Contraction Contraction

Applying FAIRness evaluation approaches to (meta)data preserved at the World Data Center for Climate (WDCC): results, lessons learned, recommendations

vEGU21, EGU21-12560

Session ESSI3.3, 27 April 2021, 13:30-15:00 (CEST)

Karsten Peters-von Gehlen*, Andrej Fast, Daniel Heydebreck, Heinke Hoeck, Amandine Kaiser, Andrea Lammert, Hannes Thiemann *peters@dkrz.de



Deutsches Klimarechenzentrum (DKRZ)

Abstract: 10.5194/egusphere-egu21-12560

K. Peters-von Gehlen et al.







Motivation (1)

Who are we and why do we care about data FAIRness?

DKRZ is the national IT service provider for simulation-based climate science in Germany and offers dedicated discipline-specific support in terms of hardware, software and services.

One of the core activities of DKRZ's data management department is hosting of the Word Data Center for Climate (WDCC). Inaugurated in 2003, the WDCC has since then been certified as trustworthy domain-specific repository and holds about 3.4PB of long-term curated data relevant for climate science – deletion of data is not planned.

We are faced with an increasing number of enquiries from scientists in need for storing their publication-relevant data along the lines of the FAIR data principles -> **Question: How FAIR are our services?**







Motivation (2)





Update of our 2019 self-assessment of WDCC FAIRness

Previous work

In 2019, the increasing community-demand for offering FAIR data services led us to <u>self-assess</u> the WDCC guidelines and practices along the lines of the published FAIR data principles.

The evaluation was NOT performed for individual datasets and revealed relatively high compliance: **F**indability (85%), **A**ccessibility (100%), Interoperability (87%), **R**eusability (85%).

Goals of this contribution

- 1) Is our self-assessment of 2019 generally in-line with other FAIRness assessment approaches?
- 2) How feasible is the application of externallydeveloped FAIRness assessment approaches to WDCC data holdings and what ?
- 3) What can we learn from applying an ensemble of FAIRness assessment approaches?

Selection of FAIRness assessment methods

So far, no globally agreed on evaluation standard exists. We took the approach of sourcing available resources to get an overview of available evaluation approaches to arrive at an ensemble of tests.

Sources for FAIRness assessment tools:

- RDA FAIR Data Maturity Model WG and its outputs (Bahim et al., 2019)
 - Manual filtering for up-to-date, maintained and fit-for**purpose** assessment tools
 - We found 2 out of 12 listed tools useful for our purposes (this and this one)
- Google search for "FAIR evaluation" -> +1 tool
- Community engagement -> +1 tool
- Previous work using our own approach -> +1 tool

vEGu21, ESSI3.3, EGU21-12560







DKRZ



Checklist for the Evaluation of Fitness for Use

WDS/RDA Assessment of Data Fitness for Use WG (Austin et al (2019))

Was not developed specifically for evaluating FAIRness, but to "...check the fitness for use (e.g. FAIRness) of a repository's holdings..." (J. Petters, pers. comm., 2021) – specifically to aid the CoreTrustSeal certification process

- Manual evaluation
- Google-form featuring 20 questions allowing for nuanced answers (Yes; Somewhat; No) (<u>Link</u>)
 - Access to form must be requested it can then be copied and used on own account.
- Specific focus on Reusability (Documentation, Provencance, Domain-specific standards,...)
- Includes visualization of input results



K. Peters-von Gehlen et al.

FAIRshake

FAIRshake – A System to Evaluate the FAIRness of Digital Objects

https://fairshake.cloud

- Partly manual, partly automatic evaluation
- Originally devised for assessment of biomedical resources, but with aim to be generally applicable (D. Clarke, pers. comm., 2021)
- Combination of automated and manual evaluation
 - 9 questions
 - Nuanced answers possible (Yes; Yes, but; No, but; No)
- The results of assessments are stored permanently on the FAIRshake platform
 - **Openly** accessible
 - Acces via the FAIRshake project "WDCC for EGU"

vEGu21, ESSI3.3, EGU21-12560

Overall FAIRness of WDCC can be displayed through a collection of Analytics







F-UJI







Devaraju et al. (2021)

https://www.f-uji.net (online test version)

https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool

- Automated evaluation
- Free web service-based service to programmatically assess FAIRness of research data objects
 - developed under the umbrella of the FAIRsFAIR project
- Accesses (meta)data using either URL or DOI
 - 17 metrics tested (Devaraju et al (2020))
 - Sourcing of external resources for additional information -(re3data.org, DataCite)
- Results cannot be stored using the web-version, the possibility to run F-UJI locally provides more flexibility
 - F-UJI on github

Reconstruction of t weather data	he 1906 Storm Tide in the Germ	aan Bright using TRIM-NP, FES2004, and DWD				
Resource PID/URL:	https://cera-www.dkrz.de/WDCC, acronym=storm_tide_1906_DWD,	/ui/cerasearch/entry? _reconstruct				
Metric Version:	metrics_v0.4	metrics_v0.4				
Metric Specification:	https://doi.org/10.5281/zenodo.40	081213				
Software version:	v1.0.6					
Summary:						
R1 33 Interces	62 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Accessible: 2 of 3				
Report:						
Report: Findable						
Report: Findable FsF-F1-01D - Data is	s assigned a globally unique iden	tífier.	C			
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i	s assigned a globally unique iden s assigned a persistent identifier	tifier.	C			
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i FsF-F2-01M - Metad publication date, su	s assigned a globally unique iden s assigned a persistent identifier data includes descriptive core ele mmary and keywords) to support	tifier. - sments (creator, title, data identifier, publisher, data findability.				
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i FsF-F2-01M - Metar publication date, su FsF-F3-01M - Metar	s assigned a globally unique iden s assigned a persistent identifier data includes descriptive core el mmary and keywords) to support data includes the identifier of the	tifier. - 				
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i FsF-F2-01M - Meta FsF-F3-01M - Metar FsF-F3-01M - Metar	s assigned a globally unique iden s assigned a persistent identifier data includes descriptive core el mmary and keywords) to support data includes the identifier of the data is offered in such a way that	tiffer				
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i FsF-F2-02D - Data i FsF-F2-01M - Meta FsF-F3-01M - Meta FsF-F3-01M - Meta Accessible	s assigned a globally unique iden s assigned a persistent identifier data includes descriptive core el mmary and keywords) to support data includes the identifier of the data is offered in such a way that	tiffer. 				
Report: Findable FsF-F1-01D - Data i FsF-F1-02D - Data i FsF-F2-01M - Meta Publication date, su FsF-F3-01M - Meta Accessible FsF-A1-01M - Meta	s assigned a globally unique iden s assigned a persistent identifier data includes descriptive core ele mmary and keywords) to support data includes the identifier of the data is offered in such a way that data contains access level and ac	tifier				

Example F-UJI online test output

FAIR Maturity Evaluation Service

FAIR Evaluation Services – Resources and guidelines to assess the FAIRness of Digital Resources

https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/

- Automated evaluation based on <u>Wilkinson et al (2019)</u>
 - 22 tests performed
 - Binary answers permetric (Yes; No)
 - Uses URI/DOI to access metadata landing page
- Assessment using a generic collection of metrics
 - Domain-specific adaptations possible
- All test results are permanently accessible on the tool's webpage
 - Search for "WDCC" here
 - Tests are associated with ORCID of the evaluator



Example FAIR Maturity Evaluation Service output





DKRZ

Self Assessment based on FAIR Principles

Our own approach to self-evaluation based on earlier work of self-assessing WDCC FAIRness.

- Manual evaluation
- Based on the published FAIR data principles
 - > 1 metric per principle, 13 in total
- Depending on the principle, answers can be either binary (Yes; No) or nuanced, e.g.
 - binary: F1 (DOI exists or not)
 - nuanced: R1.2 (level of provenance documentation)
- Test protocol is not published/documented yet (in preparation)

IF.





DKRZ

Selection of tested data and representativity (1)

Table of Contents

A subset of the WDCC-curated data was selected for evaluation to avoid the impression of "cherry picking" and to fulfill the following criteria:

- Representativeness of the sample in terms of volume
- Broad spectrum of data curation levels
 - Different granularities (see right)
 - With DOI or not
 - Data accessible or not
 - Spectrum of metadata richness
- Large-volume, high-demand as well as long-tail (small- volume, low-demand) datasets
- New and not-so-new datasets

Hierarchy of WDCC-archived data



- All levels of granularity are individually findable from the GUI
- Data access is only provided at the dataset level
- DOIs assigned at the experiment or dataset group level
 - Older data also have DOIs assigned at the dataset level

Selection of tested data and representativity (2)



DKRZ



- sum of tested projects incl. representatives: 2 PB
- sum of untested WDCC content: 1.4 PB

Some examples of tested WDCC projects

IPCC-AR5_CMIP5

GCM simulation output which formed the basis for IPCC's AR5

MILLENIUM COSMOS

Ensmble of Earth System Model simulations of the last-millenium

CARIBIC

Measurements of atmospheric constituents on commercial airliners

HDCP2-OBS

Observational data collected in the german HDCP2 project

See <u>here</u> for a full list of tested entries and detailed results

Entries for testing were chosen so that projects are represented by all levels of granularity:

13 experiments, 6 dataset groups, 13 datasets



Results (1)

Achieving comparability between tests

All tests have a different number of questions/metrics which have to be answered to perform dataset evaluation.

To achieve comparability between the tests, we normalized the result for each WDCC entry and test to obtain a FAIRness score in the range of 0..1.

Scores of 0 and 1 indicate low and high compliance with the FAIR data principles, respectively.

See <u>here</u> for a full list of detailed results

Mean scores per test





Results (2)

Agreement between tests (1)









See <u>here</u> for a full list of detailed results



Results (3)

Agreement between tests (2)









See <u>here</u> for a full list of detailed results

K. Peters-von Gehlen et al.



Results (4)

Agreement between tests (3)





See <u>here</u> for a full list of detailed results

Linear correlations between tests

	1	2	3	4	5
1		0,67	0,62	0,34	0,63
2			0,37	0,32	0,64
3				0,67	0,69
4					0,61
5					

- 1) Checklist Fitness f Use
- 2) FAIRshake
- 3) F-UJI
- 4) FAIR Maturity Evaluation Service
- 5) Self Assessment

Lessons learned



Table of Contents

All used FAIR assessment methods have individual up- and downsides. However, using just one evaluation method does not yield a thorough assessment of all FAIR criteria

Manual approaches

- Fine tuning of answers is possible and allows for nuanced evaluation
- Expert knowledge required to evaluate
 - relevant documentation and references
 - domain-specific standards and file formats
- Objectivity of the evaluator has to be assumed
 - Especially important as evaluator would be a member of repository staff
- Filling the form is **time consuming** (at least for the first time it is done)

Automated approaches

- Matching test criteria difficult because very specific standards must be met to guarantee machine-readability
 - Communication between evaluator and test provider (programmer) essential and fruitful
 O
- Data tests impossible if authentication required
 - Even if data are FAIR, tests FAIL by design
- **Domain-specific reusability is hardly testable**, as the scientific context and content of references cannot be adequately evaluated
- Testing is timely and guaranteed to be objective \odot

Recommendations for FAIRness evaluation

The main conclusion we take from our analysis of different (meta)data FAIRness evaluation approaches is that the development of hybrid systems, i.e. a combination of manual and automatic assessment, should be taken more into consideration. Furthermore, automatic evaluation approaches should allow for more pragmatic ways to fulfill certain tests.

This is because

- Machine actionability of data is sometimes not practical and cannot be strived for at any cost, e.g. for security of performance reasons
- 2) Information on the data contained in the metadata, e.g. standards and conventions, should suffice to pass automated I and R tests
- 3) Contextual information on provenance, documentation and reuse scenarios is (currently) impossible to evaluate automatically



Table of Contents



Take home for us @WDCC

Overall, the data curation approach at WDCC satisfies the requirements needed to comply with the FAIR data principles to a large part, but/and

- 1) the **granularity of the data structure** from the project level down to the dataset level **leads to less FAIR entries** at certain granularities because not all (scientifically relevant) metadata are inherited from one level to the next
- 2) machine actionability of data holdings shall not be strived for at any price due to security and performance issues
- **3) machine actionability of metadata can be easily improved** by introducing small changes to the metadata record
- Keep in touch with the tool-developers and keep pushing for the development domain-specific data standards and test frameworks, e.g. as done in <u>AtMoDat</u>







References

- Austin, C., Cousijn, H., Diepenbroek, M., Petters, J., Soares E Silva, M. (2019): WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations. DOI: 10.15497/rda00034
- Bahim, Christophe, Dekkers, Makx, & Wyns, Brecht. (2019, May 23). Results of an Analysis of Existing FAIR Assessment Tools. http://doi.org/10.15497/rda00035
- Clarke, Daniel JB, et al. "FAIRshake: Toolkit to evaluate the FAIRness of research digital resources." *Cell systems* 9.5 (2019): 417-421.
- Devaraju, A. et al. (2020, October 12). FAIRsFAIR Data Object Assessment Metrics (Version 0.4). Zenodo. http://doi.org/10.5281/zenodo.4081213
- Devaraju, A., et al., 2021. From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. *Data Science Journal*, 20(1), p.4. DOI: <u>http://doi.org/10.5334/dsj-2021-004</u>
- Wilkinson, Mark D., et al. "Evaluating FAIR maturity through a scalable, automated, community-governed framework." *Scientific data* 6.1 (2019): 1-12.
- Graphics: The Turing Way Community, & Scriberia. (2020, March 3). Illustrations from the Turing Way book dashes. Zenodo. <u>http://doi.org/10.5281/zenodo.3695300</u>



Tested WDCC entries and results

	FAIR Maturity Evaluation Service	F-UJI	FAIRshake	Fitness for Use Checklist	Self Assessment
datasets					
Entry_IDs					
2402231	0,41	54%	0,94	0,7	0,56
2437924	0,41	54%	0,92	0,625	0,56
3501559	0,45	50%	0,89	0,7	0,58
2230553	0,41	50%	0,47	0,3	0,19
3526950	0,50	54%	0,83	0,475	0,62
3878830	0,55	54%	0,86	0,575	0,56
2216468	0,45	50%	0,78	0,55	0,58
2005800	0,50	66%	0,86	0,6	0,68
3881413	0,50	54%	0,81	0,675	0,62
2164174	0,50	50%	0,83	0,525	0,58
3801057	0,50	54%	0,86	0,55	0,62
3565392	0,55	54%	0,92	0,55	0,62
2032577	0,50	54%	0,86	0,75	0,53

		FAIR Maturity Evaluation Service	F-UJI	FAIRshake		Fitness for Use Checklist	Self Assessment
dataset_gro	oups						
Entry_IDs							
	2402114	0,50	66%		0,92	0,8	0,69
	3456259	0,45	54%		0,86	0,775	0,47
	3520184	0,50	54%		0,78	0,6	0,47
	2216112	0,50	50%		0,67	0,425	0,47
	2230067	0,59	62%		0,94	0,775	0,66
	3758143	0,59	70%		0,81	0,65	0,66

Back to result plots

Table of Contents

	FAIR Maturity Evaluation Service	F-UJI	FAIRshake	Fitness for Use Checklist	Self Assessment
xperiments					
ntry_IDs					
2339666	0,41	54%	0,81	0,625	0,57
2258746	0,45	54%	0,86	0,7	0,57
3501558	0,50	66%	0,89	0,8	0,74
2230549	0,45	54%	0,64	0,55	0,32
3519345	0,55	66%	0,83	0,85	0,74
3878773	0,59	70%	0,92	0,825	0,68
2216108	0,41	54%	0,75	0,7	0,47
2005307	0,50	62%	0,78	0,8	0,76
3881144	0,59	70%	0,86	0,75	0,76
2164172	0,50	54%	0,58	0,5	0,6
3758139	0,50	54%	0,67	0,575	0,55
3565262	0,59	66%	0,92	0,8	0,74
2002555	0,50	54%	0,89	0,775	0,55

For you to check dataset characteristics:

Just go to the WDCC GUI, enter the Entry_ID into the search window and explore!! ^(C)

Login required for dataset access