Cheap and FAIR:
Building a Serverless Research Data Repository

What is FAIR Data?

Andrea Zonca & Rick Wagner
zonca@sdsc.edu & rick@sdsc.edu
SDSC

Gateways 2024, September 30, 2024, Online

SDSC
SAN DIEGO SUPERCOMPUTER CENTER
UC San Diego

The FAIR Principles

15 guidelines to promote the Findability, Accessibility, Interoperability, and Reuse (FAIR) of research datasets.

The principles are labeled as F1 or A1.2 to refer to a particular guideline.

scientific **data**

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › comment › article

Comment | Open access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie

https://doi.org/10.1038/sdata.2016.18

https://www.go-fair.org/fair-principles/

Findable

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

F1. (Meta)data are assigned a globally unique and persistent identifier.

F2. Data are described with rich metadata (defined by R1 below).

F3. Metadata clearly and explicitly include the identifier of the data they describe.

F4. (Meta)data are registered or indexed in a searchable resource.

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

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.

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles.

I3. (Meta)data include qualified references to other (meta)data.

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.

R1.1. (Meta)data are released with a clear and accessible data usage license.

R1.2. (Meta)data are associated with detailed provenance.

R1.3. (Meta)data meet domain-relevant community standards.

How is the SRDR FAIR?

Findable

- Human & machine-readable metadata
- Datasets listed in catalog
- Unique dataset identifiers

Interoperable

- Metadata based on Schema.org & DataCite
- Metadata in JSON-LD

Accessible

- Data & landing pages via HTTPS
- Access control using OAuth/OIDC
- Persistent landing pages

Reusable

- Data files in widely-used formats
- Usage guidelines defined