The International Journal of Digital Curation
Volume 7, Issue 2 | 2012

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

Constanze Curdt, Dirk Hoffmeister, Guido Waldhoff, Christian Jekel, Georg Bareth,
Institute of Geography,
University of Cologne

Abstract

The implementation of a scientific research data management system is an important task within long-term, interdisciplinary research projects. Besides sustainable storage of data, including accurate descriptions with metadata, easy and secure exchange and provision of data is necessary, as well as backup and visualisation. The design of such a system poses challenges and problems that need to be solved.

This paper describes the practical experiences gained by the implementation of a scientific research data management system, established in a large, interdisciplinary research project with focus on Soil-Vegetation-Atmosphere Data.

International Journal of Digital Curation (2012), *7*(2), 68–80. http://dx.doi.org/10.2218/ijdc.v7i2.208

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The management of research data, including sustainable storage, accurate description with metadata, easy and secure exchange and provision, back up, and visualization, is an important task within long-term, interdisciplinary research projects. For example, this is essential in research projects that focus on environmental field studies and regional modelling (Mückschel & Nieschulze, 2004). Before an implementation of a research data management system, some issues like current data and metadata standards (Rumble, 2007; Michener, 2006) and the development of a web-interface for visualization purposes should be considered (Stein, Bennett, & Scholes, 2009). The cooperation with other institutions (e.g. libraries or computing centres) might be fruitful (Downs et al., 2007).

However, the implementation of a scientific data management system poses challenges and problems that have to be settled. Due to an increasing amount of data, major data management tasks need to be solved in diverse scientific communities (Bell, Hey, & Szalay, 2009). Additionally, this applies to scientific data management of large, interdisciplinary, long-term research projects. Characteristic difficulties and challenges arise from the integration of a data management into the entire research project, the development of the systems' architecture, and the centralized delivery of data (Mückschel et al., 2007). Berman (2008) and Ailamaki, Kantere, & Dash (2010) describe approaches to tackle the problem.

This paper describes the practical experiences gained by the implementation of a data management system, the TR32DB. It was established for the inter- and multidisciplinary research project Transregional Collaborative Research Center 32 (CRC/TR 32) 'Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation' funded by the German Research Foundation (DFG). First of all, we give a short overview about the state of art of science and technology in research data management. Then we focus on the project background, the project data, and describe the project data management. Finally, the paper ends with a short conclusion.

Background

As already predicted in the past by the scientific community, the volume of scientific large and complex research data will increase and the growing demand for well deliberated research data management systems needs to be solved (Hey & Trefethen, 2003; Gray et al., 2005). Therefore, an essential duty for the whole information age is the provision of a stable, sustainable, predictable and cost-effective data infrastructure for management, organization, access, and preservation of digital data (Berman, 2008). Recently, the so called "data deluge" (Hey & Trefethen, 2003) has arrived. Scientists are increasingly overwhelmed by the volume of the data they produce doing their research. In the future, the overload of data is promised to get worse (Ailamaki, Kantere, & Dash, 2010). Special issues of *Nature* (Nelson, 2009), *Science* (Hansen, Sugden, & Alberts, 2011), *duz* (Ambruster, 2011), *d-Lib* m*agazine* (Lannom, 2011), as well as Curdt & Bareth (2010) addressed the topic and emphasized the importance of scientific data management. Consequently, management of scientific research data has become an important issue.

In addition, recent scientific articles and studies show the point of view of scientists. For example, they focus on the behaviour and motivation of scientist's handling with scientific data management with regards to publishing and sharing their data (Whyte & Pryor, 2011; RIN, 2008; Griffiths, 2009; Borgman, 2010).

Within the scientific community, most scientists agree to archive, share and publish their data to a data infrastructure or data management system. Unfortunately, major difficulties occur when scientists should deliver their data to the system. Some reasons are of technical nature, like defining useful data standards, others refer to the quality and quantity of data, and finally 'social' problems play a major role (Nelson, 2009). Furthermore, the lack of time and resources, lack of time to deal with requests for information, lack of experience or expertise in data management, or competitive factors and fear of exploitation are factors that constrain researchers from publishing their data (Griffiths, 2009).

Many national and international institutions and funding agencies have emphasized the importance of scientific research data management for publicly funded projects. For example, the Organisation for Economic Co-operation and Development (OECD) has published 'OECD Principles and Guidelines for Access to Research Data from Public Funding' (OECD, 2007), the National Science Foundation (NSF) released 'Cyberinfrastructure Vision for 21st Century Discovery' (NSF, 2007), and the Joint Information Systems Committee (JISC) has announced the report 'Managing and Sharing Data: A Best Practices Guide for Researchers' (JISC, 2009).

In Germany, the German Research Foundation (DFG) has announced three important guidelines with regard to scientific research data management that need to be considered. The 'Proposals for Safeguarding Good Scientific Practice' (DFG, 1998) describe the handling of research data, including accurate documentation of research results, as well as giving recommendations for persistent storage of primary data. For example:

> "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." (DFG, 1998).

In addition, the DFG has announced the 'Recommendations for Safeguarded storage and provision of digital scientific primary data' (DFG, 2009a). These describe the definitions of scientific primary data, recommendations about organization concepts, metadata and standards, rights management, provision of primary data, as well as quality control. Furthermore, the DFG has published a bulletin to complement the project proposal for Collaborative Research Centres (CRC) called 'Service-projects for information management and information infrastructure in CRC – INF' (DFG, 2009b). It describes the need of a service project (INF-project) that is responsible for the implementation of a scientific data management system or a data repository of all collected or created data of the project. The established system has to follow the DFG recommendations (DFG, 1998) and ensure centralized, sustainable storage, backup, achievement, and exchange of all project data, including accurate description with metadata. This system has to be implemented in cooperation with an information provider, like a local library or a computing centre.

In the framework of German research projects, several approaches exist to solve problems regarding centralized scientific research data management in the various disciplines. Due to the scientific background of the CRC/TR32-project, the next section will focus on centralized scientific research data management approaches in Germany in the field of environmental and natural sciences.

In the field of environmental and natural sciences there are several German approaches that have implemented a centralized data management system for inter- or multidisciplinary research data. Examples include:

DFG funded projects:

- Collaborative Research Centre 574[1]
- Collaborative Research Centre 299[2]
- Research Unit 816[3]
- Research Unit 402[4]
- Priority Programme 1374[5]

BMBF (Federal Ministry of Education and Research) funded projects:

- GLOWA Volta[6]
- IMPETUS[7]
- ILMS[8]

EU funded projects:

- BRAHMATWINN[9]

The afore mentioned projects are using different technologies. Within all projects it is mandatory to have central, long-term storage or archiving for/of research data. Furthermore, the linkage between research data and metadata is important, as well as access and visualization of data. The implementation of these key issues differs in the various research projects.

---

[1] Collaborative Research Centre 574: http://www.sfb574.uni-kiel.de

[2] Collaborative Research Centre 299: http://www.sfb299.de

[3] Research Unit 816: http://www.tropicalmountainforest.org/

[4] Research Unit 402: http://www.bergregenwald.de

[5] Priority Programme 1374: http://www.biodiversity-exploratories.de

[6] GLOWA Volta: http://www.glowa-volta.de/

[7] IMPETUS: http://www.impetus.uni-koeln.de/

[8] ILMS: http://ilms.uni-jena.de/

[9] BRAHMATWINN: http://www.brahmatwinn.uni-jena.de

 doi:10.2218/ijdc.v7i2.208

The physical storage and archiving of project data is realised by means of several file system and/or various database technologies (e.g. IBM DB2, MySQL). The implementation of metadata usually follows common standards like Dublin Core, or geographic standards like ISO 19115, FGDC or the INSPIRE directive (Nogueras-Iso, Zarazaga-Soria, & Muro-Medrano, 2005). In addition, a combination or enhancement of these standards is possible. For these purposes, applications like the free and open source catalogue application GeoNetworks are already available. The research data is accessible through web interfaces implemented with web technologies and programming languages like PHP, HTML, XML, CSS or AJAX or in a CMS. Moreover, web mapping applications are used for the visualization of spatial related data. Geodata is distributed by geo web servers like UMN MapServer, GeoServer or ESRI ArcGIS Server, and often served according to OGC Standards (WMS, WFS, WCS). Special features, like statistical tools or time management solutions, are also offered (Damm & Götze, 2009; Heimann, Nieschulze, & König-Ries, 2010; Kralisch, Zander, & Krause, 2009; Mückschel et al., 2007; Nauss et al., 2007; Shumilov, Rogmann, & Laubach, 2008).

Within the design of a scientific research data management system or project database it is fundamental to develop a user-friendly, sustainable and interoperable system that enables the access and usability of research data beyond the finish of the project. Therefore, it is important to follow recent standards and principles.

Implementation of a Scientific Research Data Management System for the CRC/TR32

The design of a scientific research data management system always has to be developed according to requirements of the funding agency, the project background as well as the project participant's needs. Therefore, this section gives a brief overview about the project background and the project data, and then continues with the implementation of the project data management system.

The CRC/TR32 is a joint project between the German Universities of Aachen, Bonn, Cologne, and the Research Centre Jülich. The project funding started at the beginning of 2007. Currently, the CRC/TR32 is situated in the first year of the second funding phase. There are three phases scheduled, each running for four years (Phase 1: 2007-2010; Phase 2: 2011-2014; Phase 3: 2015-2018). The research participants of the CRC/TR32 cover several disciplines like soil, plant, and agricultural sciences, (geo)hydrology, (geo)physics, meteorology, geoinformatics, remote sensing, or mathematics. The CRC/TR32 participants work on exchange processes between the soil, vegetation, and the adjacent atmospheric boundary layer (SVA). Their overall research goal is to yield improved numerical SVA models for the prediction of water-, CO2-, and energy-transfer by accounting for the patterns occurring at various scales. The project hypothesis covers the explicit consideration of patterns and structures, which lead to a common methodological framework. The study area of the CRC/TR32 is defined by the catchment area of the river Rur situated in Western Germany, parts of the Netherlands, and Belgium (CRC-TR32, 2011).

Due to the interdisciplinary background of the CRC/TR32, the project participants produce a multiplicity of heterogeneous data. The participants focus their work on the one hand on the soil, plant/landscape, and atmospheric boundary layer; and on the other hand, they work in various spatial scales ranging from the local (point) scale, via the sub-basin scale, to the regional (basin) scale. For example, this includes laboratory methods, single or sensor networks, meteorological and hydrological monitoring, as well as airborne/spaceborne remote sensing. Results of field measurement or flight campaigns include soil water content, soil CO_2 concentration, root growth, leaf characteristics, crop surface/volume models, as well as land use/land cover analyses of the Rur basin (Korres et al., 2009; Hoffmeister et al., 2009; Waldhoff & Bareth, 2009).

For modelling purposes, adequate geodata of different scales were purchased from various institutions like the National Survey Agency of North-Rhine Westphalia, the National Agency of Geology and Germany's National Meteorological Service. These include topographic or land use data, soil, elevation, climate/weather, and remote sensing data.

Furthermore, as a result of the huge amount of field/airborne measurements, modelling, and corresponding analyses within the CRC/TR32, a multiplicity of publications, presentations, reports, and pictures evolve. In addition, these data need to be handled as well.

The project participants provide their data in different data formats due to their different research background. For example, MS Excel, ASCII, NetCDF, binary format, PDF, JPEG, TIF, or GeoTIFF are provided by the project participants. In addition, the size of one dataset varies from few kilobytes to several gigabytes.

The scientific research data management system of the CRC/TR32, called the TR32DB, was established to manage all data that is created within the project. The aim is to store all project data with accurate metadata according to recent metadata standards, as well as to offer functions to share, provide, visualise, and backup the project data.

To archive a first overview of the CRC/TR32 project participants' requirements for a centralized data management system, a survey was carried out in 2007. This survey focused on general questions about the participation with a centralized data management system, the data delivery (kind, format, file size), and the operating system and software used.

The TR32DB is developed according to the requirements of the DFG as well as the needs of the CRC/TR32 project participants (outcome of the survey and experiences gained in the last years). Due to a predicted project operation time of 12 years (three four-year phases) of funding, the TR32DB must be capable of handling a huge amount of heterogeneous scientific data coming from a multiplicity of interacting sub-projects (Phase 1: 13 sub-projects; Phase 2: 23 sub-projects; Phase 3: unknown) from various disciplines. In addition, the number of scientists who work in the sub-projects need to be considered. Besides purchased geo data, collected research data (measured/modelled), publications, pictures, presentations or reports, are also provided by the project participants.

The TR32 design is basically a combination of data storage, databases, and a self-developed web-interface, including a web mapping application (Figure 1) as already described in Curdt et al. (2010). The TR32DB is implemented in cooperation with,

 doi:10.2218/ijdc.v7i2.208

and physically located at, the Regional Computing Centre (RRZK) of the University of Cologne. The TR32DB is accessible online.[10]

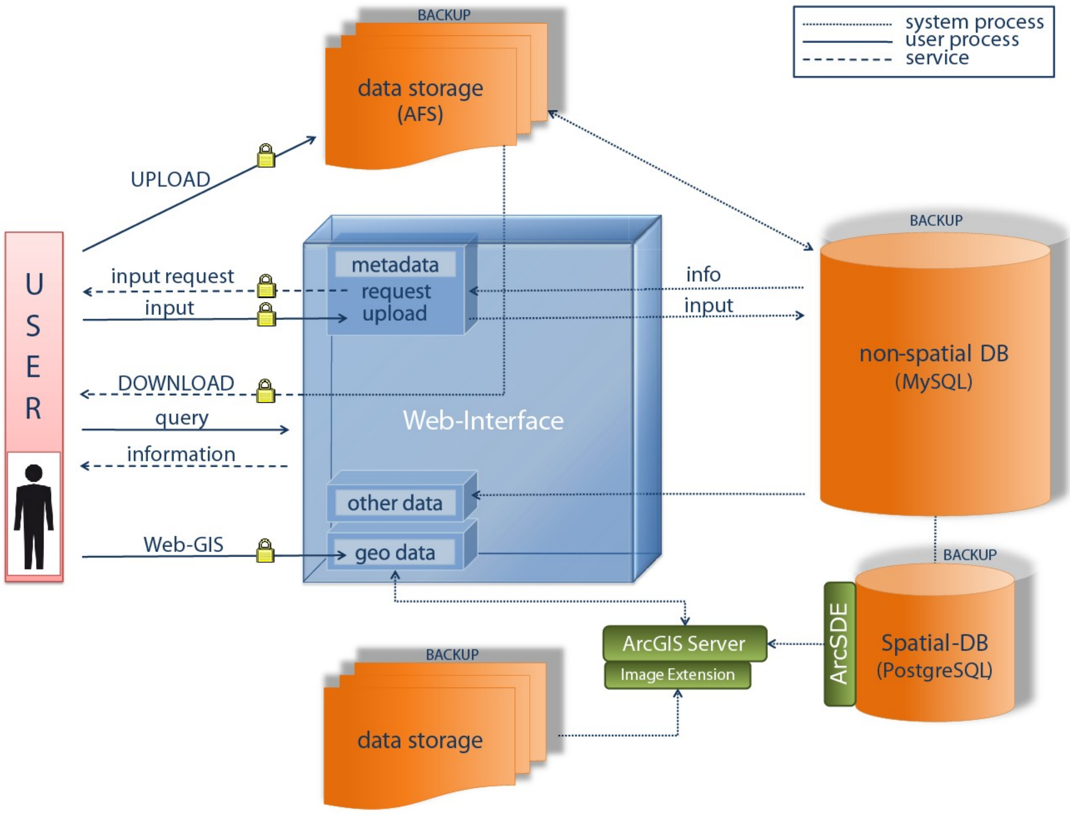

Figure 1. Structure of the TR32DB (Curdt et al., 2011).

All CRC/TR32 project data are physically stored in the Andrew File System (AFS)[11], physically located at the RRZK. The AFS is a distributed networked file system. Main reasons for choosing the AFS are scalability, security, location independence, cross platform access, sharing and synchronising of federated files, easy user administration, simple archive and backup of data. In addition, the availability of the project data during and beyond the project end is ensured as a result of the cooperation with the RRZK. The data storage is organized in a folder system that corresponds to the structure of the CRC/TR32 (funding phase, research clusters, project sections, data types).

The AFS is related to a MySQL database, also implemented and backed up at the RRZK. The MySQL database stores all descriptive metadata of physically located project data in the AFS. In addition, administrative data like user information and user rights are considered.

The CRC/TR32 metadata framework described in Curdt et al. (2009) is designed according to the needs of the project participants as well as to recent standards. To cover all data types of CRC/TR32 data, a multi-level metadata approach was chosen.

---

[10] TR32DB: http://www.tr32db.uni-koeln.de

[11] AFS: http://www.openafs.org

This approach is a combination of general, CRC/TR32 specific metadata, as well as data type specific properties. The Dublin Core Metadata Element Set is chosen as a ‘basic’ level of metadata. Thus, all types of digital project data can be at least described by the ‘basic’ level-information. In addition, CRC/TR32 specific metadata with focus on SVA can be added, such as CRC/TR32 specific SVA-keywords, measurement regions and sites. Furthermore, data type specific properties can be added to fulfil the requirements of a specific dataset. For example, a ‘publication’ requests information about a publication year, status, type, publication source, publication place, volume, issue, or page range, whereas spatial data require elements like a geographic bounding box to fulfil the requirements of a geographic metadata standard. To complete the metadata properties of a dataset, further information about the dataset (the so-called ‘metadata on metadata’) are stored automatically. These include the data file storage path, the metadata changing date, the metadata creator, as well as the project section and funding phase in which the dataset was created.

The web-interface is the connecting element between the data storage and the database, which offers a user interface for visualising and accessing the project data. Every visitor of the web-interface can use the basic functions to search for data and access the details (metadata) of a dataset. The data search is organized according to the requirements of the project participants. A quick data search is realised according to the data types, data topics, measurement regions and sites, as well as funding phases and project sections. Furthermore, a specific search form is available that combines queries regarding the dataset creator, CRC/TR32 specific keywords or measurement extent of a dataset.

Only project participants are authorized to use the internal functions of the web-interface. For example, these include: input and edit of metadata, application of DOIs for CRC/TR32 datasets, and download of datasets. In addition, authorised users are able to use the web mapping applications for geodata and the visualisation tools for weather data (Curdt et al., 2011).

Discussion and Conclusions

The importance of a well arranged research data management system should be made clear by a statement of Heidorn (2008), who notes that primary data, which is not released and archived in a public accessible infrastructure, bears a resemblance to unpublished research. He also refers to so called ‘dark data’ that is any data not easy to find for potential users.

Scientists who are responsible for the development of a centralised scientific research data management system for a large interdisciplinary research project experience different problems. These arise from the various and comprehensive requirements of the system (Mückschel et al., 2007). Furthermore, it is the experience of many research projects that often users will initially contribute to the development process of a data management system or data archive, but then make up an excuse for not uploading their own data (Nelson, 2009). However, the willingness and opportunities to share research data vary in the different scientific communities (Borgman, 2010; Griffiths, 2009).

In this paper, we have presented the centralised, scientific research data management system of the CRC/TR32 (TR32DB), a large inter- and multidisciplinary research project which focuses on regional monitoring, modelling, and data assimilation in Soil-Vegetation-Atmosphere Systems. The implementation of the TR32DB is a very important issue for the entire workflow within the CRC/TR32. The system is developed according to the requirements of the DFG, as well as to the interdisciplinary needs of the project participants (e.g. storage of various data types including corresponding metadata). In addition, the TR32DB data storage is arranged according to the expected huge file sizes (from few kilobytes to several gigabytes per single file) produced by the participants (e.g. for modelling results). In general, the realization of the TR32DB system supports the overall communication and data exchange among the different disciplines, project sections, and between the various research locations (Aachen, Bonn, Cologne, Jülich) during the entire project runtime.

The TR32DB enables easy, secure storage, backup, and archiving of all important research results. The researchers have the advantage of using this service, which is required by the DFG guidelines ‘Proposals for Safeguarding Good Scientific Practice’ (DFG, 1998). Consequently, they don’t have to take care of the management of their data themselves. Furthermore, the availability of the collected data is safeguarded beyond the end of the project funding in accordance with the DFG regulations.

To provide a fast and simple access to all project data, a user friendly web-interface was implemented. This enables the visualisation and search of all project data via metadata, as well as the data output according to security aspects. TR32 participants who store their data in the TR32DB are now able to access their data wherever they have internet access. The implemented web mapping application of spatial data for visualisation purposes is another element of the TR32DB. Users have the opportunity to visually search and explore all spatial data, including their attributes, which are purchased within the project. This is possible wherever users have internet access. In addition, it is not necessary to use certain computer software, which can handle spatial data. This enables the TR32 participants to collect information about a specific attribute of a spatial dataset at a certain location (e.g. soil or land use properties; measurement parameter, measurement extent, or measurement gaps of a specific weather station). In addition, the visualisation accelerates to discover the closest weather station and surveying stations (e.g. gauging station, Eddy Covariance station, or flux measurement tower) of interest within the project area.

The climate data tool, which is also available via the web-interface, helps project participants to query all purchased climate data. It is possible to perform a quick and easy search for climate data by choosing multiple stations, parameters and time extent. Finally, the result can be visualised in a diagram or exported to a file. By using this feature, project participants have the opportunity to directly work with the results.

A core element of the TR32DB is the self-developed metadata management, including the multi-level metadata framework. This offers the opportunity to store all kind of project data with well described metadata. The metadata input via the web-interface is easy and user-friendly. Since all basic metadata are the same for the five data type, new users get very accustomed to the input form. Besides, the metadata input form is extended with a ‘template function’. This feature enables the usage of already existing data, which can be enhanced for the new dataset. In addition, linkage between the different datasets is possible. For example, a scientific dataset with measuring results can be connected to the corresponding publication and vice versa. On the contrary, most scientific research data management systems in the environmental and natural sciences just store the research results of their scientists and consequently cannot connect these to the corresponding publication.

One main advantage of the TR32DB, compared to other scientific research data management systems, is the possibility to apply a DOI (Digital Object Identifier) to all datasets which are stored in the TR32DB. This enables TR32DB users to make their primary scientific data citable worldwide as a publication. For example, the scientific dataset 'Land use classification of 2009 for the Rur catchment' is now citable as:

> Waldhoff, G. (2010): Land use classification of 2009 for the Rur catchment. doi:10.1594/GFZ.TR32.1

A further advantage of the self-developed TD32DB is the flexible enhancement of new features and the data storage size according to the requirements of the project participants.

Acknowledgements

We gratefully acknowledge financial support by the CRC/TR32 "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" funded by the German Research Foundation (DFG).

References

Ailamaki, A., Kantere, V. & Dash, D. (2010). Managing scientific data. *Communications of the ACM 53*(6). Retrieved from http://dl.acm.org/citation.cfm?doid=1743546.1743568

Ambruster, B. (2011). Kampf um den Rohstoff des Wissens. *Duz Magazin 3.*

Bell, G., Hey, T. & Szalay, A. (2009). Beyond the data deluge. *Science 323*(5919). Retrieved from http://www.sciencemag.org/content/323/5919/1297.full

Berman, F. (2008). Got data?: A guide to data preservation in the information age. *Communications of the ACM 51*(12). Retrieved from http://portal.acm.org/citation.cfm?doid=1409360.1409376

Borgman, C.L. (2010). Research Data: Who will share what, with whom, when, and why? *China-North America Library Conference, Beijing*. Retrieved from http://works.bepress.com/borgman/238/

CRC-TR32. (2011). *Funding proposal, 2nd phase*. Retrieved from http://www.tr32.uni-koeln.de

Curdt, C. & Bareth, G. (2010). Proceedings of the data management workshop, University of Cologne, Germany. *Kölner Geographische Arbeiten 90.*

Curdt, C., Hoffmeister, D., Jekel, C., Brocks, S., Waldhoff, G., & Bareth G. (2011). TR32DB – Management and visualization of heterogeneous scientific data. Paper presented at the 19th International Conference on GeoInformatics, Shanghai, China. doi:10.1109/GeoInformatics.2011.5981092

Curdt, C., Hoffmeister, D., Waldhoff, G., & Bareth G. (2009). Implementation of a metadatabase for a regional SDI for soil-vegetation-atmosphere modelling. Paper presented at the GSDI 11 World Conference: Spatial Data Infrastructure Convergence: Building SDI Bridges to address Global Challenges, Rotterdam, The Netherlands.

Curdt, C., Hoffmeister, D., Jekel, C., Udelhofen, K., Waldhoff, G., & Bareth G. (2010). Implementation of a centralized data management system for the CRC Transregio 32 'Patterns in soil-vegetation-atmosphere-systems'. In C. Curdt, & G. Bareth (Eds.): *Proceedings of the Data Management Workshop, Kölner Geographische Arbeiten. 90*.

Damm, T. & Götze, H.-J. (2009). Modern geodata management: Application of interdisciplinary interpretation and visualization in Central America. *International Journal of Geophysics, 13*. doi:10.1155/2009/878324

Downs, R.R., Chen, R.S., Lenhardt, W.C., Bourne, W. & Millman, D. (2007). Cooperative management of a long-term archive of heterogeneous scientific data. Paper presented at the PV 2007 Conference: Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data, Oberpfaffenhofen/Munich, Germany. Retrieved from http://www.pv2007.dlr.de/Papers/Downs_CooperativeManagementOfALongTermArchive.pdf

German Research Foundation. (1998). *Proposals for safeguarding good scientific practice.* Recommendations of the Commission on Professional Self Regulation in Science. Retrieved from http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf

German Research Foundation. (2009a). *Recommendations for safeguarded storage and provision of digital scientific primary data.* Retrieved from http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901_en.pdf

German Research Foundation. (2009b). *Service-projects for information management and information infrastructure in CRC – INF.* Retrieved from http://www.dfg.de/foerderung/programme/koordinierte_programme/sfb/module/modul_inf/index.html

Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J. & Heber, G. (2005). Scientific data management in the coming decade. *SIGMOD Rec. 34*(4).

Griffiths, A. (2009). The publication of research data: Researcher attitudes and behaviour. *International Journal of Digital Curation 4*(1), 46–56. doi:10.2218/ijdc.v4i1.77

Hansen, B., Sugden, A. & Alberts, B. (2011). Making data maximally available. *Science 331*(6018). doi:10.1126/science.1203354

Heidorn, P. (2008). Shedding light on the dark data in the long tail of science. *Library Trends 57*.

Heimann, D., Nieschulze, J. & König-Ries, B. (2010). A flexible statistics web processing service: Added value for information systems for experiment data. *Journal of Integrative Bioinformatics 7*(1). doi:10.2390/biecoll-jib-2010-140

Hey, T., & Trefethen, A. (2003). The data deluge. In F. Berman, G. Fox, & A.J.G Hey, (Eds.) *Grid Computing: Making the Global Infrastructure a Reality*. New York: Wiley.

Hoffmeister, D., Bolten, A., Curdt, C., Waldhoff, G. & Bareth, G. (2009). High-resolution Crop Surface Models (CSM) and Crop Volume Models (CVM) on field level by terrestrial laser scanning. *Proceedings of SPIE 7840*. doi:10.1117/12.872315

JISC (2009). *Managing and sharing data: A best practices guide for researchers*. UK Data Archive. Retrieved from http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

Korres, W., Koyama, C.N., Fiener, P. & Schneider, K. (2009). Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. *Hydrology and Earth System Sciences 14* (751-764). doi:10.5194/hess-14-751-2010

Kralisch, S., F. Zander, & Krause, P. (2009). Coupling the RBIS Environmental Information System and the JAMS Modelling Framework. Paper presented at the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia. Retrieved from http://www.mssanz.org.au/modsim09/C3/kralisch.pdf

Lannom, L. (2011). Research data. *D-Lib Magazine 17*(1/2). doi:10.1045/january2011-editorial

Michener, W.K. (2006): Meta-information concepts for ecological data management. *Ecological Informatics 1*(1). doi:10.1016/j.ecoinf.2005.08.004

Mückschel, C.& Nieschulze, J. (2004). Editorial zum Schwerpunktthema dieser Ausgabe: Datenmanagement in interdisziplinären Umwelt-Forschungsprojekten. *Zeitschrift für Agrarinformatik 4*.

Mückschel, C., Nieschulze, J., Weist, C., Sloboda, B. & Köhler, W. (2007). Herausforderungen, Probleme und Lösungsansätze im Datenmanagement von Sonderforschungsbereichen. *eZAI, 2: 1-16.*

Nauss, T., Göttlicher, D.; Dobbermann, M., & Bendix, J. (2007). Central data services in multidisciplinary environmental research projects. *eZAI 2.*

Nelson, B. (2009). Empty archives. *Nature 461*(7261). doi:10.1038/461160a

Nogueras-Iso, J., Zarazaga-Soria, F. J., & Muro-Medrano, P.R. (2005). *Geographic information metadata for spatial data infrastructures: Resources, interoperability and information retrieval.* Berlin & Heidelberg: Springer.

NSF. (2007). *Cyberinfrastructure vision for 21st century discovery*. Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

OECD. (2007). *OECD principles and guidelines for access to research data from public funding*. Retrieved from http://www.oecd.org/dataoecd/9/61/38500813.pdf

RIN. (2008). *To share or not to share: publication and quality assurance of research data outputs*. Research Information Network. Retrieved from http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf

Rumble, J. (2007). The role of standards in managing and preserving heterogeneous scientific data. Paper presented at the PV 2007 Conference: Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data, Oberpfaffenhofen/Munich, Germany. Retrieved from http://www.pv2007.dlr.de/Papers/Rumble_Roleof%20Standards.pdf

Shumilov, S., Rogmann, A., & Laubach, J. (2008). GLOWA Volta GeoPortal: An interactive geodata repository and communication system. Paper presented at the Digital Earth Summit on Geoinformatics: Tools for Global Change Research. Retrieved from www.digitalearth-isde.org/Summit %202008/papers/63Shumilov.pdf

Stein, T. C., Bennett, K. J. & Scholes, D. M. (2009). Developing web interfaces for scientific data archives. Paper presented at the PV 2009 Conference: Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data, Madrid, Spain. Retrieved from http://www.sciops.esa.int/SYS/CONFERENCE/include/pv2009/papers/22_Stein_Interface_Design.pdf

Waldhoff G. & Bareth, G. (2009). GIS- and RS-based land use and land cover analysis: Case study Rur-Watershed, Germany. *Proceedings of SPIE 7146 (26)*. doi:10.1117/12.813171

Whyte, A. & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *International Journal of Digital Curation 6*(1), 199–213. doi:10.2218/ijdc.v6i1.182