

The FAIRest DataSet Award:

A competition to foster FAIR principles

Roman Gerlach, Kevin Lang, Nadine Neute, Jessica Rex, Annett Schröter, Volker U. Schwartz



The Idea

The Thuringian Competence Network for Research Data Management (TKFDM) is a contact point for all Thuringian universities and colleges. It provides support for all aspects of research data management (RDM). A major goal of the network is to raise awareness for RDM amongst Thuringian scientists. The FAIR principles provide key guidelines to improve Findability, Accessibility, Interoperability, and Reusability of digital assets. Following these principles promotes maximum use of research data. To appreciate the efforts it takes to make research data publicly available, TKFDM established the FAIRest Data Set Award. Thuringian scientists were invited to submit their "FAIRest" published dataset, which was evaluated by the network for how well it implemented the FAIR principles. Prize money of 2,000€ served as a participation incentive for the researchers.



Findable



Accessible



Interoperable



Reusable



General Process

In consultation with the Thuringian Ministry for Economic Affairs, Science and Digital Society (TMWWDG) conditions of participation were created and prize money of 2,000€ was determined. The award was announced in March 2020 via the network website, websites of local partners, mailing lists and Twitter.



Participation Conditions

- Submission deadline: 9 weeks after announcement
- Submission by providing link to published data set
- Scientists (authors) must be employed at one of the Thuringian universities/colleges
- One submission per author
- Usage of prize money earmarked for costs related to research data management

The award ceremony was part of the Thuringian RDM Days, a workshop series taking place once a year. By integrating the award ceremony into this event, winners were honoured in front of a wide audience. The laudatory speech honoured the winners for the high quality of their data sets and their contribution to FAIR data publication. Besides this, a consolation prize was handed over to all other data set authors. Furthermore, all data set authors were offered feedback on data set quality and possible improvements to make their data sets more FAIR.



Conclusion

The FAIRest Data Set Award represents an innovative competition to foster FAIR principles. It raises awareness for RDM and contributes to a cultural change towards FAIR and open data. However, we have noticed that evaluation of FAIRness might be challenging. Tools to assess FAIRness strongly differ in their definition and requirements that data sets must fulfil to be findable, accessible, interoperable and reusable. While some tools set fairly low bars, others place very high demands on data sets, some of which are not met even by most modern data repositories. The ARDC tool, for instance, demands an API download option, barely available in any repositories. In addition, most repositories do not offer linkage of metadata. If scientists publish their research data in a community-accepted and discipline-specific repository that is limited in its publishing options, it is challenging for authors to meet all FAIR criteria. Hence, repositories should revise their services and options, and scientist should make sure repositories comply with the FAIR principles as much as possible before publishing data sets.



Evaluation

FAIR Assessment Tool

A tool that assesses implementation of FAIR principles was to be used for a rough evaluation of submitted data sets. Table 1 provides an overview of the available FAIR assessment tools. For the FAIRest Data Set Award the ARDC FAIR self assessment tool was used as it allows a quick and easy assessment and provides weighted results.

Tool	How FAIR are your data?	FAIR data assessment tool	FAIR self assessment tool	Data Object Assessment Metrics
Provider	EUDAT	Data Archiving and Networked Services (DANS)	Australian Research Data Commons (ARDC)	FAIRsFAIR
Comment	+ quick and easy checklist - too few options for each principle - no weighting of results	- only a few options - "R" stands for result and not reusability	+ quick and easy + comprehensive options + weighted answers and percentages provided - ambiguous explanations	+ comprehensive options - too elaborate for quick evaluation - no weighting of results - published too late to be used

Rough evaluation of data sets

In total, 8 data sets were submitted by scientist of different disciplines from all Thuringian universities. All data sets were independently evaluated by two network members using the ARDC FAIR self assessment tool. To avoid subjective bias, network members only evaluated data sets that did not stem from their own institution. Additional remarks were also noted. Where evaluations differed considerably, they were discussed among all network members.

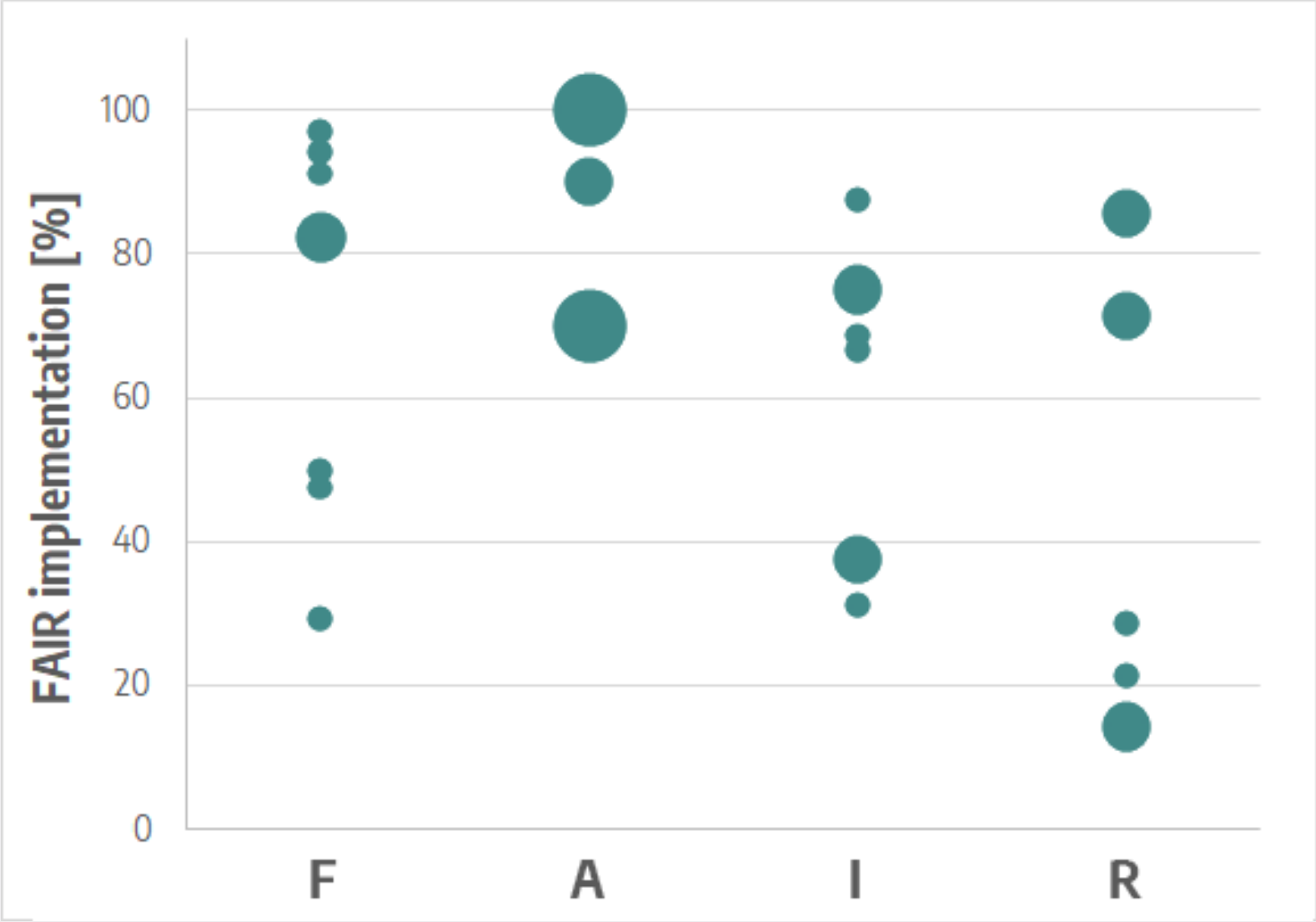


Figure 1: Percentage of FAIR implementation reached by submitted data sets (n=8) as evaluated with the ARDC FAIR self assessment tool. Results are shown individually for criteria Findable, Accessible, Interoperable and Reusable.

In general, data sets were of high quality and fulfilled most of the FAIR criteria (Figure 1). However, few data sets lacked some FAIR aspects:

Findable:

- few data sets without persistent identifier, e.g. no DOI, only URL provided
- few data sets with only little metadata
- few data sets only in registry and not findable by global search engine

Accessible:

- individual data sets had no specified access conditions
- some data sets could be downloaded by HTML or FTP but no API was provided
- some data sets missed information on whether metadata remain accessible if data set is removed

Interoperable:

- individual data set files only provided in proprietary formats
- some metadata lacked information on category or metadata scheme they belong to
- metadata are often not linked, e.g. to similar data sets by same author or institution

Reusable:

- Few data sets missed license information
- Some data sets missing a documentation or only provide scarce information

Detailed evaluation of data sets

Data sets that reached an overall FAIR evaluation of over 80% using the tool, were subjected to a detailed evaluation. Further criteria that were not queried in detail in the FAIR self assessment tool were comprised to ascertain a winner data set.

Criteria for detailed evaluation included: number of key words, linkage of metadata, linkage of publication media (e.g. scientific publication, data set, project), description of file, format and attributes, availability of different formats, existence of persistent identifier of all authors, or openness of license.