INTRODUCTION

Overview of the Conference: Human Origins – Digital Future

Andrew W. Kandel[1], Miriam N. Haidle[2] & Christian Sommer[1,3]

[1] The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities, Hölderlinstr. 12, 72074 Tübingen, Germany
[2] The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[3] Inst. of Geography, Univ. of Tübingen, Rümelinstr. 19-23, 72070 Tübingen, Germany

This volume presents the results of the conference, "Human Origins – Digital Future", organized by The Role of Culture in Early Expansions of Humans (ROCEEH), and examines current trends in the dynamic field of data management and research. The main aim of the conference was to discuss integrative aspects and approaches for developing, using, and securing large scientific databases in the future, specifically in the context of archaeological and paleoanthropological research. The conference focused on how databases with novel information technology can be used to gain new knowledge by linking, retrieving, and analyzing archaeological, paleoanthropological, paleobiological and geographical information. In addition to addressing fundamental questions of digitalization and open science, the conference examined approaches using innovative methods.

Due to the COVID-19 pandemic, the conference took place virtually over one week as a series of short, mainly prerecorded talks and interviews, daily summaries, and discussions (Haidle et al. 2020). To document the lectures and secure the results for an interdisciplinary public, we decided to present the proceedings of this event in a multi-media format.

The chapters in this volume follow the overall structure of the conference, which was organized into five main sessions about Databases, Methods, Applications, Products and Perspectives. Each chapter starts with a Mind Map created with the productivity platform Nextcloud. The first Mind Map outlines the entire conference (Fig. 1), while the Mind Maps in the following chapters (Figs. 2–6) summarize the main ideas presented in the individual sessions. Each session began with a recap of the previous day presented by Miriam Haidle using the Mind Maps as guides. As a result, those who could not attend the previous day could quickly review the essence of the talks. We hope you will also find these graphics insightful.

Within the five main chapters, the authors present their abstracts with keywords and references. Yet, it is the videos of the talks themselves which represent the heart of this volume. They provide a flavor of the broad spectrum of ideas presented. We hope you will enjoy watching and learning from them, and perhaps you will be inspired.

Kandel, A. W., Haidle, M. N., & Sommer, C. (2022). Introduction. In A. W. Kandel, M. N. Haidle, & C. Sommer (Eds.), *Human Origins – Digital Future: An International Conference about the Future of Archaeological and Paleoanthropological Databases* (pp. 9-15). Propylaeum, Heidelberg. DOI: https://doi.org/10.11588/propylaeum.882.c13431.

Since the beginning of the discipline, the science of archaeology has dealt with vast amounts and diverse types of data, which are in turn connected to other datasets. Examples include the three-dimensional location of a find within a site, the stratigraphical association with other finds, catalogs of similar finds or decoration styles from a specific period and area, to name only a few. The associations of data are used for relative dating, interpreting technologies, functions, and numerous other processes, understanding distributions within and between communities, connections, and thresholds. With growing inclusion of additional scientific studies such as botanical and faunal analyses, absolute dating, geoarchaeological and, increasingly, genetic analyses, the amount and diversity of data has expanded enormously. The potential of digital databases in gathering and combining information has given an additional push to the propensity of archaeologists to acquire ever more data.

Thus, databases have become ubiquitous in archaeology. They range from a few variables to huge datasets, from simple lists to complex relational databases; they can comprise one artifact category of a layer at a site, or cover numerous types spanning millions of years and several continents. Every project works with at least one database. Data compilation takes a large segment of a researcher’s time. Nonetheless, data are often treated as bucket finds, or even as backdirt. Unlike the more sexy finds, data are nothing to show off and thus receive limited care. At the end of a project, most of the data collected, including even newly created records, vanishes within inaccessible storage mediums and locked files, or through outdated program standards. Despite the widespread use of data, many questions remain unresolved about the development and growth of databases in archaeology, their analytical function, interoperability, and sustainability.

Databases

The first chapter highlights differences among prehistoric databases, which are heterogeneous with regard to data quality, degree of quality control, and frequency of maintenance. Uncertainties such as the imprecision of data concerning chronology, spatial location, and terminology, all impede our ability to interconnect. Databases differ in their level of granularity. Digital data tend to have a short lifespan because projects end, and the software they use may be proprietary or become outdated. Networking problems can hinder the implementation of comprehensive, large-scale evaluations and syntheses, which belong in the realm of Big Data. Following the FAIR principles – Findable, Accessible, Interoperable and Reusable – can help researchers to overcome these issues. The establishment of “prehistoric data science” is another necessary foundation to integrate elements of computer science with subject-specific competencies for archaeological, paleoanthropological, and environmental themes, among others. This makes sense because the architecture and semantics of a database are based on theoretical assumptions and specific questions arising within a scientific discipline.

The session begins with **Eric Grimm & John W. Williams** who introduce a community-curated data resource, the NEOTOMA Paleoecology Database. Based on a centralized IT-platform, shared standards, ontology and terminology, NEOTOMA provides a long-term and low-cost archive for paleobiological data and allows users to analyze multi-proxy data.

In contrast, **Andrew Kandel & the ROCEEH Team** present the ROCEEH Out of Africa Database (ROAD), which represents an integrator of multidisciplinary data about human evolution managed by one group. The talk details the different levels of decisions and actions taken to create, maintain, and implement such a diverse database as ROAD.

Franco Niccolucci describes ARIADNEplus as an example of an infrastructure that offers a very broad overview. ARIADNE provides access to archaeological datasets from excavation reports and assemblages of finds through research databases, for example, from the Archaeological Data Service (ADS) or ROAD. While he stresses the variety and diversity of archaeological data contained in ARIADNE, it can still cause problems in the application of techniques making use of Big Data. In the future, he suggests developing a specific data-centric research methodology for archaeology. **Christopher Nicholson** focuses on the Digital Archaeological Record (tDAR) as another large online repository of archaeological data, serving as a digital curator of data. Digitization can at least partially help overcome the destruction caused by excavation, but the application of the FAIR principles is necessary to hinder the fading of digital resources. Finally, **Jesús Rodríguez** highlights how different levels of uncertainty can be a major problem for archaeological and paleontological databases such as the Neogene-Quaternary Mammals Database (NQMDB). Although the collection of data seems like a neutral process, theory-driven decisions have to be made to create the architecture and semantics required of a database.

Methods

Chapter II focuses on methods which are applied to databases that make a dataset more valuable by deriving knowledge from raw information. Quantitative methods are useful in the whole workflow from data collection, quality assessment, linking databases to the analysis and interpretation of data. **Juan A. Barceló** presents the concept of reverse engineering, which aims to understand the functionality and genesis of an object by reconstructing it. Applied to archaeology, this makes it possible to reconstruct the social actions of earlier societies on the basis of the archaeological record. With the help of several examples, he points out that this reconstruction goes beyond the pure 3D recording of objects, but that human behavior can be described by functional analysis and mathematical models. **Christian Sommer** refers to quantitative methods resulting from the availability of archaeological Big Data, such as the ROAD database. Using two methods from the field of machine learning, he illustrates how the boundaries of the human niche can be inferred from site distributions and environmental data. Furthermore, he shows how the functional relationships between lithic tools can be explored using a frequent pattern algorithm borrowed from the online marketing industry.

Alice J. Williams demonstrates how targeted agent-based models can be applied to real-world archaeological data to test conceptual models. In her example, she discusses Carneiro's circumscription theory, on the basis of which she simulates the formation of social complexity within the context of geographic barriers. This approach allows her not only to highlight research gaps, but also to facilitate debates rooted in profound and logical scenarios, even where data are missing. **Christian Willmes** draws attention to how geodata can facilitate the analysis of human-environment interaction. He describes how paleo-geospatial data can be made available to the public through a web platform and web services, and how these data can be exploited through GIS analyses to provide insights into the interface between humans and the environment. Finally, he highlights the PaleoMaps concept, a framework designed to overcome the lack of interoperability of heterogeneous scientific data and enable broad exchange.

Applications

In the third chapter, we examine the ways in which research databases are currently used by archaeologists, paleoanthropologists and paleoclimate scientists. Such applications are varied and often connected with other tools, such as Geographic Information Systems (GIS) and Agent Based Models (ABM), to extract further information from existing datasets. These papers examine new considerations for databases, with several focused specifically on ROAD, while others consider paleoenvironmental data, questions about human behavior, and population dynamics.

Some of the speakers look at how archaeologists develop and use databases for different purposes. **Ewa Dutkiewicz** presents examples of databases of Paleolithic art. While these databases attempt to provide a structure for data that were previously considered unstructurable, such data are difficult to quantify digitally. She speaks about her own database, SignBase which details geometric signs on mobile objects and mentions an international project which catalogs representations of animals in European parietal and mobile art. **Shannon McPherron** asks whether a decentralized, non-standardized approach to data sharing in archaeology serves us well. He notes how methodologies in data collection and analysis have often been individualized in the past. However, new technologies for data collection and sharing may overcome some previous limitations. In the end, he argues that a variety of methods and techniques allows for a more robust understanding of the past. **Denné Reed** presents a database called Paleo Core, a platform for integrating data in paleoanthropology. Paleo Core hosts almost a dozen active research projects and includes details about more than 85,000 specimens. This integrative data management system promotes digital workflows from the start of a project up through the curation of the artifacts in museum collections.

Several talks focus on the varied uses of ROCEEH's own database, ROAD. **Rimtautas Dapschauskas & the ROCEEH Team** introduce a study that examined questions about the large-scale development of cultural behavior in human evolution. Using ROAD, his team conducted analyses of assemblages containing the red pigment ochre in Africa during the Middle Stone Age. They determined that ochre use accelerates across Africa and becomes habitual starting about 140,000 years ago. **Christine Hertler & the ROCEEH Team** discuss the development of an application in ROAD called the MapModule. This integrated GIS allows registered users to visualize the distribution of archaeological, human, faunal and plant finds stored in ROAD and also connects to external databases which contain further details about paleoecology and mammals (e.g., NEOTOMA and NQMDB). The user can also use different map backgrounds (e.g. sea level, glaciation, vegetation, biome, temperature, precipitation, and others.)

Ericson Hölzchen & the ROCEEH Team detail their experiments in simulating Neanderthal mobility (NeMo) in a virtual environment using data from ROAD in ABMs. Their models can be used to compare differences in mobility patterns between regions, quantify differences in mobility patterns during different climatic events, or differentiate subsistence strategies and raw material procurement. **Zara Kanaeva & the ROCEEH Team** demonstrate the usefulness of a new tool, the ROAD Summary Data Sheet. These PDFs can be generated automatically for any locality stored in ROAD and provide details about the site, its location, stratigraphy, dating and assemblages as well as its bibliography. Such a summary is useful for the general public interested in learning more about our human heritage and scientists who want to get an overview on the find assemblages of specific sites.

Some speakers discuss how to integrate paleoenvironmental data into questions about human behavior and population dynamics. **Ana Mateos** and her team examine the trophic resources available to humans during a warm period, MIS 11, in Europe. Using variables such as net primary productivity and ungulate carrying capacity as proxies for the abundance of plant and animal resources, they match the presence of sites with paleoenvironment to better frame the requirements established by human paleoecology. **Mika Puspaningrum** and her team aim to reconstruct the topography, hydrology, climatic patterns, and vegetation cover of the Eastern part of Java, including Sangiran, about one million years ago. They use ABM to reconstruct a background model and simulate how hominins behaved and interacted with their environment. **Manuel Will** and his team explore the evolution of human brain and body size during the Pleistocene using paleoenvironmental and paleoanthropological datasets within a quantitative statistical framework. They determine that temperature predicts body size across all human taxa – likely a direct effect of climate on human physiology. On the other hand, net primary productivity and long-term variability in mean annual precipitation work well to predict brain size in archaic hominins, but not modern humans. Multiple interacting causal mechanisms likely underlie the evolution of these key biological characteristics of humans. **Andreas Zimmermann & Isabell Schmidt** study prehistoric population dynamics and presented the geostatistical upscaling procedure named the "Cologne Protocol" which provides a consistent approach to estimate population size and density. By interpolating between estimates with a logistic equation, they model long-term population dynamics across many cultural periods in Europe. Future directions include extending the spatial dimension using regression analysis and expanding the temporal framework in conjunction with data stored in ROAD to model site distribution.

Products

The fourth chapter surveys products from and for databases. Products are aimed to reach a range of stakeholders, from field archaeologists and interdisciplinary research teams to the general public. Products include catalogs and search engines, linking of code, repositories and even researchers, as well as the communication of results. External, centralized storage options represent an important product, especially relevant for smaller projects. To provide a long-term perspective, open source approaches and standardized structures are fundamental, while the development of user interfaces for analysis is desirable. Finally, research focusing on databases and their use can in turn serve as a product, which may act retroactively on a database itself.

Wolfgang Börner introduces Wien Kulturgut as part of a digital platform for open government. It integrates data from five governmental departments to provide information on architecture, town history, art works in public space, historical maps, and archaeology. With an internet and an intranet version it aims at decision-makers as well as the local public and even tourists. The METAhub project at Frankfurt presented by **Liane Giemsch & Tanja Neumann** connects material remains in museums with performance art using digital formats. A content hub provides open information to the interested public; a planned immersive app will enrich localities with augmented reality; and discursive events such as the participative discovery of collections integrate the urban community as active partners.

Yasuhisa Kondo reports about interdisciplinary challenges and in-project action research in the Cultural History of PaleoAsia project. A meta-analysis of the database revealed distinctions in terminology and working habits between the disciplines involved. Pointing to differences in working cultures, such an approach is the foundation of efforts for more interdisciplinarity. Introducing Spacialist, **Matthias Lang & colleagues** retrace the development of a virtual open-source research environment for humanities with a strong spatial component. They show the possibilities of such an endeavor especially for small projects as well as the pitfalls such as underestimated costs and workload.

Perspectives

Chapter V ventures into the future of archaeological databases. In four interviews, influential people in scientific database management share their thoughts on the future development of databases and report the lessons learned. Viewed together, they paint a picture of diverse, interconnected databases that integrate varied stakeholders from the public, private, and scientific sectors and are openly and permanently available. However, they also address the current challenges, such as funding and institutionalization, which still need to be resolved. **Sarah Whitcher Kansa** informs about Open Context, a data publishing platform where archaeological objects can be registered with high granularity following the principle "one URL per pot sherd." The service accompanies the data of authors on its way from digitization and the creation of metadata, to annotation and publication as linked data in a permanent repository. In addition, users receive advice on important but not necessarily popular topics such as data preparation and digital ownership. **Peter McKeague** calls for the development of a unified Spatial Data Infrastructure based on existing national and international geodata standards, similar to the European Commission's INSPIRE Directive – and extending them with thematic conventions from the field of archaeology. Such a data structure allows transnational data exchange, also via already existing geodata portals, and helps make data sustainably usable and save them from fossilization.

Julian Richards provides insight into what made the Archaeology Data Service one of the most influential data providers: The Open Archival Information System as a structural basis; free accessibility of the data for everyone; and finally, free and expert advice to keep the level of data quality high. For future development, he counts on the collaboration of like-minded people to advance the FAIR principles and establish a global network of research data.

Dieta Svoboda-Baas of the Heidelberg Academy of Sciences and Humanities reports from the perspective of a research institution that supports the life cycle of data in new projects in all aspects, conceptually, structurally and financially. However, she also talks about the challenge of catching up with the preservation of digital data in existing projects, some of which have been running for many decades. She sees the future in overcoming isolated solutions through standardization and mentions the German National Research Data Infrastructure (NFDI) as a possible solution.

Closing remarks

Finally, in three appendices, we provide links to databases and other resources discussed during the conference, and list of the speakers and registered participants of the conference. In summary, we hope that this digital book helps to keep your memories of the conference alive. It was a challenge to conduct the event online during the beginning of the pandemic, but we felt it was a rewarding experience for all those involved. We greatly appreciate the interaction we had with all of our colleagues in attendance, and thank everyone for the lively discussions that ensued. We again draw your attention to a review of the conference published shortly thereafter (Haidle et al. 2020).

ACKNOWLEDGMENT

ROCEEH is located at the Senckenberg Research Institute in Frankfurt and the University of Tübingen. This long-term research project is promoted by the Joint Science Conference of the Federal Government and the state governments of the Federal Republic of Germany within the framework of the Academies' Programme. Funding comes from the Federal Ministry of Education, Science and Research and the states of Baden-Wurttemberg (Ministry of Science, Research and the Arts) and Hesse (Ministry of Science and the Arts). With additional support from the *Senckenberg Gesellschaft für Naturforschung*, the conference "Human Origins – Digital Future" took place online from July 27–31, 2020 with 113 registered participants from 19 countries ranging from the Far East to the American West. Invited researchers presented their findings in English in a series of 15 lectures, 11 poster presentations and four interviews. We are grateful to all contributors to the conference as well as to the Propylaeum team of the Heidelberg University Publishing that made this innovative publication possible.

ORCID®

Andrew W. Kandel https://orcid.org/0000-0002-9889-9418
Miriam N. Haidle https://orcid.org/0000-0002-4923-7880
Christian Sommer https://orcid.org/0000-0001-9062-0876

REFERENCES

Haidle, M.N., Bolus, M., Bruch, A.A., Hertler, C., Hochschild, V., Kanaeva, Z., Sommer, C., Kandel, A.W. (2020). Human Origins – Digital Future, an International Conference about the Future of Archaeological and Paleoanthropological Databases. *Evolutionary Anthropology, 29*, 289–292. https://pubmed.ncbi.nlm.nih.gov/33179353/