PaleoMaps: SDI for open paleoenvironmental GIS data*

Christian Willmes[1], Daniel Becker[1], Jan Verheul[1], Yasa Yener[1], Mirijam Zickel[1], Andreas Bolten[1], Olaf Bubenzer[1], Georg Bareth[1]

[1]Institute of Geography, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany. c.willmes@uni-koeln.de, daniel.becker@uni-koeln.de, jverheu1@uni-koeln.de, yyener@uni-koeln.de, m.zickel@uni-koeln.de, andreas.bolten@uni-koeln.de, olaf.bubenzer@uni-koeln.de, g.bareth@uni-koeln.de

Abstract

Paleoenvironmental studies and corresponding data are abundantly published and available in scientific records. However, paleoenvironmental data sets are comparatively rarely provided in GIS data formats. Here, we present an Open Science approach for collecting and creating GIS data, visualizing it in maps of paleoenvironments, and publishing them in a web-based Spatial Data Infrastructure (SDI), for access by the archaeology and paleoenvironment communities. The Open Science approach to the publication of data allows to properly cite the published data sets as bibliographic sources in research that builds upon these data sets. This paper has its focus on the implementation and setup of the Free and Open Source Software for Geospatial (FOSS4G)-based SDI, and on the workflow for compiling and publishing the GIS data.

Keywords: Paleoenvironment, GIS, SDI, Spatio-Temporal, Data integration, Data management, FOSS4G, Open Science, Open Access, Open Data.

1. INTRODUCTION

Paleoenvironmental information from reconstructions, simulations, or qualitative synthesis depicted and localised on maps is most often not available in Geographic Information Systems (GIS) data formats, i.e. as GIS data. This information is mostly published in written text, or in tables, figures and maps as part of scientific publications.

Where paleoenvironmental data is available in digital form, it is most often provided in non-GIS data formats, such as CSV or spreadsheets without intrinsic geographic information associated to it. However, a quite sizeable amount of paleoenvironmental information is available in the literature and also in analogue map publications.

The main aim of our project is to make this existing information available in GIS data formats, to collect, store, analyse and publish these GIS data for further use by archaeologists and paleoenvironment researchers.

The idea of "PaleoMaps" was conducted in the framework of a large interdisciplinary research project. Scientists from the Universities of Cologne, Bonn, and Aachen are researching Culture-Environment interaction in the late Quaternary in order to answer questions about the complex nature of the chronology, regional structure, and the climatic, environmental and socio-cultural contexts of major intercontinental and transcontinental events of dispersal of anatomically modern humans (AMH) from Africa to Western Eurasia, and particularly to Europe (Richter et al., 2012). The resulting maps and GIS data sets of the here presented approach are published in the Spatial Data Infrastructure (SDI) of the CRC806-Database[1] (Willmes et al., 2014), which is the web-based Research Data Management (RDM) infrastructure of the interdisciplinary and inter-institutional Collaborative Research Centre 806[2] (CRC 806). Within the Data Management and Data Services (Z2) sub-project of the CRC 806, the here presented approach is applied for providing GIS data and their visualisation in maps, according to the given research questions, for the collaborating projects, and also via online publication to the wider research community.

---

[1] http://crc806db.uni-koeln.de
[2] http://www.sfb806.de

The implementation of Open Science, Open Access and Open Data principles, as well as the application of Open Source Software is a key for the realization of PaleoMaps. Opening up closed or not easily accessible data resources by creating GIS data sets from the published information, and publishing these GIS data sets in open formats and standards, will hopefully help to advance progress in the domain of paleoenvironmental and archaeological studies. The scope of the paper is to describe the idea of and approach to provide open paleoenvironmental GIS data via an SDI that complies with the standards of the Open Geospatial Consortium (OGC).

2. PALEOMAPS WORKFLOW

The data sets are compiled and created mostly by the small team of the Z2 "Data Management and Data services" sub-project of the CRC 806, consisting primarily of the authors of this publication. It is also possible that contributions from other members of the CRC 806, or even from other institutions, can contribute to the PaleoMaps data collection.

The workflow for the creation of a PaleoMaps GIS data set from idea or demand to the published data set can be separated into the following six steps:

1. Collection of demands, ideas and suggestions for paleoenvironmental GIS data sets and maps.
2. Collection and research of data sources, applicable GIS modelling techniques, and related publications (see section 3).
3. Assessment if sufficient resources for creating the paleoenvironmental GIS data set (or map visualization) are collected, compile the resulting GIS data set (see section 4).
4. Development of individual workflow for creating the specific data - to be documented as metadata in the data set description document (see section 5.2).
5. Production of cartography for map and WebGIS publishing (see section 4.3).
6. Creation of metadata and write documentation (how was the GIS data produced, referencing data sources, provenance information; see section 5).
7. Publication of the data set (upload the GIS data, maps and metadata including documentation to the CRC806-Database data repository and publish the GIS data via the CRC806-SDI; see sections 5 and 6).

The sub-processes given in the above workflow are described in the following.

3. DATA COLLECTION

The approach for collecting data mainly consists of map digitalization (scanning, georeferencing and digitizing), or of GIS modelling from textual information or other published paleoenvironmental information, which is not yet available in GIS data formats. Thus, the focus of this paper is primarily on the acquisition of qualitative and analogue paleoenvironmental information, and on producing GIS data sets representing this information. These collected and created data sets are then combined with existing GIS data, in order to produce comprehensive paleoenvironment data sets and maps in GIS data formats. In section 3.2, some of these heterogeneous data sources are introduced in greater detail.

3.1. The Knowledge base

In order to manage and organize the different data sources and the derived or created GIS data, as well as to connect these data sets to their bibliographic, spatio-temporal and thematic contexts, a web-based Knowledgebase (KB) application was developed, called the CRC 806 Knowledgebase (CRC806-KB).

3.1.1. Technology & Architecture

The CRC806-KB is a Semantic MediaWiki (SMW) (Krötzsch et al., 2006) -based collaborative editable wiki database application. SMW (SemanticMediaWiki Contributors 2015) is an extension of the well-known MediaWiki (Wikimedia Fdn., 2015) Open Source Wiki implementation, which is known as the software basis of the Wikipedia platform. Both SMW and MediaWiki have an active developer community, mostly funded through the Wikimedia Germany Foundation which supports the active development of the application. In the standard setup of SMW, the model development, especially for larger models, can be regarded as cumbersome, because data model classes and their properties have to be edited and maintained through markup on several different wiki pages. To update or apply changes to the data model, several pages have always to be updated accordingly. In order to manage this unwieldy update process automatically, a software tool named Mobo (Heimler, 2014) was applied, which allows formulation of the model as a Domain Driven Design (DDD) pattern in text-based YAML or JSON markup.

The formulation in plain ASCII text allows for handling the model via a version control system (VCS) such as Git, which facilitates collaborative editing during the

development process of the data model. The wiki-based CRC806-KB application can also be defined as a virtual research environment (VRE). This application allows researchers of the CRC 806 to collaboratively gather and edit information, and build a knowledge base in a collaborative manner. Internal resources and data sets, which are not yet compiled, edited and published as proper GIS data sets through the SDI, can be stored in an internal shared storage system. The CRC806-KB[3] is accessible publicly, but editable only by members of the CRC 806, who own an account for the CRC806-Database.

3.1.2.   Data model

To organize the data sets, resources, and bibliography, as well as their spatial, temporal, and thematic contexts, a simple, straightforward data model was developed.

The data model of the CRC806-KB consists of the following main sub-models:

- Data set
- Bibliography
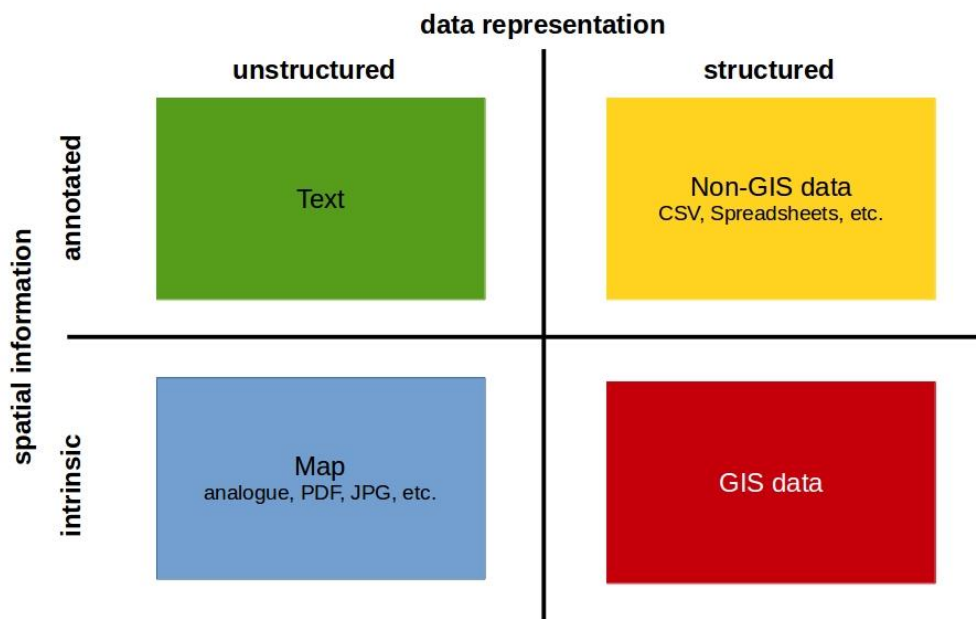- Resource
- Spatial
- Temporal
- Topic

The core of the model is the data set, which can be annotated with entities of the types Bibliography, Resource, Spatial, Temporal and Topic. To express, for example, a data set called "Inland waters of Europe during the last glacial maximum (LGM)", the data set is annotated with two GIS data resources (rivers and lakes), a spatial context (Europe), a temporal context (LGM), a topic (inland waters), and bibliographic sources, from which the river and lake geometries of the LGM were acquired.

3.2.   Data sources

The four basic types of sources for data that build the basis for the creation of the PaleoMaps (see Figure 1), are described in the following sub-section.

---

[3] http://www.sfb806db.uni-koeln.de/wiki

Figure 1: Types of data sources.

3.2.1. Unstructured data sources

The term "unstructured" refers to all data sources, which are not given in a computer processable form, in the sense of a spreadsheet, GIS data, an image produced by a sensor, or some other form of data that can be used as an input for further computational and algorithmic processing.

<u>Maps</u>

Maps published in print form or as an image bitmap in non-GIS data format (see Figure 2 for example), need to be transformed into spatial features (in raster or vector format), by georeferencing and eventually also by digitization.

Figure 2: Example of analog published maps. Source: (Pachur and Altmann, 2007).
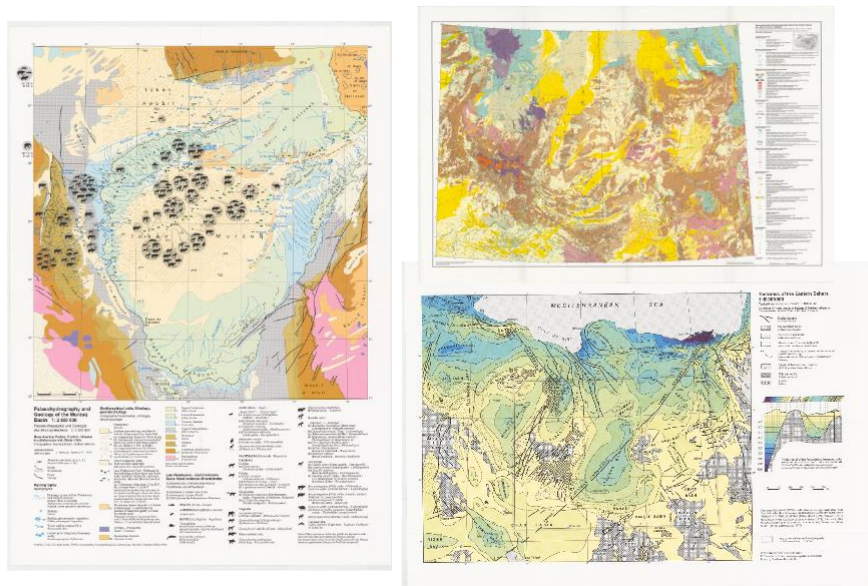
Text data

Spatial features or locations described in floating texts, for example, in a scholarly paper, can be modelled in a GIS, according to the given alphanumeric information, as spatial features. An example would be a glaciation extent, that is described as an area between well-known spatial features, e.g. "the maximum extent of an ice shield between town X and lake Y".

3.2.2.   Structured data sources

In case of digital data, the reusability is much more convenient, even if it is not available in GIS format. If a spatial context is somehow present in the data, it is possible to transform the data into GIS formats for further use in the presented context.

Non-GIS data

As non-GIS data we understand, for example, data given as CSV, spreadsheet, XML, JSON or relational data bases that contain an informal spatial reference, i.e. a field for coordinates, but without explicit definition of a spatial reference system etc., which allows creation of a GIS data set, mostly by formulating a script to read in the CSV or spreadsheet and write it out as a data set in a proper GIS data format.

GIS data

By GIS data sources we understand all kinds of data sources, which can be directly read and displayed in a desktop GIS such as QGIS (QGIS Development Team, 2016).

For example, topography data is available in abundance for the present day spatial context. With regard to paleoenvironments this is much scarcer, of course, because almost always a certain model for representing the paleotopography is required. An example would be the coastline representing the sea level during the LGM, which was 120 meters below the level of today. However, the required bathymetry data can be derived from the GEBCO data set (General Bathymetric Chart of the Oceans, 2014).

4. MAP AND GIS DATA SET COMPILATION

Data sources that are published in the here presented approach can be compiled and created in different ways.

4.1. Spatio-temporal context

An example for the compilation of a spatio-temporal context data set from the CRC806-KB would be the "LGM paleoenvironment of Europe – Map" (Becker et al., 2015), which is shown in Figure 3. For such maps, data sources that match a given spatio-temporal context are gathered and / or created. In the example map, shown in Figure 3, data sources annotated with LGM that are spatially situated within Europe are shown. These data sources are; the -120 meters sea level landmass polygon data set, the Köppen-Geiger climate classification of the LGM data set (Willmes et al., 2017), the LGM glaciation extents data set, the LGM rivers vector data set, and the LGM lakes data set, which were compiled into one GIS data set and displayed in a corresponding map. Such compilations are conducted by formulating appropriate queries on the CRC806-KB application that yield all collected and available data in the present collaborative knowledge base.

4.2. Spatial Analysis

Another way to create new GIS data sets that can be published as Open Data is through spatial analysis. In the following section, three such applications are briefly introduced. Some of these approaches are exemplarily discussed in the following section.

4.2.1. Niche-Modelling & Spatial regression analysis

In order to model possible areas, where the prehistoric tribes of Europe settled, GIS-based Regression Analyses (Anemone et al., 2011) and Niche Modelling (Becker et al., 2016) approaches are carried out within the project,-based on excavated evidence of occurrence (sites) and spatial layers of variables, which yield new GIS data that can be published as open data through the here presented SDI and Open Science approach.

Figure 3: Time Slice Map of the LGM paleoenvironment of Europe. Source: (Becker et al., 2015).

4.2.2. Site Catchment & Least Cost Path

The application of Site Catchment Analysis (SCA) and Least Cost Path (LCP) methods (Becker et al., 2017) for determining settlement areas and possible networks and links between these areas also delivers GIS data sets that are of interest for archaeological research within the CRC 806-project.

4.2.3. GIS-based PaleoDEM estimation

GIS-based modelling of landscape evolution, on the basis of Soil Erosion Modelling techniques (Mitasova et al., 2013), will be applied to derive digital elevation models of the earth's surface in prehistoric condition (PaleoDEMs). Such PaleoDEMs can serve as a very valuable input for further Paleoenvironmental modelling approaches.

4.2.4.   Climate classifications

Creating GIS data sets from published climate simulations and deriving further GIS data sets, for example by the application of a classification (Willmes et al., 2017), also yield valuable GIS data sets that are managed and published via the SDI.

4.3.   Cartography

For almost every published GIS data set a corresponding digital map will be produced and published alongside the data set. The process for conducting the cartography is not restricted to specific software and technology. QGIS and ArcGIS are applied in most cases for carrying out the cartography. The selection of the applied software for this task depends mostly on the personal preferences of the person conducting the cartography.

In most cases, the cartography for the publication, for example as WMS, via the SDI is adjusted to fit the specific application. This can also depend on the styling capabilities provided by GeoServer or MapServer, as applicable.

5. PALEOMAPS OPEN SCIENCE APPROACH

The platform that provides access to the CRC806-Database is implemented using Open Source software, and implements Open Science, Open Access and Open Data principles as described in this section. The data published via the platform is by default assigned an open Creative Commons license (Creative Commons, 2015), to implement Open Data. Additionally, the data is openly accessible, and by minting DOI's for data sets, the resources are also citable and thus properly implement the Open Science (Sitek and Bertelmann, 2014) idea.

Open Science is about making scientific research process and most importantly its output (data) accessible to all levels of an inquiring society. It is-based on the following six open principles (Whyte and Pryor, 2011):

- **Open Data** Making research data openly available, and reusable by applying open licenses.
- **Open Source** Publishing the source code, of tools applied for the conduct of scientific analysis, for review and feedback by the community.
- **Open Access** Making scientific publications accessible to the public.

- **Open Peer Review** Transparent and comprehensible and quality assurance, by openly documented review and discussion of proposed research findings.
- **Open Methodology** Detailed documentation of the application of methods including the whole process, for example software toolchains or experiment steps.
- **Open Educational Resources** Free and open resources for teaching and learning, such as MIT open courseware or Massive Open Online Course (MOOC).

Open Science is a development in the academic landscape to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional. The principles of Open Science are now also encouraged by the major research funders. For example, the European-funded project Facilitate Open Science Training for European Research (FOSTER) [4] has developed an open science taxonomy (Pontika et al., 2015) as an attempt to map the Open Science field.

In the following section, the three principles (Open Data, Open Access and Open Source) of Open Science that are implemented in the presented approach are described in detail.

5.1.   Open Data

To clarify the term 'open data', the key sentence of the Open Definition (OKFN) is quoted below:

> *"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)"* (OKFN, 2016).

To facilitate the free reuse of data, it should be made available in an open format that can be processed without proprietary software. At best, this format is even an open standard, but this is not mandatory to count as open data. This demand is successfully addressed by publication of the GIS data in the form of OGC compliant web services via the SDI (see section 6). The aspect of copyright and license is also of importance. The license must allow free reuse and modification of the data.

---

[4] https://www.fosteropenscience.eu

Attribution can be imposed, which also recognizes the good scientific practice of citing a data source, provided that it was used for a study or application. This demand is met, by applying the Creative Commons licenses (Creative Commons, 2015) for all data published through the CRC806-Database.

5.1.1. Creative Commons Licenses

As it is implemented and defined in the data publishing process of the CRC806-Database (Willmes et al., 2014), it is possible to choose between the seven Creative Commons licenses, CC-BY (Attribution), CC-BY-SA (Attribution, Share Alike), CC-BY-ND (Attribution, No Derivatives), CC-BY-NC (Attribution, Non-Commercial), CC-BY-NC-SA (Attribution, Non-Commercial, Share Alike), CC-BY-NC-ND (Attribution, Non-Commercial, No Derivatives), and CC0 (public domain).

The authors of the to-be-published data can determine the license on the basis of their own reasoning. From the CRC806-Database administration side, one of the three true open data CC licenses (CC0, CC-BY, or CC-BY-SA) is recommended, and CC-BY is preset by default, but the data set authors are not obliged to choose this license.

By the application of Tim Berners-Lee's 5 Star open data concept and implementation of the corresponding badges displayed on the data set landing pages, we try to provide an incentive for data set authors to adhere to open data principles.

If PaleoMaps data sets are published by the CRC806-Database staff the default license for these data sets is CC-BY, and because most PaleoMaps are published through the staff, almost all PaleoMaps data sets will be Open Data. Until now (late 2016), no non-open data sets are published as PaleoMaps data.

5.2. Open Access publication

To fulfil a central demand for Open Science, the resources and data sets that are published as Open Access need additionally to be accessible without access constraints, to be uniquely referenceable in a way that they can be referenced and cited in a scholarly publication. This is facilitated through the application of Digital Object Identifiers (DOI) for published resources and data sets.

5.2.1.   DOI publication for scholarly citation

The produced GIS data sets and maps are published, including a DOI minted via DOIDB (Ulbricht et al., 2016), the cooperation partner of the CRC 806 for issuing DOIs. DOIs are generally minted for web accessible resources with corresponding metadata. The metadata has to validate against a formalized metadata schema, the 'DataCite Metadata Schema for the Publication and Citation of Research Data' (DataCite, 2014).

For example, the proper scholarly citation for such an open access published GIS data set would be:

> Willmes, C., Becker, D., Brocks, S., Hütt, C., Bareth, G. (2014): Köppen-Geiger classification of MPI-ESM-P LGM simulation. CRC806-Database, DOI: 10.5880/SFB806.2

Via the DOI, it is possible to de-reference a webpage that offers the corresponding data and metadata. The above DOI can be de-referenced, for example, via the dx.doi.org service in the following form as a URL:

> http://dx.doi.org/10.5880/SFB806.2

5.2.2.   Data citation guideline

We encourage the users of our data sets to cite the original data sources that went into the used PaleoMaps data source as well, if the data sets are used for further academic research. This practice would ensure that the authors of the original paleoenvironmental research are credited and attributed for their work.

In the case of the exemplary PaleoMaps publication "LGM paleoenvironment of Europe – Map" (Becker et al., 2015), we recommend citing the use of this data set as given in the following quote:

> *The LGM Paleoenvironment GIS data applied for this research is-based on Becker et al. (2015) compiled from published data sources (Clark et al. 2001, 2009, Ehlers et al. 2011, General Bathymetric Chart of the Oceans 2014, Kelso and Patterson 2010, Verheul et al. 2015, Willmes et al. 2017).*

By including all original sources of the cited PaleoMaps data set in the cited publications bibliography, it is ensured that the original authors of the published

data sources that went into compilation of the PaleoMaps data set are appropriately credited for their contributions. See the bibliography of this publication, where all the bibliographic resources of this example are given.

The formulation of the above example can of course be adapted and changed at the discretion of the authors concerned, but the general idea of including all resources should be applied to guarantee good scientific practise.

5.2.3.   Metadata and documentation

The resulting GIS data sets and maps are published with an appended strictly formalized descriptive document (see Figure 4), containing the metadata, according to the data cite schema (DataCite, 2014), and some contextual information, proper and formal citation of the data sources, as well as advice on the further citation of the given data set.

Figure 4: Example CRC806-Database publication metadata and description document of Becker et al. (2015).

The formal required metadata can be separated into three categories:

1. Basic metadata (Datacite schema),
2. Spatial metadata (ISO19115 and further GeoServer supported formats),
3. Temporal metadata (Custom ISO8601-based schema).

Additionally, the textual descriptions of the data sets context, how it was created or derived, as well as the citations of published data sources which were the basis for the given data set, are required as metadata in the form of a bibliography.

5.3.   Open Source

The CRC806-Database and all its components (RDM, SDI, KB) are implemented completely on the basis of Open Source software (Willmes et al., 2014, Willmes 2016), and we additionally prefer to use Open Source GIS software to produce the PaleoMaps data sets.

If we produce scripts or plugins or tools for using or creating paleoenvironmental data, these products (Becker et al., 2016; Willmes et al., 2017) also will be published under Open Source terms and licenses.

6. SPATIAL DATA INFRASTRUCTURE

As introduced above, in addition to the file-based publication of GIS data in open formats, the data sets are also published via OGC Open geospatial Web Service (OWS) in order to enable the networked integration of the data into client desktop GIS and WebGIS applications. To facilitate the OGC web service-based publication, an appropriate SDI was implemented.

The SDI was implemented using free and Open Source software for geospatial applications (FOSS4G). The infrastructure consists of a GeoNode (GeoNode Contributors, 2015) -based backend, and an OpenLayers (OpenLayers Contributors, 2015)-based frontend (see Figure 5 for a screenshot of the interactive WebGIS SDI frontend, showing the Köppen-Geiger classification (Willmes et al., 2017) of an LGM climate simulation data set).

The publication of OGC compliant web services from a Desktop GIS into the CRC806-SDI is facilitated by using the OpenGeo GeoExplorer Plugin for QGIS (Boundless Inc., 2014). This plugin allows user friendly publication of maps and data sets into the GeoNode, and thus GeoServer (GeoServer Contributors, 2015) -based SDI, by entering some selected metadata and a few mouse clicks from the Graphical User Interface (GUI).

6.1. Technology and architecture

The CRC806-SDI consists of three main building blocks (see Figure 6): a GeoNode-based backend, a Mapserver and a MapProxy-based additional OpenGIS Web Services (OWS) backend, and a Typo3 Extbase & Fluid (Rau et al., 2013; Lobacher, 2014) extension to integrate the SDI into the CRC806-Database web application[5]. As the main facilitator, GeoNode (GeoNode Contributors, 2015) is installed for delivering OGC compliant geospatial data services, also known as OWS. Additionally, for services having higher performance or data amount demands, infrastructure-based on MapServer (MapServer Contributors, 2015) and MapProxy (Tonnhofer and Helle, 2015) is provided. For the frontend of the SDI, that is the web-based GUI, the services are integrated into the CRC806-Database Typo3-based web portal via a custom developed Extbase & Fluid-based extension (see Figure 5).

Figure 5: Screenshot of the CRC806-Database SDI interactive OpenLayers-based WebGIS (right), and the Typo3-based spatial data catalog GUI (left).
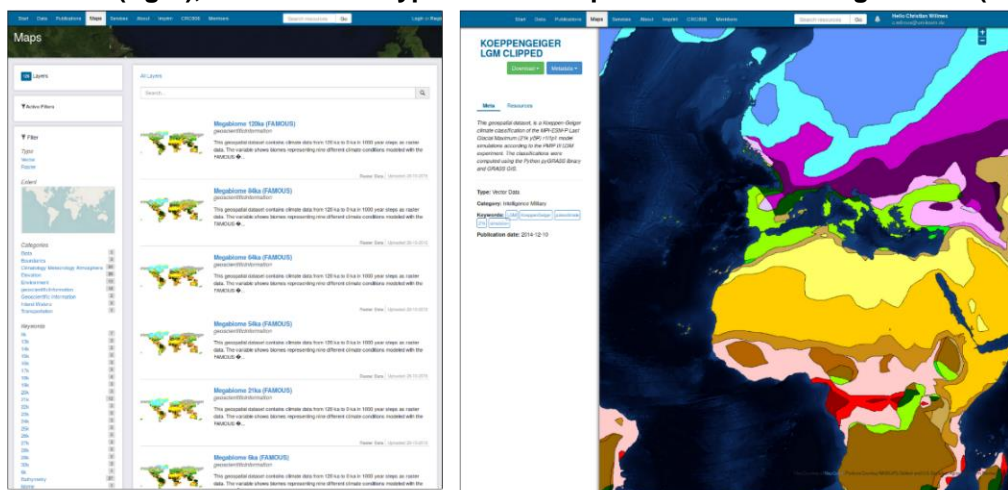
The organization of the three building blocks and interfaces between GeoNode backend, MapServer and MapProxy backend and the Typo3 web-based frontend is shown. The user is able to access the SDI via the web frontend or via OWS services, which are in case of GeoNode behind a reverse proxy, and thus have a different endpoint than the MapServer and MapProxy OWS of the SDI.

---

[5] http://crc806db.uni-koeln.de/maps

Figure 6: Architecture of the CRC806-Database SDI. Source: Willmes, C. (2016).
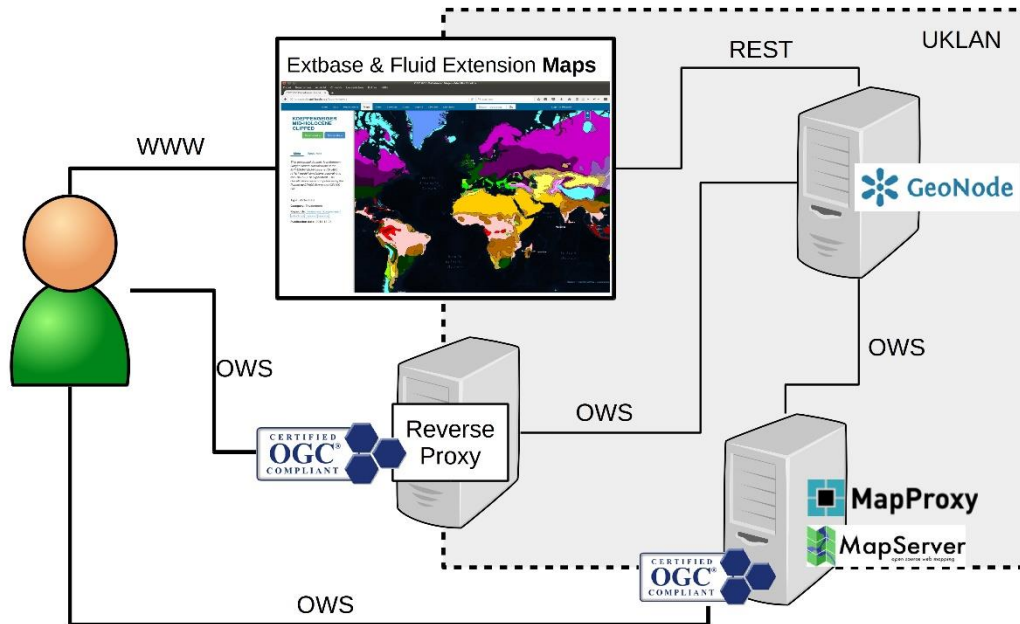
The MapProxy and MapServer instances are an artefact of a previous version of the CRC806-SDI. In 2014 GeoNode was established as the primary SDI backend (Willmes et al., 2016; Willmes, 2016). In this previous version, the SDI was relying on MapServer for serving the OGC conform web services, and MapProxy was used for caching these MapServer services, as well as some external services. The WebGIS frontend of this former CRC806-SDI version was-based on AngularJS and OpenLayers, that were later replaced by the Typo3 Extbase & Fluid implementation, as described in detail in Willmes et al. (2017).

6.2. GIS data publishing

GeoNode allows to easily publish geospatial web services. It offers two simple solutions for this task. Firstly, the standard solution is using the GeoNode web applications GUI that offer simple web-based forms to upload GIS data, define a map visualization of the data and annotate the metadata according to ISO 19115 (ISO19115-1, 2014) standard geospatial metadata. The second way is to use the desktop GIS QGIS, including the GeoNode plug-in, to create and upload geospatial web services for a GeoNode instance.

6.2.1. GeoNode as CMS for geospatial data services

GeoNode offers a feature rich web application that allows to upload, edit, style, publish, organize and manage geospatial web services. The application leverages the Django web framework, offering an extensible environment for web developers accustomed to using any modern MVC web framework. GeoNodes frontend is-based on jQuery and Bootstrap frameworks. A set of jQuery JavaScript plug-ins is used to implement the user interface, and Bootstrap CSS is used to style the pages (GeoNode Contributors, 2015).

6.2.2. Geospatial metadata

GeoNode maintains a custom metadata schema that is mapped by its underlying GeoServer to multiple state-of-the-art geospatial metadata schemas. The CRC806-SDI, and thus the PaleoMaps project, support the following export formats:

- ISO 19115
- FGDC
- ebRIM
- TC211
- Dublin Core
- DIF
- Atom

See Guiliani et al. (2016) for a detailed description of GeoServers metadata implementation, mapping, and management.

6.2.3. QGIS GeoExplorer Plug-in

Using the OpenGeo Explorer plug-in (Boundless Inc., 2015), it is possible to publish complete GIS projects (GIS data and its visualizations in maps) with some few clicks to GeoNode. A convenient metadata editor is provided in addition, in order to annotate the GIS data layers in Dublin Core, ISO 19115 and INSPIRE conforming metadata. In theory, almost all the GIS data management could be executed through the QGIS plug-in, without using the web-based frontend. However, both interfaces have their advantages and disadvantages. The most useful feature of the QGIS plug-in is the possibility of editing already published services through the plug-in; for example, adjusting the visualization or adding further GIS data to the service. This is also possible using the web interface, but the capabilities and usability appear to be more sophisticated in the QGIS plug-in.

7. CONCLUSION

The archaeological and paleoenvironmental scientific communities are demanding for data sources as described and introduced in the paper at hand. The reuse of paleoenvironmental data is often restricted in many ways. The here presented Open Science approach aims to contribute to improving this situation in the future.

Additionally, the application of open standards and formats, as well as the publication of the data under open licenses as open data and as open access, was chosen to allow the best possible dissemination of the data and thus allow for making a considerable impact on third party research applications and studies that may reuse the data provided through the CRC806-Database.

A concern, which came up early during the development of the presented project, was the possibility of copyright infringement, in some sense by just re-publishing already published information. We addressed this issue by use of scholarly published information that is according to the principles of good scientific practice free for citation, referencing and thus also reuse for reformulation of that information into another format; of course under the premise of properly citing the sources of the information. The scientific work of compiling the new GIS data set by filtering through available information and deciding which sources are incorporated and which not, is the original research conducted for creating each PaleoMaps data set. And this approach will be subject to peer-review for each new data set. This approach adds the value of utility in GIS-based applications and research projects to this reused available information, which was not accessible as GIS data before.

The project will be developed further and many improvements are already on the to-do list; for example, implementation of a proper open peer-review process. The idea is to align with the processes of existing data journals such as the IJSDIR or the Geoscience Data Journal. One of the next PaleoMaps data publications will be submitted to one of these Journals to implement the requirements and standards of these platforms. A future goal should be that all PaleoMaps data publications go through a well-defined peer-review process.

ACKNOWLEDGEMENTS

reviewers of the FOSS4G 2016 academic track and the reviewers of IJSDIR as well as the Editor of the IJSDIR for many very helpful, constructive and intelligent thoughts, recommendations, suggestions and demands for improving and correcting this manuscript.

REFERENCES

Anemone, R.L., Conroy, G.C. and C.W. Emerson (2011). GIS and paleoanthropology: Incorporating new approaches from the geospatial sciences in the analysis of primate and human evolution, *American Journal of Physical Anthropology*, 146: 19–46, DOI:10.1002/ajpa.21609.

Becker, D., Verheul, J., Zickel, M. and C. Willmes (2015). "LGM paleoenvironment of Europe – Map", in: CRC806-Database, DOI: 10.5880/SFB806.15.

Becker, D., Willmes, C., Bareth, G. and G.-C. Weniger (2016). A plugin to interface OpenModeller from QGIS for Species' potential distribution modelling, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-7: 251-256, DOI: 10.5194/isprs-annals-III-7-251-2016.

Becker, D., de Andrés-Herrero, M., Willmes, C., Weniger, G.-C. and G. Bareth (2017). Investigating the Influence of Different DEMs on GIS-Based Cost Distance Modeling for Site Catchment Analysis of Prehistoric Sites in Andalusia, *ISPRS International Journal of Geo-Information*, 6(2): 36, DOI: 10.3390/ijgi6020036.

Boundless Inc. (2014). "OpenGeo Suite plugin for QGIS", at http://qgis.boundlessgeo.com/static/docs/intro.html [accessed: April 9, 2015].

Boundless Inc. (2015). "Opengeo explorer qgis plugin", at http://qgis.boundlessgeo.com/ [accessed: August 24 2015].

Clark, P., Dyke, A., Shakun, J., Carlson, A., Clark, J., Wohlfarth, B., Mitrovica, J., Hosteler, S. and M. McGabe (2009). The Last Glacial Maximum, *Science*, 325(5941): 710-714.

Clark, P., Mix, A. and E. Bard (2001). Ice sheets and sea level of the Last Glacial Maximum, *Eos, Transactions American Geophysical Union*, 82(22): 241-247.

Creative Commons (2015). "About the licenses", at http://creativecommons.org/licenses/ [accessed: September 2, 2015].

Ehlers, J., Gibbard, P. and P. Hughes (2011). Quarternary Glaciations – Extent and Chronology. A closer look, *Developments in Quarternary Sciences*, 15.

DataCite (2014). "DataCite Metadata Schema for the Publication and Citation of Research Data. Version 3.1", DataCite e.V. Metadata Working Group. DOI: 10.5438/0011.

General Bathymetric Chart of the Oceans (2014). "GEBCO 2014 Grid – Gridded bathymetry data", at http://www.gebco.net/data_and_products/gridded_bathymetry_data/ [accessed: May 11, 2017].

GeoNode Contributors (2015). "Geonode–open source geospatial content management system", at http://geonode.org [accessed: October 23, 2015].

GeoServer Contributors (2015). "Geoserver – open source server for sharing geospatial data", at http://geoserver.org [accessed: October 23, 2015].

Heimler, S. (2014). "Semantic Mediawiki model development through object-oriented JSON schema", in: SMW CON FALL 2014, Vienna, at http://fannon.de/p/mobo-paper.pdf [accessed: May 11, 2017].

ISO19115-1 (2014). "Geographic Information – Metadata – Part 1: Fundamentals", at https://www.iso.org/standard/53798.html [accessed: May 11, 2017].

Kelso, N. and T. Patterson (2010). Introducing Natural Earth Data – NaturalEarthData.com, *Geographica Technica*: 82-99.

Krötzsch, M., Vrandecic, D. and M. Völkel (2006). "Semantic Mediawiki", in: I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold and L. Aroyo (eds.), *The Semantic Web - ISWC 2006*, Vol. 4273 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 935–942. DOI: 10.1007/11926078_68.

Lobacher, P. (2014). *Typo3 Extbase - Moderne Extension-Entwicklung für Typo3 CMS mit Extbase und Fluid*, Open Source Press, München.

MapServer Contributors (2015). "Mapserver – open source web mapping", at http://mapserver.org [accessed: December 18, 2015].

Mitasova, H., Barton, M., Ullah, I., Hofierka, J. and R.S. Harmon (2013). GIS-Based Soil Erosion Modeling, *Treatise on Geomorphology*, 3: 228-258, DOI: 10.1016/B978-0-12-374739-6.00052-X.

OKFN (2016). "Open definition 2.1", at http://opendefinition.org/od/2.1/en/ [accessed: March 15, 2016].

OpenLayers Contributors (2015). "Openlayers - a high-performance, feature-packed library for all your mapping needs", at http://openlayers.org/ [accessed: October 23, 2015].

Pachur, H.-J. and N. Altmann (2007). *Die Ostsahara im Spätquartär: Ökosystemwandel im größten hyperariden Raum der Erde*, Springer Verlag, Heidelberg.

Pontika, N., Knoth, P., Cancellieri, M., Pearce, S., (2015). "Fostering open science to research using a taxonomy and an eLearning portal", in: *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, ACM, 11.

QGIS Development Team (2016). "QGIS Geographic Information System", at http://www.qgis.org [accessed: March 27, 2015].

Rau, J., Kurfürst, S. and M. Helmich (2013). *Zukunftssichere TYPO3-Extensions mit Extbase und Fluid*, O'Reilly Verlag, Köln.

Richter, J., Melles, M. and F. Schäbitz (2012). Temporal and spatial corridors of homo sapiens sapiens population dynamics during the late pleistocene and early holocene, *Quaternary International*, 274(0): 1-4, at http://www.sciencedirect.com/science/article/pii/S1040618212004235 [accessed: May 11, 2017].

SemanticMediawiki Contributors (2015). "Semantic Mediawiki – free and open-source extension to Mediawiki", at https://semantic-mediawiki.org/ [accessed: August 17, 2015].

Sitek, D. and R. Bertelmann (2014). "*Open access*: A state of the art", in: Opening Science – The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing, Springer Open, Heidelberg.

Tonnhofer, O. and D. Helle (2015). "Mapproxy – open source proxy for geospatial data", at http://mapproxy.org/ [accessed: December 18, 2015].

Ulbricht, D., Elger, K., Bertelmann, R. and J. Klump (2016). panmetadocs, escidoc, and doidb - an infrastructure for the curation and publication of file-based datasets for GFZ data services, *ISPRS International Journal of Geo-Information* 5(3): 25.

Verheul, J., Zickel, M., Becker, D. and C. Willmes (2015). "LGM major inland waters of Europe – GIS dataset", CRC806-Database, DOI: 10.5580/SFB806.14.

Whyte, A. and G. Pryor (2011). Open Science in Practice: Researchers perspectives and participation, *The International Journal of Digital Curation*, 1(6): 199-213.

Wikimedia Fdn. (2015). "Mediawiki", at https://www.mediawiki.org/ [accessed: August 15, 2015].

Willmes, C. (2016). CRC806-Database: A semantic e-Science infrastructure for an interdisciplinary research centre. PhD Thesis, University of Cologne. urn:nbn:de:hbz:38-73811, at http://kups.ub.uni-koeln.de/id/eprint/7381 [accessed: May 11, 2017].

Willmes, C., Becker, D., Brocks, S., Hütt, C. and G. Bareth (2017). High Resolution Köppen-Geiger Classifications of Paleoclimate Simulations, *Transactions in GIS*, 21: 57–73, DOI: 10.1111/tgis.12187.

Willmes, C., Kürner, D. and G. Bareth (2014). Building Research Data Management Infrastructure using Open Source Software, *Transactions in GIS*, 18: 496– 509, DOI: 10.1111/tgis.12060.

Willmes, C., Yener, Y., Gilgenberg, A. and G. Bareth (2016). "CRC806-Database: Integrating Typo3 with GeoNode and CKAN", in: C. Curdt and C. Willmes (eds): Proceedings of the 2nd Data Management Workshop., Geographisches Institut der Universität zu Köln, Kölner Geographische Arbeiten, Vol. 96, DOI: 10.5880/TR32DB.KGA96.17.