A Taxonomy of Tools and Approaches for FAIRification

Dario Mangione
Leonardo Candela
Donatella Castelli

Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo” - Consiglio Nazionale delle Ricerche
Outline

1. Background
2. Methodology
3. Query results
4. Analysis
5. Taxonomy
6. Conclusions and Future work
Background

→ **FAIR principles**¹
  ○ 4 foundational principles
    ■ Findability
    ■ Accessibility
    ■ Interoperability
    ■ Reusability
  ○ 15 guiding principles to maximise the added-value of research objects
  ○ nothing new to the digital libraries (just the presentation format)
  ○ recently recognised at European level (REGULATION 2021/695 establishing Horizon Europe; open research data pilot 2017)
  ○ open to interpretation²

→ **FAIRification**
  ○ practical process
  ○ different implementations

¹ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), Article 1. [https://doi.org/10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Methodology

There is FAIR (Findable, Accessible, Interoperable and Reusable) and

→ FaIR (climate model)
→ FAIR (Facility for Antiproton and Ion Research)
→ ...

Need to improve precision and recall

→ filters
→ stopwords
→ OpenAIRE subjects (559)
Query results

567 publications

→ Project deliverables
→ Project milestones
→ Reports

389 software entries
Analysis

477 corpus-derived entries (277)

→ dataset published into 7 CSV

1. OpenAIRE query results
2. tool/source (duplicates)
3. tools (unique)
4. tools/classification
5. tools/FAIR principles
6. tools/FAIR resource
7. tools/domains

## FAIR principles-derived categories

<table>
<thead>
<tr>
<th>FAIR principle</th>
<th>tool/service category</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 globally unique and persistent identifiers</td>
<td>GUPRI helper</td>
</tr>
<tr>
<td>F2 rich metadata</td>
<td>Metadata helper</td>
</tr>
<tr>
<td>F3 identifier of the data</td>
<td>Metadata helper</td>
</tr>
<tr>
<td>F4 indexed in a searchable resource</td>
<td>Indexing and discovery service</td>
</tr>
<tr>
<td>I1 language for knowledge representation</td>
<td>Metadata helper Converter</td>
</tr>
<tr>
<td>I3 qualified references to other (meta)data</td>
<td>Metadata helper</td>
</tr>
<tr>
<td>R1.1 data usage licence</td>
<td>Licence helper</td>
</tr>
<tr>
<td>R1.2 provenance</td>
<td>Metadata helper</td>
</tr>
<tr>
<td>R1.3 community standards</td>
<td>Metadata helper Converter</td>
</tr>
</tbody>
</table>
Analysis

FAIR principles-derived categories

→ Accessibility as a general reference
  ○ “A1. (Meta)data are retrievable by their identifier using a standardised communications protocol”
  ○ “A1.1 The protocol is open, free, and universally implementable”
  ○ “A1.2 The protocol allows for an authentication and authorisation procedure, where necessary”
  ○ “A2. Metadata are accessible, even when the data are no longer available”
  ○ depends on F1
  ○ depends on the protocols and the policies adopted

→ I2 does not establish any relationship
  ○ “I2. (Meta)data use vocabularies that follow FAIR principles”
  ○ recursivity

→ R1 does not require a new category per se
  ○ “R1. (Meta)data are richly described with a plurality of accurate and relevant attributes”
  ○ its subdivisions do
Analysis

Entries and categories

- Indexing and discovery service: 43.3%
- Metadata helper: 22.9%
- GUPRI helper: 11.6%
- Licence helper: 1.4%
- Converter: 4.2%
- Assessment tool: 2.8%
- DMP tool: 2.1%
Analysis

Entries and FAIR principles

→ Findable 51,7%
→ Accessible 16,1%
→ Interoperable 10,3%
→ Reusable 21,9%
Analysis

Entries and scientific domains

Taxonomy

→ 7 classes
→ 15 subclasses
Taxonomy

Globally Unique, Persistent and (machine) Resolvable Identifier helper (33)

→ create GUPRIs
→ maintain GUPRIs
→ find providers
**Taxonomy**

Metadata helper (65)

- edit metadata
- extract metadata
- automatically add metadata
- validate metadata
- suggest metadata
Taxonomy

Indexing and discovery service (123)

→ index resources
→ search resources
→ access resources
→ find services
Taxonomy

Converter (39)

→ convert (meta)data
Taxonomy

Licence helper (4)

→ choose a licence
Taxonomy

Assessment tool (12)

→ evaluate FAIRness
→ find assessment tools
Taxonomy

Data Management Plan tool (8)

→ create data management plans
Taxonomy

Metadata helper (65)

→ edit metadata
→ extract metadata
→ automatically add metadata
→ validate metadata
→ suggest metadata
## Taxonomy

**FAIR principles reference**

<table>
<thead>
<tr>
<th>GUPRI helper</th>
<th>GUPRI creation and management service</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUPRI finder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metadata helper</th>
<th>Metadata editor</th>
<th>F2 F3 I1 I3 R1.2 (R1.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Metadata extractor</td>
<td>F2</td>
</tr>
<tr>
<td></td>
<td>Metadata tracker</td>
<td>F2 R1.2</td>
</tr>
<tr>
<td></td>
<td>Metadata validator</td>
<td>I1 R1.3</td>
</tr>
<tr>
<td></td>
<td>Metadata assistant</td>
<td>R1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indexing and discovery service</th>
<th>Registry</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Repository</td>
<td>F4 A</td>
</tr>
<tr>
<td></td>
<td>Indexing and discovery service finder</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Converter</th>
<th>Metadata</th>
<th>I1 I3 R1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>R1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Licence helper</th>
<th>Data</th>
<th>R1.1</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Assessment tool</th>
<th>Automatic</th>
<th>F A I R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manual</td>
<td>F A I R</td>
</tr>
<tr>
<td></td>
<td>Assessment tool finder</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DMP tool</th>
<th></th>
<th></th>
</tr>
</thead>
</table>
Conclusions and future work

→ **Findings**
  - 225 publications, 95 software entries
  - 277 tools (the majority are cross-domain)
  - taxonomy (7 classes and 15 subclasses; generally disjoint)
  - (sub)classes/FAIR principles != 1 to 1
  - no all-in-one FAIRification solutions

→ **Value**
  - state of the art of European-related FAIRification activities
  - FAIRification process support tool
  - FAIRification software development support tool

→ **Further assessment**
  - literature
  - communities of practice
  - FAIR implementation profiles

---

Thanks!

Questions?

dario.mangione@isti.cnr.it
leonardo.candela@isti.cnr.it
donatella.castelli@isti.cnr.it