

Documentation for Dataset

Evaluation of Data Repositories based on the FAIR Principles

for IDCC 2017 Practice Paper

Working Time Period: November till December 2016

Included Data-Repositories: 37

Source of Data-Repositories: <http://www.re3data.org/search?query=&countries%5B%5D=NLD>

List of evaluated Data-Repositories with URL's:

DANS-EASY	https://easy.dans.knaw.nl/ui/home
EUDAT-B2Share	https://b2share.eudat.eu/
Zenodo	https://zenodo.org
PseudoBase	http://www.ekevanbatenburg.nl/PKBASE/PKB.HTML
OpenML	http://www.openml.org/
Profiles-Registry	http://www.profilesregistry.nl/
Mendeley-Data	https://data.mendeley.com/
4TU.Centre for Research Data	http://data.4tu.nl/
CancerData.org	https://www.cancerdata.org
DHS Data Access	http://www.dhsdata.nl
WorldClim	http://worldclim.org/
World Data Centre for Soil	http://www.isric.org/
Infrared Space Observatory	http://www.cosmos.esa.int/web/iso/access-the-archive
Longitudinal Aging Study Amsterdam	http://www.lasa-vu.nl/index.htm
Southeast Asian Climate Assessment & Dataset	http://saca-bmkg.knmi.nl/
TRAILS	https://www.trails.nl/
ICOS Carbon Portal	https://www.icos-cp.eu/node/1
CESSDA	http://cessda.net/
SeaDataNet	http://www.seadatanet.org/
LISS	https://www.lissdata.nl/lissdata/
ORGIDS / RodRep	http://www.orgids.com/ / http://www.rodrep.com/
earth2observe	http://www.earth2observe.eu/
EDGAR	http://edgar.jrc.ec.europa.eu/

KNMI	https://data.knmi.nl/datasets
STITCH	http://stitch.embl.de/
ECA&D	http://www.ecad.eu/
Europeana	http://www.europeana.eu/portal/en
MycoBank	http://www.mycobank.org/
AlgaeBase	http://www.algaebase.org/
Amsterdam Cohort Studies	https://www.amsterdamcohortstudies.org/acsc/index.asp
ICTWSS	http://uva-aias.net/en/ictwss
Share ERIC	http://www.share-project.org/
LOVD3	http://databases.lovd.nl/whole_genome/genes
CARIBIC	http://www.caribic-atmospheric.com/
EIDA	http://www.orfeus-eu.org/data/eida/
Sound and Vision	http://www.beeldengeluid.nl/en
Figshare	https://figshare.com/

Traffic-Light Rating System:

	complies completely
	just about / maybe not
	fails to comply
	unclear

External view on information provided on the web-interface, help pages, data-record with metadata and about pages of the particular data-repository are source of the practice-based evaluation. The yellow coding implies that some information are provided, but it is too vague to code it red or green. In contrast the blue coding means that the evaluator could not find enough information to determine a precise colour.

Regarding I2 “(meta)data use vocabularies that follow FAIR principles” and R3 “(meta)data are associated with their provenance”, flaws in the FAIR principles were recognized and in this case all repositories were scored with the red code for these two attributes. Looking at A4 “metadata are accessible, even when the data are no longer available” and R4 “(meta)data meet domain-relevant community standards” the evaluators applied the blue scoring because they felt not in the position to determine this.

Table of Content for the Excel Spreadsheets:

Sheet 1:	Introduction
Sheet 2:	General Overview, colour coded
Sheet 3:	General Overview, Frequencies and Proportions
Sheet 4:	General Overview, Charts
Sheet 5:	DSA Holder Overview
Sheet 6:	F-A-I-R Facets Frequencies and Proportions
Sheet 7:	F-A-I-R Charts
Sheet 8:	Category Overview, colour coded
Sheet 9:	Category Overview, Frequencies and Proportions

Categorization of Data-Repositories:

Data-Archive

Provides persistent identifier, has metadata-record, provides coherent data-sets; self-proclaimed archive in information provided.

Data-Repository

Does not necessarily provides persistent identifier; metadata-record can be minimal; sometimes no coherent data-sets.

Research-Infrastructure

Offers special features and more services in addition to regular data repository traits.

Subject-based Repositories

Consists of institutional and subject based repositories, with varying qualities of metadata records, data-sets and persistent identifier, according to the needs of the particular discipline.

Online Databases

Provides an interactive (and predominantly complex) interface and dynamic information creation; sometimes difficult to determine the data-set and according metadata.

Source for FAIR Principles and subsequent Evaluation measures:

[website] FORCE 11. (2014). *The FAIR Data Principles. FORCE11*. Retrieved 18 January 2017, from <https://www.force11.org/group/fairgroup/fairprinciples>

FINDABLE

- F1** (meta)data are assigned a globally unique and eternally persistent identifier
- F2** data are described with rich metadata
- F3** (meta)data are registered or indexed in a searchable resource
- F4** metadata specify the data identifier

ACCESSIBLE

- A1** (meta)data are retrievable by their identifier using a standardized communications protocol
- A2** the protocol is open, free, and universally implementable
- A3** the protocol allows for an authentication and authorization procedure, where necessary
- A4** metadata are accessible, even when the data are no longer available

Interoperable

- 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

Re-usable

- R1** meta(data) have a plurality of accurate and relevant attributes
- R2** (meta)data are released with a clear and accessible data usage license
- R3** (meta)data are associated with their provenance
- R4** (meta)data meet domain-relevant community standards