hogan, Cynthia Love</td>
</tr>
<tr>
<td>World Data Centre for Climate (WDCC)</td>
<td>CoreTrustSeal, WDS Regular Member</td>
<td>Earth System Science</td>
<td>500 Completed</td>
<td></td>
<td>Amandine Kaiser, Andrej Fast, Hannes Thiemann</td>
</tr>
<tr>
<td>DataverseNO</td>
<td>CoreTrustSeal</td>
<td></td>
<td>500 Completed</td>
<td></td>
<td>Philipp Conzett (UiT/DataverseNO), Gustavo Durand, Julian Gautier (Harvard/Dataverse)</td>
</tr>
</tbody>
</table>

### FAIRsFAIR Object Metric

| FsF-F1-01D | Data is assigned a globally unique identifier. | OK |
| FsF-F1-02D | Data is assigned a persistent identifier. | OK |
| FsF-F2-01M | Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability. | Some datasets do not have a summary/abstract. e.g., [https://doi.org/10.1594/PANGAEA.861373](https://doi.org/10.1594/PANGAEA.861373) |
| FsF-F3-01M | Metadata includes the identifier of the data it describes. | OK |
| FsF-F4-01M | Metadata is offered in such a way that it can be retrieved by machines. | OK |
| FsF-A1-01M | Metadata contains access level and access conditions of the data. | Total score = 1, Earned = 0. NO access information is available in metadata. Possible ways to include access level metadata are: Schema.org: If public use: -> isAccessibleForFree || free [Note: isAccessibleForFree superscedes free] |

Action: Use to include access level in the metadata (Done)
Before and After

FAIR Scores of PANGAEA Datasets By Principle (Before Improvement, n=500)

FAIR Scores of PANGAEA Datasets By Principle (After Improvement, n=500)

Lessons learned

• Automatised FAIR assessment of research data objects is possible
• Supporting a very large and diverse community
  • 5 pilots
  • + CESSDA, EOSC-NORDIC, DataverseNL
• Iterative mutual improvements … ongoing process

Best use formula: (F-UJI FAIR assessment + f2f FAIR consulting)
Thank You

Task 4.5:
Anusuriya Devaraju, Robert Huber, Mustapha Mokrane, Jerry de Vries, Patricia Herterich, Linas Cepinkas, Vesa Akerman, Joy Davidson, Herve L’Hours.