FAIR

Prinzipien und Tools für die Evaluation und das Assessment

Andreas Czerniak
0000-0003-3883-4169
Agenda

● Die Prinzipien von FAIR
  ○ FAIR - Communities

● Tools für die Evaluierung und das Assessment von FAIR Indikatoren

● Ausblick
The FAIR Guiding Principles for scientific data management and stewardship (Nature, Scientific Data, 2016)

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.
FAIR - Prinzipien

F indable

R eusable
Communities zu den FAIR-Prinzipien

- **Research Data Alliance (RDA)** ([https://www.rd-alliance.org/groups/working-groups](https://www.rd-alliance.org/groups/working-groups))
  - FAIR Data Maturity Model WG
  - FAIR for Research Software WG (FAIR4RS)
  - CURE-FAIR WG
  - FAIR for Virtual Research Environments WG
  - FAIRsharing Registry: Connecting data policies, standards and databases RDA WG
  - Open Science Graphs for FAIR Data IG
  - FAIR Principles for Research Hardware IG
Projekte zu FAIR

- GO-FAIR (https://www.go-fair.org/)

<table>
<thead>
<tr>
<th>GO FAIR Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation Networks</td>
</tr>
<tr>
<td>GO CHANGE</td>
</tr>
<tr>
<td>FAIR Principles</td>
</tr>
</tbody>
</table>
Projekte zu FAIR

- FAIRsFAIR ([https://www.fairsfair.eu/](https://www.fairsfair.eu/))

- European Open Science Cloud (EOSC)
  - FAIR Working Group
  - FAIR metrics and Data Quality Task Force
Elemente der FAIR - Prinzipien

- **RDA FAIR data maturity spec.**
  - **41 Elemente**
    ([https://doi.org/10.15497/RDA00050](https://doi.org/10.15497/RDA00050), DE: [https://doi.org/10.5281/zenodo.5834115](https://doi.org/10.5281/zenodo.5834115))

<table>
<thead>
<tr>
<th>Indicator Id</th>
<th>FAIR Indicator</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDA-F1-01M</td>
<td>Metadata is identified by a persistent identifier</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F1-01D</td>
<td>Data is identified by a persistent identifier</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F1-02M</td>
<td>Metadata is identified by a globally unique identifier</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F1-02D</td>
<td>Data is identified by a globally unique identifier</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F2-01M</td>
<td>Rich metadata is provided to allow discovery</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F3-01M</td>
<td>Metadata includes the identifier for the data</td>
<td>Essential</td>
</tr>
<tr>
<td>RDA-F4-01M</td>
<td>Metadata is offered in such a way that it can be harvested</td>
<td>Essential</td>
</tr>
</tbody>
</table>

- **FAIRsFAIR spec.**
  - **15 Elemente**
    ([https://doi.org/10.5281/zenodo.6461229](https://doi.org/10.5281/zenodo.6461229))

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>FsF-F1-01D</td>
<td>Data is assigned a globally unique identifier.</td>
</tr>
<tr>
<td>FsF-F1-02D</td>
<td>Data is assigned a persistent identifier.</td>
</tr>
<tr>
<td>FsF-F2-01M</td>
<td>Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.</td>
</tr>
<tr>
<td>FsF-F3-01M</td>
<td>Metadata includes the identifier of the data it describes.</td>
</tr>
<tr>
<td>FsF-F4-01M</td>
<td>Metadata is offered in such a way that it can be retrieved by machines.</td>
</tr>
<tr>
<td>FsF-A1-01M</td>
<td>Metadata contains access level and access conditions of the data.</td>
</tr>
<tr>
<td>FsF-A1-02M</td>
<td>Metadata is accessible through a standardized communication protocol</td>
</tr>
<tr>
<td>FsF-A1-03D</td>
<td>Data is accessible through a standardized communication protocol</td>
</tr>
<tr>
<td>FsF-A1-04M</td>
<td>Metadata remains available, even if the data is no longer available</td>
</tr>
</tbody>
</table>
FAIR - Prinzipien

Findable Accessible Interoperable Reusable

FAIR ≠ Open
FAIR – Tools für die Evaluation und das Assessment

Ein Einblick
FAIR - Tools Übersicht

- **ARDC**

- **SATIFYD** – Self-Assessment Tool to Improve the FAIRness of Your Dataset (DANSK)

- **F-UJI - FAIRsFAIR**

- **OpenAIRE FAIR-Assistance**

- **…**

  [https://fairassist.org/](https://fairassist.org/)
FAIR - Tools

- auf einen einzelnen Datensatz bezogen
- umfasst das gesamte Repositorium/FIS

F-UJI

OpenAIRE FAIR-Assistance
F-UJI (FAIRsFAIR)

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

- Begutachtung von einem ausgewählten Datensatz
F-UJI (FAIRsFAIR)

FAIR assessment

F-UJI is a web service to programatically assess FAIRness of research data objects (aka data sets) based on metrics developed by the FAIRsFAIR project.

Please use the form below to enter an identifier (e.g. DOI, URL) of the data set you wish to assess. Optionally you also can enter a metadata service (OAI-PMH, SPARQL, CSW) endpoint URI which F-UJI can use to identify additional information.

Research Data Object (URL/PID):

Enter a valid PID or URL of the dataset's landing page (e.g. a DOI)

Start FAIR Assessment
F-UJI (FAIRsFAIR)

Assessment Results:
Evaluated Resource:

OpenAIRE Research Graph Dump

FAIR level: ②  advanced
Resource PID/URL: 10.5281/zenodo.5801283
DataCite support: enabled
Metric Version: metrics_v0.4
Metric Specification: https://doi.org/10.5281/zenodo.4081213
Software version: 1.4.6
Download assessment results: (JSON)
Save and share assessment results:
## F-UJI (FAIRsFAIR)

### Findable
- 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 programmatically.

### Interoperable
- FsF-F1-02M - Metadata uses semantic resources
- FsF-F1-03M - Metadata is represented using a formal knowledge representation language.
- FsF-F3-01M - Metadata includes links between the data and its related entities.

### Accessible
- FsF-A1-01M - Metadata contains access level and access conditions of the data.
- FsF-A1-03D - Data is accessible through a standardized communication protocol.
- FsF-A1-02M - Metadata is accessible through a standardized communication protocol.
- FsF-A1-02M - Data is available in a file format recommended by the target research community.
- FsF-A1-03M - Metadata includes license information under which data can be reused.
- FsF-A1-04M - Metadata includes provenance information about data creation or generation.
- FsF-A1-05M - Metadata follows a standard recommended by the target research community of the data.
## F-UJI (FAIRsFAIR)

### Report:

**Findable**
- FsF-F1-01D - Data is assigned a permanent identifier.
- FsF-F1-02D - Data is assigned a persistent identifier, publisher, publication date.
- FsF-F2-01M - Metadata includes a permanent identifier, creator, publication date.
- FsF-F3-01M - Metadata includes an identifier, publisher, publication date.
- FsF-F4-01M - Metadata is offered programmatically.

**Accessible**
- FsF-A1-01M - Metadata contains a description.
- FsF-A1-03D - Data is accessible to the public.
- FsF-A1-02M - Metadata is accessible to the public.

### Metric tests:

<table>
<thead>
<tr>
<th>Test</th>
<th>Test name</th>
<th>Score</th>
<th>Maturity</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>FsF-I1-02M-1</td>
<td>Vocabulary namespace URIs can be identified in metadata</td>
<td>1</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>FsF-I1-02M-2</td>
<td>Namespaces of known semantic resources can be identified in metadata</td>
<td></td>
<td>?</td>
<td>✔️</td>
</tr>
</tbody>
</table>

### Debug messages:

- **INFO** Number of vocabulary namespaces extracted from all RDF-based metadata: 13
- **INFO** Default vocabulary namespace(s) excluded: 
  - http://www.w3.org/1999/xhtml/vocab#
  - http://purl.org/dc/elements/1.1/
  - http://schema.org/
  - http://ogp.me/ns#
- **INFO** Check the remaining namespace(s) exists in LOD: []
- **WARNING** NO vocabulary namespace match is found
- **WARNING** Vocabulary namespace(s) or URIs specified but no match is found in LOD reference list (examples): 
  - http://www.openarchives.org/OAI/2.0/oai_dc.xsd
  - http://schema.datacite.org/meta/kernel-4.1/metadata.xsd
  - https://creativecommons.org/licenses/by/4.0
  - http://creativecommons.org/licenses/by/
  - https://zenodo.org/record
  - http://datacite.org/schema/
  - http://www.w3.org/1999/xhtml
  - http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd
Problem: Nur ein Datensatz wird evaluiert!
OpenAIRE Validator – FAIR-Assistance

- https://provide.openaire.eu

- Begutachtung der gesamten Metadaten eines Repositoriums/FIS
Run compatibility test

1. Select datasource
2. Select guidelines
3. Select parameters
4. Finish

Select base URL from one of your registered repositories
-- none selected --

or enter new

https://pub.uni-bielefeld.de/oai
OpenAIRE Validator – FAIR-Assistance

![Bar Chart](image-url)

For content

Highcharts.com
### OpenAIRE Validator – FAIR-Assistance

<table>
<thead>
<tr>
<th>Rule Name</th>
<th>Rule Description</th>
<th>Rule Weight</th>
<th># of Records</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDA-A1-01M</td>
<td>Metadata contains information to enable the user to get access to the data.</td>
<td>9</td>
<td>1000/1000</td>
<td></td>
</tr>
<tr>
<td>RDA-A1-05D</td>
<td>Data can be accessed automatically (i.e. by a computer program)</td>
<td>9</td>
<td>1000/1000</td>
<td></td>
</tr>
<tr>
<td>RDA-F1-01M</td>
<td>Metadata is identified by a persistent identifier.</td>
<td>5</td>
<td>1000/1000</td>
<td></td>
</tr>
<tr>
<td>RDA-F1-01M</td>
<td>Metadata is offered in such a way that it can be harvested and indexed.</td>
<td>5</td>
<td>1000/1000</td>
<td></td>
</tr>
<tr>
<td>RDA-F1-01M</td>
<td>Metadata uses knowledge representation expressed in standardised format.</td>
<td>9</td>
<td>982/1000</td>
<td></td>
</tr>
<tr>
<td>RDA-R1-1-02M</td>
<td>Metadata refers to a reuse license.</td>
<td>5</td>
<td>661/1000</td>
<td></td>
</tr>
</tbody>
</table>

#### Chart:

- **For usage**
  - Lithiv Use of OAI-ORE
  - Use of DataCite Metadata Schema (M)
  - Use of OAI-DC (M)
  - Use of OpenAIRE CRIS-CERIF Metadata Schema (M)

- **Score for content:** 99
- **Score for usage:** 88
Ausblick
Communities

● Etablierung neuer Working Groups u.a. bei RDA
  ○ FAIR Principles for Research Hardware

● Diskussion um Standardisierung der Interpretation der FAIR-Elemente.

● Unterstützung durch Training oder Zertifizierungen von Repositorien.
Repositories/Evaluator - Tools

- FAIR-SignPosting (https://signposting.org/FAIR/)

- Standardisierung der Ergebnisse zwischen den FAIR-Evaluatoren.
Vielen Dank für Ihre/Eure Aufmerksamkeit und seid FAIR.