

FAIRization: What qualifies a search engine for distributed research data?

Dr. Timo Borst

A. Kazakova / Dr. A. Latif / F. Limani

*ZBW Leibniz Information Centre
for Economics*



FAIR data principles and their (possible) implementation...

Findable: Data and metadata are easy to find by both humans and computers. Machine readable metadata is essential for automatic discovery of relevant datasets and services, and for this reason are essential to the FAIRification process.

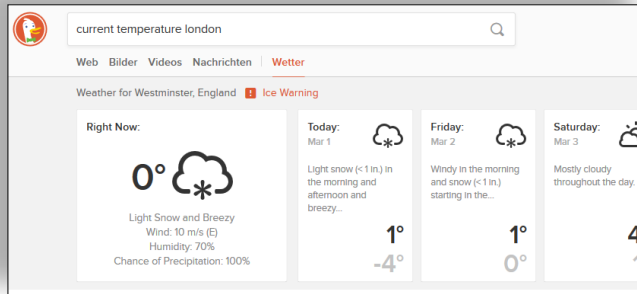
- F1: (meta) data are assigned globally unique and persistent identifiers
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data it describes
- F4: (meta)data are registered or indexed in a searchable resource

<https://www.dtls.nl/fair-data/fair-principles-explained>

- Derived from web pages exposing their content in a machine-readable way
- Search engines constantly crawling / harvesting / indexing these pages
- What could this mean in case of research data?

What could indexing & retrieval of research data mean?

Some fundamental differences between document and data search...



„current temperature in London“

Not covered by classic document-based web search

- Only a minority of research data is text-based
- „data doesn't need documents“ – but we still rely on document-oriented representations / surrogates / proxies of data to be indexed & to become discoverable in compliance with general web search

Suggestions from the GeRDI project



Generic Research Data
Infrastructure

- GeRDI = **Generic Research Data Infrastructure** (including search & discover capabilities)
- Fundamental distinction: **discovering vs. exploring data**
- How to make data **Findable** (in a distributed environment)?
 - Automatic metadata management (including harvesting & indexing)
 - Introducing proxies resp. their textual representations as facets (such as time span, location, data provider, topics/vocabulary,...)
 - Supporting basic exploration & processing of (accessible) data