

Implementierung eines Projektdatenmanagements für das interdisziplinäre Forschungsprojekt TR32 „Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation“

Constanze Curdt, Dirk Hoffmeister, Guido Waldhoff, Georg Bareth

Geographisches Institut, Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln
(c.curd, dirk.hoffmeister, guido.waldhoff, g.bareth)@uni-koeln.de

Abstract: Data storage and data exchange is a key issue in interdisciplinary research projects which focus on environmental field studies and regional modelling. The overall success of such projects depends on a well organized data management and data exchange system between all involved project partners. This includes the organization of data, the implementation of a database and the maintenance of such a system for intensive data exchange between the project sections. This paper introduces the data management design of the interdisciplinary research project SFB TR 32 „Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation“.

1 Einleitung

In interdisziplinären Forschungsprojekten, die sich mit umweltbezogenen Feldstudien und regionaler Modellierung befassen, ist ein strukturiertes Management von (räumlichen) Daten essentiell [MN04]. Gerade in transregionalen, interdisziplinären DFG-Sonderforschungsbereichen (SFB/TR) ist eine besonders hohe Menge an unterschiedlichen Daten zu erwarten. Im Sinne der „guten wissenschaftlichen Praxis“ ist für den intensiven Austausch zwischen den einzelnen Teilprojekten und die nachhaltige, dauerhafte Verwaltung und sicherere Speicherung der gesammelten Messdaten und Forschungsergebnisse ein Datenmanagement-Teilprojekt vorgesehen. Ebenfalls empfiehlt die DFG die Umsetzung der Aufgaben in Kooperation mit einem lokalen Rechenzentrum.

In diesem Beitrag wird das Datenmanagement des SFB/TR 32 „Patterns in Soil-Vegetation-Atmosphere Systems (SVA): Monitoring, Modelling, and Data Assimilation“ (TR32) vorgestellt. Dieses besteht im Wesentlichen aus einer TR32-Projektdatenbank (TR32DB, <http://www.tr32db.uni-koeln.de/>), in der alle Projektdaten mit dazugehörigen Metadaten gesammelt, gespeichert und verwaltet werden. Der TR32 beschäftigt sich inhaltlich mit der Wechselwirkung von Mustern und Strukturen im Boden-Pflanzen-Atmosphärensystem im Einzugsgebiet der Rur (<http://www.meteo.uni-bonn.de/projekte/tr32-wiki/>). An der Erarbeitung der Ergebnisse sind verschiedene wissenschaftliche Disziplinen der Universitäten Aachen, Bonn und Köln sowie des Forschungszentrum Jülichs beteiligt.

Das TR32-Projektmanagement wird in Kooperation mit dem Regionalen Rechenzentrum (RRZK)/dem Zentrum für Angewandte Informatik (ZAIK) (<http://www.uni-koeln.de/rzrk/>) der Universität zu Köln implementiert und im Folgenden erläutert.

2 Aufbau der TR32DB

Eine wesentliche Aufgabe des TR32-Datenmanagement-Teilprojektes ist die Bereitstellung eines Systems zur Speicherung von Projektdaten sowie deren Austausch innerhalb und zwischen den einzelnen Teilprojekten. Weiterhin sind die Datensicherheit und die Speicherung von dazugehörigen Metadaten von besonderer Bedeutung. Des Weiteren bietet das Datenmanagement-Teilprojekt Dienstleistungen aus den Bereichen GIS und Fernerkundung für das gesamte Projekt an [WB08].

Resultierend aus dem interdisziplinären Hintergrund des TR32 werden in der TR32DB-Projektdatenbank unterschiedliche Daten mit den dazugehörigen Metadaten entsprechend den jeweiligen Normen und Standards gespeichert (z. B. von ISO und OGC). Hierzu gehören neben Messdaten, Berichten, Fotos, Präsentationen usw. auch lizenzierte räumliche Daten. Diese Geodaten sind aus den Bereichen Topographie, Landnutzung, Klima und Fernerkundung sowie verschiedene Umweltdaten (u. a. hydrologische, geologische, Wald- und Bodenkarten), die größtenteils für das gesamte TR32-Untersuchungsgebiet in unterschiedlichen Maßstäben vorliegen.

Wie in Abb. 1 dargestellt, besteht die Struktur der TR32DB aus den Komponenten (i) Datenspeicher, (ii) Datenbank und (iii) Web-Oberfläche mit (iv) integriertem Web-GIS.

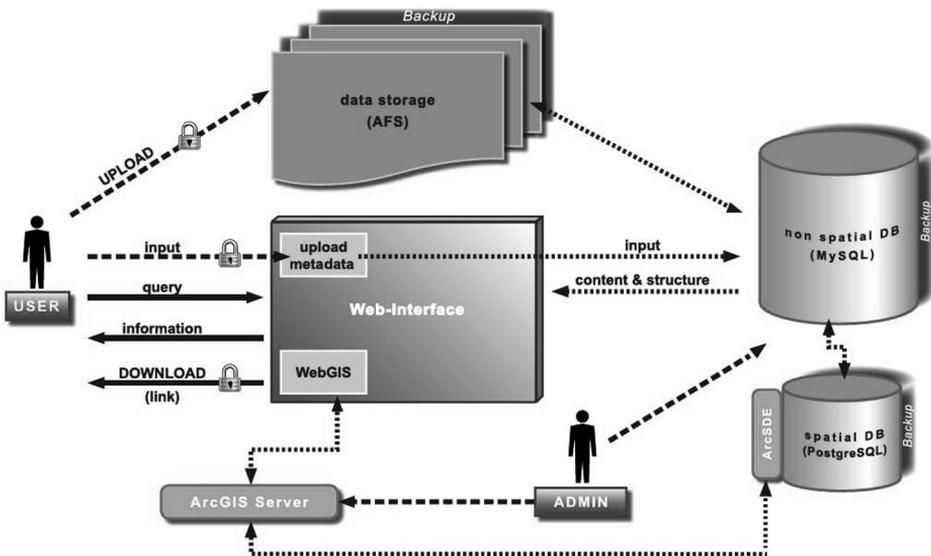


Abbildung 1: TR32DB – Struktur [CHW08]

TR32DB-Datenspeicher

Der TR32DB-Datenspeicher (vgl. Abb. 1) ist der physische Speicher- und Sicherungsort der Projektdaten. In Kooperation mit dem ZAIK/RRZK erfolgt die Speicherung der Daten im Andrew File System (AFS). Die wesentlichen Vorteile vom AFS liegen in einer leichten Dateiverwaltung, Benutzerzugriffsrechten, Skalierbarkeit, Plattformunabhängigkeit sowie einfachen Archivierungs- und Back-up-Möglichkeiten. Das AFS ist innerhalb der Struktur am RRZK zusammen mit dem High-Performance Computing-File System (HPC) Teil des Storage-Area-Network (SAN) (vgl. [Ka05]).

Im AFS sind die einzelnen Daten in einer Ordnerstruktur organisiert und entsprechend der Struktur des TR32 sortiert nach den einzelnen Clustern, Teilprojekten und Datentypen (z.B. gemessene Daten, Bilder oder Vorträge) gespeichert. Authentifizierte Projektmitarbeiter haben die Möglichkeit ihre Daten direkt in den für sie freigegebenen Bereich der TR32DB zu verschieben. Über ein Skript wird ein automatisierter Eintrag in die TR32DB-Datenbank vorgenommen, der Informationen wie Dateiname, -format, -typ und Speicherort beinhaltet.

TR32DB-Datenbank

Eine Datenbank bietet die Möglichkeit, eine Vielzahl von Daten beständig, nachhaltig und effizient zu verwalten. Weitere Vorzüge bestehen u.a. in der Zentralität, Datenunabhängigkeit, Mehrbenutzerbetrieb, Datensicherheit, Datenintegrität sowie Anfragebearbeitung (vgl. [Br08]).

Die TR32DB-Datenbank besteht aus zwei Teilen (vgl. Abb. 1). Der erste Teil wird durch eine MySQL-Datenbank implementiert, in der beispielsweise administrative TR32-Daten wie Anwender, deren Zugehörigkeit zu Teilprojekten oder dessen Rechte gespeichert werden. Abhängig vom jeweiligen Datentyp werden außerdem die entsprechenden Metadaten zu den Daten im Datenspeicher vorgehalten. Ihre Struktur sollte sich an aktuelle Standards und Normen richten (vgl. [NZM05]). Die TR32DB-Metadaten sind an die aktuelle INSPIRE-Richtlinie (<http://inspire.jrc.ec.europa.eu/>) angepasst. Die Darstellung der Informationen aus der Datenbank sowie die Metadateneingabe erfolgt über eine Web-Oberfläche. Der zweite Teil ist durch eine PostgreSQL-Datenbank umgesetzt. In dieser werden räumliche Daten oder gemessene Projektdaten mit räumlichem Bezug vorgehalten.

TR32DB-Web-Oberfläche mit integriertem Web-GIS

Die TR32DB-Web-Oberfläche mit integriertem Web-GIS greift auf Daten im Dateisystem und in der Datenbank zu (vgl. Abb. 1). Ihre wichtigste Aufgabe ist die Visualisierung der unterschiedlichen Daten und deren Bereitstellung. Anwender können abhängig von ihren Rechten (und den Datenlizenzen) u.a. Daten suchen, herunterladen oder Metadaten zu ihren Daten eingeben. Aufgrund von lizenzierten und gemessenen Daten ist die Bereitstellung einer stabilen, sicheren Web-Oberfläche (vgl. Abb. 2) erforderlich.

Die Umsetzung der TR32DB-Web-Oberfläche erfolgt u. a. mit der Auszeichnungssprache HTML/XHTML, der Skriptsprache PHP sowie der Stylesheet-Sprache CSS. Eine Kombination aus PHP und SQL ermöglicht die Interaktion zwischen Datenbank und Web-Oberfläche. Die Sicherheit der Web-Oberfläche wird mittels SSL-Technologie umgesetzt.

Im integrierten Web-GIS werden räumliche Projektdaten, die in einer Geodatenbank

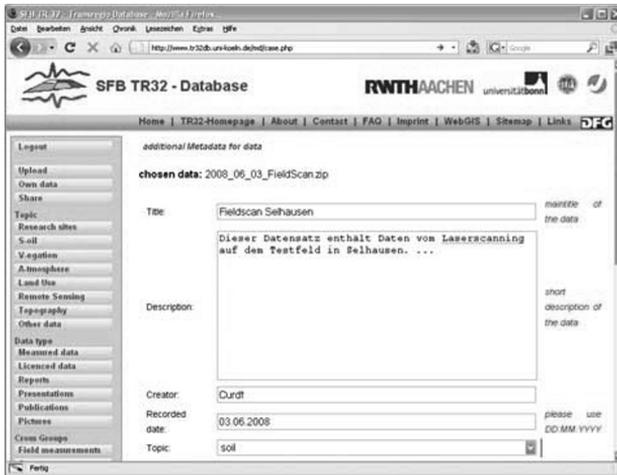


Abbildung 2: TR32DB-Web-Oberfläche mit Metadateneingabe

vorgehalten werden, visualisiert und verwaltet. Ebenfalls werden u. a. Analysefunktionen bereit gestellt und Web-Mapping Services für Projektmit-arbeiter angeboten. Die Umsetzung erfolgt mit ESRI ArcGIS Server, der mit Hilfe der ESRI ArcSDE-Technologie auf die objektrelationale Datenbank PostgreSQL mit der Erweiterung PostGIS zugreift.

3 Zusammenfassung und Ausblick

Neben dem Management von Daten ist für den Forschungsansatz des TR32 ebenfalls die Vorhaltung von räumlichen Daten in verschiedenen Maßstabebenen für die räumliche SVA-Modellierung notwendig. Daher ist eine Geodateninfrastruktur (GDI) entsprechend zu gestalten. Beim Aufbau dieser GDI für das Untersuchungsgebiet der Rur wird die aktuelle INSPIRE-Richtlinie berücksichtigt. Im Hinblick auf eine prognostizierte Vielzahl auch von großen Dateien in den unterschiedlichen Maßstabebenen wird eine sichere, stabile und gut strukturierte TR32DB Struktur innerhalb der Umgebung des Rechenzentrums entwickelt. Dieses ermöglicht auch nach dem Ende der Projektlaufzeit den Zugriff auf Projektdaten in der TR32DB für TR32 Mitarbeiter und andere Interessierte.

Literaturverzeichnis

- [Br08] Brinkhoff, T.: Geodatenbanksysteme in Theorie und Praxis, Wichmann, Heidelberg, 2008.
- [CHW08] Curdt, C., Hoffmeister, D., Waldhoff, G., Bareth, G.: Spatial Data Infrastructure for Soil-Vegetation-Atmosphere Modelling: Set-up of a spatial database for a research project (SFB/TR32).-Proc. XXI ISPRS Kongress, Peking, China, Juli 2008.
- [Ka05] Kalle, C.: Kolloquium Ausgewählte Themen der Datenverarbeitung „Datenhaltung am ZAIK/RRZK“, Köln, 2005. http://www.uni-koeln.de/rrzk/multimedia/kolloquium/ws0506/_pdf/Datenhaltung-am-ZAIK.pdf (20.10.2008).
- [MN04] Mückschel, C., Nieschulze, J.: Editorial zum Schwerpunktthema dieser Ausgabe: Datenmanagement in interdisziplinären Umwelt-Forschungsprojekten. Zeitschrift für Agrarinformatik, Heft 4, Jahrgang 2004.
- [NZM05] Noguerras-Iso, J., Zarazaga-Soria, F. J., Muro-Medrano, P. R.: Geographic Information Metadata for Spatial Data Infrastructures, Springer, Berlin, 2005.
- [WB08] Waldhoff, G., Bareth, G.: GIS- and RS-based land use and land cover analysis-case study Rur-Watershed, Germany. - SPIE Proc. Geoinformatics 2008, China, 2008.