# CAT4KIT: A cross-institutional data catalog framework for the FAIRification of environmental research data

Mostafa Hadizadeh[1], Sabine Barthlott[1], Romy Fösig[1], Uğur Çayoğlu[2], Robert Ulrich[3], Felix Bach[4], Christof Lorenz[1]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Garmisch-Partenkirchen & Karlsruhe, Germany
[2]Karlsruhe Institute of Technology, Steinbuch Centre for Computing, Karlsruhe, Germany
[3]Karlsruhe Institute of Technology, KIT Library, Karlsruhe, Germany
[4]Leibniz Institute for Information Infrastructure, Karlsruhe, Germany

# Research data – Good practice vs. reality

## FAIR data via open repositories



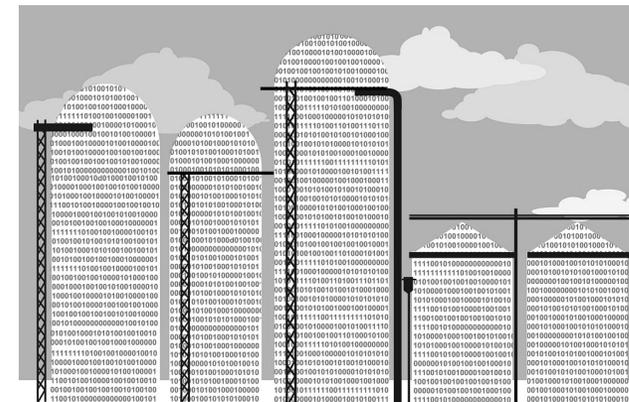Open and reproducible science further requires data on remotely accessible storage systems and repositories

## But even today…

- highly heterogeneous data–formats
- inconsistent and limited metadata
- exchange via Emails / simple cloud storage
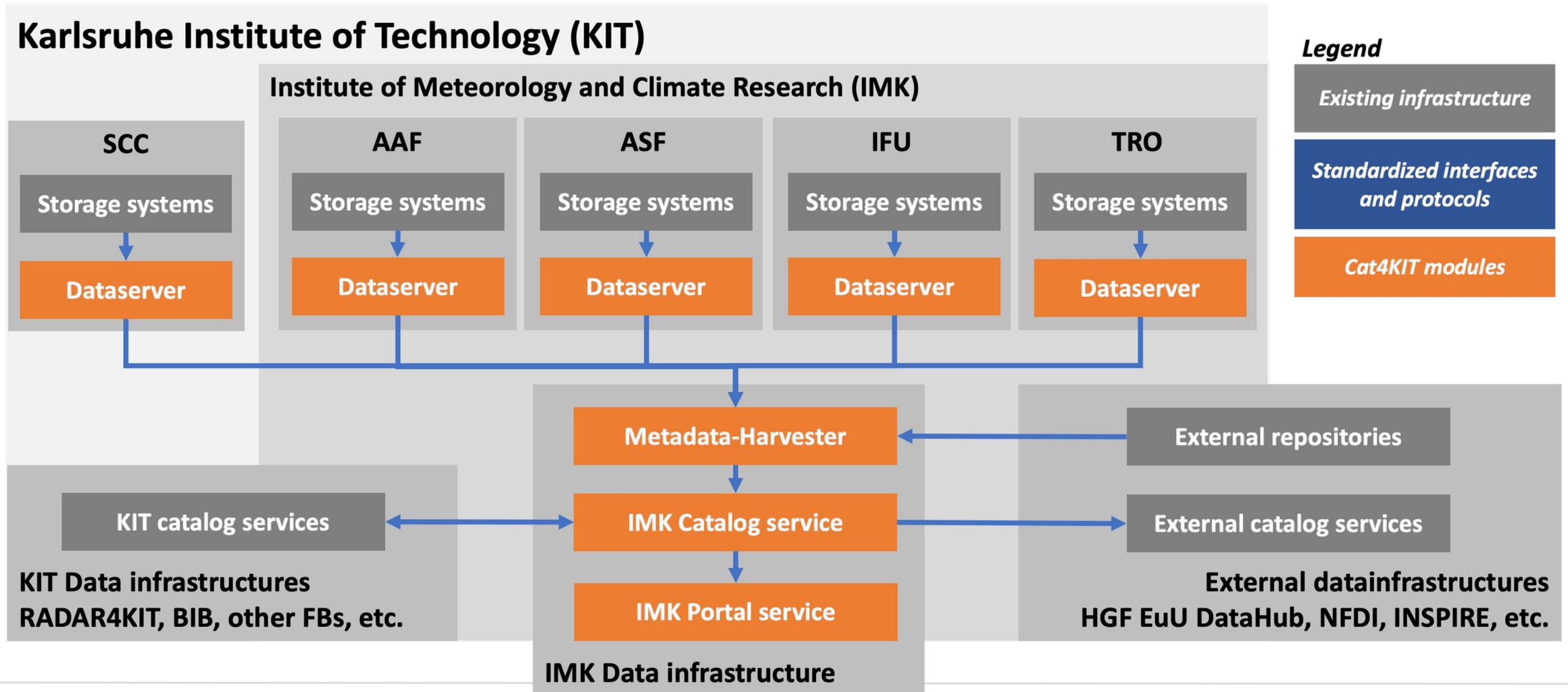- data only „locally" available and not indexed



**A lot of day-to-day research data is not FAIR!**
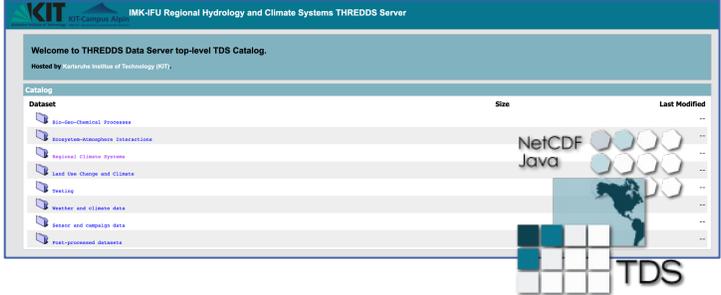
# The research project Cat4KIT

- Funded by KIT ExU-Initiative Research Data Management

- Project duration: 2 years (June 2022 - May 2024)

- Participation of 4 departments (KIT-IMK), the KIT Library and the KIT SCC

- Extended consortium consists of researchers, data managers, data producers and software developers

- **Aim: Development of a user-friendly software stack with wich researchers can make their day-to-day data FAIR!**

# Cat4KIT – Components and links



**Karlsruhe Institute of Technology (KIT)**

**Institute of Meteorology and Climate Research (IMK)**

| SCC | AAF | ASF | IFU | TRO |
|---|---|---|---|---|
| Storage systems | Storage systems | Storage systems | Storage systems | Storage systems |
| Dataserver | Dataserver | Dataserver | Dataserver | Dataserver |

Metadata-Harvester ← External repositories

KIT catalog services ↔ IMK Catalog service → External catalog services

IMK Portal service

**KIT Data infrastructures**
RADAR4KIT, BIB, other FBs, etc.

**IMK Data infrastructure**

**External datainfrastructures**
HGF EuU DataHub, NFDI, INSPIRE, etc.

**Legend**

Existing infrastructure

Standardized interfaces and protocols

Cat4KIT modules

Institute of Meteorology and Climate Research – WG Datamanagement

# Data provider and server for different data worlds

| Medium raster data | Large raster data | One-dimensional data |
|---|---|---|
| • **THREDDS-Server** <br><br> • **Interaction with NetCDF data through various interfaces (OpenDAP, WMS, WFS, etc.)** <br><br> • **Directory-like catalog infrastructure** | • **Cloud-optimised data formats, storage systems and interfaces** <br><br> • **Community-driven libraries and tools (Pangeo-Stack!)** | • **Database and SensorThings API (STA) from Fraunhofer IOSB** <br><br> • **STA reference implementation** <br><br> • **Widely used in E&E Helmholtz-community** |

Institute of Meteorology and Climate Research – WG Datamanagement

# Community standards for (meta)-data & catalogs

| Raster Data | Time-Series Data | Catalog Framework |
|---|---|---|
| • **Quasi-standard for NetCDFs from environmental sciences**<br><br>• **Catalog of standardized attributes and variable names**<br><br>• **Provision of self-explanatory datasets** | • **Lightweight and flexible data model and interface for sensor data**<br><br>• **Certified by OGC as the successor of SOS**<br><br>• **Application in environmental sciences → HMC STAMPLATE-project** | • **Community-driven geospatial catalog infrastructure (collections, catalogs, items) based on GeoJSONs**<br><br>• **Highly modular due to extensions**<br><br>• **Rapidly growing ecosystem** |

CF-Conventions

Open Geospatial Consortium

**SensorThings API**

STAC SpatioTemporal Asset Catalog

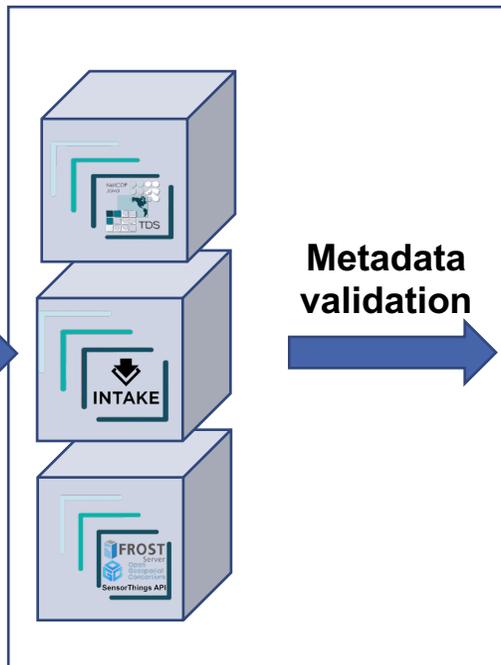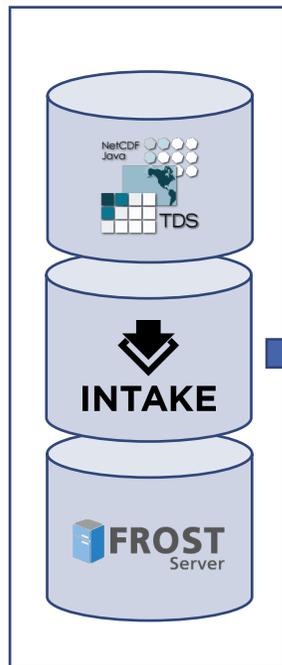# Simple integration, interaction and access



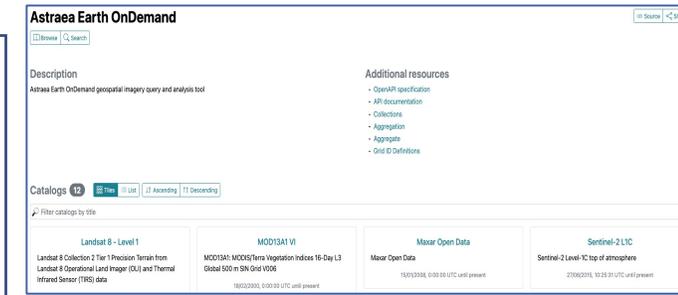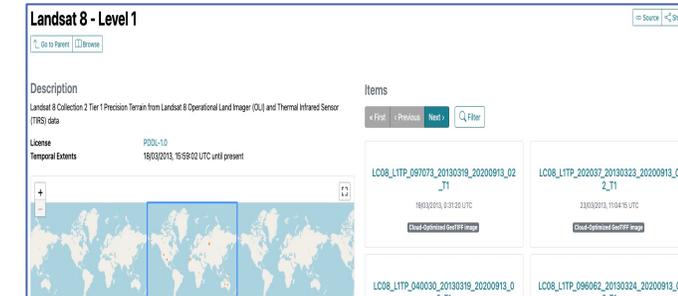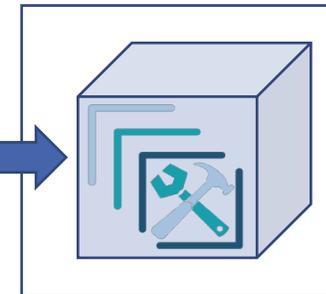**Data provider**   **Harvester & Ingester & DB**   **Exposer**   **Frontend**

Metadata validation

DS2STAC   PgSTAC   STAC FastAPI   STAC Browser

Institute of Meteorology and Climate Research – WG Datamanagement
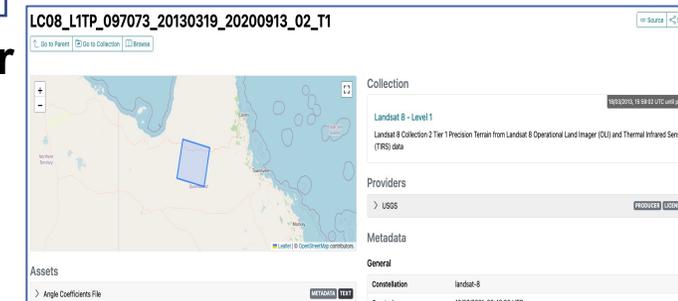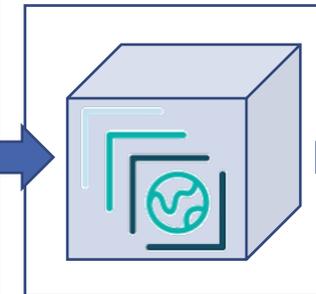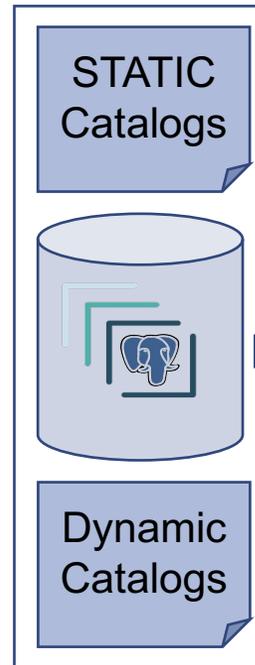
# Status and outlook

- Identification of suitable software / tools / libraries finished

- DS2STAC-modules ready, but not yet connected

- Validator is still missing

- Frontend for each data source has been tested; linkage still WIP

- Infrastructure needs to be filled with data!!!

- Connection of modules within the next 6 months

# Thank you very much for your attention

**Contact: Mostafa.hadizadeh@kit.edu, Christof.Lorenz@kit.edu**

Institute of Meteorology and Climate Research – WG Datamanagement