

World Data Center for Climate (WDCC) - Improvement of FAIRness of an established repository

Eileen Hertwig, Andrea Lammert, Heinke Höck, Andrej Fast, Hannes Thiemann
contact: hertwig@dkrz.de

WDCC: World Data Center for Climate

long-term archive hosted by DKRZ

- focus on climate model output relevant for climate and Earth system research
- service for data providers and re-users
- highly standardized – following FAIR principles
- certified as Trustworthy Data Repository by CoreTrustSeal and regular member of World Data System
- DOIs for data collections, PIDs for datasets



WDCC: <https://www.wdc-climate.de/>

FAIRness at WDCC:

different FAIRness evaluations (2022):

Tool	Method
Checklist for Evaluation of Dataset Fitness for Use (CFU)	manual
FAIR Maturity Evaluation Service (FMES)	automated
FAIRshake	hybrid
F-UJI	automated
Self Assessment	manual

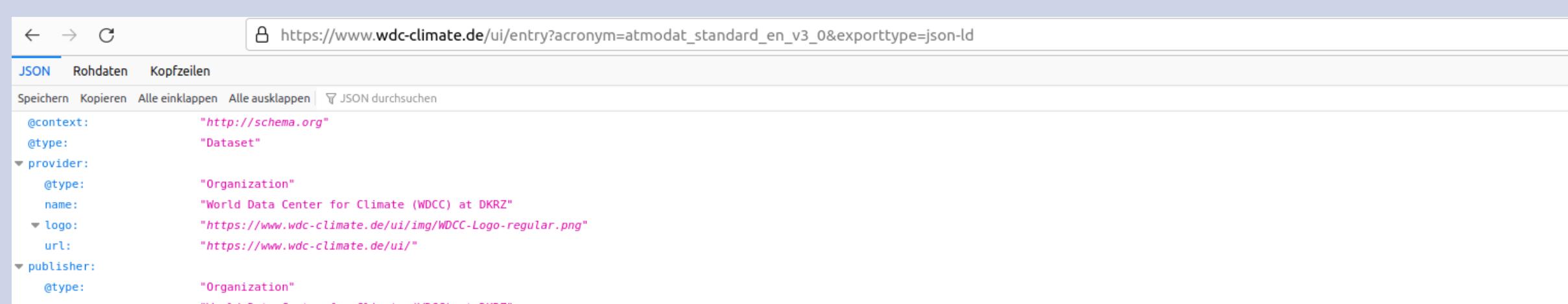
→ mean FAIR score of data at WDCC: 0.67 (range 0.5-0.88)

- *manual*: contextual aspects of FAIRness
- *automated*: strictly standardized aspects of machine actionability
- *hybrid*: combines advantages and eliminates deficiencies of manual & automated approaches
- manual scores higher than automated, hybrid shows highest score

FAIRness at WDCC – Improvements:

JSON-LDs:

- all metadata in WDCC is machine-readable by using JSON-LD format (with schema.org vocabulary included)
- enhances interoperability across data catalogs and makes the data easier to find



FAIRness at WDCC – Plans for the Future:

continue to improve FAIRness at WDCC

- work on problems of automated FAIRness checker (F-UJI)
 - but not just to improve F-UJI score, really become FAIRer
- standardize metadata about funder including CrossRef Funder and project IDs
- focus further on machine-readability and -actionability of (meta)data
- challenge: implementing FAIR principles for great amount of data from very high resolution models



Findable

→ easy to find and assigned globally resolvable identifiers

Accessible

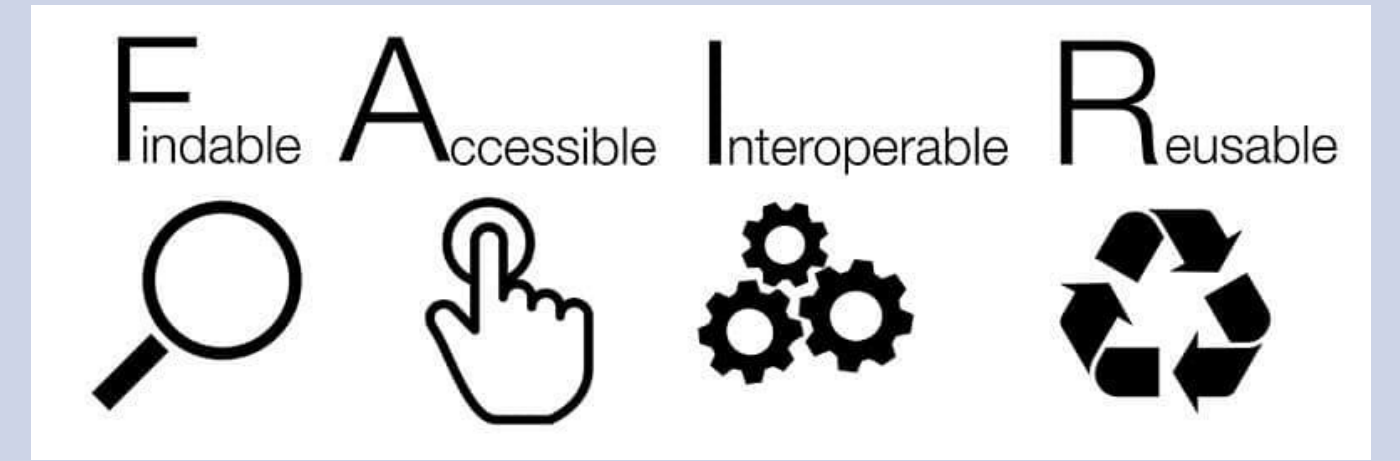
→ licenses and access mechanisms are made explicit

Interoperable

→ (meta)data comply with community specific file formats and standards such that they can readily be combined with other data

Reusable

→ data are associated with ample metadata, documentation and uncertainty estimates such that the scope of the re-use becomes immediately clear



Assessment of FAIRness at WDCC – current state:

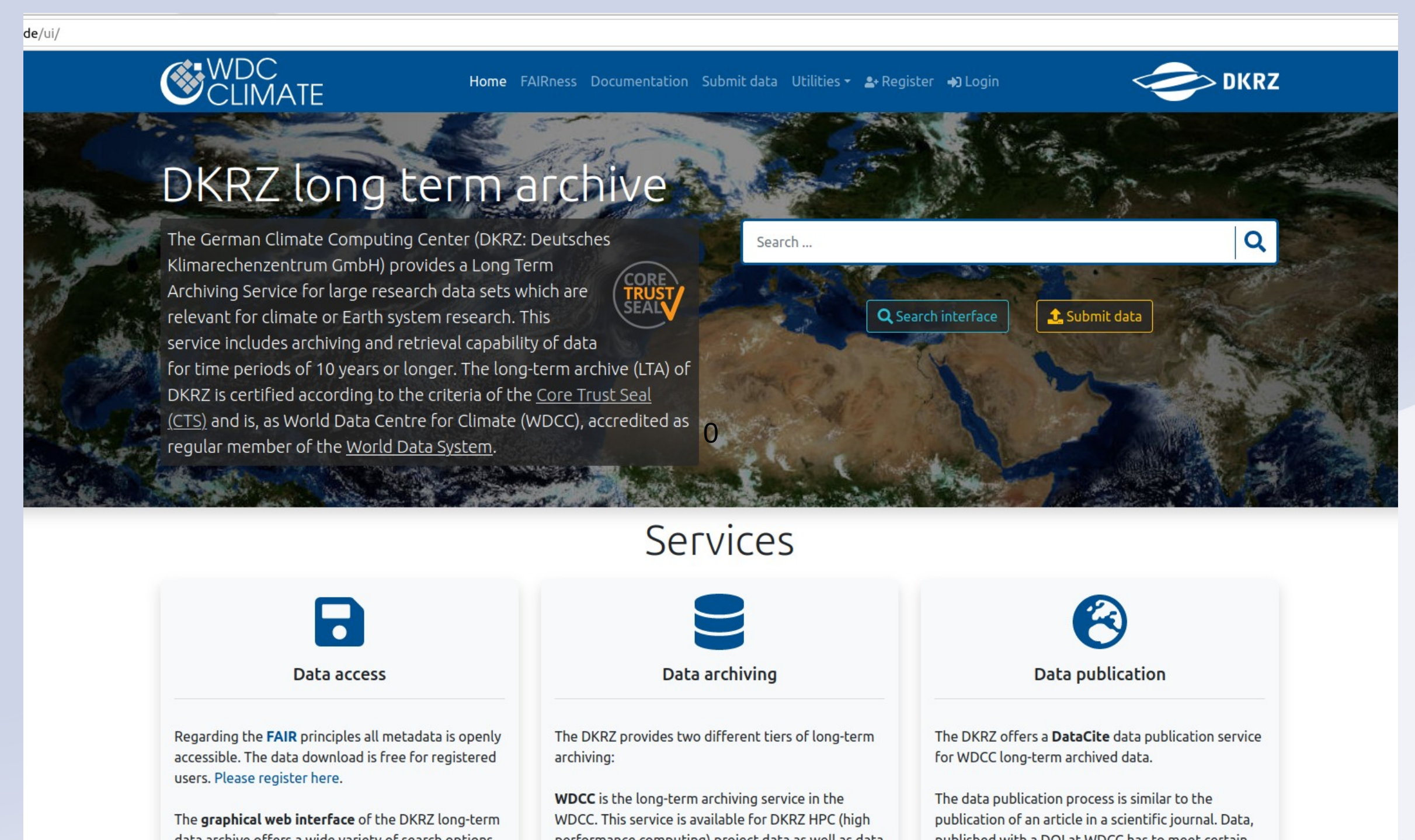
- *F-UJI test* is performed for all new dataset collections with DOI
 - results improve over time, e.g. because of PIDs and improvements in JSON-LDs
- some problems remain with automated approach of F-UJI, e.g.:
 - users need to be logged in to download data
 - F-UJI wants download information (file size, file format, links), which WDCC as a tape archive cannot fully provide
 - WDCC has hierarchical structures and not all information/metadata is available on all levels

details in Peters-von Gehlen et al. „Recommendations for Discipline-Specific FAIRness Evaluation Derived from Applying an Ensemble of Evaluation Tools“. In: Data Science Journal 21 (2022). DOI: 10.5334/dsj-2022-007.



PIDs:

- **Persistent Identifier (PID)**: long-lasting reference to a dataset (or other digital object)
- citable data
- DataCite DOIs (**D**igital **O**bject **I**dentifier): type of PID
- first data DOI ever assigned in WDCC (2004)
- WDCC dataset collections can be published with DOI
- since 2022 individual WDCC datasets are assigned a PID



www.dkrz.de