

Digital Data and Tools in Archaeology: The ROCEEH Out of Africa Database (ROAD)

Zara Kanaeva¹, Volker Hochschild², Michael Märker¹ and Christian Sommer¹

¹Heidelberg Academy of Sciences and Humanities, Tübingen, Germany

²Eberhard Karls University of Tübingen, Germany

Abstract

The ROCEEH Out of Africa Database (ROAD) is a multidisciplinary database with archaeological and paleobiological content. It was developed to support the interdisciplinary research project "The Role of Culture in Early Expansions of Humans" (ROCEEH), whose main aim is to study the cultural aspects of human expansions over the last three million years. The concept of the database is to bring together archaeological, paleoanthropological and paleoenvironmental content to allow quantitative and statistical analysis of the information stored. The data entered into ROAD include new results produced by researchers involved in the project, as well as published data from previous excavations. The data rely on a standardized vocabulary that is applied across the project and its publications; the data can be retrieved from the database in tabular form or using SQL queries. In addition to the archaeological and paleobiological information, ROAD contains spatial data (e.g., vector and raster data), generated by ROCEEH and other projects. One important part of ROAD is its web-based application, which allows the user to easily insert, update, review, query, export, visualize and analyse data.

Keywords:

databases, digital archaeology, SQL queries

1 Introduction

The rapid development of applied computer science offers many new perspectives for digital archaeology, although the methods of digital information technology have long been used in archaeology. One of these methods is the structured storage and management of data using databases and database management systems (DBMS).

Archaeologists use a variety of DBMS, from single-user systems (like Access) to client-server, multi-tier systems. Most database applications used in the field of archaeology are based on a relational DBMS such as SQLite, MySQL or PostgreSQL, which can store and manage many types of data. Archaeologists use databases in a variety of ways, during excavations and afterwards, in order to answer specific research questions. In Reed et al.

(2015), we can find a good overview of the modern technical possibilities for storing and using field data, while Ramsey et al. (2014) describe a database that supports a research project studying the “Response of Humans to Abrupt Environmental Transitions (RESET)”. The RESET database focuses on the methodology used in the project. It contains information on the occurrence and chemical composition of glass shards from tephra and cryptotephra deposits found across Europe.

The “The Role of Culture in Early Expansions of Humans” (ROCEEH) project investigates the role of cultural evolution for expansions of archaic humans between three million and 20,000 years ago in Africa and Eurasia. Understanding the interaction of cultural evolution, biology and changing environmental conditions is one research objective of the project. The ROCEEH Out Of Africa Database (ROAD), one application of which (the graphical user interface) is presented in this article, is a web-based information system providing archaeological, paleoanthropological and paleoenvironmental data, visualization of the saved data, SQL querying and spatial visualization of the query results, and various maps and tools (Märker et al., 2009, 2013). As of February 2018, ROAD contains information about 1,663 localities and 7,606 assemblages of finds.

2 Development of ROAD

The development of ROAD includes the development of the database and its application, ROADWeb. ROAD’s aims and functionalities were specified from the outset: ROAD must (1) integrate heterogeneous data; (2) be easily accessible for storing and querying of data; (3) allow import and export of data in various formats (csv, Bibtex)¹; (4) provide data visualization tools (for example visualization of geological profiles); (5) provide simple spatial visualization of data.

For the physical storage of the heterogeneous data, we decided to use a relational database. The development of a relational database follows a well-known procedure: we have our reality and this reality must be mapped in two steps (logical and physical) onto a database. The mapping result after the first step is the logical database model, which describes which aspects of the data will be stored and manipulated in the database. The mapping result after the second step is the physical model of the database, which describes how the data will be stored and manipulated in the database.

ROAD model

The logical model of ROAD was developed by means of Entity/Relationships diagrams. It includes more than 50 entities, their descriptive attributes, and the relationships between entities. In Figure 1, showing the simplified logical model of ROAD, the ovals represent the general subjects contained in the database, while the central rectangle and the lines connecting it to the ovals show relationships. Because the ovals represent one or many entities, most of the connecting lines represent more than one relationship. All of the lines

¹ Later it became clear that we also needed JSON format.

joining the central rectangle represent many-to-many (m:n) relationships. The line linking locality and assemblage, and those linking assemblage and archaeological finds/paleo-faunal finds/paleo-botanical finds represent one-to-many (1:n) identifying relationships. The other lines represent one-to-many (1:n) non-identifying relationships.

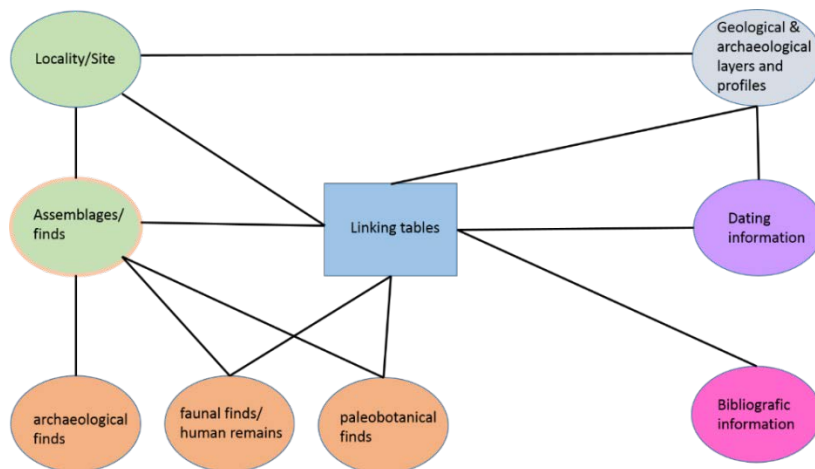


Figure 1: The logical structure of ROAD

The logical model of ROAD consists of six main parts: (1) localities and assemblages, (2) layers and profiles, (3) dating results, (4) assemblage descriptions, (5) bibliographic information, and (6) lookup and linking tables (Figure 1). The first part of the data includes general information about localities and assemblages; the second, information about geological layers and their corresponding geological profiles, and archaeological layers and the corresponding archaeological profiles of localities. The dating-results part provides information about the ages of the layers and assemblages. Detailed assemblage descriptions include information based on the interpretation and analysis of finds. ROAD categorizes assemblages of finds into several types – archaeological finds, faunal remains, human remains and botanical remains. The analysis of finds (i.e. assemblage descriptions in ROAD) yields more detailed information about the types of finds. For example, archaeological finds might be stone artefacts, and their description include information about their technology, typology and functional traces. Archaeological finds can also be symbolic artefacts – i.e. artefacts that convey an abstract meaning. The symbolic artefacts in ROAD are objects that clearly represent art, music and ornaments. The tables with bibliographic information include information about publication resources, author name(s), publication title, publisher name, journal title, volume, issue, publication year and an abstract. The lookup tables in ROAD include information about taxonomical classification of plants and animals (they are used to look up the taxonomical classification for plant and animal remains). The linking tables contains the linking information. Additionally, ROAD includes tables with analytical results about habitat and climate.

The physical model of ROAD was implemented using the PostgreSQL database management system.

ROAD implementation

On its front-end, ROAD application (ROADWeb) uses .php, .javascript and .html scripts. On its back-end, it uses php scripts, Apache Webserver, and the PostgreSQL Database Management System with PostGIS extension and MapServer. Figure 2 demonstrates the interplay of the software used.

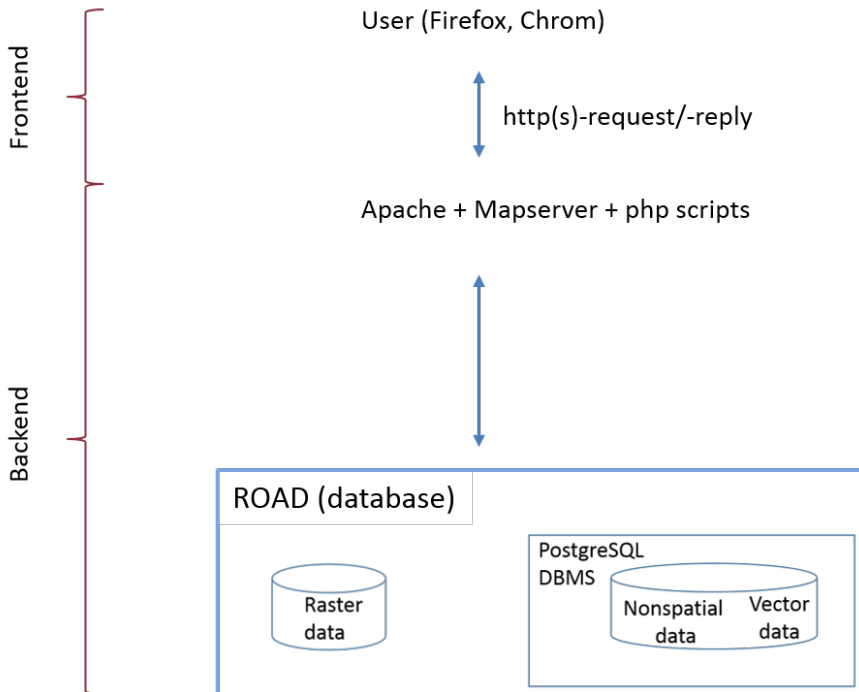


Figure 2: The infrastructure of ROAD

The database is available for search online at:

http://www.roceeh.uni-tuebingen.de/roadweb/map_modul.

3 ROAD: database application

SQL Querying

One of the key functionalities of ROAD is data querying. Using the ROADWeb SQL query tool, (shown in Figure 3), the user can compose basic SQL SELECT queries. For each clause of the SELECT query (basic syntax), the query tool offers one or more graphic elements.

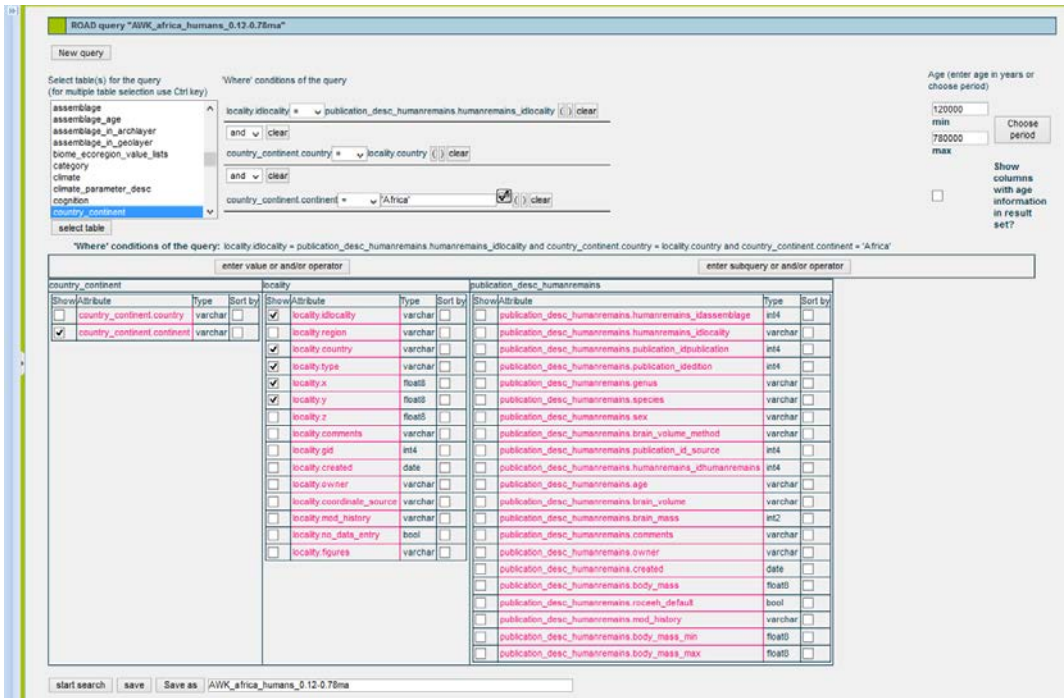


Figure 3: SQL query tool showing a query to find all African hominids from 780,00 to 120,000 years before the present

The first graphic element to be shown is a multiple selection box listing ROAD tables. The user can select one or more tables for the FROM clause of the query. After the table selection, columns of each selected table are displayed and the SELECT and WHERE clauses can be specified. When the specification of the query clauses is finished, the query execution can be started using the “start search” button. The query results are always displayed as a table. Additionally, the query results can be spatially visualized with aid of the ROAD WebGIS module. Figure 4 shows the ROAD WebGIS module displaying the spatial distribution of the search results of the query shown in Figure 3.

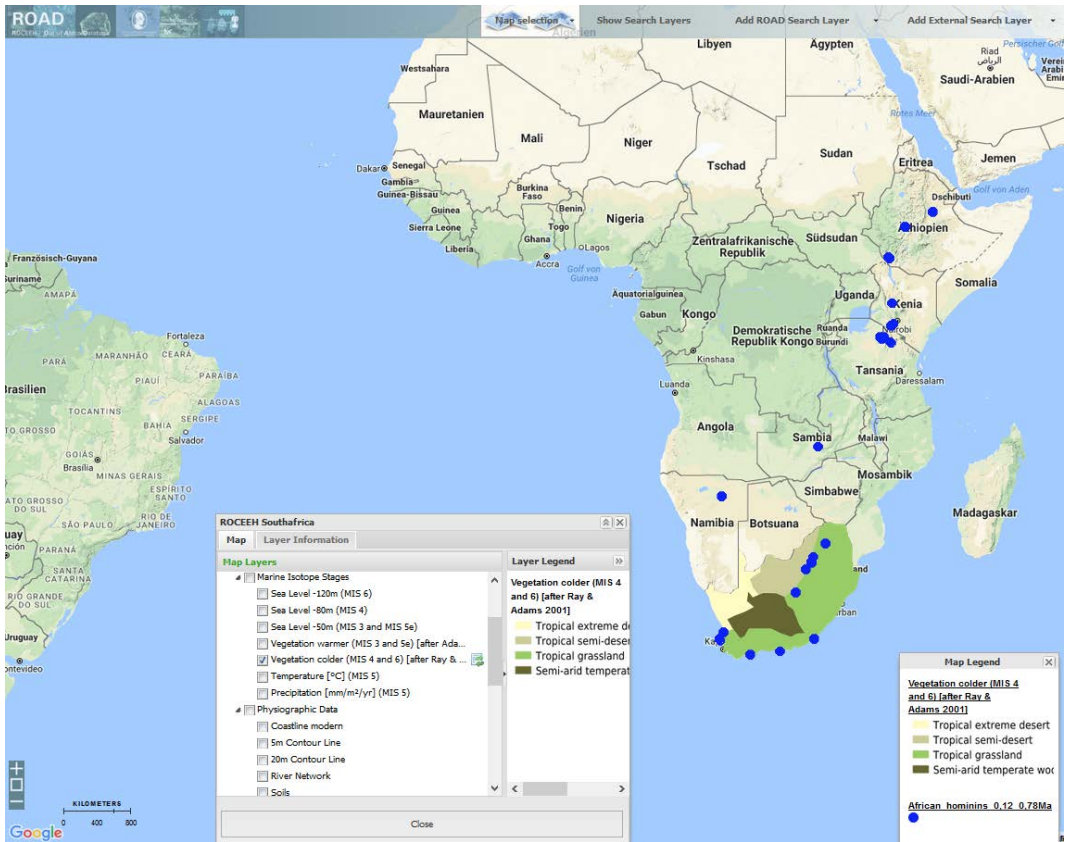


Figure 4: The ROAD WebGIS module maps all ROAD localities with hominid finds dating from 780,000 to 120,000 years before the present. In addition to the basic background layer, the map shows spatial data for vegetation during a cold phase (MIS 6: 191,000 to 130,000 years before the present).

The ROAD WebGIS module is a simple visualization tool which does not require any additional user skills. It allows the user to request remote spatial data, perform a simple ROAD search, map the search results coming from the query tool, or search external databases. If the user needs further spatial analysis, there is the possibility of exporting the search data in csv format in order to import the data into more comprehensive tools like ArcGIS or QGIS. For example, with an appropriate query the user can compute the number of associated ROAD entries of every ROAD locality, and then export the results and import them into the Spatial Analyst module of ArcGIS Pro². This module calculates as well as visualizes the density of the imported query results. Figure 5 shows the calculated spatial data density of ROAD localities.

² <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/an-overview-of-the-density-tools.htm>

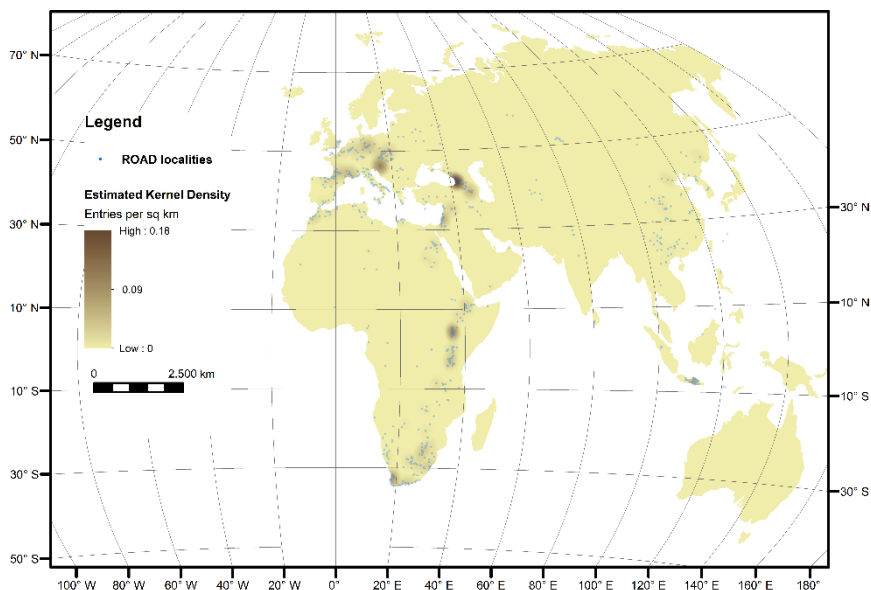


Figure 5: The map shows the ROAD entry density per square kilometre

ROAD: quantitative analysis of the data

The amount of data in ROAD allows us to perform quantitative or statistical data analysis. Some subprojects in ROCEEH use time-series analysis. The data for this time-series analysis are generated in ROAD by multiple runs of one query for various assemblage ages over a particular time interval. For the generation of the time-series data, ROADWeb includes the ‘time slice’ tool, which computes the distribution of localities or assemblages over a selected historical period. Figure 6 shows the distribution of all ROAD assemblages against the age over the time interval 1,000,000–20,000 years.

With the ‘time slice’ tool, the user can vary the step size, the size of the step overlap, as well as the subject of the analysis. In Figure 6, the size of one step is 10,000 years, the overlap is 0 years, and the analysis subject is all ROAD assemblages.

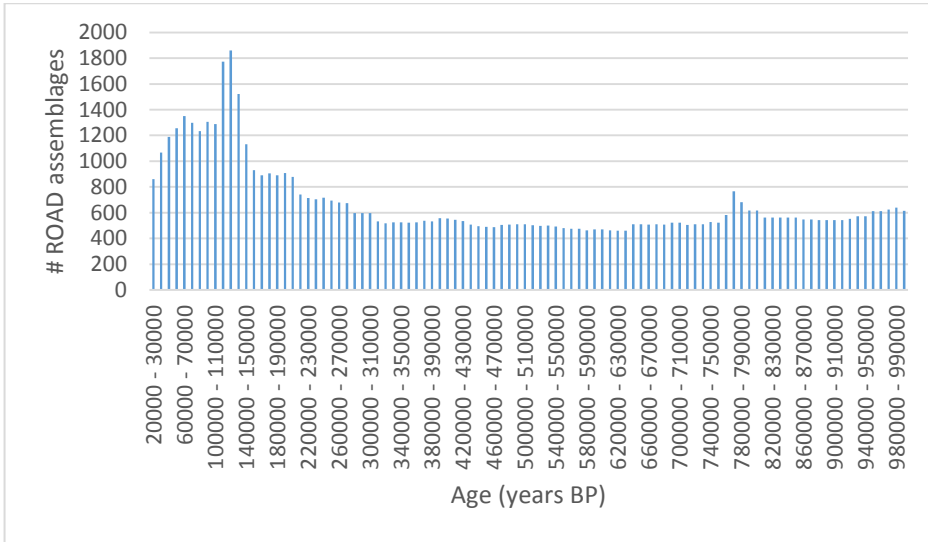


Figure 6: A simple Excel plot for ROAD assemblage counts against the age over the time interval 1,000,000–20,000 years before the present. Because most of the assemblages have as their age a time interval (for example 20,000–40,000 years), the same assemblage may be counted several times (for example the assemblage of the age 20,000–40,000 years before the present will be counted for the queried ages 20,000–30,000 years and 31,000–40,000 years)

With a small modification of the ROADWeb ‘time slice’ tool, more statistical computations can be done. For example, for each ROAD locality we can compute its stone tools and, using Excel, visualize the distribution of ROAD localities over the stone tools counted (Figure 7).

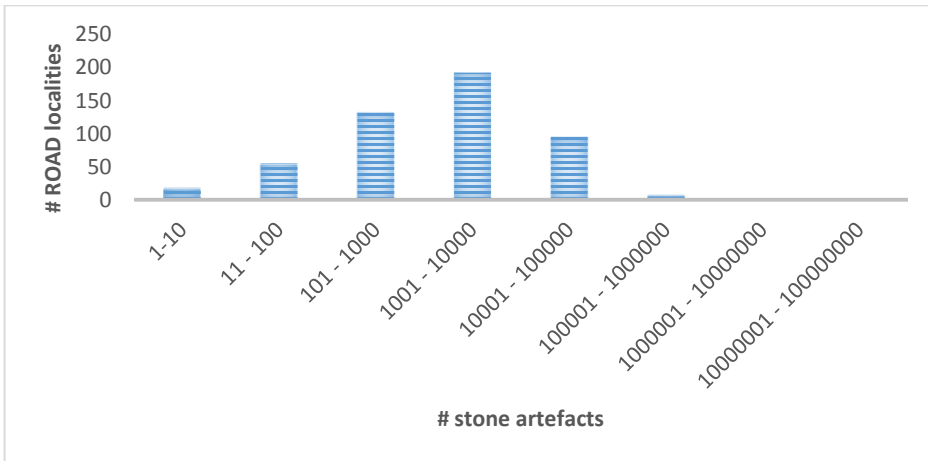


Figure 7: The diagram shows number of localities over counts of their stone artefacts. The average number of stone artefacts per locality is 12,085. For some inexperienced users, this will seem a remarkably high number

Many aspects of archaeological information are numerical. If we apply the appropriate statistical method, we can perform many types of statistical analysis on the ROAD data. However, the choice of method is not easy and depends on the research questions. A good introduction to statistics in the context of archaeology is given in Drennan (2010).

4 Linking ROAD with other databases

The WebGIS module of ROAD allows spatial visualization of the search results from ROAD. Additionally, ROAD search results can be combined with data from other databases. The ROAD WebGIS module currently allows searching two external databases: Neotoma and Neogene Quaternary Mammals Database (NQMDB). Neotoma covers the Pliocene-Quaternary part of Earth's history, during which humans evolved and modern ecosystems developed. NQMDB hosts lists of micro- and macro-mammals for Pleistocene fossil localities in western Eurasia.

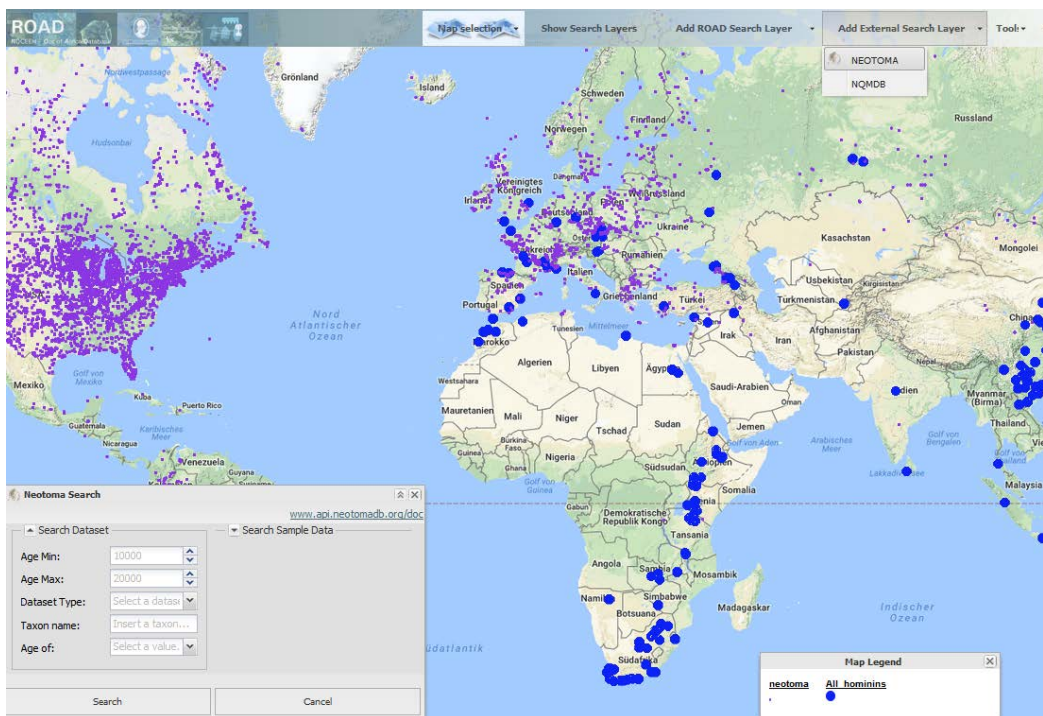


Figure 8: Linking Neotoma search results to ROAD search results

5 Conclusions

ROAD integrates interdisciplinary work from the fields of archaeology, paleobiology and paleoanthropology. This data integration is very useful for understanding patterns of human expansions over time and space, and for hypothesis building or verification by means of statistical analysis, not just within the ROCEEH project, although the statistical analysis could not be performed without additional knowledge of statistics or without training on quantitative analysis of archaeological data.

ROAD is useful for all archaeologists, and especially for those archaeologists and scientists who are interested in the discussion of how quantitative analysis is done in archaeology and how it could be done in archaeology.

Acknowledgments

The manuscript has benefited from the contribution of many ROCEEH team colleagues. Financial resources were provided by the research project “The Role of Culture in Early Expansions of Humans” sponsored by the Heidelberg Academy of Sciences and Humanities.

References

- Drennan, D. Robert (2010). *Statistics for archaeologists: A common sense approach*. Springer
- Märker, M., Kanaeva, Z., Quenéhervé, G., Hochschild, V., Bolus, M., Bruch, A., Conard, N., Haidle, M.N., Hertler, C., Kandel, A., Mosbrugger, V., Schrenk, F. (2013). ROAD: A georelational database with WebGIS functionalities for the assessment of the Role of Culture in early expansion of humans. *Journal of Archaeological Science*.
- Märker, M., Kanaeva, Z. & ROCEEH Team (2009). ROAD: The Role of Culture in early expansion of humans data base. In Curdt, C. & Bareth, G. (eds.): *Proceedings of Data Management Workshop, Cologne, Germany, 29.-30.10.2009. Kölner Geographische Arbeiten*, 90: 93-97.
- Ramsey, C., Housley, R. A., Lane, C. S., Smith, V. C., Pollard, M. A. (2014). The RESET tephra database and associated analytical tools. *Quaternary Science Reviews*, 118, 33-47
- Reed, D., Barr, A., McPherron, S. P., Bob, R., Geraads, D., Wynn, J. G., Alemseged, Z. (2015). Digital Data Collection in Paleoanthropology. *Evolutionary Anthropology*, 24, 238-249