Central data management in environmental research projects – selected problems and solutions

Claus Mückschel, Cornelia Weist, Wolfgang Köhler

Biometrie und Populationsgenetik
Justus-Liebig-Universität
Heinrich-Buff-Ring 26-32
D-35392 Giessen
claus.mueckschel@agrar.uni-giessen.de

Abstract: Collection and provision of all data, which may differ in spatial and time scales, and ensuring their long-term availability and readability are the primary goals of a central data management in a research project. One approach is setting up a central information system, which visualizes the manifold data of all subprojects in a uniform format and makes them publicly available. The setup and the successful operation of an information system faces technical and thematic challenges.

The information system is based on a classic client/server-architecture. Several routine tasks of the data management are managed by the Open Source Content-Management-System (CMS) TYPO3. This has proved to be a good tool to improve effectively interdisciplinarity and exchange of research results.

This paper introduces the concept as well as the system architecture of the central data management of the interdisciplinary research project ”Land use concepts for peripheral regions” (SFB 299) supported by the German Research Foundation (DFG), which is now in its 12th year of funding. Several adaptations of the database concept as well as customizations of system components and functions for visualization and analyses are discussed. The adaptations have become necessary due to technical or thematic aspects. Elementary problems of data management are shown and different approaches are discussed and in conclusion evaluated in respect to each other.

1 Introduction

Research projects with environmental focus, like the SFB 299, involve different scientific disciplines which produce large datasets. Thus an interdisciplinary data management is necessary to provide effective mechanisms for storage, investigation, data transfer and archiving. Furthermore the system should support communication between the different disciplines. A central information system tackles these tasks.

Usually the amount of data poses no technical problems. Challenges result from the need to structure heterogeneous datasets following strict rules, gather and filter information and provide scientists and institutions with the necessary data on the long term.

2 System requirements and solution strategies

2.1 System architecture

Due to constantly changing requirements and focuses of the supported SFB subprojects and to follow the information technological advance the information system has to be under constant development and customization. This leads inevitably to a high dynamic within the data management. Thus an easy way of extending the data model despite the complexity of system and data is a pivotal precondition for the successful maintenance of an information system. Furthermore it must be possible to enter content from any location but govern the data centrally. Since the large amount of data and because of changing co-workers the implementation of structures for the long term interpretability of all datasets is of special importance.

The system developer has to decide on the use of commercial or open source products. Besides the costs for implementation and maintenance the selection criteria in the SFB 299 are scalability, stability, compatibility with other already used or planned products and modifiability. [AFH06].

The open source strategy with their mostly platform independent systems and the open source code offers advantages. Especially missing functionalities can be added or existing tools can be modified through self programmed extensions. Project specific adaptations and further developments can be integrated into the existing projects without major problems. The use of standardized interfaces largely guarantees universality, reusability, exchangeability, expendability as well as the straightforward usability of the information system and its modules.

2.2 System acceptance

Besides thoroughly planned data storage, conception must consider usability and acceptance of the (complete) system. The user's direct involvement in the development process is by far the best way of quality control. Furthermore the users get familiar with the system early and it allows them to make the comprehensible connection of the stored data.

During the course of the whole project only the constant cooperation and communication between experts from applied disciplines and developers from the field of information technology ensures the continuous adaptation to the user's needs. Special attention has to be paid to the system usability. Ease of use is an indispensable precondition for system acceptance. Simple, intuitive and preferably familiar interfaces facilitate the use of complex systems and increase the motivation for central (meta) data storage. In this context browser based graphical user interfaces are to mention which need no additional client software. Another example is the implementation of a central browser based WebGIS-component into the existing information system. The set up of a project wide WebGIS offers the advantage that even SFB co-workers with little or no GIS experience can perform simple GIS based analyses.

2.3 Data dynamic

The requirements of the subprojects concerning data storage differ strongly depending on their project specific needs and underlie a temporal dynamic. This makes it difficult to determine the metadata frame a priori. Already in the beginning of the SFB 299 in 1997 a first recommendation was given to the subprojects concerning data transfer and data quality. During the course of the project the metadata concept has been adjusted and customized several times, because the data complexity grew stronger as it could have been conceived beforehand. Therefore the metadata was concretized and extended repeatedly. The further development of the database system had to be regarded as well. Due to this progression the metadata file has been structured in a way which allows extending it but still assures an easy handling on any platform.

3 Current system architecture of the SFB information system

The information system of the SFB 299 is conceived as an open system, which comprises of different system modules and services. They are a collection of comfortable query tools with respect to thematic, temporal, spatial and context aspects. The setup is classic client server architecture [Mü04]. Typo3, a content management system on the basis of PHP and MySQL, is a central component (http://typo3.org). The strict separation of content, layout and function, enabled by Typo3, has optimized the use and management of the SFB information system and has increased the automation grade of several workflows. Typo3 fits well into the existing data structures, which were realized by LAMP systems. A WebGIS-application serves for evaluation, analysis and presentation of spatial data. It is based on the open source software UMN Mapserver and Mapbender. These products allow for the use of OGC compatible standards like Web Map Service (WMS), Web Feature Service (WFS) and Web Map Context (WMC) [Op07]. The database PostgreSQL serves as data server. The spatial language extension PostGIS allows the storage of spatial data. It contains GIS like analysis functions and supports the transformation of coordinates into different coordination systems. Fig. 1 gives an overview over the implemented system components.

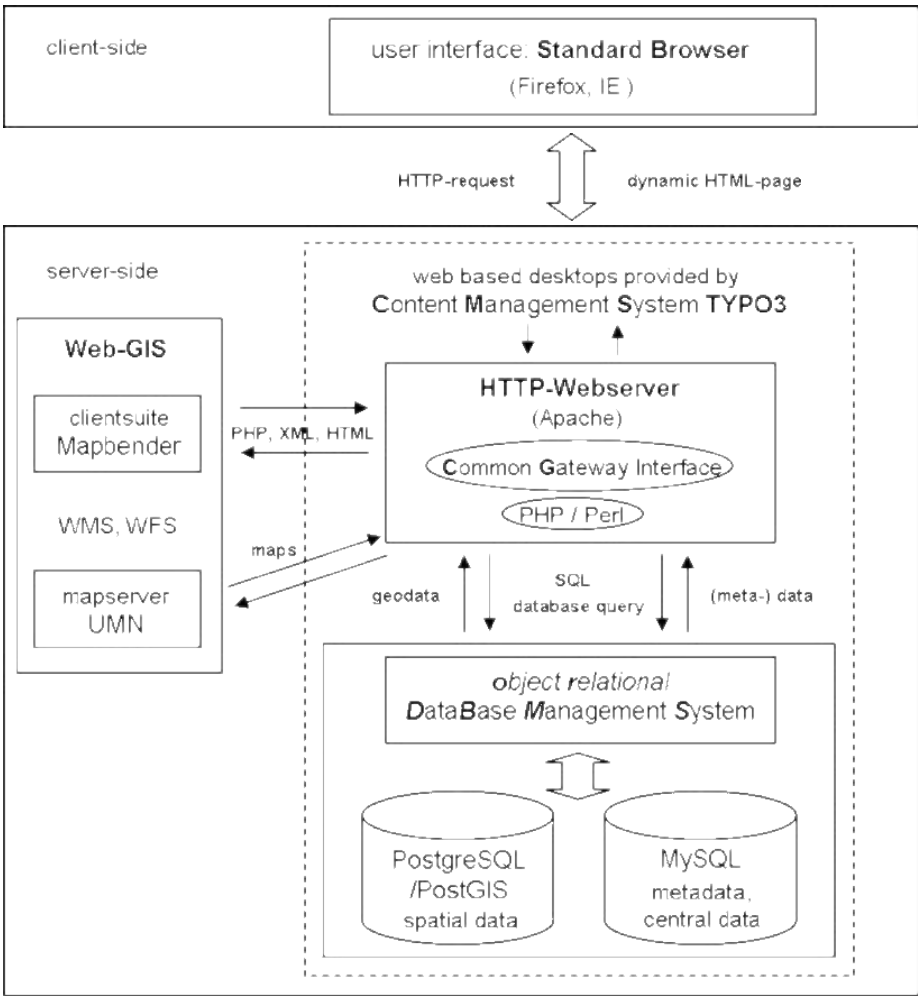

Fig. 1: Schema of the architecture of the SFB information system and implemented components

4 Literature

[AFH06] Altmann, W., Fritz, R., Hinderink, D.: Typo3 Enterprise Content Management. Open Source Press, München, 2006.

[Mü04] Mückschel, C. et al.: Web-basierte Informationssysteme in interdisziplinären Umwelt-Forschungsprojekten – am Beispiel der beiden DFG-Sonderforschungsbereiche 299 (Gießen) und 552 (Göttingen/Kassel). Zeitschrift für Agrarinformatik, 3, 2004; S. 46-55

[Op07] Open Geospatial Consortium, Inc.: OpenGIS Specifications. http://www.opengeospatial.org/standards/ - access August 2007.