

A semantic wiki as repository to review published palaeo-data in East Africa

F. A. Viehberg¹, C. Willmes², S. Esteban¹, R. Vogelsang³

Institute of Geology and Mineralogy¹, Institute of Geography², Institute of Prehistoric Archaeology³, University of Cologne
e-mail: {finn.viehberg, c.willmes, s.estebanlopez, r.vogelsang}@uni-koeln.de

Introduction

East Africa is deemed a centre of origin and dispersal of anatomically modern humans from Africa to Western Eurasia. Since several decades archaeologists and geoscientists explore suitable sites in this region and along potential pathways to evaluate this research hypothesis. At the same time, analytical methods applied to the archives improved in measuring accuracy or spatial/time resolution. The amount of published scientific data is enormous, but has to be carefully checked on their robustness compared to modern standards. Therefore, it is necessary to compile datasets and excerpt the given data that are source of scientific interpretations (e.g., palaeoenvironment, palaeoclimate, evolution patterns, time models etc.). In addition, the names of the study sites in East Africa are often transcribed from different languages or hold several synonyms for various reasons. Thus, we decided to use a semantic wiki to have the advantage of queryable structured data combined with the ability of web based frontend for collaborative editing of the content.

Screenshots of the Afriki web application

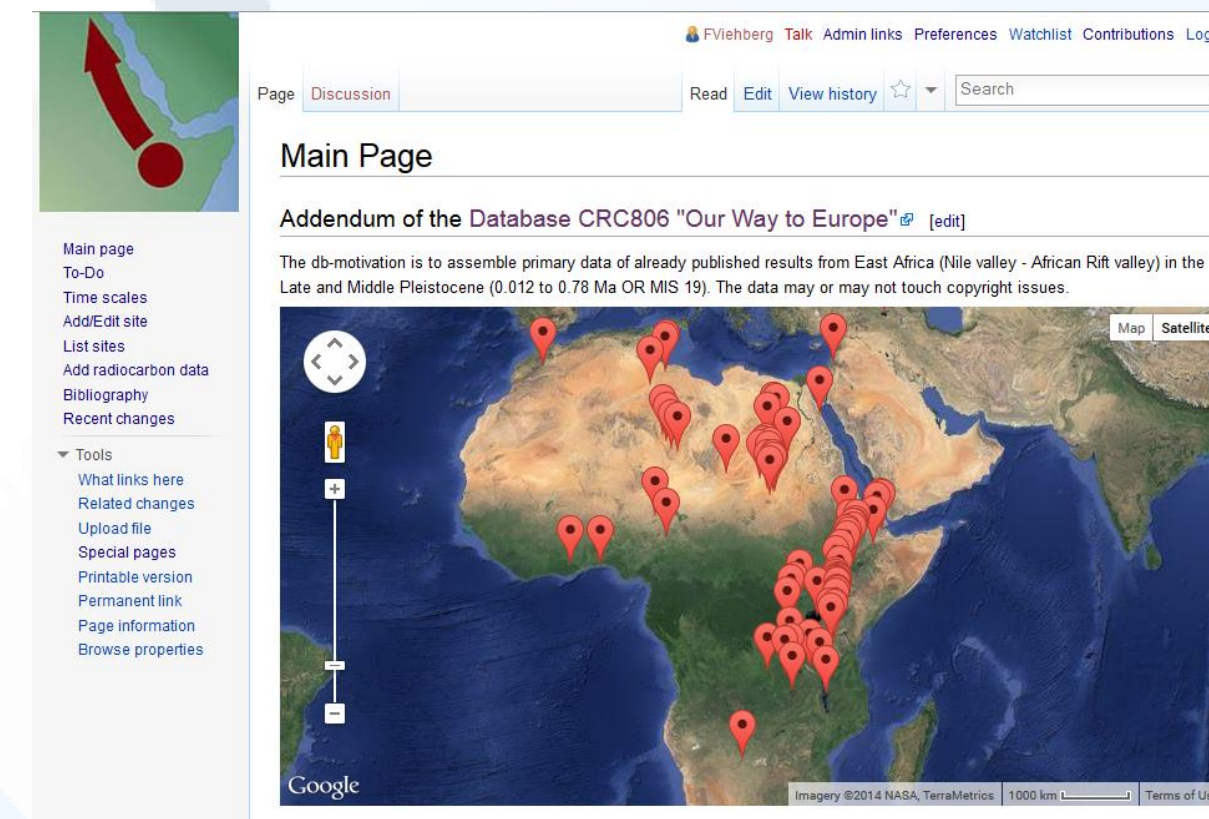


Figure 1A: Start page including Map

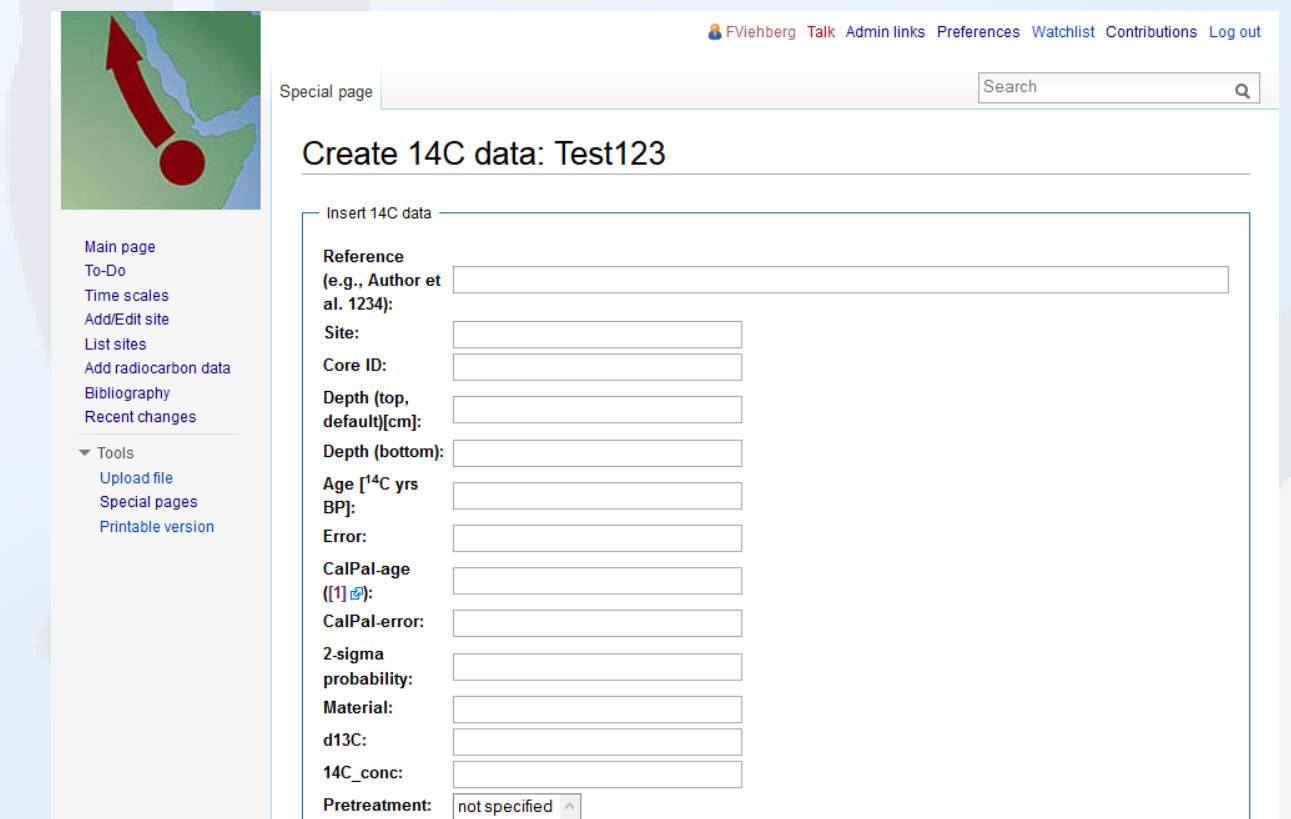


Figure 1B: Data entry form



Figure 1C: Bibliography

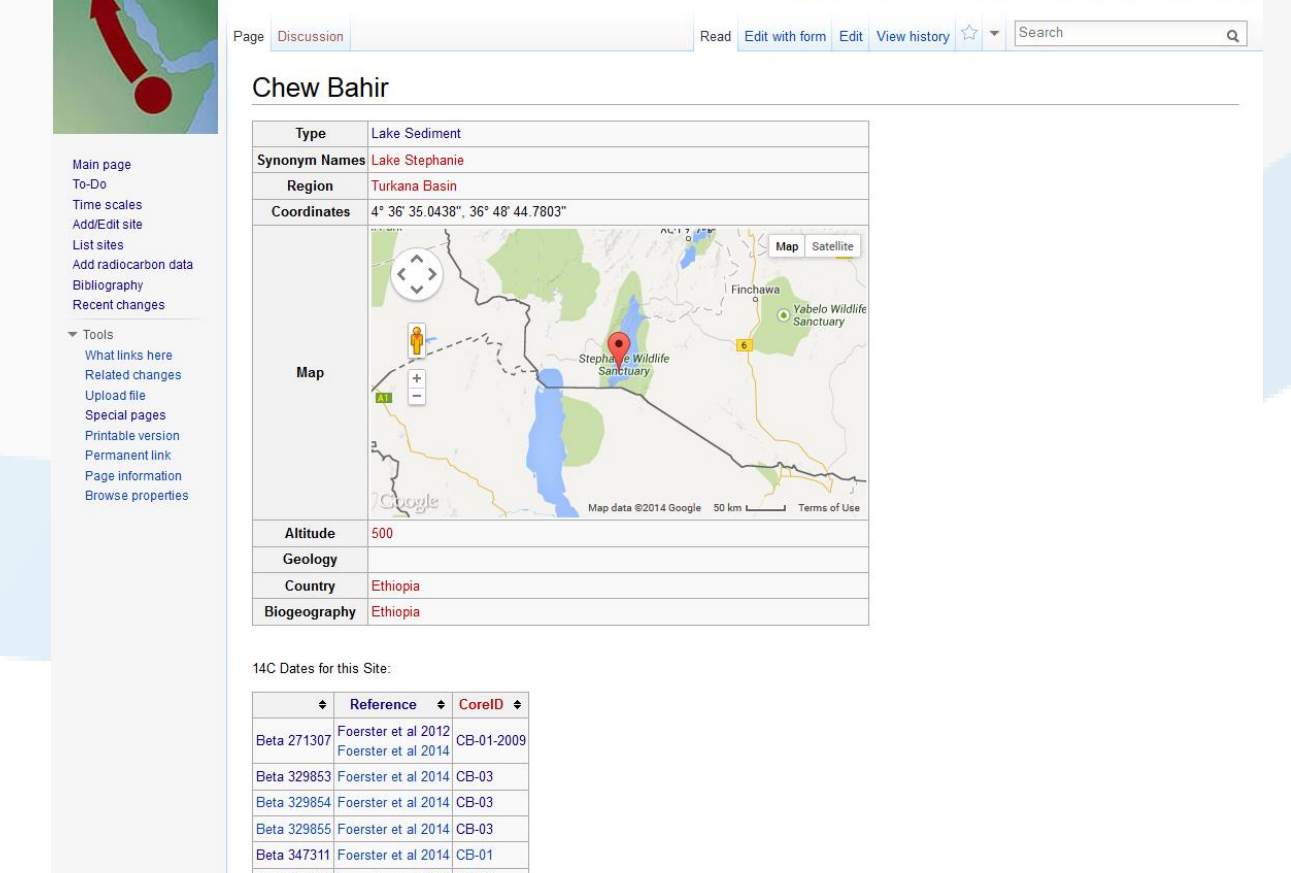


Figure 1D: Data display (map + related content)

Technology

The application is based on Semantic Mediawiki (SMW, KRÖTZSCH et al. 2006), an extension of the famous Mediawiki software, the widely used wiki system mostly developed and maintained by the Wikimedia foundation as the software base of Wikipedia. The SMW extension allows to enter structured semantic data on wiki pages. This data can then be queried through several interfaces within the wiki and the Mediawiki API as well as an SPARQL endpoint for access from external applications. Query results can be exported in several well known formats, such as CSV, XML, JSON, and more (see fig. 2). It is also possible to display query results directly in the wiki, using a number of provided display formats, like tables, data graphs or maps.

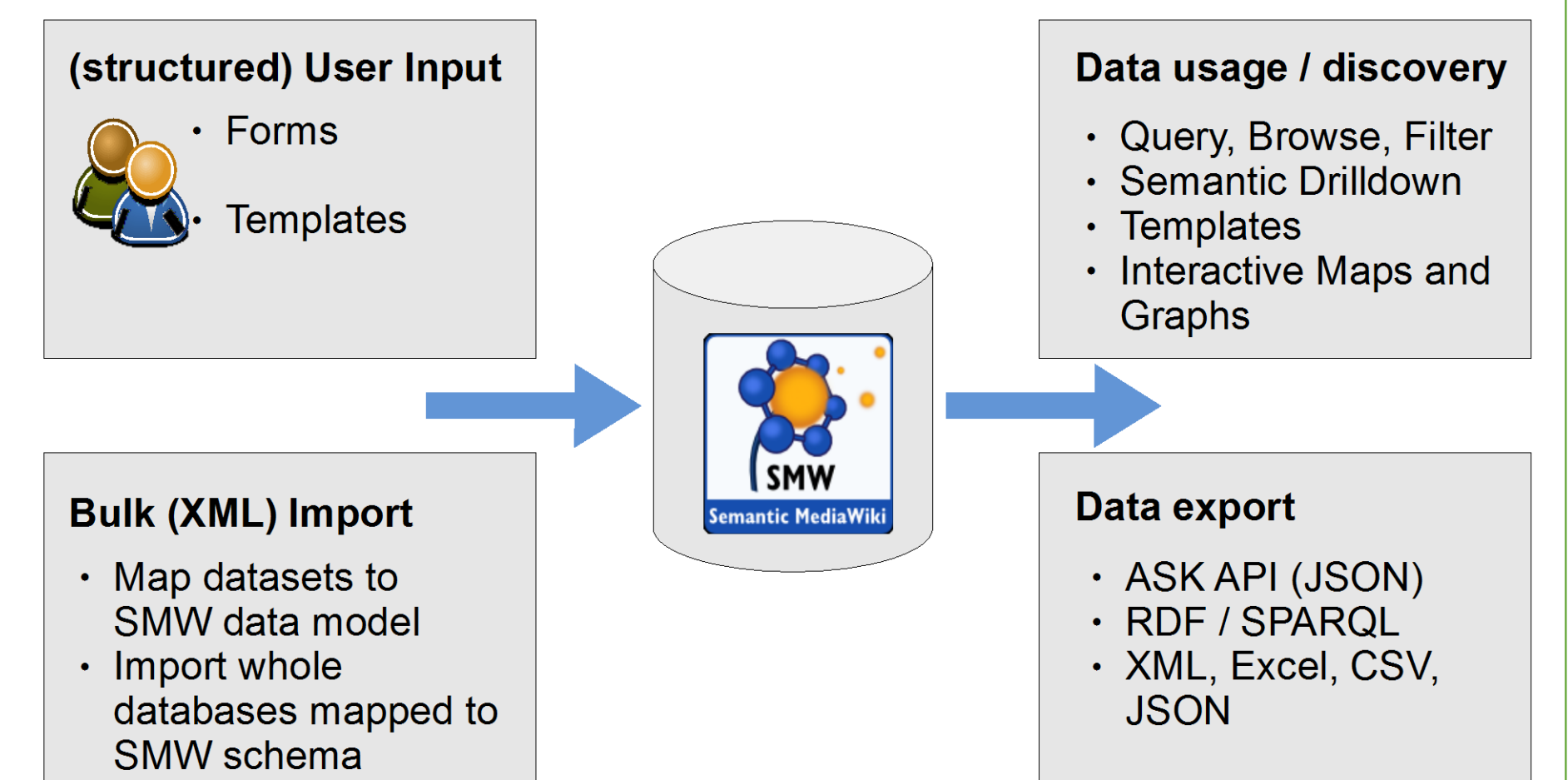


Figure 2: SMW interfaces

Data base development

The SMW system allows to collaboratively develop and maintain a data basis and thus implements a collaborative research environment (CRE). The CRE is developed applying a prototyping (NAUMANN & JENKINS, 1982) approach (see fig. 3). An outcome of this approach is a domain data model of the structured information stored in the system, which can be formalised and mapped to existing data bases to allow data integration between applications.

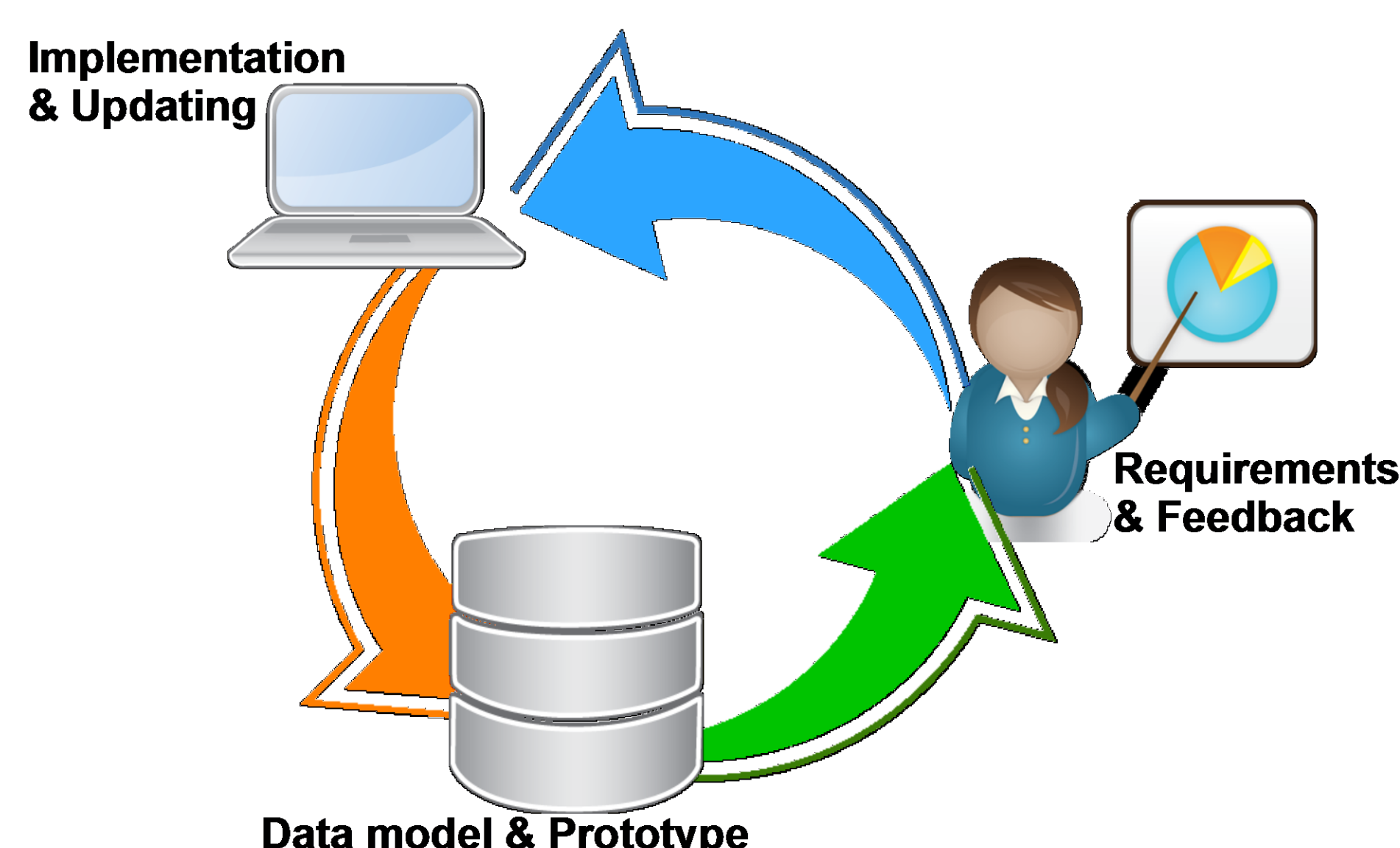


Figure 3: Prototyping

Applications

The data can be exported in common data formats, e.g. CSV (see fig. 4), for use in software packages such as statistical or GIS analysis, or plotting diagrams for re-evaluation purposes in scientific publications. Used references in the database are easily imported in BibTeX-format to assure the bibliographical entries. To guarantee further accessibility the data may also be queried via an API and consumed by other web applications.

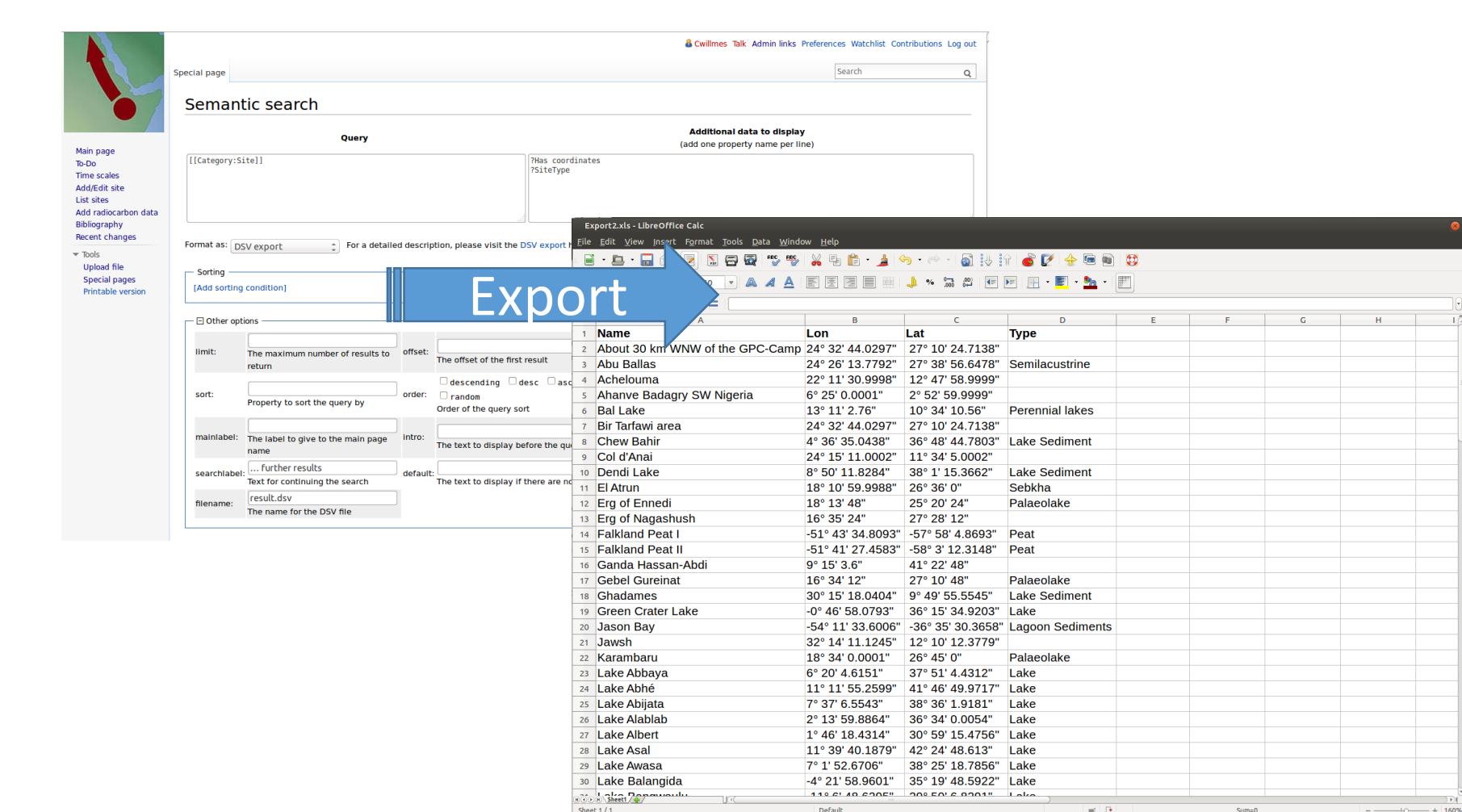


Figure 4: Query result export in common data formats

Outlook

Future development of the database will concentrate on three topics.

- (1) Import of archaeological data from reference sites within the CRC 806 to complement the palaeo-records.
- (2) Implementation of advanced queries to increase the usability of the web application. This will include e.g. comparison charts of reference sites, list of fossil assemblages and X-Y plots of age models.
- (3) Establishing the CRE and the use of SMW in more projects of the CRC 806. Work on similar applications within B and C cluster has already started.

References

- KRÖTZSCH, M., VRANDEIC, D., VÖLKELE, M., (2006): Semantic MediaWiki. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L. (Eds.), The Semantic Web - ISWC 2006. Vol. 4273. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 935–942.
- NAUMANN, J. D., & JENKINS, A. M. (1982): Prototyping: the new paradigm for systems. MIS Quarterly, Vol. 6 Issue 3, pp. 29-44, doi: 10.2307/248654.