

Towards a FAIR-compliant Harmonised AI-based Automatic Metadata for Climate Research



Leibniz ScienceCampus
Digital Transformation
of Research

Felix Bach¹

with Anna Jacyszyn¹, Tobias Kerzenmacher², Sabine Barthlott², Sven Hertling¹, Sibylle Haßler², Christian Bonatto Minella¹

¹  **FIZ Karlsruhe**
Leibniz Institute for Information Infrastructure

²  **KIT**
Karlsruhe Institute of Technology



STS Netherlands Conference, University of Twente, 15-17 April 2026

DiTraRe

Leibniz Science Campus *Digital Transformation of Research*

- Growth core to **establish new research branch**.
- Long-term collaboration:



+



- **4+4 years** (start: September 2023).
- Analyse the process of **digital transformation of research**.
- **Multilevel interdisciplinary approach**.

More on DiTraRe **tomorrow 9:30-10:00** Session 18A (T19) by **Anna Jacyszyn**
*Challenges of Coordinating an Interdisciplinary Consortium: Leibniz
ScienceCampus DiTraRe*

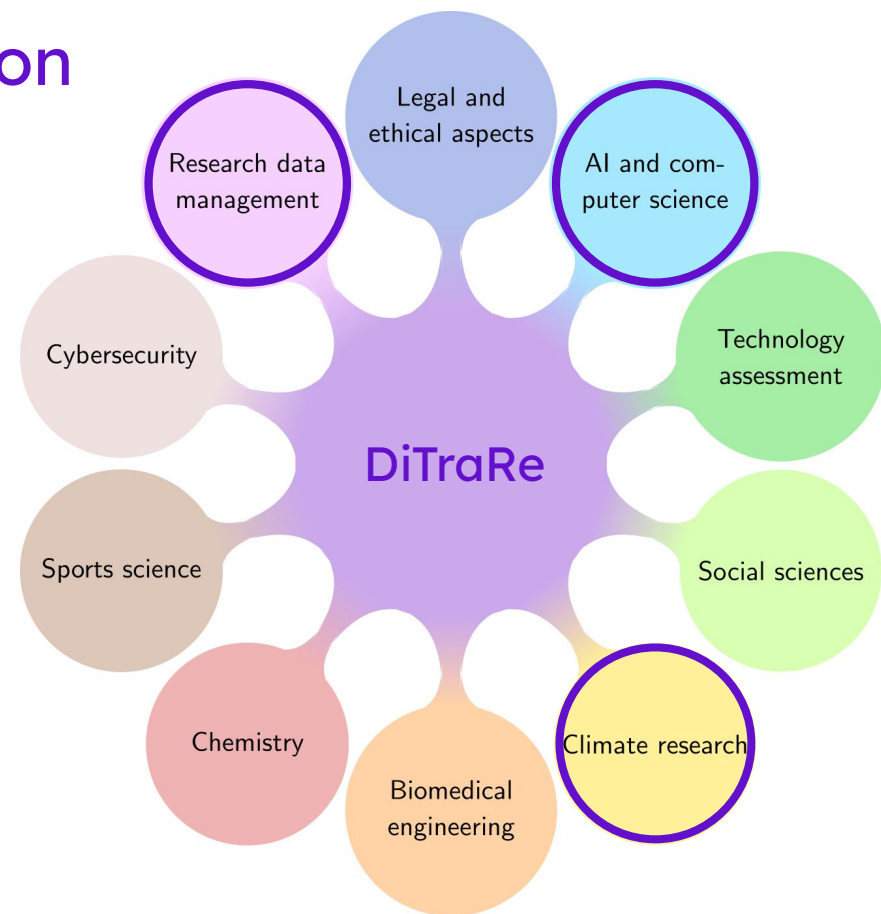
Interdisciplinary collaboration

DiTraRe working groups:

- Climate research
- AI and computer science
- Research data management

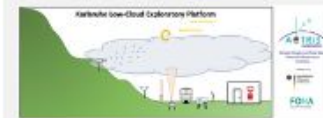
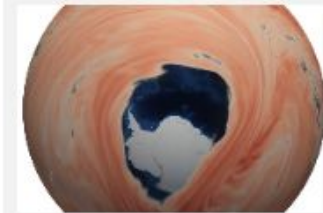
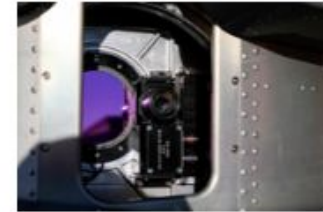
→ Project idea:

Automatic metadata for climate research.



Use case climate research

- Beyond MIPAS/ENVISAT: Placing observational constraints on our understanding of the middle atmosphere (more information)
- GLORIA: This unique infrared imaging spectrometer is central to various aircraft, balloon and satellite activities (more information)
- ICON-ART: The next generation of an atmospheric-chemistry model system (more information)
- COCCON: The "COLlaborative Carbon Column Observing Network" (mehr Informationen)
- KLOCX: The "Karlsruhe Low-Cloud Exploratory Platform" (mehr Informationen)



Problem: metadata as invisible labour

- Metadata creation is distributed, often invisible work
- Fragmented practices across disciplines and repositories
- Legacy datasets lack contextual richness
- Limits reproducibility, comparability, and reuse

A socio-technical perspective

Metadata sits at the intersection of:

- Human practices
- Technical systems
- Epistemic standards

Researchers
(practices)

AI Systems
(extraction)

Ontologies
(standards)

Approach: from text to structured knowledge

- Hybrid workflow combining automation and curation
- AI-assisted extraction + semantic alignment
- Confidence-aware metadata suggestions

Documents &
Data

AI Extraction

Ontology Mapping

FAIR Metadata

Infrastructure & impact

- Integration into repositories (RADAR, PANGAEA, TS4NFDI)
- Enables harmonisation across domains
- Supports new epistemic practices of data reuse
- Towards evolving, AI-assisted metadata ecosystems

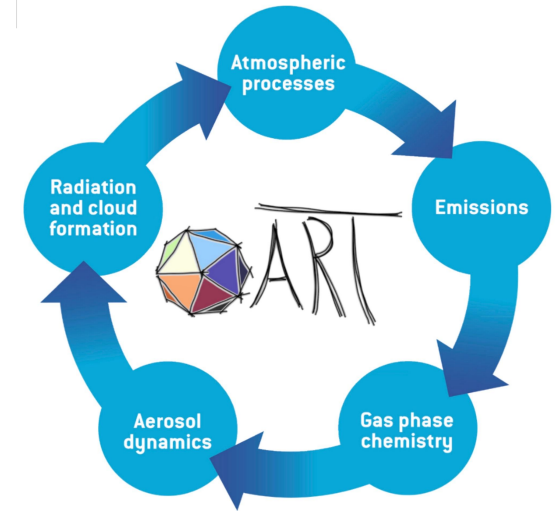
Repositories

AI Layer

Semantic Layer

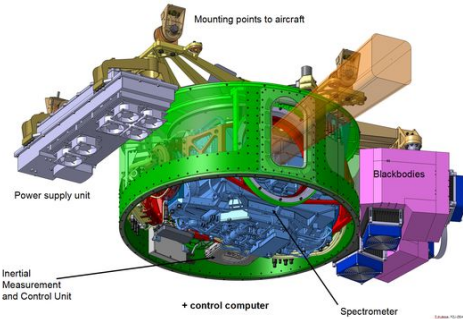
IMKASF — Impact of Climate Research

- Atmospheric trace gases & chemistry
- Satellite & airborne observations
- Data-driven climate modeling
- Part of KIT research ecosystem
- Understanding climate change
- Extreme weather prediction
- Policy guidance
- Sustainability strategies



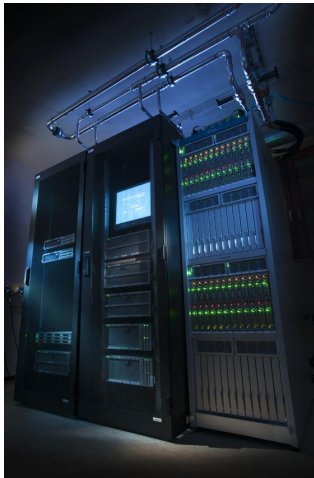
Atmospheric Measurement Systems

- Satellite missions (ENVISAT, etc.)
- Aircraft & balloon campaigns
- Ground networks (COCCON)
- Spectroscopy & remote sensing



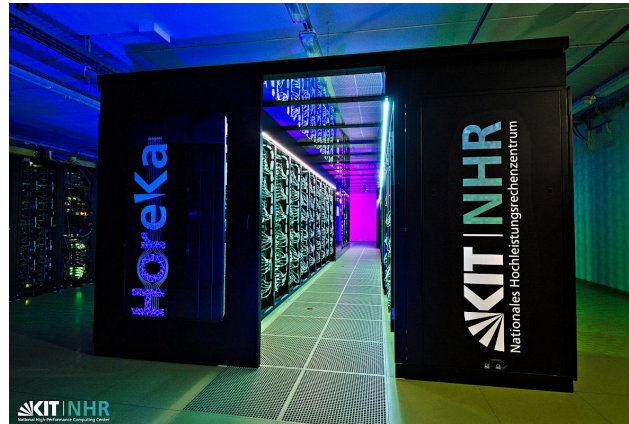
Climate Data & Challenges

- Petabyte-scale datasets
- Heterogeneous formats
- Metadata gaps
- Reproducibility issues



*The IMKASF
satellite data
processor at KIT
(Markus Breig,
KIT)*

*The HoreKa supercomputer at KIT (Simon
Raffeiner, KIT/SCC)*



Climate Data at KIT (IMK-ASF)

Multi-platform data generation:

Satellites, aircraft, ground stations, models

Data pipeline:

Measurement → metadata → QA → storage → publication

Data volumes:

- MB–GB per dataset
- TB–PB total scale

Key challenges:

- Heterogeneity
- Missing metadata

Solution:

AI-supported metadata & FAIR data (Automated metadata extraction, Pattern recognition, Prediction & simulation, FAIR data integration)

The IMK-ASF Atmospheric Research Ecosystem (KIT)

Core Research Pillars

The institute operates as a closed loop between Global Observation and Numerical Modelling to understand the Earth's atmosphere.

- Satellite & Aircraft (SAT/FFB): Large-scale remote sensing of trace gases and dynamics (MIPAS, GLORIA).
- Ground-Based & In-Situ (BOD/TOP/SKL): High-precision long-term trends (FTIR/COCCON) and aircraft-based sensors for validation.
- Modelling Group (MOD/MSK): Theoretical frameworks (ICON-ART, KASIMA, EMAC) that simulate aerosols, reactive gases, and climate feedbacks.

Atmospheric Data & Modelling at IMK-ASF (KIT)

High-Resolution Observational Platforms

IMK-ASF operates sophisticated remote-sensing instruments that capture a "fingerprint" of the atmosphere's chemical composition.

- **MIPAS (Satellite & Balloon versions):** Michelson Interferometer for Passive Atmospheric Sounding.
 - **Focus:** Global vertical profiles from the troposphere to the mesosphere.
 - **Species Captured:** T, H₂O, O₃, CH₄, N₂O, NO₂, HNO₃, ClONO₂, N₂O₅, CFC-11, CFC-12, HNO₄, CCl₄, CF₄, C₂H₆, SF₆, OCS, SO₂, and BrONO₂
- **GLORIA (Aircraft-borne):** Gimballed Limb Observer for Radiance Imaging of the Atmosphere.
 - **Focus:** 2D and 3D "tomographic" imaging of the atmosphere with high spatial resolution.
 - **Extended Species:** Adds pollutants and biomass burning markers like **PAN**, **HCN**, **NH₃**, **C₂H₂**, **HCOOH**, and **COF₂**.

Data Scale & Infrastructure

Source	Typical Data Volume	Key Feature
GLORIA Campaigns	~10+ Terabytes per campaign	High-res 3D "snapshot" of the air.
MIPAS Archive	Hundreds of Terabytes	Global trend data.
ICON-ART Runs	Multiple Petabytes per simulation	Coupled chemistry-weather feedback.

Thank you for your attention!

- Stay **connected** with DiTraRe
 - Website: www.ditrare.de/en
 - Newsletter: www.ditrare.de/en/newsletter
 - Email: ditrare@fiz-karlsruhe.de
 - LinkedIn: www.linkedin.com/company/ditrare
 - Mastodon: social.kit.edu/@DiTraRe
 - YouTube: www.youtube.com/@DiTraRe
 - Zenodo: zenodo.org/communities/ditrare
- Interdisciplinary **Colloquium**: monthly
- **Symposium**: second edition April 2027 @ KIT!



www.ditrare.de/en

