

CRC806-Database: Implementation of a data management for an interdisciplinary research project



C. Willmes, D. Hoffmeister, C. Hütt, D. Kürner, K. Volland,
A. Bolten and G. Bareth

Institute of Geography, University of Cologne

Introduction

For the data management of the Collaborative Research Centre 806 „Our way to Europe“ (CRC806), funded by the German Research Agency (DFG), at the Universities of Cologne, Bonn and Aachen, the here presented CRC806-Database is in development.

Within the 21 sub-projects of the CRC806 about 30 Post-Docs, 60 Ph.D. candidates and more than 100 graduates from the archaeological-, geo- and cultural- sciences working together and producing diverse data, to answer the questions of how modern humans (*homo sapiens sapiens*) migrated from Africa to Europe and why and how the Neanderthals (*homo neanderthalensis*) disappeared.

The CRC is planned for an overall period of twelve years, and is subdivided into three four-year project terms which are underlying an evaluation at the end of each term. In the sense of the DFG proposals for safeguarding good scientific practice (Deutsche Forschungsgemeinschaft, 1998) it is required for any CRC to provide a sustainable data management policy and its implementation, which guaranties the accessibility to the produced research data for at least the duration of the project plus ten years, which results in an up to 22 years period.

Data management

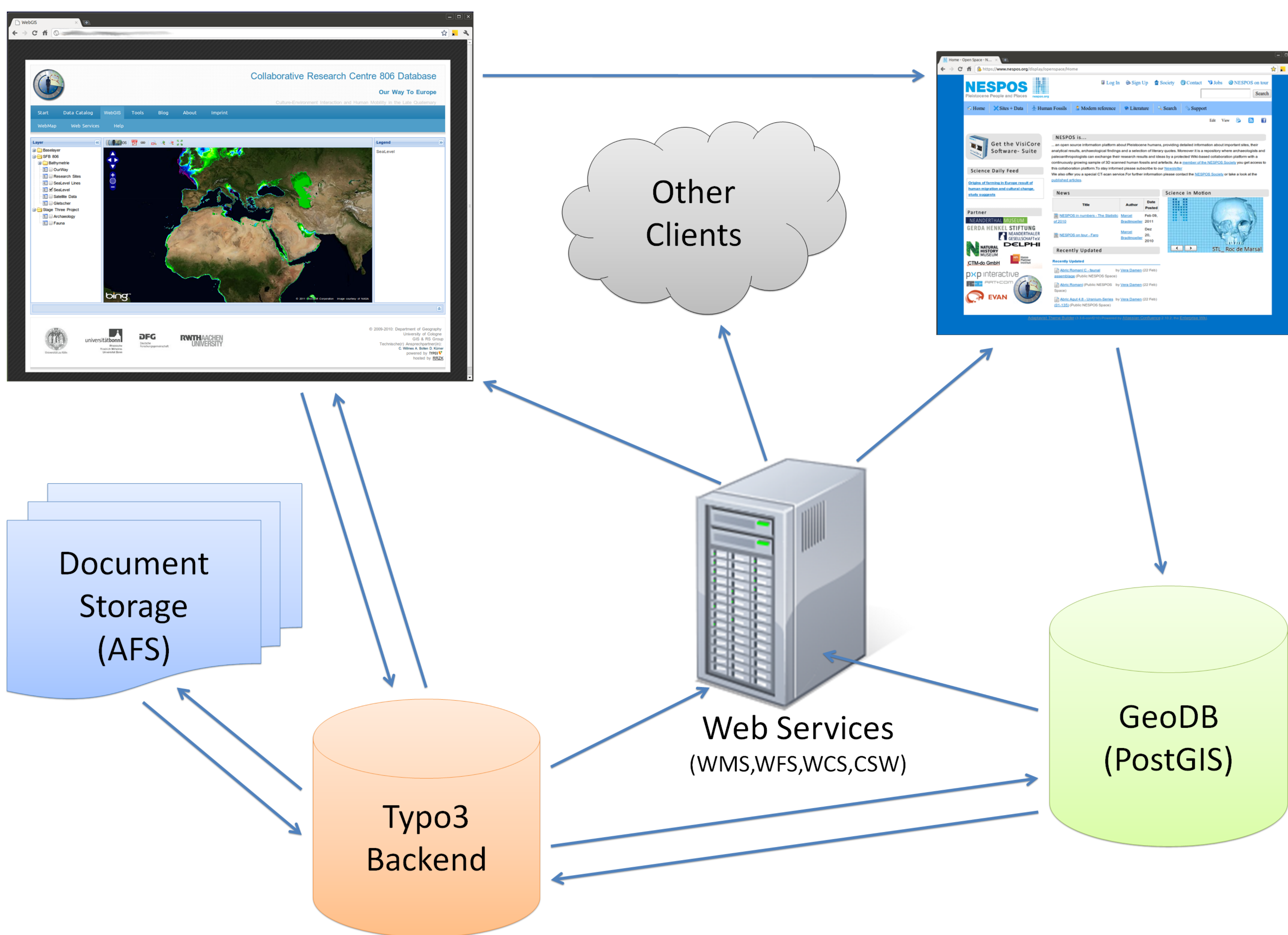
The aim of the project is to provide a data integration, data sharing and collaboration hub for the researchers of the CRC, and as well an information platform about the project for the interested public.

Data integration means to provide the possibility to aggregate and combine different data sets of the database into one visualization, processing and/or catalog interface.

Data sharing means to provide an intuitive interface for data contribution or upload into the data base and to provide access using well known (preferably open) data formats and to provide standard interfaces to the data for automated access by other (web-) applications.

Collaboration means to provide tools like centralized calendars, ticketing systems, wikis, mailing lists and anything that enhance the collaboration between the researchers of the project.

Additionally the data management project offers several services, like data purchase, data preparation, analysis/processing and visualizations as well as general technical support to help the researchers work with the data and tools of the CRC806-Database.



System architecture of the CRC806-Database data management infrastructure.

System architecture

All functionality of the CRC806-Database is accessible through the web portal, the geodata is additionally accessible through standard based OGC [1] interfaces like CSW, WMS, WFS or WCS for use with desktop applications or for automated integration with other (web-) applications.

The implementation of the web portal and the underlying data management infrastructure is built on several open source software applications, such as Typo3 [2], Geoserver [3], deegree [4], PostgreSQL/PostGIS [5], Openlayers [6] and GeoExt [7] as well as ESRI's ArcGIS Server [8] technology.

The user rights and web content management is facilitated through Typo3, utilizing the Typo3-based multi-level user access control system for fine grained access to each dataset granted by the dataset authors.

All automated data management tasks are controlled from the Typo3 backend, this covers the data upload and download, data synchronization between the different storages (AFS, GeoDB). Two extensions for the Typo3 framework are developed, one to manage the data workflow (upload, metadata handling and synchronization of the GeoDatabase and the catalog interface) and one to synchronize user rights between the Typo3 user database and the deegree owsProxy, which manages the access to the OGC services (WMS, WFS, WCS, CSW).



www.sfb806.de



www.sfb806db.uni-koeln.de

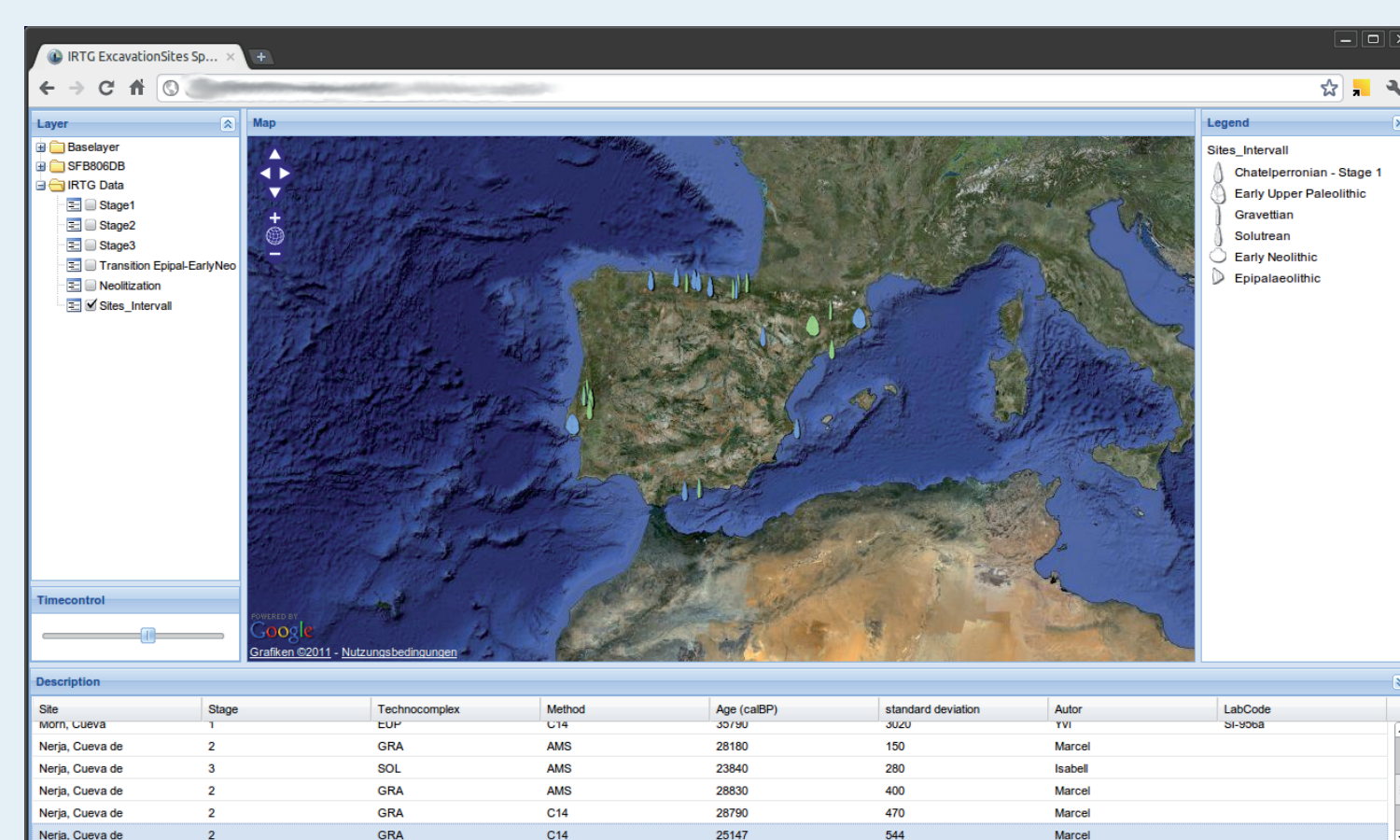
WebGIS

An important aim of the web portal is to provide functionality and tools for data analysis, visualization and integration to gain additional insight into the different datasets and help to abstract maybe new patterns by combining, analyze and filtering the data through the WebGIS interface in its spatial and temporal context.

The CRC produces a heterogeneous variety of data. There are records for archaeological excavations sites and archaeological artifact findings. There are data records from the geoscientists which describe climate archives, and there are data sets produced by the ecosystem-, socialsystem- and climatesystem- modelers.

All those data sets have a spatial and temporal extent, which allows to integrate those data sets in a spatio temporal WebGIS application for further analysis and visualization (see the screenshot of the example application).

The WebGIS application is built using the OpenLayers [6] and GeoExt [7] Libraries for building rich web mapping clients. The temporal visualizations and the geo-computation functionalities are implemented using the i2maps [9] framework.



Screenshot of an spatio temporal WebGIS for archaeological excavation sites.

Data integration

An other aim of the CRC806-Database project is to offer available relevant data for the project to the researchers. The CRC806-Database integrates many freely available geo datasets, like OpenStreetMap, Landsat Satellite Images, SRTM and ASTER DEMs, and several other freely available data sets from sources like NaturalEarth, GeoCommons and others.

The CRC806-Database maintains a collaboration with the NESPOS project [10]. NESPOS is an open source information platform about Pleistocene humans, providing detailed information about important sites, their analytical results, archaeological findings and a selection of literary quotes. The CRC806-Database provides the technology and infrastructure to extract the spatio temporal informations about the data form the NESPOS project and maintain spatio temporal data sets, which will be accessible through the CRC806-Database WebGIS and will be used to provide a webmap application for the NESPOS platform.

The interface to the NESPOS Database is implemented by an implemetaion of a PlugIn into the confluence wiki software on which the NESPOS database is based. The PlugIn writes the spatio temporal coordinates and some parameters for each dataset entered into NESPOS data base into a table of the GeoDatabase of the CRC806DB. This allows to extract an always up to date spatio temporal dataset out of the comprehensive NESPOS database which is then available for further research within the CRC806 and NESPOS.

References

Deutsche Forschungsgemeinschaft (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft". Denkschrift: Prposals for safeguarding good scientific practice, Wiley-VCH, Weinheim.

Internet References

- [1] OGC - Open Geospatial Consortium: <http://www.opengeospatial.org/>.
- [2] Typo3 Enterprise CMS: <http://typo3.org/>.
- [3] Geoserver: <http://geoserver.org/>.
- [4] deegree Framework: <http://www.deegree.org>

- [5] PostGIS: Spatial Extension for PostgreSQL: <http://postgis.refractions.net/>.
- [6] Opnlayers - JavaScript Web Mapping Framework: <http://openlayers.org/>.
- [7] GeoExt - JavaScript Toolkit for Rich Web Mapping Applications: <http://geoext.org/>.
- [8] ESRI ArcGIS Server: <http://www.esri.com/software/arcgis/arcgisserver/index.html>.
- [9] i2maps - The GeoComputation Platform: <http://ncg.nuim.ie/i2maps/>.
- [10] NESPOS - Pleistocene People and Places: <http://www.nespos.org/>.