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
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

- **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)

---

**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.*
Demands on CRC/TR32 Scientific Data Management

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

- **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

Temperature, Dew Point, Pressure, Wind, CO, CO₂, H₂O

Leaf measurement
Wind temperature radar

Hyperspectral
Plant probe

Atmospheric

CO₂-fluxes from the soil

Laboratory

Eddy CO₂-fluxes

Gas exchange of plants

(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)

- **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](http://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
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
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
TR32DB web-interface (Search)

[Image of the TR32DB web-interface]

For using full-text search you have to select a data type!

If you don’t select a phase, all phases are selected for your search.

Selected Keyword from selective list. To receive selective list, enter a word and wait until list pops up.

Selected Keyword from selective list. To receive selective list, enter a word and wait until list pops up. Please consider that this can take some seconds. You will find a list of all GEMET Keywords [here](#) and further information about GEMET Thesaurus [here](#).

Only data, basic geodata and pictures are associated with regions & sites. Other data types will not be considered if you select this option.
TR32DB web-interface (DOI)

Citation: Waldhoff, Guido (2010): Land use classification of 2009 for the Rur catchment. DOI: 10.1594/GFZ.TR32.1.

DOI: 10.1594/GFZ.TR32.1

Related DOI(s): DOI: 10.1117/12.813171

Title: Land use classification of 2009 for the Rur catchment

Abstract: This data set contains the land use classification of 2009 for the study area of the CRC/Transregio 32: "Patterns in Soil-Vegetation-Atmosphere Systems: monitoring, modelling and data assimilation", which is the catchment of the river Rur. The study area is mainly situated in the western part of North Rhine-Westphalia (Germany) and parts of the Netherlands and Belgium, covering an area of approximately 2365 square kilometres. The land use classification is derived from a supervised, multi temporal remote sensing data analysis using "Advanced Spaceborne Thermal Emission and Reflection Radiometer" (ASTER) and RapidEye data. ASTER is a multispectral satellite sensor, which has three bands in the visible and near infrared (VNIR) with 15 m spatial resolution, six bands in the shortwave infrared (SWIR) with 30 m, and five bands in the thermal infrared (TIR) with 90 m. For the land use classification the VNIR data acquired on July 27, 2009 were used. Each sensor of the RapidEye earth observation satellite