



TR32DB - Management of Research Data in a Collaborative, Interdisciplinary Research Project

Constanze Curdt (1), Dirk Hoffmeister (1), Guido Waldhoff (1), Ulrich Lang (2), and Georg Bareth (1)

(1) University of Cologne, Institute of Geography, Germany (c.curd@uni-koeln.de), (2) University of Cologne, Chair of Computer Science, Germany

The management of research data in a well-structured and documented manner is essential in the context of collaborative, interdisciplinary research environments (e.g. across various institutions). Consequently, set-up and use of a research data management (RDM) system like a data repository or project database is necessary. These systems should accompany and support scientists during the entire research life cycle (e.g. data collection, documentation, storage, archiving, sharing, publishing) and operate cross-disciplinary in interdisciplinary research projects. Challenges and problems of RDM are well-known. Consequently, the set-up of a user-friendly, well-documented, sustainable RDM system is essential, as well as user support and further assistance.

In the framework of the Transregio Collaborative Research Centre 32 'Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation' (CRC/TR32), funded by the German Research Foundation (DFG), a RDM system was self-designed and implemented. The CRC/TR32 project database (TR32DB, www.tr32db.de) is operating online since early 2008. The TR32DB handles all data, which are created by the involved project participants from several institutions (e.g. Universities of Cologne, Bonn, Aachen, and the Research Centre Jülich) and research fields (e.g. soil and plant sciences, hydrology, geography, geophysics, meteorology, remote sensing). Very heterogeneous research data are considered, which are resulting from field measurement campaigns, meteorological monitoring, remote sensing, laboratory studies and modelling approaches. Furthermore, outcomes like publications, conference contributions, PhD reports and corresponding images are regarded.

The TR32DB project database is set-up in cooperation with the Regional Computing Centre of the University of Cologne (RRZK) and also located in this hardware environment. The TR32DB system architecture is composed of three main components: (i) a file-based data storage including backup, (ii) a database-based storage for administrative data and metadata, and (iii) a web-interface for user access. The TR32DB offers common features of RDM systems. These include data storage, entry of corresponding metadata by a user-friendly input wizard, search and download of data depending on user permission, as well as secure internal exchange of data. In addition, a Digital Object Identifier (DOI) can be allocated for specific datasets and several web mapping components are supported (e.g. Web-GIS and map search).

The centrepiece of the TR32DB is the self-provided and implemented CRC/TR32 specific metadata schema. This enables the documentation of all involved, heterogeneous data with accurate, interoperable metadata. The TR32DB Metadata Schema is set-up in a multi-level approach and supports several metadata standards and schemes (e.g. Dublin Core, ISO 19115, INSPIRE, DataCite). Furthermore, metadata properties with focus on the CRC/TR32 background (e.g. CRC/TR32 specific keywords) and the supported data types are complemented. Mandatory, optional and automatic metadata properties are specified.

Overall, the TR32DB is designed and implemented according to the needs of the CRC/TR32 (e.g. huge amount of heterogeneous data) and demands of the DFG (e.g. cooperation with a computing centre). The application of a self-designed, project-specific, interoperable metadata schema enables the accurate documentation of all CRC/TR32 data. The implementation of the TR32DB in the hardware environment of the RRZK ensures the access to the data after the end of the CRC/TR32 funding in 2018.