Scientific Research Data Management for Soil-Vegetation-Atmosphere Data – The TR32DB

Introduction – Project Background – Scientific Research Data Management – Conclusion

Constanze Curdt, Dirk Hoffmeister, Guido Waldhoff, Christian Jekel, Georg Bareth
GIS & RS Group, Institute of Geography, University of Cologne, Germany

University of Cologne
Institute of Geography

Introduction

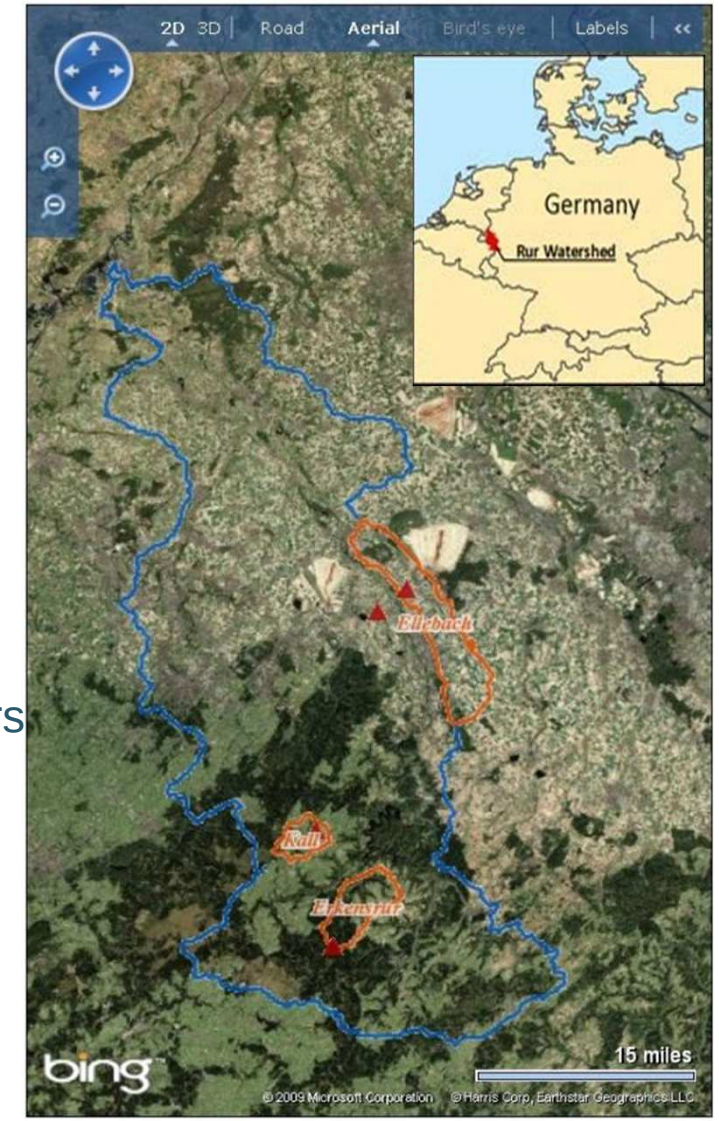
- **Scientific Research Data Management is an important task in every interdisciplinary, long term research project**
  - collection, storage, backup, archiving of all heterogeneous data including descriptive metadata
  - development of a sustainable, stable system
  - support overall communication and data exchange
- **Important for research project with focus on regional monitoring, environmental field studies, and regional modelling**
- **Challenges and problems need to be solved**

In our case:

- **Transregional Collaborative Research Centre (CRC/TR):**
  - funded by the German Research Foundation (DFG)
  - operation time up to 12 years
  - based at up to three separate research locations
  - combination of cross-disciplinary research fields
  - research interests are essential, complementary, and synergetic in nature

CRC/TR 32

- **Interdisciplinary research project**
  - between the Universities of Aachen, Bonn, Cologne, and the Research Centre Jülich
- **Several research areas:**
  - geophysics, soil and plant science, hydrology, remote sensing, meteorology, mathematics
- **Research goal:**
  - research on exchange processes between the soil, vegetation, and atmosphere (SVA)
  - development of improved numerical models for the prediction of water, CO_2, and energy transfers by accounting for the patterns accounting at various scales
- **Project area: Catchment of the River 'Rur'**
  - 3 sub regions



DFG Framework Recommendations

- **Proposals for 'Safeguarding Good Scientific Practice' (DFG, 1998):**
  - documentation of research results
  - secure, persistent storage of primary data for 10 Years
    → CRC/TR32 project data storage for at least 22 years

> **Recommendation 7**
> *Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.*

- **Recommendations for 'Safeguarded storage and provision of digital scientific primary data' (DFG, 2009):**
  - definitions: scientific primary data, organization concepts, metadata and standards
  - rights management and provision of primary data
  - quality control

- **Bulletin 60.06: 'Service-projects for information management and information infrastructure in CRC – INF' (DFG, 2009):**
  - implementation of a scientific data management system / data repository
  - according to 'Good Scientific Practice' (e.g. data storage, backup, achievement)
  - in cooperation with information providers (e.g. libraries, computing centres)

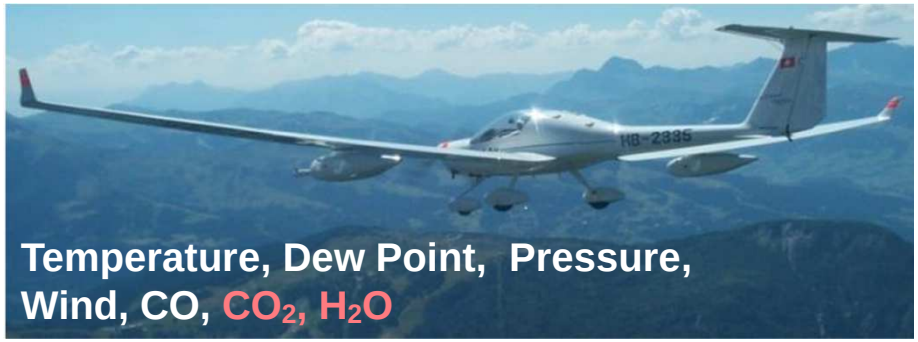
Demands on CRC/TR32 Scientific Data Management

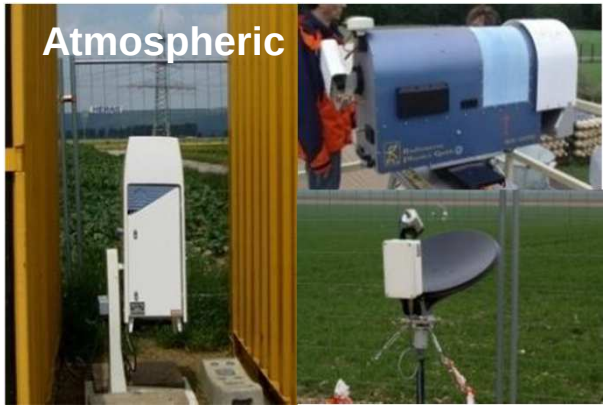
- **Planned project operation time:**
  - 3 Phases á 4 years
    - → **12 years of reseach**

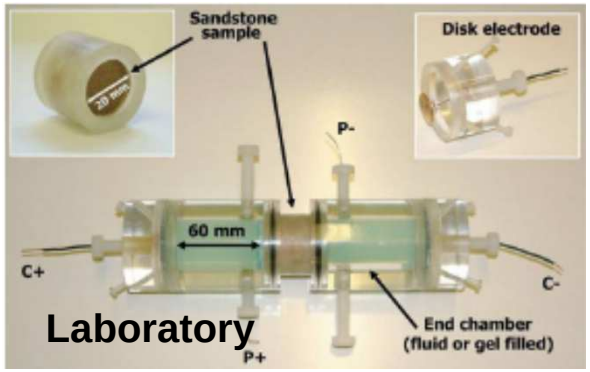
- **Structure of the CRC/TR32:**
  - 4 scientific clusters (A, B, C, D) + central cluster (Z) → each with sub-projects (1. Phase: 15 SP; 2. Phase: 23 SP; 2. Phase: ?)
    - → **ca. 60 project sections**

- **Number of participating scientists:**
  - each project phase: ca. 8 PostDocs, 20 PhDs, and 20 Master students
    - → **total: 24 PostDocs, 60 PhDs, 120 Masters**

CRC/TR 32 - Data







(Kollet et al., 2008; TR32 2011)

Motivation for CRC/TR32 Scientific Data Management

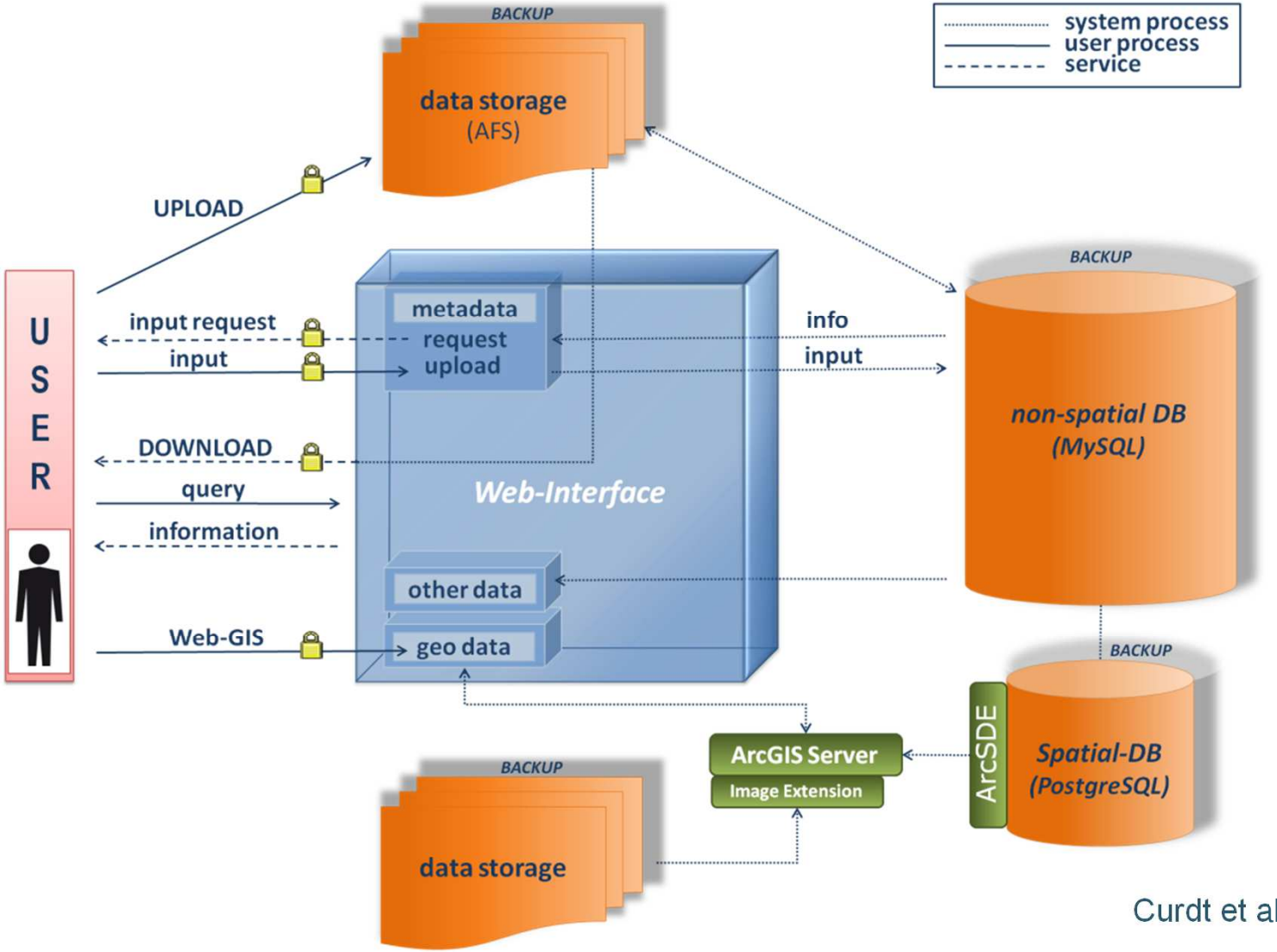
- **Therefore, we have to ensure:**
  - sustainable storage and backup of all project data
  - accurate description of all project data with metadata
  - easy and secure provision, visualisation, and exchange of project data
  - development according to user needs and project demands
  - compliance with DFG recommendations (e.g. cooperation with information provider)

- **Consequently, we have to avoid:**
  - redundant data collection and data storage

Status Quo – CRC/TR32 Data Management

- **Following the state of art**
  - CRC 299, CRC 564, CRC 574, RU 816, RU 402, etc.
- **Combination of**
  - DBMS
  - file management
  - web-Interface including web mapping application
- **According to recent standards and principles**
  - programming standards (Java, JavaScript, XHTML, PHP, CSS)
  - metadata standards (Dublin Core, ISO, INSPIRE directive)
  - etc.
- **Cooperation with and implementation at Regional Computing Center of the University of Cologne (RRZK)**

CRC/TR32 Data Management Structure (TR32DB)



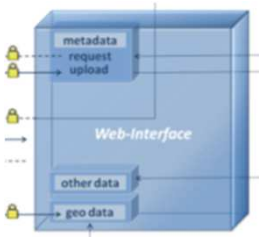
Curdt et al., 2011

CRC/TR32 Data Management Structure (TR32DB)

- **Project data storage:**
  - AFS (Andrew File Systems), a networked file management system
  - cooperation with, support of, and physically located at RRZK
  - hierarchical data storage (project funding phase, research cluster, project section, data type)

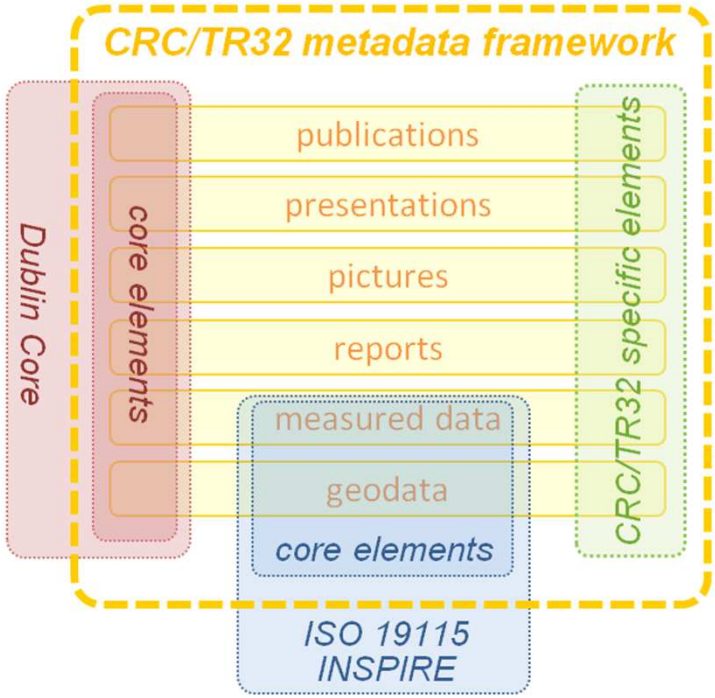
- **MySQL-Database:**
  - Uploaded data files and corresponding metadata
  - administrative data (user information, rights, etc.)

- **Web-Interface:**
  - Located at: **www.tr32db.de**
  - Every user: Representation and search of project data
  - Authorized user:
    - Download and exchange project data
    - Upload corresponding metadata to data files
    - Visualisation of purchased geodata and climate data
    - Apply CRC/TR32 DOIs for data files



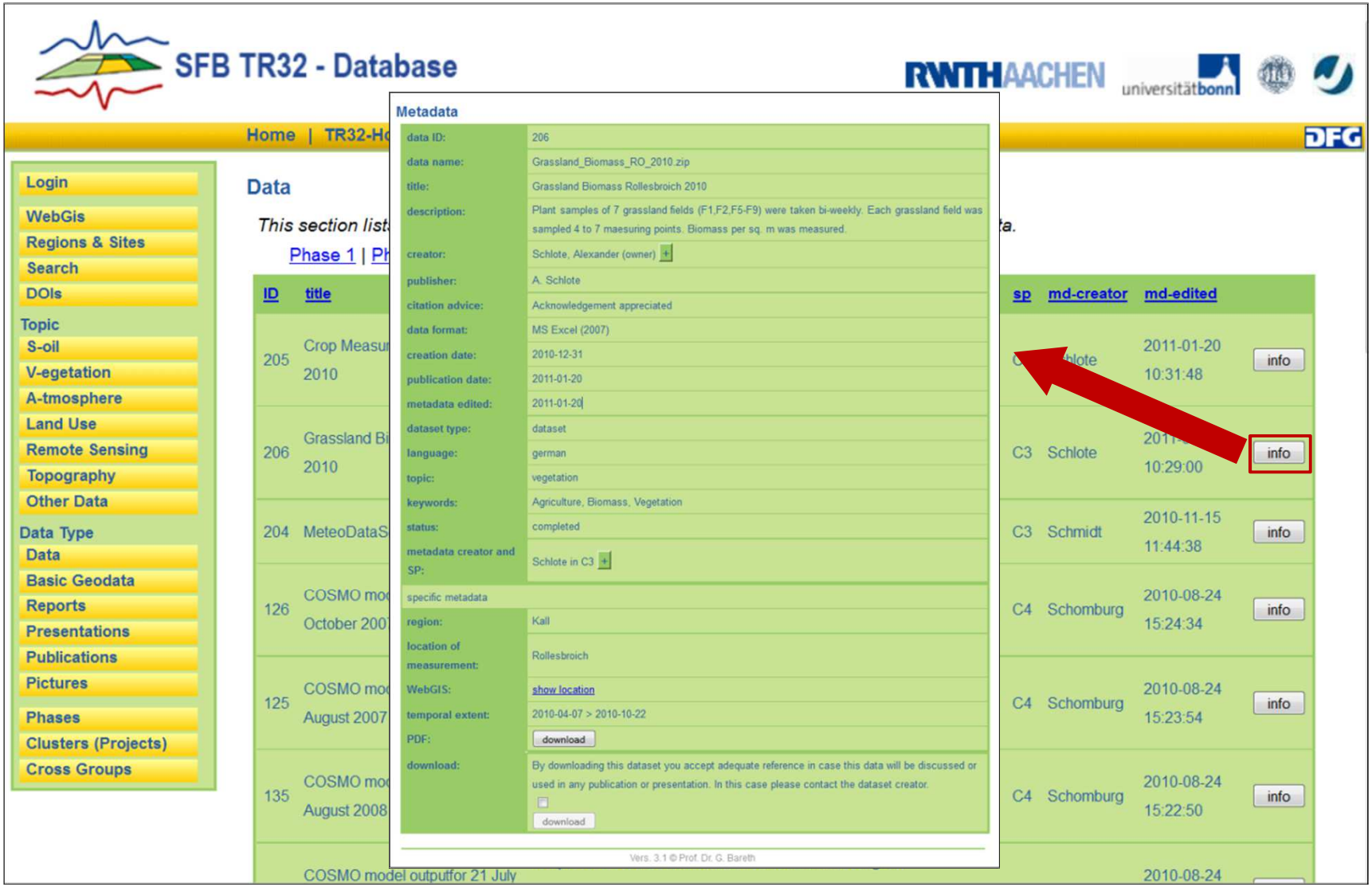
CRC/TR32 Metadata Framework

- **Metadata quantity depends on data type**
- **Development according to:**
  - project participants needs
  - data type demands
  - recent metadata standards
- **multi-level approach:**
  - general properties (Dublin Core)
  - CRC/TR32 specific properties
  - data type specific properties (e.g. ISO, INSPIRE)



Curdt et al., 2009

TR32DB Web – Interface



Conclusion: CRC/TR32 – Data Management

- **Implementation of a running system**
  - sustainable storage, and backup of all project data
  - accurate description of project data with metadata according to data types
  - handling of heterogeneous research data files with huge file sizes (KB to GB)
  - linkage of project data
  - search, provision, access, and exchange of data via Web-Interface
  - application of CRC/TR32 – DOIs for project data
  - visualisation and search for geodata and climate data including attributes via web-interface

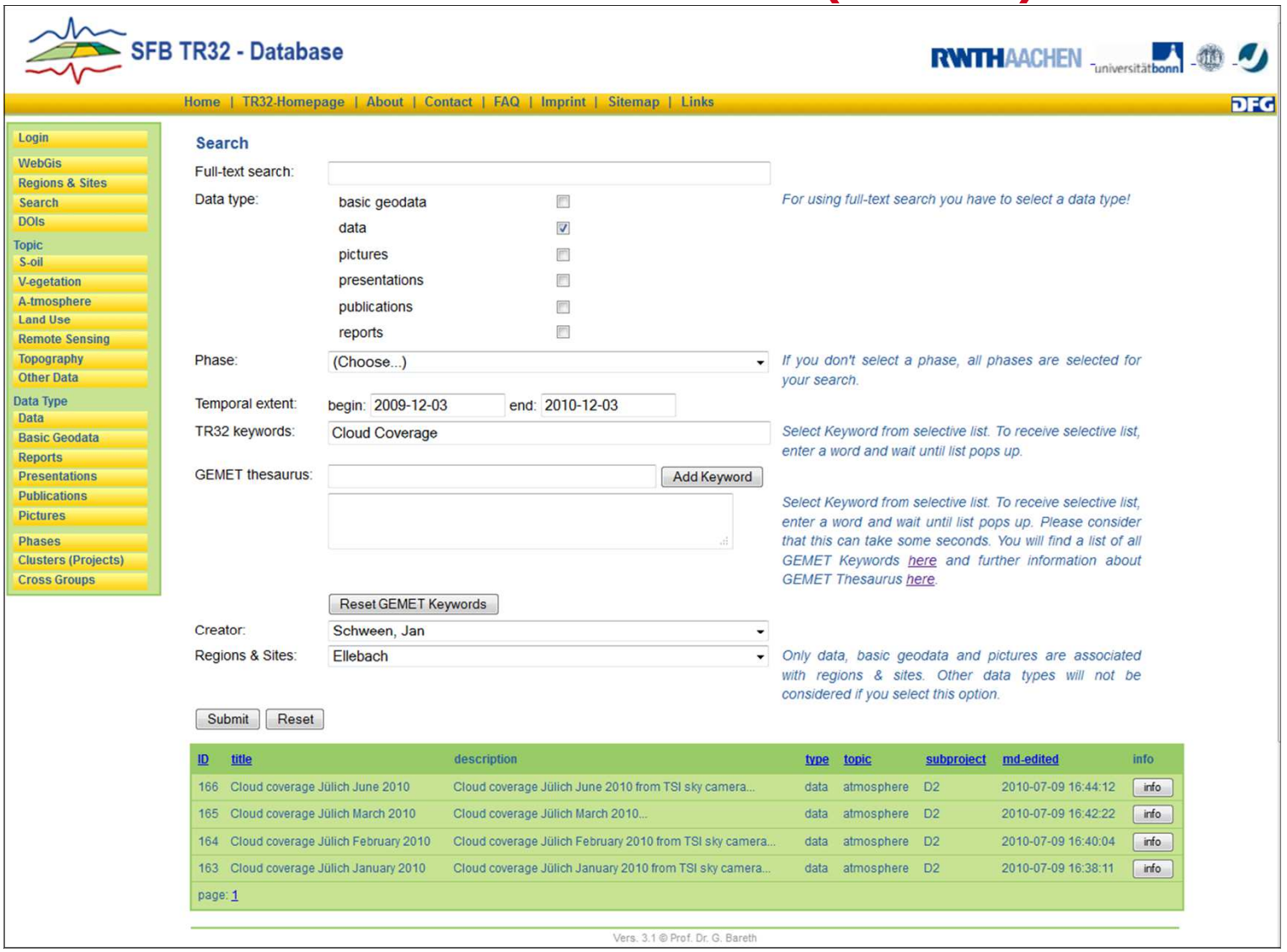
- **Development according to**
  - requirements of DFG
  - demands and needs of research partners
  - recent standards and principles

Scientific Research Data Management for Soil-Vegetation-Atmosphere Data – The TR32DB

Constanze Curdt, Dirk Hoffmeister, Guido Waldhoff, Christian Jekel, Georg Bareth
GIS & RS Group, Institute of Geography, University of Cologne, Germany

University of Cologne
Institute of Geography

TR32DB web-interface (Search)



TR32DB web-interface (DOI)