FAIR Data

Fernando Aguilar Gómez (CSIC)
What is “FAIR data”?

- FAIR Data proposes a certain philosophy to communities to enable the production of quality scientific data and encourage knowledge to be accessible in a clearer and simpler way.

- **Findable** - The first step to (re)use data is to find it. Metadata and data must be easy to find for both humans and machines. Machine-readable metadata is essential for the automatic discovery of datasets and services and is therefore an essential component of the FAIRification process.

- **Accessible** - Once the user finds the required data, they need to know how they can access it, which may include authentication and authorisation.

- **Interoperable** - Typically, data must be integrated with other data. In addition, data must interoperate with applications or workflows for analysis, storage and processing.

- **Reusable** - The ultimate goal of FAIR is to optimise the reusability of data. To achieve this, metadata and data must be well described so that they can be reproduced and/or combined in different environments.

This facilitates the reproducibility of the data generated.
Principles and sub-principles - Findable

The first step in (re)using data is to find it. Metadata and data must be easy to find for both humans and computers. Machine-readable metadata is essential for the automatic discovery of datasets and services and is therefore an essential component of the FAIRification process.

- F1: (Meta)data are assigned globally unique and persistent identifiers
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource
**FAIR principles**

- **Transform principles into indicators**
- **Technically:** How can we implement them?
- RDA Indicators, FAIRsFAIR indicators, etc

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F1</td>
<td>F1-01M</td>
<td>Metadata is identified by a persistent identifier</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>F1-02M</td>
<td>Metadata is identified by a universally unique identifier</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>F1-01D</td>
<td>Data is identified by a persistent identifier</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>F1-02D</td>
<td>Data is identified by a universally unique identifier</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>F2-01M</td>
<td>Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>F2-02M</td>
<td>Metadata is provided for the discovery-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>F3-01M</td>
<td>Metadata includes the identifier for the data</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>F4-01M</td>
<td>Metadata or landing page is harvested by general search engine</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>F4-02M</td>
<td>Metadata is harvested by or submitted to domain/discipline-specific portal</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>F4-03M</td>
<td>Metadata is indexed in institutional repository</td>
</tr>
</tbody>
</table>

- **Generic tests Vs. domain specific**
Implications and analysis tools

- Which elements, components are affected?
- According to “FAIR Data principles”:
  1. **Data**: use a proper format
  2. **Metadata**: community standard. Machine-actionable (JSON, XML, RDF...)
  3. **PID**: Persistent Identifier (e.g. DOI). Provided by an accepted authority.
  4. **Repository/Data service**: indexed and machine-actionable, specific technical features
  5. **Added value**: data as research product

- **FAIR EVA**: Evaluator, Validator & Advisor - fair.csic.es
FAIR metrics and Data Quality TF

- Working group to define how to **assess** data **quality** and **FAIRness** within EOSC.
- Taking into account the current state of the art and the generic and community context.
- **FAIR**
  - Different **tools** available, different implementations
  - **Ambiguous results** that can confuse funders, researchers, publishers...
  - Proposed **governance** model
  - Flexible perspective: FAIR **Principles**
  - Adapted to communities and possible types of users
FAIR Data

Fernando Aguilar Gómez (CSIC)