MAX PLANCK INSTITUTE OF GEOANTHROPOLOGY PRESENTS THE

Big Historical Data Conference

Meeting Programme & Book of Abstracts

http://bhdc.earth

22–25 Nov 2023

Table of Contents

Big Historical Data committees

Organising committee

Ricardo Fernandes

Systems Archaeology Research Group Leader, Department of Archaeology, Max Planck Institute of Geoanthropology

Carlo Cocozza

Systems Archaeology Research Group, Department of Archaeology, Max Planck Institute of Geoanthropology

Alison Harris

Systems Archaeology Research Group, Department of Archaeology, Max Planck Institute of Geoanthropology

Sean Hixon

Systems Archaeology Research Group, Anthropogenic Ecology Research Group, Max Planck Institute of Geoanthropology

Technical and administrative staff

Anja Schatz

Department Assistant, Department of Archaeology, Max Planck Institute of Geoanthropology

Anja Hannawald

Department Assistant, Department of Archaeology, Max Planck Institute of Geoanthropology

Thomas Baumann

Head of IT Department, Max Planck Institute of Geoanthropology

Karam Kayali

IT Department, Max Planck Institute of Geoanthropology

Michelle O'Reilly

Graphic design and media production, Max Planck Institute of Geoanthropology

Session organisers

S1 Big databases: How large-scale archives can boost archaeological research

Martina Farese (La Sapienza University, IT)

Giulia Formichella (La Sapienza University, IT), Noemi Mantile (Università della Campania "Luigi Vanvitelli, IT)

S2 Big Zooarchaeological Data: Challenges and potentials for multi-scalar analysis across millenia

Angela Trentacoste (British School at Rome, IT), Jesse Wolfhagen (Purdue University, USA), Sarah Whitcher Kansa (Open Context)

S3 Environments of big cultural heritage data integration

Mike Fisher (Max Planck Institute of Geoanthropology, DE), Dovydas Jurkenas (Max Planck Institute of Geoanthropology, DE)

S4 Archaeological time-series: The quest for robust ^{14}C-dated proxies for the intensity of prehistoric activity

John Meadows (Christian-Albrechts-Universität, DE), Peer Kröger (Christian-Albrechts-Universität, DE)

S5 Data and people in interaction: Network analysis for everyone

Vera Klontza (Masaryk University, CZ),

Barbora Ruffíni (Masaryk University, CZ)

S6 Leveraging Big Data, GIS, and machine learning in remote sensing

Manuel J. H. Peters (Max Planck Institute of Geoanthropology, DE), Amina Jambajantsan (Max Planck Institute of Geoanthropology, DE)

S7 AI for historical and archaeological Big Data analysis

Jochen Büttner (Max Planck Institute for the History of Science, DE)

S8 Insights from reusing large prehistoric and interdisciplinary databases

Christian Sommer (University of Tübingen, DE), Angela A. Bruch (Senckenberg Research Institute and Natural History Museum, DE), Nicholas J. Conrad (University of Tübingen, DE), Christine Hertler (Senckenberg Research Institute and Natural History Museum, DE), Mirium N. Haidle (Senckenberg Research Institute and Natural History Museum, DE), Volker Hochschild (University of Tübingen, DE), Zara Kanaeva (University of Tübingen, DE), Andrew W. Kandel (University of Tübingen, DE), ROCEEH Team

S9 Palaeoclimatic and palaeoenvironmental databases: Exploring the dynamics of human-environmental systems

Achim Brauer (GFZ German Research Centre for Geosciences, DE)

S10 History unleashed: Harnessing the knowledge stored in historical documents

Adam Izdebski (Max Planck Institute of Geoanthropology, DE), Carlo Cocozza (Max Planck Institute of Geoanthropology, DE)

S11 Modelling expansions in South America: Integrating archaeology and linguistics

Fabrício Ferraz Gerardi (Universität Tübingen, DE), Bruno de Souza Barreto (University of São Paulo, BR)

Welcome

Dear colleagues,

We are delighted to welcome you all to the *Big Historical Data Conference*. It is an honour to have such a diverse and distinguished group of scholars and researchers gathered to discuss the latest advancements and challenges in the field of historical and archaeological data analysis.

We are motivated to explore the potential of Big Data in uncovering hidden insights about our past. With the increasing availability of digital resources and innovative technologies, we are now able to tackle complex questions and gain a deeper understanding of our history and cultural heritage. By bringing together experts from various disciplines, we will foster fruitful discussions and collaborations that will further enhance our analysis and interpretation of historical data.

In the 11 conference sessions, we will cover a wide range of topics, such as AI and machine learning, big databases, data integration, and network analysis. Each session has been carefully designed to provide a platform for exchanging knowledge, sharing experiences, and exploring new methodologies in tackling the challenges of big historical data.

It is our hope that this conference will serve as a stepping stone towards a better understanding of the past and its relevance in shaping our present and future. Let us use this opportunity to engage in meaningful conversations, build connections, and ignite new ideas that will have a lasting impact on our field.

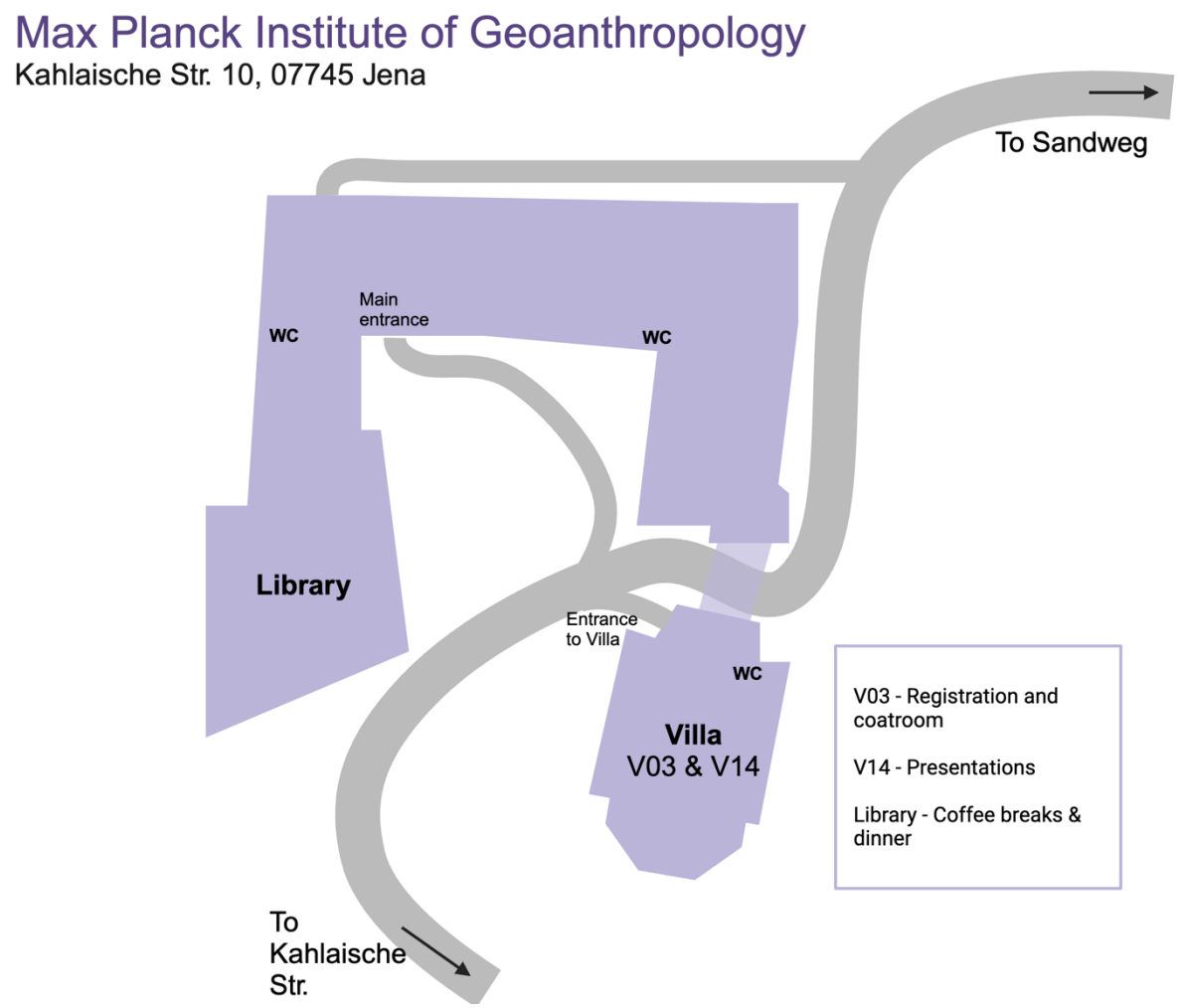
Once again, we extend a warm welcome to all of you and wish you a productive and enlightening conference.

Yours sincerely,

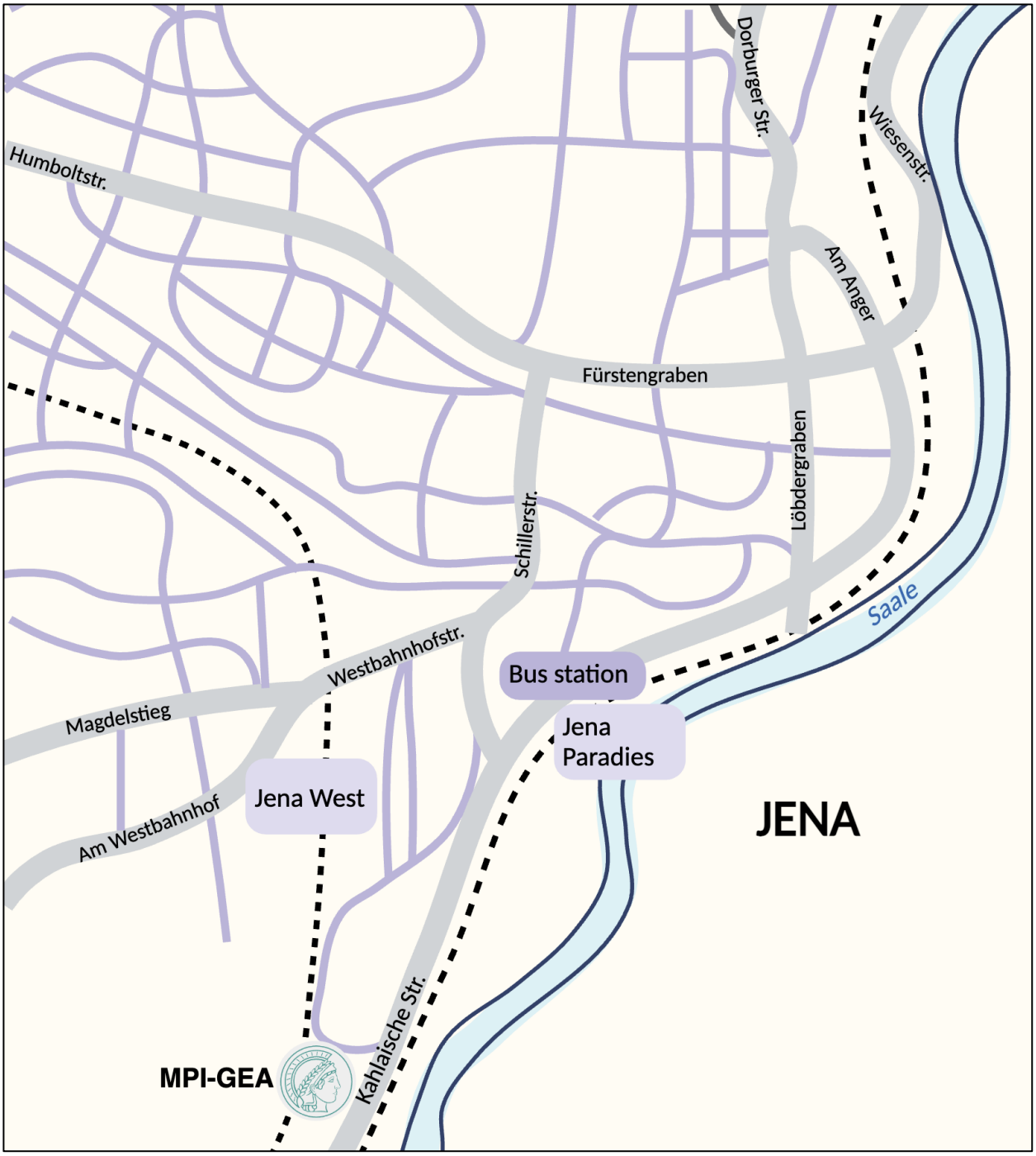
Ricardo Fernandes
Carlo Cocozza
Alison Harris
Sean Hixon

Venue

The conference is held at the Max Planck Institute of Geoanthropology, located at Kahlaische Strasse 10, 07745 Jena. Participants can register and deposit their outerwear in the registration room V03 on the ground floor of the villa building. The presentations will be given in the meeting room V14 on the first floor of the villa building. Refreshments (water and juice) will be available at all times in V14. Coffee breaks, lunch, and the conference dinner can be found in the Institute library.



Arriving in Jena



By train

Jena can be reached by train from the Frankfurt International Airport, Berlin Brandenburg Airport Willy Brandt, Albrecht Dürer Airport Nürnberg, and Leipzig/Halle International Airport. Train connections can be found at the Deutsche Bahn website. We recommend arriving at Jena Paradies Station or Jena West Station for easy access to the city centre.

By coach / bus

Jena can also be reached by long-distance coach, Flixbus. Schedules and tickets are available here. The bus depot is located next to the Paradies train station.

Getting to MPI-Geoanthropology

The Max Planck Institute of Geoanthropology can be reached on foot (~15 min) from the city centre, or using city public transport (line 2 tram). Public transport tickets can be purchased on city buses and trams using cash or credit card or by downloading the FAIRTIQ app to your smartphone. Public transit timetables are available here.

Taxis

Jena has good taxi service from the train stations. If calling a taxi, be sure to give the complete address as there are three MPIs in Jena! A number of car services are listed here, and more can be found online. Taxigenossenschaft Jena e.G. +49 3641458888; City Taxi Jena e.V. +49 3641 55660; Taxi Keinert +49 170 1802960; Franz Braschel Taxibetrieb +49 3641 604826; Taxi Hermann +49 3641 368644; Taxi David Geitner +49 3641 238876; Taxi Mike Zenker +49 3641 448634

The Villa at Kahlaische Str. 10

Daily schedule

Time	Wednesday
8:45-9:00	Welcome
9:00-10:30	Session 1, Part 1
10:30-11:00	Coffee
11:00-12:30	Session 1, Part II
12:30-13:30	Lunch
13:30-14:00	Session 1, Part III
14:00-15:00	Session 2, Part I
15:00-15:30	Coffee
15:30-17:00	Session 2, Part II

Time	Thursday
9:00-10:30	Session 3, Part I
10:30-11:00	Coffee
11:00-11:30	Session 3, Part II
11:30-12:30	Session 4
12:30-13:30	Lunch & Photo
13:30-14:30	Session 5
14:30-15:00	Session 6, Part I
15:00-15:30	Coffee
15:30-16:00	Session 6, Part II
16:00-17:45	Session 7
18:00-19:00	Dinner
19:00-19:30	Plenary

Time	Friday
9:00-10:30	Session 8
10:30-11:00	Coffee
11:00-12:30	Session 9
12:30-13:30	Lunch
13:30-15:00	Session 10
15:00-15:30	Coffee
15:30-16:45	Session 11
16:45-17:00	Closing remarks

Time	Saturday
10:00-12:00	A tour into the history of Jena

Library

Programme in brief

Wednesday, November 22

8:45-9:00	Welcome address		
9:00-10:30	Session 1: Big Databases: How large-scale archives can boost archaeological research I Session organisers: Martina Farese, Giulia Formichella, and Noemi Mantile		
	1-1	Ricardo Fernandes	The Pandora Initiative: promoting fairer data collaboration
	1-2	Martina Farese	From the Neolithic to the Middle Ages: a big data view of foodways in ancient Italy
	1-3	Antonio Caruso	STARC OSTEOARCH: An open access resource for recording and sharing human osteoarchaeological data
	1-4	Claire Ebert	Tracking human diet and migration using the Caribbean and Mesoamerica Biogeochemical Isotope Overview (CAMBIO)
	1-5	Natalia Riabogina	Building archaeobotanical and archaeozoological database in the BIAD: structure, creation experience and comparative analysis perspectives
	1-6	Victor Yan Kin Lee	Bridging human paleogenetics and archaeology: problems with existing practices and prospects for overcoming them
10:30-11:00	Coffee break		
11:00-12:30	Session 1. Big Databases: How large-scale archives can boost archaeological research II Session organisers: Martina Farese, Giulia Formichella, and Noemi Mantile		
	1-7	Andrew Kandel	The ROAD Database as a Research Tool for Exploring Human Evolution
	1-8	Felix Riede	Cultural evolutionary trends as revealed by a novel, expert-sourced dataset on lithic technology, toolkits, and artefact shapes for the Final Palaeolithic/earliest Mesolithic of Europe (15-11ka BP)
	1-9	David Laguna-Palma	An open and interdisciplinary approach to modelling movement along the ancient Lybian Sea

	1-10	Efthymia Nikita	MetaMobility Databases: Promoting Greco-Roman bioarchaeology
	1-11	Antonio Campus	Big Archaeological (Legacy) Data. Collection, dissemination and reuse for large scale analysis
	1-12	Maite I. Garcia-Collado	Exploring large-scale historical processes using IsoIberMed: a new isotope database for bioarchaeological samples from Medieval Iberia
12:30-13:30	Lunch		
13:30-14:00	Session 1: Big Databases: How large-scale archives can boost archaeological research III Session organisers: Martina Farese, Giulia Formichella, and Noemi Mantile		
	1-13	Sean Hixon	Presenting a compilation of isotopic data from Madagascar
	1-14	Carlo Cocozza	The IsoMemo Initiative: Enhancing archaeological and historical research through isotope databases
14:00-15:00	Session 2. Big Zooarchaeological Data: Challenges and potentials for multi-scalar analysis across millenia I Session organisers: Angela Trentacoste, Jesse Wolfhagen, and Sarah Whitcher Kansa		
	2-1	Sarah Whitcher Kansa	From small to big data in zooarchaeology
	2-2	Kitty Emery	Crosswalking, linking, and verbatim interpretations: adventures in legacy zooarchaeological biodiversity open data publishing
	2-3	Canan Çakırlar	Cross-disciplinary reuse potential of big zooarchaeology data: What do we need?
	2-4	David Orton	Assessing approaches to chronological uncertainty in large-scale zooarchaeological data synthesis
15:00-15:30	Coffee break		
15:30-16:30	Session 2. Big Zooarchaeological Data: Challenges and potentials for multi-scalar analysis across millenia II Session organisers: Angela Trentacoste, Jesse Wolfhagen, and Sarah Whitcher Kansa		
	2-5	Matthew Hill Jr.	Tracking long-term subsistence among the Native North American Great Plains hunters

2-6	Max Price	Wool they won't they: Combing the zooarchaeological evidence for wool production in Northern Mesopotamia c. 4500-1500 cal. BC
2-7	Anthony King	The RomAniDat network: a zooarchaeological Big Data initiative for the ancient Roman world
2-8	Roberto Ragno	Understanding variation in zooarchaeological NISP data: Bayesian betabinomial models for Italian contexts in the 1st millennium CE

Thursday, November 23

9:00-10:30 Session 3: Environments of big cultural heritage data integration I

Session organisers: Michael Fisher and Dovydas Jurkenas

3-1	Michael Fisher	Heritage-environment interaction: Theorising Big (Palaeo)Data integration across the nature-culture divide
3-2	Julian Richards	ARIADNE: aggregating archaeological datasets at an international scale
3-3	Pablo Barruezo-Vaquero	Transdisciplinary data integration into a computational ontology: some reflections from the DataARC Project on modelling human–non-human entanglements
3-4	Thomas Huet	"Shared heritage": management and integration of cultural heritage data across Arches-based platforms in the Global South
3-5	Jonas Gregorio de Souza	Mapping the archaeological Pre-Columbian heritage in South America
3-6	MAHSA and MAEASaM Teams	Modelling heritage information in two diverse regions: reflections on the integration of heterogeneous big data from the MAHSA and MAEASaM projects

10:30-11:00 Coffee break

11:00-11:30 Session 3. Environments of big cultural heritage data integration II

Session organisers: Michael Fisher and Dovydas Jurkenas

3-7	Dovydas Jurkenas	Hordes of Data: Multilayered documentation and collection of archaeological heritage and environmental data across the Mongolian landscape

	3-8	Ankhsanaa Ganbold	Creating the Mongolia cultural heritage database (immovable historical and cultural heritage)

11:30-12:30 Session 4. Archaeological time-series: The quest for robust ^{14}C-dated proxies for the intensity of prehistoric activity

Session organisers: John Meadows and Peer Kröger

4-1	Jan Kolar	Large radiocarbon datasets in relation to archaeological context and metadata
4-2	Fiona Laviano	Building a database of dated Neanderthal fossil remains to estimate the spatiotemporal pattern of Neanderthal extirpation
4-3	Jiyoung Park	Investigating long-term population dynamics in the Han River Basin using radiocarbon dating
4-4	Thomas Huet	NeoNet, an app for radiocarbon modelling of the Late Mesolithic / Early Neolithic transition in South-Central and South-Western Europe

12:30-13:30 Lunch and group photo

13:30-14:30 Session 5: Data and people in interaction: Network analysis for everyone

Session organisers: Vera Klontza and Barbora Ruffini

5-1	Tamás Czuppon	Hidden information or misuse? Data and network analysis on burial customs (6-9 c. AD)
5-2	Freg J. Stokes	Mapping networks of capital and resistance in tropical South America
5-3	Britt Davis	Neutron activation analysis and networks: Possibilities and constraints
5-4	Christopher Carleton	'Symbolic scaling' in ancient and contemporary cities

14:30-15:00 Session 6. Leveraging Big Data, GIS, and machine learning in remote sensing I

Session organisers: Manuel Peters and Amina Jambajantsan

6-1	MAESaM Team	An experiment with automated detection of diverse archaeological features in an African context using deep-learning and free satellite multispectral imagery
6-2	Ahmed Mahmoud	Automated change detection monitoring of archaeological sites using machine learning and Google Earth Engine

15:00-15:30 Coffee break

15:30-16:00 Session 6. Leveraging Big Data, GIS, and machine learning in remote sensing II

Session organisers: Manuel J. H. Peters and Amina Jambajantsan

6-3	Manuel J. H. Peters	Multi-scale automated feature detection in the Argentinean Andes: A comparative analysis of satellite and UAV
6-4	David Stott	Positive false positives: Searching for annular landforms in Northern Europe

16:00-17:45 Session 7. AI for historical and archaeological Big Data analysis II

Session organiser: Jochen Büttner

7-1	Jakob Hauser	Enhancing the Seshat databank: A computationally readable mapping of historical facts to scholarly references
7-2	Hassan El-Hajj	Prompt me a dataset: Large-scale ML dataset creation in the humanities
7-3	Mathieu Aubry	Discovering and analyzing patterns in large historical image databases
7-4	Amina Jambajantsan	Leveraging satellite imagery and Deep Learning in Mongolian archaeology
7-5	Jianyin Roachell	Deep Learning and Bayesian analysis of historical time series data
7-6	Adam Izdebski	Applying Bayesian machine learning to causality modelling in historical social-ecological systems: case studies from early modern Greece and Poland
7-7	Raphael Schlattmann	Trajectories of change: Using text embeddings to track knowledge evolution

18:00-19:00 Dinner

19:00-19:30 Plenary

Friday, November 24

9:00-10:30 Session 8: Insights from reusing large prehistoric and interdisciplinary databases

Session organisers: Christian Sommer, Angela Bruch, Nicholas Conrad, Christine Hertler, Miriam N. Haidle, Volker Hochschild, Zara Kanaeva, Andrew Kandel, and the ROCEEH Team

8-1	Samantha Scott Reiter	Columns and rows: publishing, preparing, harmonizing and preserving protocols, guidelines and resources from the Big Interdisciplinary Archaeological Database
8-2	Anaïs Vignoles	Shortcomings related to archaeological bibliographic databases used for macro-scale modeling: the SIGMENT database for estimating Middle and Recent Gravettian lithic technical tradition distributions in Western Europe
8-3	Giulia Marciani	Beyond labels. Exploring patterns of formal description of lithic variability in Late Pleistocene technologies
8-4	Rimtautas Dapschauskas	Using big data to gain new insights about large-scale developments of human behavioral evolution: ROAD and Middle Stone Age ochre use
8-5	Christian Sommer	Mapping archaeological cultures and periods with network cartography
8-6	Jesús Rodríguez	Modelling the adaptations of hominins to climate in Europe from MIS14 to MIS11

10:30-11:00 Coffee break

11:00-12:30 Session 9. Palaeoclimatic and palaeoenvironmental databases: Exploring the dynamics of human-environmental systems

Session organiser: Achim Brauer

9-1	Pedro Andrade	Long-term arsenic exposure evidence in northern Chile revealed through analysis of compiled osteological and elemental data
9-2	Michele Abballe	Modelling the evolution of the hinterland of Ravenna: human-environment relationships in an extremely dynamic alluvial landscape
9-3	Michela Leonardi	Palaeoecology made easy: pastclim and tidysdm, two new R packages to make the best use of palaeo big data

	9-4	Margherita Colucci	A case study: modelling human-malaria coevolution through time using paleoclimatic and palaeoenvironmental databases
	9-5	Ümit Cevher Elmas	Quantity over quality? Unraveling the impact of chronological uncertainties in paleoenvironmental databases: A case study from Lake Burdur
	9-6	Thomas Giesecke	The Neotoma Palaeoecology Database as a research tool to assess human landscape interactions during the Quaternary
12:30-13:30	Lunch		
13:30-15:00	Session 10: History unleashed: Harnessing the knowledge stored in historical documents Session organisers: Adam Izdebski and Carlo Cocozza		
	10-1	Luke Anderson-Trocmé	Unraveling French Canadian ancestry: An interplay of genetics, genealogies, and geography
	10-2	David Max Findley	Harmonizing big datasets: PANTROPOCENE, Land use modelling, and the historical tropics
	10-3	Daniel Baránek	From historical sources to Wikibase
	10-4	Etienne Wasmer	A cross-verified database of locations visited by notable individuals
	10-5	Clément Gorin	A convolutional network approach to mining historical map data
	10-6	Sonia Medina Gordo	[POSTER] Script-based automation practices: some experiences for managing early medieval society through written evidence
15:00-15:30	Coffee break		
15:30-16:45	Session 11. Modelling expansions in South America: Integrating archaeology and linguistics Session organisers: Fabrício Ferraz Gerardi and Bruno de Souza Barreto		
	11-1	Paul Heggarty	Combining archaeology, language and genetics into a single human (pre)history through time, space, and ecology
	11-2	Rafael de Almeida Lopes	Using spatial and chronological modelling to understand the Polychrome Tradition Expansion in Central Amazon
	11-3	Bruno de Souza Barreto	Demic diffusion, cultural transmission and change: assessing the spread of the Koriabo pottery style as an

			archaeological correlate of the Cariban-languages expansions in northern Amazonia
	11-4	Fabrício Ferraz Gerardi	A phylogenetic study of the Cariban language family
	11-5	Kathrin Nägele	Beyond broad strokes – the future of ancient DNA in the Americas
16:45-17:00	Closing remarks		

Saturday, November 25

10:00-12:00	Talk: A tour into the history of Jena

http://bhdc.earth

Presentation Abstracts

22–25 Nov 2023

Session 1: Big databases: How large-scale archives can boost archaeological research

S1-1 The Pandora Initiative: promoting fairer data collaboration

Ricardo Fernandes[1]
[1]Max Planck Institute of Geoanthropology, Germany
Contact: fernandes@gea.mpg.de

Historical and archaeological databases are of paramount importance for research, providing a wealth of information that can be used to understand our past and inform our future. However, the move towards Big Historical Data faces several challenges. These include the dispersal of data across different repositories, the concentration of major data initiatives in affluent western countries leading to a form of data colonialism, and the difficulty in accessing expensive modelling tools necessary for data analysis. The Pandora data initiative has been developed to address these issues, providing a platform where data communities can freely set up, self-manage, and self-curate their electronic resources. This is achieved in accordance with FAIR (Findable, Accessible, Interoperable, and Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) principles, including the assignment of unique identifiers and the provision of various types of open-access licenses. Pandora also allows users to deposit not only files but also links to resources stored externally, with all data in Pandora, or linked via Pandora, made available through an API. One of the API links is to a software platform that includes tools for data visualization, analysis, and advanced modelling. The platform is increasingly incorporating artificial intelligence, including AI tools to query data and create open linked data formats. Some of the features of Pandora will be illustrated in the presentation, demonstrating its potential to revolutionize the way we manage and use historical and archaeological data.

S1-2 From the Neolithic to the Middle Ages: A Big Data view of foodways in ancient Italy

Martina Farese[1], Giulia Formichella[1], and Noemi Mantile[2]
[1]La Sapienza University, Italy
[2]Università della Campania "Luigi Vanvitelli", Italy
Contact: m.farese@uniroma1.it

The Mediterranean diet, an intangible cultural heritage since 2013, is a distinctive trait of the countries facing the Mediterranean basin. Its convivial nature, focused on sharing food and passing down traditions, is tightly linked to society and culture. We want to explore how and whether this peculiar diet has changed through time in Italy specifically. To do so, a Big Data approach has been applied to the stable isotope data produced for this area. Human, animal and plant remains recovered from archaeological sites can be used to investigate the lifestyle of past populations. In particular, carbon, nitrogen and sulphur stable isotopes ($\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$) extracted from bone collagen and plant material reflect the protein portion of the organism's diet. Here we propose an investigation of the evolution of the Mediterranean diet in Italy from the beginning of the Neolithic to the end of the Middle Ages (ca. 6000 BCE - 1500 CE). A Bayesian approach has been applied to the isotopic data available in the scientific literature for these millennia to reconstruct the Italian diet over the ages. The emerging dietary trend reflects a general reliance on plant proteins rather than animal ones. Stable carbon isotopes suggest an almost exclusive consumption of C_3 plants before the Bronze Age and during

the Roman times, while C_4 plants were part of the human diet during the Bronze and Iron Ages and the Middle Ages. This contribution illustrates the potential of archaeological, anthropological and historical databases to observe how dietary signatures have changed according to major historical events: from the start of farming economies in the Neolithic and the rise and fall of italic civilisations in the Metal Ages to the territorial and cultural unity of the Roman Empire and the new fragmentation of peoples during the Middle Ages.

S1-3 STARC OSTEOARCH: An open access resource for recording and sharing human osteoarchaeological data

Antonio Caruso[1]

[1]Science and Technology in Archaeology and Culture Research Centre (STARC), The Cyprus Institute, Cyprus

Contact: a.caruso@cyi.ac.cy

This work introduces an open access resource for recording and sharing macroscopic data on archaeological human skeletal remains: STARC OSTEOARCH. The resource was developed using a cloud-based platform, Airtable, and it accommodates data collection on individual skeletons as well as disassociated commingled remains. The attributes covered include key information on demography (sex and age-at-death), taphonomy (zonation, anatomical preservation, weathering, fragmentation, other types of post-mortem alteration), diet (dental diseases, dental wear), mechanical stress (entheseal changes, osteoarthritis, Schmorl's nodes, vertebral arthritis), biodistances (cranial metrics and nonmetrics, dental metrics and nonmetrics), and various pathological lesions. Given the flexibility required in data collection from different assemblages with different research questions, but also the constant development of new methods and recording standards, STARC OSTEOARCH was designed so that each user may add, remove, and edit tables and attributes. References and key descriptions of the categories per attribute have been incorporated in the resource to facilitate raw data sharing, which can be achieved using Airtable's sharing functionality. This initiative aims at promoting standard practices in the field and facilitating data collection, analysis and sharing.

S1-4 Tracking human diet and migration using the Caribbean and Mesoamerica Biogeochemical Isotope Overview (CAMBIO)

Claire Ebert[1], Sean W. Hixon[2], Gina M. Buckley[3], Richard J. George[4], Sofía Pacheco-Fores[5], Juan Manuel Palomo[6], Ashley E. Sharpe[7], Óscar R. Solís-Torres[2], Douglas J. Kennett[4], and Ricardo Fernandes[2]

[1]University of Pittsburgh, Department of Anthropology, PA, USA
[2]Max Planck Institute of Geoanthropology, Germany
[3]University of Algarve, ICArEHB, Portugal
[4]University of California, Department of Anthropology, CA, USA
[5]Hamline University, Department of Anthropology, MN, USA
[6]University of Arizona, Department of Anthropology, AZ, USA
[7]Smithsonian Tropical Research Institute, Center for Tropical Paleoecology and Archaeology, Panama
Contact: c.ebert@pitt.edu

Calls for Open Science have been voiced across Latin America and the Caribbean for well over a decade. An Open Science approach to studying the past that emphasizes data stewardship, accessibility, and transparency allows archaeologists to provide long-term perspectives on overcoming contemporary issues ranging from inequality to mass migration and adaptations to climate change. In this talk, we present CAMBIO (Caribbean & Mesoamerica Biogeochemical Isotope Overview), a collaborative effort led by Early Career Researchers based in Latin America, the US, and Europe to systematically curate and contextualize published biogeochemical data from the Caribbean, Mesoamerica, and Central America in an open-access and multilingual format (Spanish, French, and English). Individuals in the human dataset (n=5035) represent a broad time span, from the Archaic (~7000 BC) to the Colonial/Historic (~AD 1500-1800). Because it is the largest repository of biogeochemical data for the region, the broad spatial and temporal scope of CAMBIO means that the database holds great potential to address a range of anthropological and archaeological questions about past human behavior. We provide two examples. The first synthesizes stable carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotopic data demonstrating dietary change over time related to plant domestication (particularly maize) and agricultural intensification. The second uses strontium isotope data ($^{87}Sr/^{86}Sr$) to document population movement, highlighting the rise of large Classic period (AD 1-500) cities such as Teotihuacan in central Mexico and the kingdoms of the Maya lowlands. While CAMBIO currently focuses on isotopic measurements from human skeletal tissues, the initiative is expanding to include measurements from animal, plant, and geological samples, which will allow researchers to address questions about dynamic interactions among humans, animals, and the environment in the Caribbean, Mesoamerica, and Central America over the past 10,000 years.

S1-5 Building archaeobotanical and archaeozoological database in the BIAD: structure, creation experience and comparative analysis perspectives

Natalia Riabogina[1], Jelena Bulatović[1], Claudia Speciale[1,2], Adrian Timpson[3], Elisabetta Canteri[4], Jan Kolář[3], and Stephen Shennan[3]

[1]University of Gothenburg, Sweden

[2]IPHES Catalan Institute for Human Palaeoecology and Social Evolution, Spain

[3]University College London, Institute of Archaeology, UK

[4]University of Copenhagen, Copenhagen, Denmark

Contact: natalia.riabogina@gu.se

The Big Interdisciplinary Archaeological Database (BIAD) is an initiative that standardizes and consolidates a wide range of archaeological data, providing a pan-European perspective on prehistoric processes from the Carpathians and Balkans to the Scandinavian peninsula and from the French Atlantic coast to the Urals. The structure of BIAD's collected data adheres to a specifically designed template, integrating many modern and older sources. Qualitative and quantitative information is meticulously gathered from original publications, incorporating coded taxon identifications and links with occupation phase, chronology and other archaeological site data to ensure accuracy and consistency over time and space. Here, we focus on archaeobotanical and archaeozoological data within BIAD, marking a significant advance in their collection, harmonization, interpretation and potential for extensive analytical research.

To date, BIAD has synchronized approximately 1380 archaeozoological and 850 archaeobotanical phases from numerous sources, and data collection efforts are ongoing. Collecting archaeozoological and archaeobotanical data has led to a variety of challenges, such as integrating information from different scientific schools of thought into a unified database, dealing with misidentifications or lack of contextual information, addressing sampling and preservation biases and tackling interpretive complexities.

The great availability of plant macro remains and faunal assemblages from archaeological contexts linked to other scientific data within BIAD now offers new analytical possibilities for facilitating dietary reconstructions, examining farming practices, exploring cultural and exchange networks and comparing regional variations in economies. Additionally, these datasets provide novel opportunities for modeling to simulate past agricultural systems and predict potential interactions with factors like climate, land-cover changes and demographic fluctuations. However, extracting meaningful insights from the data requires interdisciplinary collaboration and contextual analysis. In our paper, we advocate for the publication of primary archaeozoological and archaeobotanical data following the standardized protocol developed in BIAD, which maximizes the potential for robust data and enables comprehensive analysis.

S1-6 Bridging human paleogenetics and archaeology: problems with existing practices and prospects for overcoming them

Victor Yan Kin Lee[1], Robert Staniuk[2], Adrian Timpson[2], Stephen Shennan[2], Mark G. Thomas[3], and Fernando Racimo[1]

[1]Section for Molecular Ecology and Evolution, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[2]University College London, Institute of Archaeology, UK

[3]University College London, Genetics Institute, Department of Genetics, Evolution and Environment, UK

Contact: yan.kin.lee@sund.ku.dk

The recent increase in available ancient DNA (aDNA) data has provided new opportunities to understand our past, but also imposed challenges for data recording and sharing. Several resources have been developed to curate and harmonize large amounts of data, e.g. the Allen Ancient DNA Resource (AADR) and the Poseidon framework. Though widely used, these efforts have largely focused on the genetic aspect of data associated with human remains, and lack easy-to-use tools for relating it to non-genetic data (e.g. material culture, isotopes, or associated faunal or botanical remains). Additionally, inconsistent naming conventions limit cross-referencing with other archaeological databases. Thus, investigating the relation between genetic patterns and other historical processes remains a persistent challenge. Making progress on these issues requires taking a step back to re-evaluate what type of data we collect, how we collect it, and which standards we use to report it. Here, we present a meta-analysis of current standards in publishing and collecting aDNA data to evince issues that limit our ability to tackle interdisciplinary research questions. Specifically, we focus on attempts to relate two large aDNA datasets to a large archaeologically-focused database: the Big Interdisciplinary Archaeological Database (BIAD). Throughout this process, we have found numerous practical and conceptual challenges for interoperability across databases, which we think will be best addressed via community efforts. We also propose a specific data collection and structuring methodology for co-analysing genetic and archaeological data which, we argue, will open the way for more systematic efforts at exploring how genes, culture and other historical processes have interacted in the past.

S1-7 The ROAD database as a research tool for exploring human evolution

Andrew W. Kandel[1], Michael Bolus[1], Angela A. Bruch[2], Claudia Groth[2], Miriam N. Haidle[2], Christine Hertler[2], Julia Heß[2], Volker Hochschild[1], Zara Kanaeva[1], Maria Malina[1], Friedemann Schrenk[2], Christian Sommer[1] and Nicholas J. Conard[1]

[1]The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities at the University of Tübingen, Germany

[2]The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities at the Senckenberg Research Institute and Natural History Museum, Germany

Contact: a.kandel@uni-tuebingen.de

ROAD, the ROCEEH Out of Africa Database (https://www.roceeh.uni-tuebingen.de/roadweb/), serves as a valuable resource for the study of human evolution. ROAD collects information about prehistoric sites in Africa, Europe and Asia ranging from three million to 20,000 years ago. The database unifies geographical data with stratigraphic and chronological information. ROAD assimilates data about archaeological assemblages, including human fossils, faunal and botanical remains as well as climate and bibliography.

Starting in 2008, the multidisciplinary team of the research center "The Role of Culture in Early Expansions of Humans", known as ROCEEH (https://www.roceeh.net), has thus far integrated 2,400 localities containing more than 23,000 assemblages from over 5,300 publications written in most major languages. Thus ROAD serves as a storehouse of information that can be easily accessed and explored using various methods of data science.

ROAD enables user interaction through its web-based application called ROADWeb, written in php, javascript and html. The SQL query builder tool helps users formulate questions to the database. The resulting lists can be exported as CSV files, or visualized directly in ROAD using the Map Module. The ROAD team assists users in formulating complicated queries and provides the results through php scripts that perform live queries of ROAD.

In this talk, we discuss case studies that illustrate some of the ways in which the data from ROAD can boost large-scale studies related to human evolution. One study used archaeological data in ROAD to examine how continent-wide settlement patterns, as shown through the "human niche", relate to large-scale climatic trends in Africa. Another study used faunal data from Europe to create a spatiotemporal species distribution model which was then used to estimate abundances of species during those times. Finally, one study examined the chronological and spatial development of ochre use in Africa over the last 500,000 years.

S1-8 Cultural evolutionary trends as revealed by a novel, expert-sourced dataset on lithic technology, toolkits, and artefact shapes for the Final Palaeolithic/earliest Mesolithic of Europe (15-11ka BP)

Felix Riede[1], David N. Matzig[1], Miguel Biard[2], Philippe Crombé[3], Javier Fernández-Lopéz de Pablo[4], Daniel Groß[5], Thomas Hess[1], Mathieu Langlais[6,7], Federica Fontana[8], Ludovic Mevel[9], William Mills[10], Martin Moník[11], Nicolas Naudinot[12], Caroline Posch[13], Tomas Rimkus[14], Damian Stefański[15], Hans Vandendriessche[3], and Shumon T. Hussain[1]

[1]Department of Archaeology and Heritage Studies, Aarhus University, Denmark

[2]Institut National de Recherches Archéologiques Préventives, Technologie et Ethnologie des Mondes Préhistorique, France

[3]Department of Archaeology, Ghent University, Belgium

[4]I.U. de Investigación en Arqueología y Patrimonio Histórico, University of Alicante, Spain

[5]Museum Lolland-Falster, Nykøbing F, Denmark

[6]CNRS UMR 5199 PACEA, University of Bordeaux, France

[7]SERP University of Barcelona, Spain

[8]Dipartimento di Studi Umanistici – Sezione di Science Preistoriche e Antropologiche, University of Ferrara, Italy

[9]CNRS UMR 7 8068 Technologie et Ethnologie des Mondes Préhistoriques, Paris-Nanterre, France

[10]Centre for Baltic and Scandinavian Archaeology (ZBSA), Schleswig, Germany

[11]Department of Geology, Faculty of Science, Palacky University Olomouc, Czech Republic

[12]Université Côte d'Azur, CNRS-CEPAM, France

[13]Natural History Museum Vienna, Vienna, Austria

[14]Institute of Baltic Region History and Archaeology, Klaipêda University, Lithuania

[15]Archaeological Museum, Kraków, Poland

Contact: f.riede@cas.au.dk

Archaeological systematics, together with spatial and chronological information, are commonly used to infer cultural evolutionary dynamics in the past. For the study of the Palaeolithic, and particularly the European Final Palaeolithic and earliest Mesolithic, proposed changes in material culture are often interpreted as reflecting historical processes, migration, or cultural adaptation to climate change and resource availability. Drawing on recent developments in computational archaeology, we here present a data-driven assessment of the existing Final Palaeolithic/earliest Mesolithic cultural taxonomy in Europe. Our dataset consists of a large expert-sourced compendium of key sites, lithic toolkit composition, blade and bladelet production technology, as well as formal lithic artefact shapes. The dataset comprises 16 regions and 86 individually named archaeological taxa ('cultures'), covering the period between ca. 15,000 and 11,000 years ago (cal BP). Using these data, we use geometric morphometric and multivariate statistical techniques to meta-analytically explore to what extent the dynamics observed in different lithic data domains (toolkits, technologies, armature shapes) correspond to each other and to the culture-historical relations of taxonomic units implied by traditional naming practice. Our analyses support the widespread conception that some dimensions of material culture became more diverse towards the end of the Pleistocene and the very beginning of the Holocene. At the same time, cultural taxonomic unit coherence and efficacy appear variable, leading us to explore potential biases introduced by regional research traditions, inter-analyst variation, and the role of disjunct macroevolutionary processes. In discussing the implications of these findings for narratives of cultural change and diversification across the Pleistocene-Holocene transition, we emphasize the increasing need for cooperative research and systematic archaeological meta-analyses.

S1-9 An open and interdisciplinary approach to modelling movement along the ancient Lybian Sea

David Laguna-Palma[1] and Pablo Barruezo-Vaquero[2]

[1]Department of Prehistory and Archaeology, University of Granada, Spain

[2]University of Granada, Spain

Contact: dlaguna@ugr.es

This paper is based on the efforts made by the PERAIA Project, which gathers an international team to trace movement and understand the historical connectivity within the regions bordering the Libyan Sea during the transition from the Late Bronze Age to the Iron Age (ca. 1400-700 BC). The study is thus diachronic in scope and considers different scales of analysis. Its theoretical framework blends notions from Historical Ecology, human ecodynamics, and Landscape Archaeology, as well as approaches from Digital Archaeology. In other words, we aim to model movement whilst accounting for the systemic and historical interaction between humans and non-humans across this vast region. The complexity of this research has required the application of different digital approaches (e.g., geospatial analysis and archaeological network research) in a methodology that maps archaeological sites by tying together various strands of evidence, from satellite imagery, historical sources, and environmental data. Such a large quantity of multivariate data is managed through a geospatial dataset, the backbone upon which our interdisciplinary knowledge is built. Likewise, the project seeks to attain ethical practices based on Open Science. Hence, digital technologies and the LOUD+FAIR data principles have been at the core of the project in terms of data development and management. This paper seeks to briefly explain–from theory to practice–our approach and workflow to modelling past mobility while enabling open research. We believe such an approach might be of interest to other researchers working in interdisciplinary environments.

S1-10 MetaMobility databases: Promoting Greco-Roman bioarchaeology

Efthymia Nikita[1], Antonio Caruso[1], Panagiotis Koullouros[1], Mahmoud Mardini[1], and Evi Margaritis[1]

[1]The Cyprus Institute, Cyprus

Contact: e.nikita@cyi.ac.cy

The MetaMobility project aims at exploring human mobility and its impact on health, diet, mechanical and physiological stress, and economy in the Greco-Roman world. Its more specific geographic and temporal focus is the Hellenistic and Roman Central and Eastern Mediterranean and Middle East. In the context of this project, our team has been building a series of online open access databases focused at the promotion of bioarchaeology in the region. Some of these databases are bibliographic; they compile the publication details of published works on human osteoarchaeology, zooarchaeology and archaeobotany in the region from prehistory to Medieval times, acknowledging that many of these works are available as gray literature, diminishing their visibility to international scholarship. Currently available bibliographic databases include BiblioArch and BiblioArch-Italia, and we are currently completing Zoo-BiblioArch and BiblioArch-Archaeobot. In addition, we have created the SrIsoMed database in which we have compiled all published strontium isotopic values for organic and inorganic materials across the Mediterranean. The aim here was to facilitate palaeomobility and provenance studies, particularly through the interactive map of the database and the different search functionalities. Importantly, we are currently compiling all published data on Hellenistic and Roman human osteoarchaeology, zooarchaeology and archaeobotany and we are going to share them openly in a new database with an interactive map and appropriate search functionalities. In this paper, we briefly present the above initiatives and the challenges we have faced/are facing in their development, update and long-term curation. Finally, we see this conference as an ideal setting to exchange opinions with other participants on the above challenges, find ways to overcome them, and potentially come up with future collaborations.

S1-11 Big Archaeological (Legacy) Data. Collection, dissemination and reuse for large scale analysis

Antonio Campus[1] and Gabriele Gattiglia[1]

[1]University of Pisa, Italy

Contact: antonio.campus@cfs.unipi.it

In a post-digital discipline (and world), it is becoming more and more common for archaeologists to reuse data collected by other scholars at different times, with different methods, tools, and technologies, and this aware reuse of data is bringing new challenges to archaeological practice.

Since 2011, the MAPPA Laboratory of the University of Pisa has been engaged in the systematic digitisation of (all available) legacy archaeological and palaeoenvironmental data, starting from the urban area of Pisa (MAPPA project – 2011-2013) and expanding the boundaries of research over the years (ager Pisanus – 2018-2021 and MAGOH 2020-2022 projects). The MAPPA/MAGOH site-less archiving system currently collects more than 7950 intervention records (i.e. every single archaeological activity performed per unit of time and space), covering a large area of northern Tuscany (Italy), and preserves and disseminates data and metadata. Digitisation required significant efforts to overcome the issues related to different degrees of analysis, terminology, and chronological or topographical approximation. This increased availability of complex and variable data prompts archaeologists to think in terms of 'Big Data', accepting the use of large volumes of data from different and heterogeneous sources, as well as approximation and messiness. The aggregation of

all available data (or at least as much of it as possible) may indeed provide insight into large-scale phenomena, enhancing even details invisible to limited data sets and mitigating sampling bias.

This approach has been followed in the Ager Pisanus project to outline human-environmental inter-relationships in the Hellenistic, Roman, and Late-Antique periods by means of GIS and R-based spatial analyses (e.g. Point Pattern Analysis, Geostatistical Analysis, Regression Models). The interoperability between qualitatively and quantitatively different data thus offered a more defined understanding of the overall pattern at a landscape scale.

S1-12 Exploring large-scale historical processes using IsoIberMed: A new isotope database for bioarchaeological samples from Medieval Iberia

Maite I. Garcia-Collado[1], Carlo Cocozza[2], Ricardo Fernandes[2], and Michelle Alexander[3]

[1]GIPYPAC, Department of Geography, Prehistory and Archaeology, University of the Basque Country (UPV/EHU), Spain

[2]Max Planck Institute for Geoanthropology, Germany

[3]University of York, UK

Contact: maiteirisgarcia@gmail.com

The Medieval period was a transformative time in the history of Iberia, which witnessed multiple significant socio-political transitions. Suevi, Vandal and Visigothic kingdoms emerged after the disarticulation of the Roman Empire, which themselves disappeared after the Islamic conquest in 711. For around 800 years Iberia harboured a multifaith society, where the co-existence of Christian, Jewish and Muslim communities led to a uniquely blended culture. New Christian kingdoms arose in the north and slowly expanded south throughout the later Medieval period, consolidating feudal regimes and taking the last Islamic rulership of Granada in 1492.

In this complex setting isotopic proxies offer relevant insights on the socio-economic organisation and environmental context of Medieval populations. The now significant accumulation of isotopic measurements allows for broad scale comparative studies that promise to offer novel insights into the history of the Middle Ages. IsoIberMed was developed to explore the potential of this research.

IsoIberMed is a database fully accessible online and continuously updated which collates publicly available isotopic measurements on bioarchaeological samples from continental Spain, the Balearic Islands, Portugal and Andorra dated to between 5th and 15th c. CE. Currently it comprises data from approximately 3000 human, 1000 faunal and 300 plant samples. This mostly consists of carbon and nitrogen stable isotope measurements on collagen extracted from bones, but it also includes other sampled materials (e.g. dental enamel, dentine, seeds) and from different isotopic proxies (i.e. sulphur, oxygen, strontium isotopes).

To illustrate the research potential offered by IsoIberMed we conducted a Bayesian meta-analysis of the compiled data. This was used to explore spatiotemporal patterns in human diet and spatial mobility, as well as their link to intricate socio-economic and cultural backgrounds in the context of the mosaic of climatic and environmental conditions that characterise this dynamic region between the Atlantic and the Mediterranean.

S1-13 Presenting a compilation of isotopic data from Madagascar

Sean Hixon[1], Ricardo Fernandes[1], Andrea Baden[2], Marina Blanco[3], Tim Eppley[4], Bruno Frédérich[5], Jörg Ganzhorn[6], Steve Goodman[7,8], Mitch Irwin[9], Lisa Kelley[10], Michel Loïc[5], James Loudon[11], Laurent Mittelheiser[5], Yedidya Ratovonamana[12], Matt Sponheimer[13], and Brooke Crowley[14]

[1]Max Planck Institute of Geoanthropology, Germany

[2]City University of New York, NY, USA

[3]Duke University, NC, USA

[4]San Diego Zoo, CA, USA

[5]University of Liege, Belgium

[6]University of Hamburg, Germany

[7]Field Museum, IL, USA

[8]Association Vahatra

[9]Northern Illinois University, IL, USA

[10]Saint Louis Zoo, MO, USA

[11]Eastern Carolina University, NC, USA

[12]University of Antananarivo, Madagascar

[13]University of Colorado, Boulder, CO, USA

[14]University of Cincinnati, OH, USA

Contact: hixon@gea.mpg.de

Isotopic research with modern biological materials from Madagascar has helped both to characterize endemic biodiversity and to identify ecological interactions that affect the spatial distribution and abundance of species. These data can also be integrated with similar data from ancient material to learn about the resource use of extinct taxa and past changes in the diet of extant animals. We present the first open-access, island wide compilation to include the full extent of isotopic data gathered from biological materials on the island and in the near coastal marine environment. The metadata structure of our compilation 1) makes the data easily searchable and accessible to users who are relatively unfamiliar with the literature and 2) helps encourage full data reporting in future publications. Consequently, the database should facilitate the stable isotope research by making it easy to identify both research gaps and opportunities for using various isotope systems to answer a given research question.

S1-14 The IsoMemo Initiative: Enhancing archaeological and historical research through isotope databases

Carlo Cocozza[1] and Ricardo Fernandes[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: cocozza@gea.mpg.de

Isotopes have become indispensable tools in archaeological and historical research, providing critical insights into past climates, dietary practices, and migration patterns. The utility of isotopes, however, is significantly enhanced when they are organized into databases, enabling researchers to compare and contrast findings across different studies and geographical locations. This paper introduces the IsoMemo initiative, a network of autonomous databases designed to facilitate the use of isotopes in archaeological and historical research. IsoMemo implements cutting-edge technologies to ensure data access aligns with the FAIR (Findable, Accessible, Interoperable, and Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) data principles, promoting efficient and ethical data usage. The presentation will

include illustrative examples demonstrating the value of this data collaborative model in advancing our understanding of the past.

Session 2: Big Zooarchaeological Data: Challenges and potentials for multi-scalar analysis across millennia

S2-1 From small to Big Data in zooarchaeology

Sarah Kansa[1], Angela Trentacoste[2], and Jesse Wolfhagen[3]

[1]Open Context

[2]British School at Rome, Italy

[3]Purdue University, IN, USA

Contact: sarahkansa@gmail.com

This paper introduces the session "Big Zooarchaeological Data: Challenges and potentials for multi-scalar analysis across millennia." Zooarchaeological data analysis leverages standard descriptions and ontologies more regularly than other areas of archaeology, making zooarchaeology "low hanging fruit" for archaeological data integration. While the potential for addressing "Big Data" research topics in zooarchaeology is great, there are still many challenges in moving from "small data" collected by individuals or small teams to "Big Data" brought together by many researchers. We highlight past and ongoing collaborations that bring together small data to address big picture topics. We discuss the challenges of attempting to integrate legacy data with newly collected data, as well as working with multiple analysts across many projects with different research agendas. Improving data documentation, sharing of methods, and positioning data for future integration are all steps individual practitioners can take to improve the longevity and relevance of their data for "Big Data" research.

S2-2 Crosswalking, linking, and verbatim interpretations: Adventures in legacy zooarchaeological biodiversity Open Data publishing

Kitty Emery[1]

[1]Florida Museum of Natural History, University of Florida, FL, USA

Contact: kemery@flmnh.ufl.edu

Zooarchaeological data is essential to accurate biodiversity research, providing vital Anthropocene links between paleontology and neontology. For example, conservation biology depends on human-related data to quantify pre-human baselines and measure differing levels and types of human activity, while ecological/evolutionary research requires human-related data to determine causality for observed variations. Luckily, zooarchaeological analyses should be an ideal source of detailed, verifiable, spatially and temporally discrete biodiversity data that shares taxonomic, anatomic, and life-stage terminology to share with open data repositories that publish animal biodiversity research from paleontological (such as the PaleoBiology DataBase) and neontological (such as the Global Biodiversity Information System) specimens.

However, as this paper will discuss, the hurdles to publishing zooarchaeological data in biodiversity repositories are more numerous and complex than we might expect. Even though we study biological data, which data we collect, our basic terminologies, our interpretive methodologies, and the contextual depth that we consider vital vary widely across the disciplines. The difficulties in "fitting" zooarchaeological data into other biodiversity datasets is problematic, but more troubling is the possible misinterpretation of zooarchaeological data within Big Data biodiversity research. For example, what is an individualCount – an MNI or a NISP? Does the neontologist count 10 femoral segments as 10 deer or five or one? Do 10 jaguars in a multi-use ossuary reflect population size?

Here I explore solutions (and some continuing puzzles) to these hurdles: when and how to conserve original data, how to cross-walk terms among various repositories, how and where to provide essential methodological information, and how to link across disciplinary databases to ensure that future researchers can find original data, replicate analyses, and link "extended data". Zooarchaeological data is indeed biological data, but it is also cultural data, and our data-types, methods, and data interpretations must reflect the cultural contexts and meanings of the remains.

S2-3 Cross-disciplinary reuse potential of big zooarchaeology data: What do we need?

Canan Çakırlar[1], Rob Lenders[2], and Youri van den Hurk[3,1]

[1]University of Groningen, Netherlands

[2]Radboud University, Netherlands

[3]Norwegian University of Science and Technology, Norway

Contact: c.cakirlar@rug.nl

No single zooarchaeologist would be able to generate data sufficiently large enough to tackle supraregional and diachronic issues. Large zooarchaeological data are usually collated from a variety of sources using different means. Which means lead to more impact, if we define impact as cross-disciplinary reuse and knowledge utilization in non-academic settings.

In this paper we assess the impact of Dutch BoneInfo as a top-down initiative to collect all big zooarchaeology data vs. question-driven academic initiatives that collate data relevant to research question. BoneInfo is a government-funded, online database of taxonomic absence/presence data from commercial archaeology. We suggest pathways to mobilize the reuse potential of big zooarchaeological data in cross-disciplinary and heritage settings.

S2-4 Assessing approaches to chronological uncertainty in large-scale zooarchaeological data synthesis

David Orton[1], Sarah Everett[1], and Mathilda Holmes[2]

[1]University of York, UK

[2]University of Leicester, UK

Contact: david.orton@york.ac.uk

One major obstacle to large-scale zooarchaeological syntheses is the challenge of integrating data from numerous sites or contexts with diverse, overlapping, and variably resolved dating information. Traditional approaches place data into predefined chronological bins, typically based on the mid-points of reported date ranges. This forces a trade-off between resolution, accuracy, and inclusion: larger bins reduce the resolution at which trends can be detected; smaller bins increase the risk of incorrect assignment unless poorly/inconveniently dated entities are excluded. Moreover, binning inevitably pushes apparent changes towards predefined points, potentially making gradual trends appear sudden or vice versa. This is particularly problematic where the chosen divisions are based on researchers' initial hypotheses regarding the timing of changes.

This paper explores the application of three alternative approaches that seek to incorporate dating uncertainty in the analysis: unweighted running 'snapshots', aoristically weighted means, and Monte Carlo methods. We consider the appropriateness of each method to different classes of zooarchaeological data, paying particular attention to the nature of archaeological date ranges as a conflation of duration and uncertainty. We argue that Monte Carlo methods are most appropriate where date ranges primarily represent uncertainty, while deterministic approaches may be more suitable where duration is a major component. In general, the former will be true for specimen-level data (e.g. metrics, stable isotopes), while the latter may often apply at the assemblage level (e.g. relative taxonomic frequencies) — though much depends on levels of data aggregation and reporting. We illustrate the discussion with examples of taxonomic and metrical data drawn from ongoing data synthesis projects in European historical zooarchaeology.

S2-5 Tracking long-term subsistence among the Native North American Great Plains hunters

Matthew Hill, Jr.[1], Jesse L. Wolfhagen[2], and Erik Otárola-Castillo[2]

[1]University of Iowa, IA, USA

[2]Purdue University, IN, USA

Contact: matthew-e-hill@uiowa.edu

The economies of Native North American hunters of the Great Plains are often depicted as depicting as being almost exclusive reliant on large-bodied prey, such as bison. However, this simplified perspective ignores temporal and spatial variability in environmental conditions, including Native economies encompassing hunting and gathering, horticulture, and eventually agriculture over a more than 13,000-year history. Here we employ a large dataset of faunal remains from more than 500 archaeological sites to refine the accuracy of long-term dietary patterns in the Great Plains and investigate the broad patterns of specialization and diversification of faunal use among Great Plains indigenous people through time. This study uses Bayesian Multilevel Modeling (BMM) provides a framework to balance the drawbacks of the current data structure and improve predictive power. Our findings suggest that even though prehistoric indigenous people of the Great Plains maintained a way of life associated with bison over time, bison was not

the sole focus nor the dominant species in their diet. In addition, we observe that hunting strategies and dietary variation through time does not support a model of progressive resource intensification from hunting and gathering to farming. Instead, Great Plains indigenous hunters exploited a variable set of faunal resources contingent on their environmental and socio-demographic settings.

S2-6 Wool they won't they: Combing the zooarchaeological evidence for wool production in Northern Mesopotamia c. 4500-1500 cal. BC

Max Price[1] and Jesse Wolfhagen[2]

[1]Durham University, UK

[2]Purdue University, IN, USA

Contact: max.d.price@durham.ac.uk

Wool was one of the most important commodities in ancient Mesopotamia, a fact amply demonstrated by texts that document the large-scale production and processing of wool. Zooarchaeologists have devised a number of methodological techniques for examining intensive wool production. Here, we offer a survey and critical examination of the zooarchaeological data, focusing on kill-off patterns and biometrical data from sites dating to the Chalcolithic through Middle Bronze Age (c. 4500-1500 BC) in northern Mesopotamia/eastern Anatolia. We synthesize the existing data and offer new Bayesian statistical methods to estimate the retention of older animals (percentage of mandibles in stages G, H, or I), the ratio of males to females, and the size of sheep. We corroborate the shift in sheep size over time, a finding of earlier studies, with a larger population becoming dominant in most sites in the 4th millennium BC. But we find no clear pattern in the ratio of males to females due to a high level of uncertainty in mixture models of biometrical data, despite using an up-to-date statistical approach based on Bayesian modeling. This highlights the challenges of utilizing mixture modeling to estimate sex composition in caprine assemblages. There are, however, important — though not radical — shifts over time in age-at-death data, with the number of older caprines, matching expectations of an increasing regional economic focus on wool production from the late 4th through early 2nd millennia BC.

S2-7 The RomAniDat network: a zooarchaeological Big Data initiative for the ancient Roman world

Anthony King[1], Dominika Schmidtová[2], Maaike Groote[3], Ricardo Fernandes[4]

[1]University of Winchester, UK

[2]Department of Archaeology and Museology, Masaryk University, Czech Republic

[3]Institut für Prähistorische Archäologie, Freie Universität Berlin, Germany

[4]Max Planck Institute of Geoanthropology, Germany

Contact: tony.king@winchester.ac.uk

Zooarchaeological research serves as a compelling methodology for the reconstruction of past human dietary habits, hunting methods, domestication of animals, and environmental changes among other purposes. This enables researchers to delve into the intricate interactions between human and animal life over time, thereby unveiling crucial facets of ancient human societies, their evolutionary process and impact on biodiversity. As zooarchaeological data is accumulated opportunities arise to investigate aforementioned phenomena at various scales. This, in turn, requires that a collective effort is made amongst the experts

within the field of zooarcheology to devise and execute strategies for effective data collection, aligned with Open Science, FAIR, and CARE principles.

This paper outlines the RomAniDat initiative, a collaborative network amassing zooarchaeological data pertaining to the ancient Roman world. It consists of an organized network of independent databases, assembled according to spatial parameters and dedicated to accumulating data relating to faunal abundance. Data, synthesized from the different databases, is disseminated via the Pandora data platform, which also supports methods for data visualization, summarization, and analysis.

The first publication of the RomAniDat initiative featured faunal data from Italy, covering the period from 500 BCE to 500 CE (Schmidtová et al. 2023). Since then, the RomAniDat network has broadened its scope to include additional databases focused on Roman Netherlands and Great Britain. This presentation aims to provide an overview of the RomAniDat initiative, highlighting its current status and illustrating its research potential through modelling case studies. Furthermore, it will briefly outline potential future trajectories for the initiative.

Schmidtová, D., King, A. C., Klontza-Jaklová, V., Reese, D. S., Rizzetto, M., Fernandes, R. 2023: Presenting the RomAniDat data community and an archaeofaunal dataset for Roman Italy. Journal of Archaeological Science: Reports 47, 103764."

S2-8 Understanding variation in zooarchaeological NISP Ddta: Bayesian betabinomial models for Italian contexts in the 1st millennium CE

Roberto Ragno[1]

[1]University of Bari, Italy

Contact: roberto.ragno@uniba.it

This study investigates a comprehensive dataset of 465 zooarchaeological NISP samples from 228 sites across mainland Italy, spanning from the 1st century BCE to the 11th century CE. Large NISP datasets can be challenging to analyse, as they often exhibit overdispersion, influenced by contextual factors like historical and depositional influences. Consequently, a more nuanced approach is required, moving beyond simplistic calculations of mean values for individual animal species. In this study, I employ betabinomial models within a Bayesian statistical framework to address overdispersion and generate credible distributions. The models are developed using non-informative priors for four distinct chronologies (Roman, late Roman, early Medieval, and 11th century), considering different context types, macroregions (Northern, Central, and Southern Italy), and site elevations. Compared to alternative distributions, the betabinomial models more effectively handle variability between samples and within groups, capturing overdispersion and accurately representing complex patterns of variation in the counts. Importantly, these models also account for differences in sample and group sizes. The findings underscore the significance of considering multiple factors when assessing the likelihood of farmed and wild animal occurrence in historical contexts. This study investigates a comprehensive dataset of 465 zooarchaeological NISP samples from 228 sites across mainland Italy, spanning from the 1st century BCE to the 11th century CE. Large NISP datasets can be challenging to analyse, as they often exhibit overdispersion, influenced by contextual factors like historical and depositional influences. Consequently, a more nuanced approach is required, moving beyond simplistic calculations of mean values for individual animal species. In this study, I employ betabinomial models within a Bayesian statistical framework to address overdispersion and generate credible distributions. The models are developed using non-informative priors for four distinct chronologies (Roman, late Roman, early Medieval, and 11th century), considering different context types, macroregions (Northern, Central, and

Southern Italy), and site elevations. Compared to alternative distributions, the betabinomial models more effectively handle variability between samples and within groups, capturing overdispersion and accurately representing complex patterns of variation in the counts. Importantly, these models also account for differences in sample and group sizes. The findings underscore the significance of considering multiple factors when assessing the likelihood of farmed and wild animal occurrence in historical contexts.

Session 3: Environments of big cultural heritage data integration

S3-1 Heritage-environment interaction: Theorising Big (Palaeo)Data integration across the nature-culture divide

Mike Fisher1 and Dovydas Jurkenas[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: michael.fisher@gea.mpg.de

The preservation of cultural heritage today relies on the ability of communities and institutions to work together to integrate a variety of analog and digital methodologies, including conservation, education, and documentation. Digital documentation affords a range of methods that support broad geographic coverage and dissemination of results, potentially better contributing to cultural heritage conservation and education efforts. A key component of documentation is identifying disturbances and risks to cultural heritage resources, particularly economic development and environmental change.

As data-rich methods of recording cultural heritage such as LiDAR, photogrammetry, and Very-High-Resolution satellite imagery expand in use, alongside the increasing number of documented resources globally, both vertical and horizontal scalability become essential concerns. However, growing computational and storage requirements enlarge the technospheric volume of non-biodegradable matter, impacting the environment and potentially contributing to anthropogenic climate change. Similarly, duplication of digital documentation endeavours can produce redundant datasets and unnecessary strain on the natural environment. In turn, increasingly extreme climatic activity presents further long-term risk to immovable cultural heritage.

Linked Open Data (LOD) principles offer one partial solution to this conundrum by enabling canonical dissemination of data and integration of datasets from related fields and beyond. This can enrich cultural palaeodata with information from historical, regional, and environmental sources. Furthermore, knowledge of the palaeoenvironments in which cultural resources were originally formed not only deepens our understanding of the past, but also tells us about climatic change over the longue durée and forms a baseline for estimating further environmental risk to immovable cultural heritage resources. Insights drawn from integrating these types of datasets can help prepare local communities, institutions, and first responders for mitigating threats and disturbances to cultural properties. This paper explores modes of digitalisation that engender Big (palaeo)Data integration for a better understanding of the dynamic relationship between cultural heritage and its natural environment.

S3-2 ARIADNE: aggregating archaeological datasets at an international scale

Julian Richards[1]

[1]Archaeology Data Service, University of York, UK

Contact: julian.richards@york.ac.uk

The archaeological research community was an early adopter of digital tools for data acquisition, organisation, analysis, and presentation of research results of individual projects. However, the provision of e-infrastructure and services for data sharing, discovery, access, and re-use has lagged behind. This situation has been addressed by the ARIADNE Research Infrastructure. Developed with European funding, ARIADNE has created an e-infrastructure that enables data providers to register and provide access to their digital resources through the ARIADNE data portal, facilitating discovery, access, and research. ARIADNE has aggregated resources from over 45 data providers, spanning over 40 countries and 4 continents. The portal (https://portal.ariadne-infrastructure.eu) now provides online access to some 4 million research resources. It already provides an accessible showcase for specialised environmental data sets, such as SEAD, allowing them to be cross-searched alongside more generalised archaeological datasets. The ARIADNE Knowledge Base is based upon Linked Open Data technologies and it is underpinned by a flexible and extensible architecture, enabling multiple combinations and presentations of the same underpinning data. Learning from previous data aggregation projects, we have defined a subset of the CIDOC CRM to be used as a strict ontology and paid close attention to data standards and controlled vocabularies to achieve a high degree of interoperability. This paper will discuss some of the challenges of large-scale data integration and describes the approaches adopted to ensure that the ARIADNE Knowledge Base is an effective tool for researchers. Our experience provides a case study based on an advanced user community for European-wide cross-disciplinary developments in e-infrastructures including EOSC and ECCCH.

S3-3 Transdisciplinary data integration into a computational ontology: some reflections from the DataARC Project on modelling human–non-human entanglements

Pablo Barruezo-Vaquero[1]

[1]MEMOLab (Biocultural Archaeology Lab), Department of Medieval History, University of Granada, Spain

Contact: pablobva@ugr.es

This paper presents a case study based on the DataARC Project, an NSF-funded international project aimed to integrate transdisciplinary biocultural heritage data. DataARC harnessed many concepts brought by Historical Ecology and Human Ecodynamics and put them to work in a digital environment. The Project drew substantially on the work of NABO, an international and interdisciplinary research community focused on the historical biocultural legacies of the North Atlantic. Active for 30 years, NABO has generated an impressive amount of multivariate data (archaeological, historical, ecological, geological, etc.). The DataARC Project integrated most of such legacy data in a computational ontology aiming to represent the human ecodynamics of the North Atlantic in a rigorous and efficient way. The project followed LO(U)D and FAIR standards for developing such an ontology. The CIDOC CRM framework was moreover operationalised to model the interlinkages between multiple data(sets)—and, consequently, of human–non-human agents.

Intending to reflect on the modelling of the biocultural heritage of landscapes, this paper presents some insights from my experience working at DataARC. It thus explores different theoretical and methodological aspects of interest for integrating multivariate data in a Big Data framework guided by a transdisciplinary ethos. I will specifically focus on a case that mapped archaeological data and integrated them with other datasets in the ontology. Hence, this will serve to ponder the challenges and opportunities of semantic modelling for legacy data, with the aim of modelling the historical ecologies of landscapes. Furthermore, this case also considers the importance of developing scalable approaches to interlink 'low-level' and 'medium-level' data with broader narratives. Thus, the paper delves into the creation of knowledge using multivariate datasets from grey sources, which might be of help for representing the complexity of biocultural heritage.

S3-4 "Shared heritage": management and integration of cultural heritage data across Arches-based platforms in the Global South

Thomas Huet[1], Alan Tapscott[2], Jonas Gregorio de Souza[2], Junaid Abdul Jabbar[3], Manoj Lokare[3], Ash Smith[4], Mike Fisher[5], Mahmoud Abdelrazek[6], Bryan Alvey[7], Marco Nebbia[8], and Krisztina Baranyai[9]

[1]University of Oxford, School of Archaeology, UK

[2]Universitat Pompeu Fabra, Spain

[3]McDonald Institute for Archaeological Research, University of Cambridge, UK

[4]Maritime Endangered Archaeology, University of Southampton, UK

[5]Max Planck Institute of Geoanthropology

[6]Advanced Research Computing Centre, University College London, UK

[7]Central Asian Archaeological Landscapes, Institute of Archaeology, University College London, UK

[8]Institute of Archaeology, University College London, UK

[9]Maritime Asia Heritage Survey, Kyoto University, Japan

Contact: thomas.huet@arch.ox.ac.uk

Arches is a FOSS geospatial semantic web-based purpose-built platform for cultural heritage management. Several university projects on the heritage of the Global South are using Arches as an information system to assess the cultural heritage of large continuous geographical areas (sub-Saharan Africa, North Africa and the Middle East, Eurasian steppes, Amazon-Andes, etc.). The use of this software, in such contexts, tackles many of the challenges posed by digital humanities: large datasets, open access and open data, capacity building, digital gap, etc. Interoperability and sustainability are among the main challenges faced by Arches-based projects. To meet these challenges, we develop computer routines, in the form of computer scripts, and host them with reference data on a collaboration and version control platform (https://github.com/achp-project). Among these shared reference data sets, the so-called Arches Resource Models are CIDOC-CRM-compliant classes for describing and organising information related to cultural heritage in Arches: Places, Buildings, Geoarchaeological corings, etc. These graph structures can be parsed and compared with standardised metrics, such as the number of common subgraphs, to understand the similarities and differences in the assessments of cultural heritage made by the different Arches-based projects.

We will present this FAIR approach to cultural heritage management, and discuss current and future challenges.

S3-5 Mapping the archaeological Pre-Columbian heritage in South America

Jonas Gregorio de Souza[1], Marco Madella[1], Marco Moderato[1], Alan Tapscott Baltar[1], Shaddai Heidgen[1], Sebastian Fajardo Bernal[2], Juan Carlos Vargas Ruiz[3], Eduardo Góes Neves[4], and André Strauss[4]

[1]CASES, Universitat Pompeu Fabra, Spain

[2]Delft University of Technology, Netherlands

[3]Universidad del Magdalena, Colombia

[4]Museum of Archaeology and Ethnology, University of São Paulo, Brazil

Contact: jonas.gregorio@upf.edu

This presentation explores how the Mapping the Archaeological Pre-Columbian Heritage in South America (MAPHSA) project integrates transregional datasets from Brazil and Colombia, encompassing a broader range of cultural heritage and enabling cross-cultural comparisons. MAPHSA aims to create an integrated database of legacy and new information for the pre-Columbian archaeological heritage of South America, providing an open access digital resource based on the FAIR data standard sharing philosophy. This groundbreaking initiative offers the first-ever mapping resource and data repository for Brazil and Colombia. The database benefits researchers in the surveyed countries, and also the international archaeological community and national heritage agencies, contributing to heritage protection and cross-border studies.

MAPHSA's multidisciplinary approach combines archival work, remote sensing, machine learning, and on-ground surveys to identify and assess the state of preservation of archaeological sites. The workflow integrates different datasets into a Postgres/PostGIS geodatabase, and the Arches information management platform that features a CIDOC CRM ontology data model, drawing on the experience of similar-scale Arcadia-funded projects.

This initiative raises international awareness among academics, heritage professionals, and general citizenry of the vulnerability of South American heritage to threats such as deforestation, agricultural expansion, and fire. By highlighting the heritage value of areas like the Amazon and the (often overlooked) Cerrado savanna, MAPHSA fosters a change in public perception and promotes further actions for their protection.

Through its integrated database and multidisciplinary approach, MAPHSA serves as a powerful model for transregional data integration and the enhancement of heritage data accessibility. This presentation contributes to the broader discourse on heritage preservation, cross-cultural comparisons, and the importance of interdisciplinary collaboration in safeguarding South America's rich cultural heritage.

S3-6 Modelling heritage information in two diverse regions: reflections on the integration of heterogeneous big data from the MAHSA and MAEASaM projects

Mapping Archaeological Heritage in South Asia (MAHSA) Project[1] and Mapping Africa's Endangered Archaeological Sites and Monuments (MAEASaM) Project[1]

[1]McDonald Institute for Archaeological Research, University of Cambridge, UK

Contact: mahsaadmin@arch.cam.ac.uk

The MAHSA and MAEASaM projects work within different geographic regions of South Asia and Africa, dealing with large numbers of published (and unpublished) archaeological sites and monuments records.

Much of the available documentation is not available in a digital format, with paper publication being the default. Information about archaeological heritage varies in both the level of detail and the way in which sites are documented. This paper explores the work and challenges of the two projects in digitising, modelling, re-interpreting, linking, and re-using non-digital archaeological data into a structured and standardised digital format using both common and unique controlled vocabularies, so that it may become findable, accessible, interoperable and reusable (FAIR).

Both projects, working with different partners, are compiling existing data through legacy records (archival records, published documents, reports, and surveys) as well as newly created data in the form of field documentation, remote sensing surveys, historical map analysis and automated site detection methods. All these data follow the same or similar data standards and recording methodologies, and will be published in publicly accessible databases using the Arches platform.

Retaining context while delivering standardised usable data is a key challenge to this type of data modelling, and the meanings of the original source and context are inevitably subject to re-interpretation during this process. The projects are working collaboratively with heritage stakeholders, each bringing their own expectations of data use, management, and accessibility, in addition to navigating the post-colonial spaces in which they operate. To ensure usability, different use case scenarios have been identified and are being developed with collaborators and database end-users. Drawing on examples from each project, we explore a number of interventions used to grapple with some of the challenges in dealing with heterogeneous archaeological big data.

S3-7 Hordes of Data: Multilayered documentation and collection of archaeological heritage and environmental data across the Mongolian landscape

Dovydas Jurkenas[1] and Mike Fisher[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: jurkenas@gea.mpg.de

Mongolia is home to one of history's most extensive empires, as well as some of the earliest-known nomadic states, and immovable archaeological heritage covers much of the country's landscape. While Mongolia is the world's least-densely populated nation, rapid industrialisation and urbanisation threaten to destroy this heritage before it is thoroughly identified and documented. Anthropogenic climate change further threatens these historical and prehistoric remains through melting permafrost and increasingly erratic aeolian and fluvial activity.

The Mongolian Archaeology Project: Surveying the Steppes (MAPSS) deploys an integrated digital methodology to document these archaeological and ethnographic landscapes, producing the first open-access, semantically interoperable, reproducible, and comprehensively populated geospatial database of Mongolian cultural heritage resources. MAPSS combines archival field data with the results of manual remote sensing, Machine Learning detection, ground truthing survey, and digital imaging to record tens (and potentially hundreds) of thousands of sites and their environments and conditions. Our research uses these data to explore topics such as palaeohydrology, human-environment interaction, diachronic patterns of nomadic mobility, and heritage resource vulnerability.

In this presentation, we describe the modalities of MAPSS palaeodata creation, recording, and storage, with particular emphasis on collecting and integrating environmental data. We also explore the means by which MAPSS prepares its large cultural heritage datasets for broader integration with other Big Data sources.

These include SPARQL integration of domain-standard vocabularies such as the Getty AAT (Art & Architecture Thesaurus) and semantic interoperability through RDF (Resource Description Framework) expression of domain-standard ontologies such as CIDOC-CRM (Comité Internacional de Documentación-Conceptual Reference Model). Our discussion revolves around the relationship between "big" cultural heritage data and its twin environments–one from which they come and one in which they are generated.

S3-8 Creating the Mongolia cultural heritage database (Immovable Historical and Cultural Heritage)

Ankhsanaa Ganbold[1] and Bayarsaikhan Jamsranjav[2,3]

[1]Preventive Conservation and Risk Studies, Mongolian National Center for Cultural Heritage, Mongolia

[2]National Museum of Mongolia, Mongolia

[3]Max Planck Institute of Geoanthropology, Germany

Contact: ganhaa@gmail.com

In Mongolia, the beginning of protection and registration of cultural heritage on the basis of modern law was established in 1921. At first, it was carried out by research institutions. The first Law on the Protection of Cultural Heritage was adopted in 1970, and in 1990 it joined the UNESCO Convention on the Protection of Natural and Cultural Heritage.

Currently, the protection of cultural heritage is governed by a law approved in 2015, which is in interrelated with about 50 laws in force in Mongolia. According to the law, the cultural heritage database is organized by administrative units, and the national cultural heritage register and database is centralized in the National Center of Cultural Heritage.

As of 2023, about 80,000 historical and cultural immovable monuments of 10,000 historical and cultural sites have been compiled in 8 types in the State Unified Register and Database of Cultural Heritage on paper and electronic basis.

Here we would like to introduce our experiences and challenges for collecting, structuring and managing the cultural heritage data in Mongolia.

Session 4: Archaeological time-series: the quest for robust ^{14}C-dated proxies for the intensity of prehistoric activity

S4-1 Large radiocarbon datasets in relation to archaeological context and metadata

Jan Kolář[1], Adrian Timpson[1], Robert Staniuk[1], Jelena Bulatović[2], Claudia Speciale[2,3], Natalia Ryabogina[2], Samantha S. Reiter[4], Helene Agerskov-Rose[2], and Stephen Shennan[1]

[1]University College London, Institute of Archaeology, UK

[2]Gothenburg University, Sweden

[3]IPHES Catalan Institute for Human Palaeoecology and Social Evolution, Spain

[4]National Museum of Denmark, Denmark

Contact: j.kolar@ucl.ac.uk

Archaeology is witnessing a boom in the use of large scientific datasets via interdisciplinary approaches. In fact, novel narratives of the human past at diverse spatial and temporal scales cannot be achieved without such datasets and analytical tools rooted outside the borders of a single discipline. Although radiocarbon dates create one of the most important categories of archaeological data (crucial for dating human activities and building palaeodemographic models), archaeologists often insert them into extant databases without contextual information and references, thereby disassociating them from other archaeological information. This effectively prevents the use of these dates together with other data types (e.g. archaeobotanical, archaeozoological or isotopic) which are usually collated separately in specific (subdisciplinary) databases at different spatio-temporal scales.

In this paper, we present the interdisciplinary approach of the "Big Interdisciplinary Archaeological Database" (BIAD) to large-scale data collection, one originating from the combined efforts of several ERC-funded projects with a common focus on European prehistory 6000-500 BCE. The existence of common aims across the involved subdisciplines enables us to demonstrate that concurrent data collection with harmonised quality standards is a crucial prerequisite of all interdisciplinary research. As a result, we can directly relate tens of thousands ^{14}C dates and the demographic processes inferred from them with scientific data on many socio-economic aspects of prehistoric European societies. Our approach is illustrated via two case studies where we relate demographic changes to shifts in subsistence strategies and cultural practices. We also show how the archaeological data from regions with diverse interdisciplinary data coverage can be used for the creation of large-scale and long-term perspectives on the human past.

S4-2 Building a database of dated Neanderthal fossil remains to estimate the spatiotemporal pattern of Neanderthal extirpation

Fiona Laviano[1] and Frédérik Saltré[1]

[1]Global Ecology Laboratory, Flinders University, Adelaide, SA, Australia

Contact: fiona.laviano@flinders.edu.au

Changes in ecosystem resources distribution driven by major and local climate shifts have influenced the mobility and dispersal of Neanderthal and Homo sapiens hunter-gatherers. The Neanderthals disappeared and the early *H. sapiens* migrated through Eurasia during the Marine Isotope Stage 3 interglacial (57 to 29,000 years BP). Recent attempts to estimate timing and causes of Neanderthal extinction suggest an extinction around 40,000 years BP with regional variability and claim for a multifactorial process with probable local combination, suspecting variability in food resource acquisition due to climate turnover and human competition according to region. However, chronological inconsistencies are at the root of most of the debates about the last Neanderthal occurrences and the absence of a generalized, accurate and precise spatiotemporal pattern of Neanderthal extirpation (i.e., local extinction) prevents inter-regional comparison with plausible external ecological factors. Here we build a new database accounting for reliably dated Neanderthal fossils, also including dated Mousterian industries from strata from the same sites, to enable comparing regional chronologies in Western Europe. We compile the sites accounting for Neanderthal fossils across Eurasia and radiometric age estimates from Western European sites, along with reliability criteria focusing on dating methods. We then applied a spatiotemporal statistical method to these Neanderthal fossil records (i.e., suitable dates on securely identified fossils) showing that Neanderthal extinction was a long-term process in Europe. We map a spatial and taphonomic bias-corrected extinction chronology of the Neanderthals aiming to facilitate interregional comparison at continental scale with ecological factors that could have been involved in the extirpation.

S4-3 Investigating long-term population dynamics in the Han River Basin using radiocarbon dating

Jiyoung Park[1] and Sejin Kim[1]

[1]Seoul National University, Department of Archaeology and Art History, South Korea

Contact: geiyoung@snu.ac.kr

The Han River basin, situated in the central part of the Korean Peninsula, played a central role in the pre, proto and historic periods of Korea, maintaining high population densities since early periods. Despite a long history of research showing intensive settlement and farming from the Bronze Age to the Baekje Kingdom (15th century BC to 7th century AD), which imply dynamic changes in population size characterized by patterns of aggregation and dispersion, there has been little focused research into the long-term population change patterns in the region due to the disconnect between prehistoric and historical studies.

In order to redress this problem, this study focuses on examining the spatio-temporal changes in population using radiocarbon data. First, population growth patterns will be analyzed using the Summed Probability Distribution (SPD) method, while addressing aforementioned problems through a data cleansing process. Raw counts of sites and houses, along with simulations, will assess errors caused by sampling bias in carbon dating and their potential impact on population dynamics analysis. Second, alternative methods such as the aoristic weight will be used to examine changes in settlement sizes, duration, and spatial patterns.

The results reveal an overall population increase from the Bronze Age to the Three Kingdoms period, with distinct demographic processes become apparent in each region. This highlights the complexity of demographic changes when considering both temporal and spatial patterns of human populations. By shedding light on the long-term population dynamics in the Han River basin, this study contributes to a better understanding of the region's socio-cultural transformations and population movements.

S4-4 NeoNet, an app for radiocarbon modelling of the Late Mesolithic / Early Neolithic transition in South-Central and South-Western Europe

Thomas Huet[1], Niccolò Mazzucco[2], Miriam Cubas Morera[3], Juan F. Gibaja[4], F. Xavier Oms[5], António Faustino Carvalho[6], Ana Catarina Basílio[7], and lías López-Romero[8]

[1]University of Oxford, School of Archaeology, UK

[2]University of Pisa, Dipartimento di Civiltà e Forme del Sapere, Italy

[3]Universidad de Alcalá, Spain

[4]Institució Milà i Fontanals (IMF-CSIC), Spain

[5]University of Barcelona, Spain

[6]CEAACP, University of Algarve, Portugal

[7]FCT, ICArEHB, University of Algarve, Portugal

[8]Instituto de Arqueologia de Merida (IAM), CSIC-Junta de Extremadura, Spain

Contact: thomas.huet@arch.ox.ac.uk

The NeoNet app is an online open source interactive and user-friendly R Shiny application for mapping radiocarbon dates from the Late Mesolithic / Early Neolithic transition in the European river basins of the north-central Mediterranean (https://tinyurl.com/neonet-app). The application offers a mobile, multi-scalar geographic window for selecting dates using various filters in addition to location, such as chronology, cultural periods, and date quality. A calibration window allows aggregation (SPD) of selected dates by archaeological layers, sites and periods. The dataset – with over 2,500 curated radiocarbon dates and their archaeological contexts – is published as open data, will be extended to the Atlantic river basins of south-western Europe (more than 1,000 new dates), and will be made harvestable by radiocarbon gazetteers (https://github.com/ropensci/c14bazAAR). The app is developed and open to contributions on GitHub (https://github.com/zoometh/neonet), and will allow on-the-fly Bayesian and Harris matrix modelling. The transition from hunter-fisher-gatherers groups to early farmers communities has been one of the crucial research questions in Late Prehistory. In this sense, the NeoNet application, by offering an easy and almost reproducible way of working with radiocarbon dates, makes it possible to model, test and mitigate different research hypotheses.

Session 5: Data and people in interaction: Network analysis for everyone

S5-1 Hidden information or misuse? Data and network analysis on burial customs (6-9 c. AD)

Tamás Czuppon[1] and Pázmány Péter[2]
[1]Hansagi Museum, Hungary
[2]Catholic University, Budapest, Hungary
Contact: czuppon.tomi@gmail.com

In the last decades use of statistical analysis and networking methods in archaeological research became very common. Numerous theories can help to gain more and more information about finds but several misuses have also appeared. On this way researches on Avar Age (6-9 century AD) finds are not an exemption as well. This presentation is about how and what can we gain with the use of networking methods about the change of burial customs in Avar Age cemeteries on the southern part of Transdanubia in Hungary. With the research on 6000 graves in this area these methods show us new, formerly hidden information and connections but cannot give answers about all of our archaeological questions. How can we use these methods? What are the limits? How big database we need? How much of plus information can we get through networks from the graves?

S5-2 Mapping networks of capital and resistance in Tropical South America

Freg J. Stokes[1]
[1]Max Planck Institute of Geoanthropology, Germany
Contact: stokes@gea.mpg.de

This talk addresses the challenges and opportunities that come with cataloguing and visualising capital flows and political conflicts in the tropical rainforests of South America. I am currently working on a QGIS database documenting the five-hundred-year history of Indigenous resistance to colonisation in the Amazon and neighbouring forests, drawing on archival and contemporary Indigenous accounts. In order to analyse how Indigenous resistance has affected deforestation rates, this database will also incorporate deforestation statistics from across these biomes during the same time period, predominantly utilising archival sources combined with recent multidisciplinary studies. Simultaneously, I will also be mapping out the flows of capital between South America and the Global North that have driven this deforestation wave. Such a project faces numerous challenges. How can this research catalogue and visualise historical events in areas outside of strong state control, where uncertainty is the norm? How can fluid processes such as the movement of capital and the resistance of mobile Indigenous groups be depicted? Considering the history of colonisation being documented, what ethical and logistical guidelines need to be in place to ensure that Indigenous groups can both access the information in the database and maintain ownership of the histories that they share? Tune in to hear some possible answers, and plenty more questions.

S5-3 Neutron activation analysis and networks: Possibilities and constraints

Britt Davis[1] and Matthew A. Peeples[1]

[1]Arizona State University, AZ, USA

Contact: jbdavi18@asu.edu

Archaeologists have successfully used site-based material attributes, such as ceramic proportions, to examine interaction networks in prehistory. Fewer studies have explicitly used chemical sourcing data for this purpose. In many parts of the world where specific artifact production locations are unknown, chemical sourcing can only provide generalized areas of production based on theoretical assumptions, such as the criterion of abundance. This research uses both existing and simulated Neutron Activation Analysis (NAA) data obtained from ancient Maya pottery from the Belize River Valley region in western Belize to examine how the scale and scope of these data may affect our ability to use NAA data for regional network analysis. The analysis is aimed at understanding the degree to which centrality in a NAA-based network where there are overlapping areas of geological similarity is influenced by this geological context. Furthermore, we examine to what extent the directionality of pottery can be observed given chemical sourcing constraints. Broadly this research examines how geological similarity affects our ability to find connections in the past using chemical sourcing data.

S5-4 'Symbolic scaling' in ancient and contemporary cities

Chris Carleton[1] and Patrick Roberts[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: carleton@gea.mpg.de

Quantitative models predicting scaling relationships between population size, infrastructure, and economic output have been applied to present, and more recently, past urban contexts. However, they have yet to be deployed to explore the relationship between urban growth and ideological expression. Here, we apply Settlement Scaling Theory to a sample of ancient and modern cities to explore whether symbolic infrastructure follows the same scaling patterns identified for other urban attributes and whether those patterns were the same for the ancient and modern cases. Our analysis reveals differences in 'symbolic scaling' between urban areas and different types of monuments that we argue can be used to explore the relationship between society, settlement and ideology. We highlight the contemporary relevance of archaeology and history in providing unique, long-term observations of urban-socioeconomic-ideological interactions.

Session 6: Leveraging Big Data, GIS, and machine learning in remote sensing

S6-1 An experiment with automated detection of diverse archaeological features in an African context using deep-learning and free satellite multispectral imagery

Stefania Merlo[1] and MAEASaM team[1]

[1]McDonald Institute for Archaeological Research, University of Cambridge, UK

Contact: sm399@cam.ac.uk

Remote sensing data has demonstrated to be of invaluable use for both the detection and the monitoring of archaeological sites globally. New developments in artificial intelligence (AI) technology, in particular deep learning computer vision techniques, present opportunities for the automated detection of potential archaeological features in less explored contexts, for example in Africa, where the high diversity of landscapes and sites and the relative paucity of ground data are a challenge when using visual inspection and traditional machine learning techniques.

This paper presents an innovative computer vision deep-learning approach that uses remote sensing data and derivatives for the identification of a diverse range of archaeological sites in a selection of countries in Africa. Different Convolutional Neural network (CNN) algorithms are compared for performance and accessibility in the detection and segmentation of features using freely available remote sensing imagery (ie Sentinel), starting from the knowledge of a wide variety of sites derived from the database of the Mapping Africa's Endangered Archaeological Sites and Monuments (MAEASaM) project. This consists of legacy and remote sensing based records of thousands of sites of different morphology and size, across a variety of geographical settings.

S6-2 Automated change detection monitoring of archaeological sites using Machine Learning and Google Earth Engine

Ahmed Mahmoud[1] and Nichole Sheldrick[1]

[1]University of Leicester, UK

Contact: amam9@leicester.ac.uk

This paper will present recent developments in a remote sensing workflow developed by the Endangered Archaeology in the Middle East and North Africa (EAMENA) project using machine learning algorithms for automating the detection of threats to archaeological sites. The automated change detection (ACD) system uses free satellite imagery and high-performance computing power available via Google Earth Engine to compare a series of Sentinel-2 images to highlight areas of change and identify where, when, and what types of change have occurred within the proximity of known archaeological sites. Previous versions of EAMENA's ACD methodology produce a binary classification of change/no change (Rayne, et al. 2020). The updated workflow uses machine learning classifiers (Random Forest) to carry out supervised classification of land cover in the study area and analyse time series of satellite images to classify changes in land use and categorise the type of change which has occurred. By analysing sequential classified images, we can identify

where archaeological sites are within proximity to significant land cover changes and at risk of being affected by these changes. This workflow can improve efficiency in monitoring archaeological sites by rapidly identifying which sites are most at risk and helping heritage professionals to target resources where they are needed most urgently. In addition, we developed a user interface for the EAMENA ACD which allows users; experts or with no-prior knowledge in the field of remote sensing and GIS to use the system and easily adapt it for archaeological monitoring. Case studies undertaken in North Africa will be presented which demonstrate the workflow and outputs, as well as the advantages and challenges that this methodology presents. References Rayne, L., et al. 2020. "Detecting change at archaeological sites in North Africa using open-source satellite imagery." Remote Sensing 12(22):3694. https://doi.org/10.3390/rs12223694

S6-3 Multi-scale automated feature detection in the Argentinean Andes: A comparative analysis of satellite and UAV

Manuel J.H. Peters[1], Amina Jambajantsan[1], and Veronica Zuccarelli[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: manuel.peters@gea.mpg.de

In the rugged terrain of the Argentinean Andes, the Cusi Cusi micro-region presents a unique challenge for archaeological research. Located within the Puna of Jujuy, at an elevation ranging from approximately 3800 to 4200 meters above sea level, this area is marked by plains interrupted by sharp elevation changes. Long-term human occupation traces include archaeological, historical, and modern features, such as habitational structures, burial mounds, agricultural terraces, retaining walls, perimeter walls, and cattle enclosures.

In the context of the 2023 provincial constitution reforms, the dynamics concerning underground resource ownership and indigenous rights are increasingly complex. The Puna of Jujuy is situated in the wider region within Argentina, Bolivia, and Chile that holds 60% of the lithium in the world, known as the "Lithium Triangle". New extractive politics are resulting in mounting pressures on indigenous landownership and environmental stability. This context underscores the increasing importance of documenting the region, focusing not only on its archaeological heritage but also on the contemporary circumstances of indigenous communities.

This research integrates satellite and UAV imagery with machine learning for automated feature detection. Specifically, we undertake a comparative analysis of satellite data and high-resolution UAV-based photogrammetry to assess their effectiveness in detecting and analysing various features. The validation process included on-site ground-truthing visits, providing insights into the heritage and challenges of the Cusi Cusi micro-region.

S6-4 Positive false positives: Searching for annular landforms in Northern Europe

David Stott[1], Peter Jensen Maring[1], and Søren Munch Kristiansen[2]

[1]Moesgaard Museum, Denmark

[2]Department of Geoscience, Aarhus University

Contact: ds@moesgaardmuseum.dk

This presentation explores the results from the Ring Project, which used convolutional neural networks (CNNs) to search for ramparted ring-shaped landforms in national lidar datasets comprising 9 countries and 5 German Bundesland. While the primary focus of the project was on the detection of the broadly contemporaneous early medieval Scandinavian, Slavic and Frisian ring fortresses, we also presumed that this would produce a wide variety of false positives relevant for researchers in other disciplines, as one person's false positive is another's object of interest. These include fluvio and peri glacial landforms such as rimmed kettleholes and pingo ruins, and meteor impact craters.

This was performed by preprocessing the lidar derived DTMs as multi-band residual relief images and using these to perform training and detection by integrating the PyTorch implementation of Mask R-CNN with the OSGeo software stack. QGIS was used for data visualization and evaluation. This was compared to our earlier approach using morphological image feature detection using Hough transforms and template matching.

In this presentation we'll outline our data preparation, augmentation, training and classification workflows. We'll then discuss the challenges inherent in detecting these features over such a wide area, including those presented by sparse training sets, heterogeneous lithologies and landscape histories, and feature classifications overlooked during training the CNN. We'll also explore the main difficulty of evaluating our results: How can we be sure our results are novel given the disparate levels of cultural and geological heritage records available to us in the different administrative areas we explored?

Session 7: AI for historical and archaeological Big Data analysis

S7-1 Enhancing the Seshat Databank: A computationally readable mapping of historical facts to scholarly references

Majid Benam[1], Enrico Cioni[2], Maria del Rio-Chanona[1], Pieter Francois[3], Jakob Hauser[1], Daniel Hoyer[4], Dániel Kondor[1], Hannah Lawrence[3], Peter Turchin[1,3,5], Federica Villa[3], and Harvey Whitehouse[3]

[1]Complexity Science Hub, Austria

[2]Seshat: Global History Databank, Evolution Institute, FL, USA

[3]Centre for the Study of Social Cohesion, University of Oxford, UK

[4]Evolution Institute, FL, USA

[5]Department of Ecology and Evolutionary Biology, University of Connecticut, CT, USA

Contact: hauser@csh.ac.at

Understanding the dynamics of socio-cultural evolution on a global scale necessitates data sources that encompass well-structured, verifiable, historical facts, over long-time scales. The Seshat: Global History Databank answers this call, covering over 1500 different variables and 160,000 respective values, that describe the political and cultural fabric of 600 polities (political systems governing an area at a time) from the Neolithic to the Industrial Revolution. Seshat's strength lies in its data collection process, which involves the manual collection of information, followed by a review from domain-level experts.

However, no such data source currently provides a mapping between the historical facts it provides and their sources that would allow further high-throughput analysis on the sources themselves. Addressing this limitation opens significant research opportunities at the juncture of history and natural language processing.

In this study, we extend the Seshat Databank to include mappings from data points to their corresponding scientific sources in a way that allows automated text-based analysis methods to be applied to this relationship.

A single data point in this new dataset comprises a variable, a corresponding value, the related polity, its period of existence, and the scientific reference that substantiates the data point. Additionally, we provide expert commentary for both the value and the reference document, and, when possible, we utilize information within Seshat to narrow the mapping down to document sections.

This dataset's key strength lies in its capacity to provide structured access to historical facts buttressed by evidence.

It will be a valuable resource for natural language processing research, such as the development of specialized language models and the creation of recommender systems for historical literature. Thus, this work can bring significant value to the wider scientific community beyond historians and computer scientists.

S7-2 Prompt me a dataset: Large-scale ML dataset creation in the humanities

Hassan El-Hajj[1,2]
[1]Max Planck Institute for the History of Science, Germany
[2]Berlin Institute for the Foundation of Learning and Data (BIFOLD), Germany
Contact: hhajj@mpiwg-berlin.mpg.de

The intensive digitization efforts of the last couple of decades have led to the creation of big-humanities-data, which in turned opened the door toward corpus level analyses and hypotheses formulation within the humanities. This development, coupled with the advent of Machine Learning (ML), has also led to an increased demand for not only digitized sources, but annotated digitized sources that can be used to train ML models for down-stream research tasks (e.g., style analysis, image similarity, image interpretation). It is at the annotation level that many of the digital humanities projects aiming to use ML within their pipeline get bogged down, as this process is both time and resource intensive.

In this paper, we present and evaluate a dataset creation pipeline that aims to alleviate this problem and lower the threshold to ML adoption within digital humanities research. The proposed pipeline for historical documents dataset creation is based on multi-modal models (e.g., CLIP, GLIP, GroundingDINO), where researchers could simply "prompt" the model in natural language for either image classes, but more importantly, image regions and classes. Such zero-shot inference pipelines can be used (and optimized) to iteratively extract semantically relevant data from digital collections (e.g., scans) in order to quickly create question and hypothesis relevant datasets with minimal resources. We believe that it is through such an iterative process that we can transition from big-humanities-data to big-annotated-humanities-data.

S7-3 Discovering and analyzing patterns in large historical image databases

Mathieu Aubry[1]
[1]École des Ponts ParisTech, France
Contact: mathieu.aubry.2006@polytechnique.org

In this presentation, I will discuss how a diversity of historical document analysis problems can be approached with a common methodology, which is at the core of the ERC DISCOVER project on 'Discovering and Analyzing Visual Structures'. I will first introduce the problem and present the key elements of the proposed approaches, based either on finding links between documents [1] or on learning to decompose them in a structured way [2]. I will then give an overview of problems we are currently tackling with these approaches, including paleographical analysis [3], copied details identification in artworks [1], historical scientific illustration circulation [4], astronomical diagram analysis and book ornaments recognition. Finally, I will discuss how interfaces for enabling historians to compute and navigate these results on their own data and to provide feedback are essential, and their current state of development.

[2] Learning Co-segmentation by Segment Swapping for Retrieval and Discovery X. Shen, A. Efros, A. Joulin, M. Aubry ArXiv 2021

[2] Unsupervised Layered Image Decomposition into Object Prototypes T. Monnier, E. Vincent, J. Ponce, M. Aubry ICCV 2021

[3] The Learnable Typewriter A Generative Approach to Text Line Analysis Y. Siglidis, N. Gonthier, J. Gaubil, T. Monnier, M. Aubry ArXiv 2023

[4] Image Collation: Matching illustrations in manuscripts R. Kaoua, X. Shen, A. Durr, S. Lazaris, D. Picard, M. Aubry ICDAR 2021

S7-4 Leveraging satellite imagery and deep learning in Mongolian archaeology

Amina Jambajantsan[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: jambajantsan@gea.mpg.de

Mongolia has witnessed the rise and fall of numerous empires and has a culturally rich heritage. However, the vast expanse of the country presents challenges in discovering, documenting, and supervising archaeological sites for the limited number of archaeologists available. The MAPSS project, funded by Arcadia, is dedicated to constructing an extensive archaeological database, encompassing established sites from various Mongolian databases and newly uncovered ones.

Machine learning introduces a new era in archaeology, offering a transformative approach to address the constraints associated with traditional methods. By performing Convolutional Neural Networks (CNNs) on high-resolution optical satellite imagery, using Python and Jupyter notebooks, further integrated with ArcGIS Pro, we discern potential sites and features.

To enhance the precision of documenting specific sites, we employ drone flights over the selected areas, capturing highly detailed Digital Elevation Models (DEMs) and orthomosaic maps.

Our discoveries not only contribute to archaeological research but also emphasize the wider significance of data-driven exploration for preserving and comprehending the cultural heritage of varied regions.

S7-5 Deep Learning and Bayesian analysis of historical time series data

Jianyin Roachell[1] and Ricardo Fernandes[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: roachell@gea.mpg.de

Historical and archaeological research frequently employs time series data, a critical tool in complex system studies. Time series analysis is instrumental in assessing the impact of interventions and events on system dynamics. The derived inferences can evaluate human agency's influence on historical trajectories, enhance socio-economic system design, and pinpoint warning signs for potential tipping points. Our presentation will delve into two analytical methods for historical time series data: Bayesian structural time-series and deep learning algorithms.

The Bayesian structural time-series offers a counterfactual approach to assess the impact of interventions, such as policy decisions, and events like extreme disasters (e.g., natural disasters, pandemics). On the other hand, deep learning algorithms are utilized to detect tipping points, identify bifurcation types, and predict the system status post-event.

We have developed innovative software that seamlessly integrates these modelling approaches with the Pandora data platform. This integration facilitates easy access to these tools, along with data and model instances stored at Pandora. Our presentation will demonstrate the analysis of historical time series data using these novel methods.

S7-6 Applying Bayesian machine learning to causality modelling in historical social-ecological systems: case studies from early modern Greece and Poland

Adam Izdebski[1], Ricardo Fernandes[1], Georgios Liakopoulos[1], Piotr Guzowski[2], Alessia Masi[3], Cristiano Vignola[3] and Lucrezia Masci[3]

[1]Max Planck Institute of Geoanthropology, Germany

[2]University of Białystok, Poland

[3]La Sapienza University, Italy

Contact: izdebski@gea.mpg.de

With the advent of machine learning and big data, new opportunities are emerging for the quest for history's Holy Grail: elucidating causality, and, ultimately, developing a predictive science of the past. In my talk, I will present a new algorithm developed in MPI Geoanthropology, Bayesian Model Selection under Constraints (BMSC), designed to help in exploring and evaluating past causal relationships based on historical, archaeological and palaeoenvironmental data. I will discuss two case studies, both focused on understanding the interplay of environmental (physical, climatic, epidemic, etc.), economic and cultural factors as drivers of human niche construction strategies in Greece (Peloponnese and Macedonia) and Poland (Pomerania).

S7-7 Trajectories of change: Using text embeddings to track knowledge evolution

Raphael Schlattmann[1], Malte Vogl[1], Bernardo S. Buarque[1], Aleksandra Kaye[1], and Jochen Büttner[1]

[1]Max Planck Institute for the History of Science, Germany

Contact: rschlattmann@mpiwg-berlin.mpg.de

The presented research explores the evolution of knowledge systems through a framework of so-called socio-epistemic networks (SEN). This framework comprises three interconnected layers: social, semiotic (or material), and semantic, proposing a comprehensive and multilayered approach to understanding structural developments of knowledge. For recent periods in the history of science, the primary input for these layers usually consists of (semi-)unstructured texts such as scientific publications, but also letter data, administrative records, and texts of other provenance. To examine their diachronic development and specifically analyse (language-based) change on the semantic layer, we are exploring local vs global variations of document embedding "densities", indicating the concentration of vectors in different regions of embedding space. In the context of individual documents, this enables us to track how each document's neighbourhood changes over time and how the concentration of surrounding documents (similar semantic content) accelerates, stays consistent, or declines. This is done for all documents of the respective corpus, facilitating a global comparison between individual documents and/or regions over time. By doing this, we can follow trajectories of individual documents or groups of documents focusing on either content (topic combinations) or metadata (documents of certain author(s), institutions, etc.). One application of this approach is to examine the relationship between the development of individual scholars' written works and the

development of the disciplinary fields in which they work. As a case study, we explore the connection between the works of physicist Hans-Jürgen Treder – a prominent GDR physicist - and the field of General Relativity and Gravitation (GRG) research. GRG as a field gained traction only after World War II, following a stagnation period in general relativity research, coinciding with Treder's most productive years. We examine this development from 1945-2000, compare it with other individuals from the same period, and discuss limitations, problems, and potential other applications.

Session 8: Insights from reusing large prehistoric and interdisciplinary databases

S8-1 Columns and rows: publishing, preparing, harmonizing and preserving protocols, guidelines and resources from the Big Interdisciplinary Archaeological Database

Samantha S. Reiter[1], Robert Staniuk[2], Jan Kolář[2], Adrian Timpson[3], Jelena Bulatović[4], Natalia Ryabogina[4], Claudia Speciale[4,5], and Emily J. Kate[6]

[1]National Museum of Denmark, Denmark

[2]University College London, UK

[3]University College London, Institute of Archaeology, London, UK

[4]University of Gothenburg, Sweden

[5]IPHES Catalan Institute for Human Palaeoecology and Social Evolution, Spain

[6]University of Vienna, Austria

Contact: samantha.scott.reiter@natmus.dk

Archaeology is increasingly reliant on large datasets, including information from a range of disciplines across large geographic and temporal ranges. Several consortia have proposed data management solutions (i.a. ADS, ARIADNEplus, DANS-EASY, IsoArcH, and tDAR). Although excellent guidelines exist for the planning, production and maintenance of large datasets in relation to FAIR (Findable, Accessible, Interoperable and Reusable) and CARE (Collective benefit, Authority to Control, Responsibility and Ethics) principles, archaeological research lacks guidelines for publishing data which facilitate the interoperability of results between different topical, regional or temporal specialisms for transdisciplinary research. This particular lacuna is the subject of this paper.

BIAD (Big Interdisciplinary Archaeological Database) records a variety of bioarchaeological (botanical, faunal, human) and cultural data from archaeological contexts. In pursuit of that goal, BIAD has created strategies and standards to unify different scales and types of data within a relational database, and which applies to databases across the board. The lessons and techniques learned by data inputters are presented here and include: (1) concrete publication recommendations for "Big Data" formats, (2) merged taxa for archaeozoological and botanical data and reassessed datasets for correlation with occupation phases, (3) the advantages of the metadata assembled for BIADwiki (national resources and equivalency tables, e.g. for various skeletal and dental recording systems) and (4) new systems for integrating age/sex categories from different anthropological recording standards.

S8-2 Shortcomings related to archaeological bibliographic databases used for macro-scale modeling: the SIGMENT database for estimating Middle and Recent Gravettian lithic technical tradition distributions in Western Europe

Anaïs Vignoles[1], Arnaud Caillo[2], Laurent Klaric[1], William Banks[2]

[1]UMR 8068 TEMPS, Université Paris X-Nanterre, France

[2]UMR 5199 PACEA, Université de Bordeaux, France

Contact: anais.l.vignoles@gmail.com

Modern technical improvements in computational approaches and the growing availability of large archaeological and environmental datasets has allowed the development of macro-scale (regional, continental, global) studies of human-environment relationships during prehistory. These studies often rely on modeling procedures that aim to quantify the potential roles that environmental factors played in influencing macro-scale spatial and temporal patterns of cultural diversity or population dispersal. Like all modeling procedures, though, the quality of input data is crucial to ensure the pertinence of model outputs and their interpretation. It is thus paramount to critically assess chronological, cultural and/or biological information for the employed occurrence datasets. It is usually impossible to personally conduct assessments at such scales, and such work typically relies on reviews of published literature, which has inherent biases.

Here, we present the « Sites du Gravettien Moyen et récENT » (SIGMENT) database, which aims to reliably estimate the geographic distribution of Middle and Recent Gravettian lithic technical traditions in France and neighboring regions. We compiled contextual and cultural data from 177 published references on 247 collections from 165 sites in France, Spain, Italy and Germany. We then established a qualitative grid to classify the presence reliability of archaeological traditions at each site. We found that many of these occurrences are unreliable and tagged them as not suitable for integration into larger scale models focused on the geographic distribution of archaeological traditions. Building on the historiographic and archaeological study of several assemblages from southwestern France, we discuss the factors that may have influenced the traditions' over- or under-estimation from a strict bibliographic approach, including varying analytical approaches, excavation and conservation methodologies. and inconsistencies in how assemblages are described in the published literature.

S8-3 Beyond labels. Exploring patterns of formal description of lithic variability in Late Pleistocene technologies

Giulia Marciani[1,2], Eugenio Bortolini[1], Alice Leplongeon[3,4], Leonardo Vallini[5], Andrew W. Kandel[6], Elham Ghasidian[7], Saman Heydari-Guran[7], Adriana Moroni[3], Fabio Negrino[8], Luca Pagani[5,9], Julien Riel-Salvatore[10], Annamaria Ronchitelli[2], and Stefano Benazzi[1]

[1]University of Bologna, Dipartimento di Beni Culturali, Italy

[2]University of Siena, Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, U. R. Preistoria e Antropologia, Italy

[3]Department of Archaeology, KU Leuven, Geo-Institute, Belgium

[4]UMR 7194 HNHP, Muséum national d'Histoire naturelle – CNRS – UPVD, Institut de Paléontologie Humaine, France

[5]University of Padova, Department of Biology, Italy

[6]The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities at the University of Tübingen, Germany

[7]DiyarMehr Institute for Palaeolithic Research, Iran

[8]University of Genova, Dipartimento di Antichità, Filosofia, Storia, Italy

[9]Institute of Genomics, University of Tartu, Estonia

[10]Département d'anthropologie, Université de Montréal, Canada

Contact: giulia.marciani@unibo.it

Traditional taxonomies used for studying lithic assemblages are usually context-specific which creates limitations in disentangling broad-scale, long-term scenarios of technological change. To overcome these issues, this study introduces a formal descriptive approach focused on technological traits whose interpretation is consistent across different classifications. We apply it to published data from approximately 160 archaeological layers dating between 60-30 ka BP, specifically investigating the Middle to Upper Palaeolithic transition across Eurasia and comparing it with technological changes occurring during the same timeframe in Africa. This context is particularly challenging due to the mechanisms of biocultural interaction, population movement, and functional convergence.

By shifting the scale of analysis from the techno-complex or assemblage level to a more flexible attribute level, this study explores the relevance and associations of technological traits using various exploratory analyses and pattern recognition techniques. The analysis reveals changes in the frequency, diversity, and meaningful associations among technological traits and clusters across time and space. Moreover, the proposed approach allows for the quantification of similarities and differences between individual archaeological layers, considering technological and typological variables. In this framework, lithic assemblages can be directly compared, regardless of the original methodological and theoretical frameworks used in their initial description. Although it remains constrained by the extent to which attributes were consistently recorded in published reports, this approach helps enhance higher-level reproducibility by removing existing dependence on disparate formal taxonomic systems and increasing awareness of research biases. It can also be 'scaled up' by the virtually unlimited addition of assemblages. The formal descriptive approach presented in this study represents a significant advance in lithic analysis, enabling researchers to untangle complex scenarios of technological change and human dispersals, ultimately enriching our knowledge of prehistoric human societies.

S8-4 Using big data to gain new insights about large-scale developments of human behavioral evolution: ROAD and Middle Stone Age ochre use

Rimtautas Dapschauskas[1]

[1]Heidelberg Academy of Sciences and Humanities, The Role of Culture in Early Expansions of Humans (ROCEEH), University of Tübingen, Germany

Contact: rimdap@icloud.com

The palaeosciences contribute to our understanding about human behavioral evolution by providing empirical information on the timing, geographical origins and expansions, population dynamics with spatio-temporal fluctuations, and the material, ecological, geographical, and climatic factors involved. Today, this knowledge is not solely derived from the discovery of new archaeological sites or fossils. Long-term developments are increasingly explored through interdisciplinary research databases that can handle vast amounts of data beyond the scope of conventional methods. This approach unveils long-term patterns that would otherwise be challenging to discern. In this presentation, I will discuss the methods and outcomes of the most comprehensive meta-analysis to date on ochre use in the African Middle Stone Age, conducted using the ROCEEH Out-of-Africa Database (ROAD). To address data variability resulting from different dating methods, we applied statistical concepts such as time averaging and finite mixture distribution to the mined data.

In conjunction with a comprehensive theoretical framework, supported by empirically grounded bridging arguments that are, nevertheless, independent of the analyzed data (to mitigate the risk of circular reasoning), we have gained fresh insights into the long-term development of non-utilitarian ritual behavior during human evolution. Our findings hold intriguing implications for comprehending the cultural factors influencing the demographic expansion of *Homo sapiens* as an ""ultra-social species,"" as well as the concurrent acceleration of cultural evolution during the Middle Stone Age in Africa. Our study stands as an excellent example of how data mining in big georelational databases can unveil patterns of chronological development, cultural evolution and geographic expansion that would remain hidden using traditional methods.

S8-5 Mapping archaeological cultures and periods with network cartography

Christian Sommer[1], Andrew W. Kandel[1], and Volker Hochschild[1]

[1]The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities at the University of Tübingen, Germany

Contact: christian.sommer@uni-tuebingen.de

Network maps or flow maps are an established method in human geography to visualize and analyze movement or relationships between places. They show centers, peripheries and (invisible) borders of the mapped phenomena, with examples ranging from the service area of airlines to the spread of social media.

In an experimental study we applied this technique to draw an image of prehistoric culture as seen through the lens of archaeological science. Our dataset derived from the ROCEEH Out of Africa Database (ROAD) contains information from >1,300 sites with >15,000 assemblages. We constructed a network linking temporally overlapping assemblages and estimated their similarity using an index based on 32 artifact categories. As a result, we created a map of the likelihood of material cultures covering Africa and Eurasia

between 3,000,000 and 20,000 years ago. This network can be subdivided into eleven cultural periods and more than 100 cultures and technocomplexes.

Firstly, our results show that the distribution of archaeological sites is heavily uneven, with dense clusters of sites in some regions and an underrepresentation of samples in other regions. This is the results of several layers of bias that obscure our view on the original ranges of cultures: e.g. the effect of taphonomic processes in various environments, the intensity of scientific research and their representation in the literature. Secondly, we observed that spatial autocorrelation, i.e. the decay of cultural similarity with distance, could not be detected in our dataset, which indicates that our sample size or categorical granularity might still be under-sampled, despite using one of the largest databases available. Thirdly, our maps illustrate how the naming of archaeological cultures and periods is influenced by conventions – and subject to change of conventions.

S8-6 Modelling the adaptations of hominins to climate in Europe from MIS14 to MIS11

Jesús Rodríguez[1], Christian Willmes[2], Christian Sommer[3], and Ana Mateos[1]

[1]National Research Centre on Human Evolution (CENIEH), Burgos, Spain

[2]Institute of Geography, University of Cologne, Germany

[3]The Role of Culture in Early Expansions of Humans (ROCEEH), Heidelberg Academy of Sciences and Humanities at the University of Tübingen, Germany

Contact: jesus.rodriguez@cenieh.es

The time interval from 560 ka and 360 ka (MIS14 to MIS11) was critical for human evolution in Europe. In this period, humans spread throughout Western Europe during both cold and warm stages, exposing themselves to harsh environmental conditions which include very low ambient temperatures. A dataset of 68 archaeological assemblages and the Oscillayers Database of Palaeoclimate Maps were used to model the distribution of humans during the MIS14 to MIS11 interval and their adaptation to cold conditions. The areas of Western Europe with suitable environmental conditions to sustain human population were identified using a species distribution model. Eventually, the maximum sustainable human population in those areas was estimated based on the relationship between hunter-gatherer population density and net primary productivity. The estimated total sustainable population of Western Europe from MIS14 to MIS11 oscillated between 13,000 and 25,000 individuals and it was not relegated to the Mediterranean peninsulas during the cold stages, as usually assumed. Estimates of several temperature parameters were obtained for the archaeological localities and a simple heat-loss model was used to evaluate the efficacy of different physiological and behavioural adaptations to prevent hypothermia during the night. The results show that physiological and anatomical adaptations alone, were not enough to tolerate the low winter temperatures of this part of the Middle Pleistocene. Nevertheless, the use of a simple fur cover was an extremely effective response to cold, even without using fire.

Session 9: Palaeoclimatic and palaeoenvironmental databases: Exploring the dynamics of human-environmental systems

S9-1 Long-term arsenic exposure evidence in northern Chile revealed through analysis of compiled osteological and elemental data

Pedro Andrade[1], Bernardo Arriaza[2], Consuelo Huerta[2], Yamil Neira[3], and Lucas Romer[3]

[1]Department of History, Universidad de Concepcion, Chile

[2]Universidad de Tarapacá, Chile

[3]Universidad de Concepcion, Chile

Contact: pandrademd@udec.cl

Potable water is scarce in the coast of northern Chile. In addition, the periodic presence of earthquakes and ENSO events make this a harsh environment for humans. Yet, the region has been home to permanent human occupations for at least 12000 years.

For almost 15 years, scientific analyses have been carried out in the Arica region of northern Chile revealing that local populations were exposed to different metals, including arsenic. The compilation of results shows that arsenic concentrations in humans may exceed by two orders of magnitude the threshold considered as non-toxic (1 ug/g). While studies of modern populations from Arica suggest a historical adaptation that led to genetic resistance to arsenic.

We will present the compiled human osteological and elemental data for the coastal region of northern Chile. The analysis of the results reveals how adaptive processes explains the resilience of local populations.

S9-2 Modelling the evolution of the hinterland of Ravenna: human-environment relationships in an extremely dynamic alluvial landscape

Michele Abballe[1]

[1]Ghent University, Belgium

Contact: michele.abballe@gmail.com

While Italian archaeology is still struggling to find a way to embrace open data policies, reports from rescue/preventive archaeology are increasing day by day. Unfortunately, these datasets remain largely unpublished and are usually "locked" in physical archives, even though the recent release of the Geoportale Nazionale per l'Archeologia (GNA) in July 2023 may further promote their exploitation. Despite the difficulties in accessing these reports, they contain novel archaeological evidence that helps update our historical knowledge. Furthermore, they provide a wealth of palaeosols attestations, precious geological markers widespread within floodplains subject to sudden changes in the river network and relevant rates of vertical accretion. To exploit these data, a digital workflow has been developed to include several hundreds of

stratigraphies collected from both unpublished archaeological reports, published geological records and legacy geological data, as well as sites from archaeological maps, to model the evolution of the palaeolandscapes of a floodplain subject to intense morphological changes such as the hinterland of Ravenna. These models gave us new insights into human-environment relationships, especially in terms of settlement patterns, changes in the river network and reclamation processes, more precise representation of the physical environment in the past as well as a better understanding of the geomorphological biases affecting our current archaeological knowledge.

S9-3 Palaeoecology made easy: pastclim and tidysdm, two new R packages to make the best use of palaeo Big Data

Michela Leonardi[1], Margherita Colucci[2], and Andrea Manica[1]

[1]University of Cambridge, Evolutionary Ecology Group, Dept. of Zoology, UK

[2]Max Planck Institute of Geoanthropology, Germany

Contact: ml897@cam.ac.uk

The release of large datasets of archaeological occurrences (e.g. ROAD) and palaeoclimatic data series covering hundreds of thousands of years (e.g. Beyer et al. 2020, Krapp et al. 2021) opened the door to new exciting possibilities in the field of palaeoecology.

In this talk we will present two new R packages released by our lab that facilitate the use and analyses of palaeo Big Data in the context of palaeoecology.

pastclim (Leonardi et al. 2023a) allow easy access and manipulation of several sets of palaeoclimatic reconstructions covering the whole world between 5 million years ago and future climate projections. It can also work on custom data, if properly formatted. It contains a set of functions to recover the climate for time periods of interest either for the whole world or specific areas, extract data from locations scattered in space and/or time, retrieve time series from individual sites, and manage the ice or land coverage.

tidysdm (Leonardi et al. 2023b) takes advantage of the tidymodels framework in R to perform species distribution/habitat suitability/ecological niche modelling based on present-day or palaeo data. This is the first software specifically designed to work with diachronic data, a task that with other packages/pieces of software is not always possible or requires extensive tweaking. The integration with pastclim allows hassle-free access and handling of climatic data.

Finally, we will present some examples of their applications, to show how much they can offer to scholars studying the past (e.g. archaeologists, archaeozoologists, palaeoecologists, evolutionary biologists).

S9-4 A case study: modelling human-malaria coevolution through time using paleoclimatic and palaeoenvironmental databases

Margherita Colucci[1], James Blinkhorn[1], Michela Leonardi[2], Andrea Manica[2], and Eleanor Scerri[1]

[1]Max Planck Institute of Geoanthropology, Germany

[2]University of Cambridge, Evolutionary Ecology Group, Dept. of Zoology, UK

Contact: colucci@gea.mpg.de

Malaria is a major widespread disease that led to the rise of resistance mutations in the human genome (e.g. sickle-cell disease) and its worldwide incidence is strongly driven by factors such as increasing human population density and urbanisation, with significant impact from environmental conditions. All these aspects may favour an increase of malaria transmission, yet mutations in the human genome indicate that selection pressures pre-date urbanisation.

Using high-resolution continuous paleoclimatic datasets (last 120,000 years) and spatially explicit database of anthropogenic land use estimates (HYDE v3.2), this paper investigates the impact of climatic and environmental changes on malaria incidence around 25-22 thousand years ago (KYA) in West and Central Africa. It also considers possible anthropogenic alterations of the environment that may have created the conditions for the Anopheles vector species to thrive and, therefore, may have shaped and catalysed the spread of malaria.

We model the distribution of several Anopheles vectors, examining environmental and climate changes and signatures of human activity. This project combines several aspects to model the interactions between the environment, malaria vector, human factors and the consequent selective pressures on the human genome.

We demonstrate how bringing all these elements together (paleoclimate and environmental data, epidemiological information, human genetic data, as well as human intervention on the environment) enabled us to test the conditions that favoured the vector-human interaction and shaped a major global health burden.

S9-5 Quantity over quality? Unraveling the impacts of chronological uncertainties in paleoenvironmental databases: A case study from Lake Burdur

Ümit Cevher Elmas[1] and Berk Gazanfer Süleyman[1]

[1]İstanbul Technical University, Eurasia Institute of Earth Sciences, Turkey

elmasu22@itu.edu.tr

The growing tendency to incorporate natural, computational, or data sciences into examining human-environment relationships enables unique methodological approaches for investigating the human past. The association of paleopalynological studies with understanding complex human-environment interactions has become a cornerstone of such research, as reflected in the data shared for further analysis. However, along with the common criticisms regarding assumptions about the temporal order of sedimentation, pollen provenance, and spatial resolution, paleopalynology also faces challenges in constructing the 'best' chronological order, a problem shared with other paleoenvironmental studies. Considering the necessity of

comparing pollen proxies with other proxies and models, an issue most researchers agree on today, this problem leads to further discrepancies.

Albeit improvements in calibrating dates and error envelopes within age-depth models, researchers often rely on traditional models or accept previous chronologies without much scrutiny. However, gaining insight into the complexities of archaeological/historical periods necessitates an enhanced chronological record. As data accumulates in the 'big data era', it becomes crucial for researchers to critically question and modify older datasets when required. Here, we propose a chronological data-handling approach to encourage researchers from various disciplines that deal with human-environment relationships. We will examine the dating records of sediment core taken from Lake Burdur, which spans approximately 4600 years BP, as a case study to better understand climatic and vegetational changes in southwestern Anatolia with a robust chronology. By doing so, we aim to emphasize the significance of haphazardly constructed chronologies in shaping divergent historical interpretations. Based on preliminary results, five different models utilized in constructing age-depth relationships yield highly variable chronological results for making historical and archaeological inferences. Among these techniques, employing Bayesian age-depth models that enable various inputs from historical knowledge to absolute dating records, researchers can enhance their ability to reconstruct climate and vegetation dynamics.

S9-6 The Neotoma Palaeoecology Database as a research tool to assess human landscape interactions during the Quaternary

Thomas Giesecke[1], John W. Williams[2], Simon J Goring[2], Walter Finsinger[3], Suzanne Pilaar Birch[4], Graciela Gil-Romera[5], Steffen Wolters[6], and Vojtěch Abraham[7]

[1]Department of Physical Geography, Utrecht University, Netherlands

[2]Department of Geography and Center for Climatic Research, University of Wisconsin-Madison, WI, USA

[3]ISEM, University of Montpellier, CNRS, IRD, EPHE, France

[4]Department of Anthropology and Department of Geography, University of Georgia, GA, USA

[5]Department of Geoenvironmental processes and Global Change , Pyrenean Institute of Ecology-CSIC,Spain

[6]Lower Saxony Institute for Historical Coastal Research, Germany

[7]Department of Botany, Faculty of Science, Charles University, Czech Republic

Contact: t.giesecke@uu.nl

More than 100 years of pollen analysis provide accounts of past vegetation cover on temporal scales of years to millennia, documenting the development of vegetation globally. With the recognition of human-induced vegetation change in fossil pollen assemblages, more than 80 years ago, European archaeology and vegetation history started to support each other in the understanding of human-landscape interactions. While initially archaeological knowledge helped dating pollen diagrams, radiocarbon dating allowed for independent conclusions to be drawn from palaeoecology. The ever-growing number of pollen records and the development of personal computers with increasing computing power called for the digital collection and analysis of pollen analytical results. This led to the establishment of databases starting in the late 1980s developing into a Global Pollen Database, while keeping their continental identities. As a pioneer in open data, Eric Grimm stimulated and facilitated this progress and devised the next step in the development: The Neotoma Database, which is designed to hold many different types of palaeoecological and palaeoenvironmental data, such as botanical and zoological finds or sediment geochemistry, supporting their joint analysis. An essential feature of Neotoma is that it stores the different data types as well as the relationship between sample positions along the archive. Neotoma is not only a database infrastructure, but also a community bringing researchers from diverse disciplines together. It consists of constituent databases

that curate the data while benefiting from shared visualisation and access tools such as the interactive Neotoma Explorer and the neotoma2-R-package. The coverage of pollen data in Neotoma is global, while the distributions of many other proxies, such as vertebrate fauna, isotopes, or diatoms, continue to expand. We invite palaeoecological data communities to join Neotoma, by contributing their data or serving as stewards, particularly where the data type already exists or can be easily implemented.

Session 10: History unleashed: Harnessing the knowledge stored in historical documents

S10-1 Unraveling French Canadian ancestry: An interplay of genetics, genealogies, and geography

Luke Anderson-Trocmé[1], Dominic Nelson[2], Shadi Zabad[2], Alex Diaz-Papkovich[2], Nikolas Baya[3], Ben Jeffery[3], Ivan Kryukov[2], Christian Dina[4], Hélène Vézina[5], Jerome Kelleher[3], and Simon Gravel[2]

[1]University of Chicago, IL, USA

[2]McGill University, Canada

[3]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK

[4]Nantes Université, CNRS, INSERM, l'institut du thorax, France

[5]BALSAC Project, Université du Québec a Chicoutimi, QC, Canada

Contact: landersontrocme@uchicago.edu

The tapestry of human genetic history is woven through the ages across geographic landscapes through waves of migrations that have shaped the lineages of our ancestry. In this context, our study delves into the intricacies of the French-Canadian population, which largely descends from approximately 8500 French settlers in the 17th and 18th centuries.

Our research draws on an extensive database of over 4 million Catholic parish records from Quebec dating back four centuries, an invaluable resource for studying historical demography and understanding the influence of geography on human migrations and settlements.

By integrating this spatially resolved population scale genealogy with genomic data of French-Canadian individuals, we have traced the appearance of regional population structure, and the influence of geographical features such as river networks on migration rates and genetic similarities. This approach has revealed a series of sequential migrations up the tributaries of the St. Lawrence River, indicating the role of natural waterways in population expansion.

Furthermore, we show that the population scale genealogy encodes the structure of the present-day population as is evident from our novel genome simulation methods that provide a detailed spatiotemporal model of genetic variation at scales ranging from tens to thousands of kilometers. We have released this as a freely accessible simulated whole-genome sequence dataset with spatiotemporal metadata for 1,426,749 individuals reflecting the complex French-Canadian population structure.

Our study not only sheds light on historical demography and the influence of geography on human migrations and settlements but also underscores the value of historical documents in enriching cutting-edge genetic research. We believe this research exemplifies the potential of integrating historical databases with genomic data, paving the way for a deeper understanding of the intricate fabric of human history.

S10-2 Harmonizing big datasets: PANTROPOCENE, land use modelling, and the historical tropics

David Max Findley[1] and Patrick Roberts[1]

[1]Max Planck Institute of Geoanthropology, Germany

Contact: findley@gea.mpg.de

Existing historic land use models are largely based upon simplistic, reductionist, or environmentally deterministic presumptions about human land use. To be of use in historically-oriented disciplines, ranging from archaeology to climatology, land use models must be made sensitive to human agency and patterns of political, economic, social, and cultural organization ("Lifestyles"). They may then inform ongoing debates over humans' historic impact on environments and the global climate, the environmental consequences of industry and colonialism, and even the unequally distributed preludes of the Anthropocene. Constructing such models is an inherently multidisciplinary pursuit, and they must be built, from the bottom up, using archaeological, archival, and palaeoecological data. Harmonizing these empirical datasets to produce historic land use models quantifying Iberian colonialism's impact on tropical forests has been one of the defining challenges of the PANTROPOCENE project, which studies linkages between colonial exploitation in the pan-tropics and the environmental crises of the Anthropocene epoch. Here, we detail how (relatively) simplistic land use models were applied in the Philippines between 1564 and 1896 CE to create template Lifestyles and a basis for mapping land use change, forest cover, and human-induced atmospheric change, over three centuries of colonial expansion and resistance. These models simultaneously provide us with a framework for assessing and incorporating multi-disciplinary data sources while also proving highly adaptable to empirical data gathered by historians, archaeologists, paleozoologists, paleobotanists, and anthropologists. Their simplicity, we argue, is intrinsic to their success in this multi-disciplinary project, a first-of-its-kind effort.

S10-3 From historical sources to Wikibase

Daniel Baránek[1]

[1]Institute of History, Czech Academy of Sciences, Czech Republic

Contact: baranek@hiu.cas.cz

Research in the historical sciences has four classic phases: heuristics, criticism, interpretation, and narrative/presentation. The heuristic phase is usually the most time-consuming, although this phase, unlike the other three phases, requires more ""craft"" skills and less creativity and intellectual input. In recent years, developments in computer science have reached a point where much of the ""craft"" work can be transferred to computers through machine learning and human intellectual capacity can be devoted to the other three more creative phases. In reality, however, the delegation of much of the heuristic ""craft"" to computing in historical science still represents an ideal that has yet to be achieved. Although memory institutions (archives and museums) are digitizing massively, usually only in the form of simply taking pictures of archival material without any further processing.

The aim of this paper is not only to briefly introduce the digital humanities tools that I use in the processing of historical documents (OCR/HTR: eScriptorium, machine translation of texts, semantic annotation using machine learning, Wikibase, SPARQL), but especially to show the workflow and data conversion between the individual stages.

S10-4 A cross-verified database of locations visited by notable individuals

Etienne Wasmer[1], Palaash Barghava[2], Minda Belete[1], Morgane Laouenan[3], and Olivier Gergaud[4]

[1]New York University Abu Dhabi, United Arab Emirates

[2]Columbia University, NY, USA

[3]UMR 8589 CNRS - Université Paris 1 Panthéon-Sorbonne, France

[4]Kedge

Contact: ew1555@nyu.edu

History makes individuals and individuals make history. Famous individuals contribute to the visibility of cities, and vice-versa. The production of historical data on notable individuals has expanded in recent years. Databases such as Pantheon 1.0 (Yu et al. 2016), Pantheon 2.0 (Jara-Figueroa et al. 2019), and more recently our BHHT (Laouenan et al., 2022, https://www.nature.com/articles/s41597-022-01369-4) are widely used by researchers in the social sciences to document important topics. Geographical information on birth and date for a maximum of 2.29 million individuals (BHHT) over more than 3500 years of human history has been gathered, but a consolidated database of places visited by these individuals is still missing.

We fill this gap by systematically collecting and then sorting all the geographical links contained in the various biographies in the BHHT database. Using information from the text in a structured way, we assign a reasonable range of years for each location associated with an individual to identify their locations of residence and work over their lifetime. We currently focus on the English Wikipedia along with the geographic information contained in Wikidata for these individuals. Using Wikipedia editions for cross-verification as in Laouenan et al. (2022) has several advantages. In particular, using multiple Wikipedia editions simultaneously allows us to further assign a confidence and intensity measure of association for each location to an individual. This metric may prove particularly useful for measuring the impact of the presence of notable individuals on city growth from a historical perspective (with various focuses such as the development of global cities – e.g 20th century in Americas, 21st century in the Global South, or the Middle-Age and Industrial Revolution in Europe).

We show strength and weaknesses of our database and recommend selecting a subsample for greater accuracy.

S10-5 A convolutional network approach to mining historical map data

Clément Gorin[1], Pierre-Philippe Combes[2], Laurent Gobillon[3], Gilles Duranton[4]

[1]Université Paris 1 Panthéon-Sorbonne, France

[2]Sciences Po, France

[3]Paris School of Economics, France

[4]University of Pennsylvania, PA, USA

Contact: gorinclem@gmail.com

Historical maps contain rich information about buildings, transport networks, and land-use at different points in time. However, this information is encoded into pixel values and is not readily exploitable for analysis at scale. This paper develops a fully-convolutional neural network to extract pixel-level information from multiple collections of digitized historical maps covering mainland France in the middle of the 18th, 19th, and 20th centuries. Our model performs with remarkable accuracy and proves robust to considerable

representation heterogeneity, both within maps and across collections, as well as severe class imbalance. Our approach is efficient, scalable, and readily transferable to other historical maps with minimal manual labelling. The extracted information is used to document and analyse the development of urban areas and the changes in land-use from a historical perspective.

POSTER PRESENTATION

S10-6 Script-based automation practices: some experiences for managing early medieval society through written evidence

Sonia Medina Gordo[1] and Alfred Mauri Martí[2]

[1]Universitat de Barcelona, Spain

[2]Centre d'Estudis Martorellencs, Spain

Contact: sonia.medina@ub.edu

Knowledge generation requires careful reflection on how we manage information. This is true in many past-constructing disciplines, given the interpretative nature of data we work with. In recent years, there has been an increasing recognition of the value in investing time to unravel the complexity of our datasets. This data engagement not only allows us to actively participate in the data lifecycle but also ensures its applicability in other contexts. Similarly, scholars are becoming more aware of the inefficiency of studies that handle large quantities of data without ensuring their qualitative richness. In this regard, the availability of new evidence should help us to challenge or extend the observations generated on certain facts, instead of contributing to information overload. As a result, these practices of data introspection are emphasizing the importance of considering factors such as the speed of data gathering alongside the volume of information required for study development.

This contribution aims at reflecting on these issues through our experience managing written evidence from the early medieval period. Among other things, these vestiges are relevant because they allow us to define the logic of production and reproduction of a place through the human agency of this period. The volume of information available to address this reality has led us to implement management strategies that guarantee both the traceability of the process and data quality preservation. To do so, we build on the potential offered by FileMaker Pro DBMS and its functionality for script-based data manipulation. Simultaneously, we rely on a conceptual model through which we identify and define data. The combination of both elements has allowed us to develop a semi-automated labelling process for the units extracted, aiming to strike a balance between the amount of data we can work with and the speed of its retrieval.

Session 11: Modelling expansions in South America: Integrating archaeology and linguistics

S11-1 Combining archaeology, language and genetics into a single human (pre)history through time, space, and ecology

Paul Heggarty[1]

[1]Pontificia Universidad Católica del Perú, Peru

Contact: Paul.Heggarty@gmail.com

South America offers especially fertile ground for combining multiple disciplines' complementary records of the human past, across highly diverse ecological contexts. Alongside material culture in archaeology, on a more intangible dimension of culture, too, we can reconstruct another window on our past, from the rich panorama of indigenous languages here. Further perspectives, fast being enriched, are offered by human genetics, especially ancient DNA, and palaeoclimatology.

That linguistic contribution to this cross-disciplinary mix is often misconstrued, however. So this talk first clarifies what linguistics *can* reliably say about the past — and what it cannot. It summarises basic principles for a cross-disciplinary methodology, and dispels common misunderstandings. Language families attest to major expansions — but of people, culture, or both? And much less well known are linguistic convergence areas, which attest to converse processes, of contact and interaction.

Linguistics has various methods, old and new, to set language lineages in their real-world contexts in time, space, culture and environment. I briefly survey these methods, from the most traditional (e.g. linguistic palaeontology) to the most recent (e.g. Bayesian phylogenetics), and set out, for other disciplines, their strengths, weaknesses and limitations. In South America, where ecological and societal contrasts can be so stark, language histories too can be unexpectedly informative also on human-environment relationships through time.

On all of these issues, this talk proposes how new modelling software might best combine Big Data, from each discipline, into a more coherent, holistic vision. I set this out first for the big-picture patterns in the linguistic panorama of South America, and what they mean for other disciplines. I then focus on a case-study that ideally illustrates all issues: a new hypothesis on the origins and expansions of Quechua, the most widely-spoken surviving language legacy of the native Americas, informed by linguistics, archaeology, genetics — and now also ecology.

S11-2 Using spatial and chronological modelling to understand the Polychrome Tradition Expansion in Central Amazon

Rafael de Almeida Lopes[1], Philip Riris[2], Fábio Silva[2], Fernando Ozorio de Almeida[3], Eduardo Góes Neves[1], Eduardo Kazuo Tamanaha[4], Jaqueline Belletti[5], Erêndira Oliveira[6]

[1]Museu de Arqueologia e Etnologia da USP (MAE-USP), Brasil

[2]Bournemouth University, UK

[3]Universida de Estadual do Rio de Janeiro, Brasil

[4]Instituto de Desenvolvimento Sustentável Mamirauá, Brasil

[5]Independent Researcher

[6]Museu Paraense Emílio Goeldi

Contact: rafael.cardoso.lopes@usp.br

Around a thousand years ago, Amazonian history was marked by general social transformation for which the causes are still little known. Foremost among those transformations figures the expansion of occupations associated with Polychrome pottery that reshape most of Central and Western Amazonia. Studying the dimensions of the Polychrome Expansion is a key part in understanding Amazonian Deep History which is profoundly intertwined with the history of the Amazonian Forest itself. Varied tools have been applied for that, but here we propose that spatial and chronological modelling are of paramount use. To follow on this path, we start by presenting current debates about the period of 1,000 AD in Amazonia in general and about the Polychrome Expansion particularly, taking notice of critical gaps of knowledge that can be approached with spatial and chronological methods. Then, we describe methods used — such as SPDs and FastMarching — as well as the database used, AmazonArch, and analysis subsequent results. Finally, discussion will focus on the three identified moments of expansion of Polychrome occupations and its possible meanings.

S11-3 Demic diffusion, cultural transmission and change: assessing the spread of the Koriabo pottery style as an archaeological correlate of the Cariban-languages expansions in northern Amazonia

Bruno de Souza Barreto[1]

[1]University of São Paulo, Brasil

Contact: barretobrunosb@gmail.com

Between AD 1200 and 1500, the Koriabo pottery style emerged across various regions of Northern Amazonia, encompassing the Guianas and the Lower Amazon. Notably, these ceramics were also found in sites dating from the 16th and 17th centuries in the Lesser Antilles, associated with the Kalinago peoples. Despite its wide distribution, Koriabo ceramics exhibit remarkable homogeneity in decorative forms and vessel shapes, with most radiocarbon datings ranging from AD 1000 to 1500. Some scholars propose that these ceramics are an archaeological correlate of the Carib-speaking peoples, given their co-occurrence in regions historically occupied by them. Conversely, others contend that Koriabo ceramics are more closely tied to the development of long-distance exchange networks. In this study, I propose a two-folded explanation for the Koriabo phenomenon, involving the making and reproduction of exchange networks, as well as demic diffusion resulting in significant population movements. These migrations likely occurred at a rapid pace,

contributing to the creation of more homogeneous archaeological patterns. A prevailing archaeological model for Carib-language expansion suggests that the last phase of expansion commenced around AD 1000, driven by population growth and strengthened exchange networks, culminating in a diasporic movement to regions encountered during the colonial era. The spread of Koriabo ceramics into the Lower Amazon, coupled with local interaction processes, led to the emergence of scenarios characterized by cultural diversity and style hybridism. This is evident in the ceramic cache features and deposition pits found in sites along the estuarine coast of Amapá. In this paper, I aim to discuss these hypotheses by employing Geographic Information Systems (GIS) and time-series modeling of radiocarbon dates. By utilizing these analytical tools, I hope to gain deeper insights into the dynamics of the Koriabo pottery style's dispersal and its implications for understanding the cultural interactions and population movements within Northern Amazonia.

S11-4 A phylogenetic study of the Cariban language family

Fabrício Ferraz Gerardi[1], Fernando Orphão de Carvalho[2], Tiago Tresoldi[3]

[1]Universität Tübingen, Germany

[2]National Museum/UFRJ, Brasil

[3]Uppsala University, Sweden

Contact: fabricio.gerardi@uni-tuebingen.de

This study presents a preliminary linguistic phylogenetic analysis of the Cariban language family, a group of indigenous languages in Northern South America and Central Brazil. The family comprises around three dozen closely related languages, mostly spoken by small communities with a few hundred speakers. However, much of the family's history remains unknown, leading to numerous theories on its homeland and expansion routes. Following best practices in computational historical linguistics, we conducted initial analyses using neighbournets and Bayesian MCMC inference with different evolutionary models. Data was collected from reliable sources and organized into an independent, normalized, and open-access database in CLDF format, with cognate assignments made by experts. We discuss our classification results in the context of earlier classifications, examining statistical support for consensual and nearly consensual clades, such as Parukotoan and Pekodian, and other clades proposed in the literature, including the "Venezuelan" branch, Wayana-Apalai, and Panare-Pemongan. As part of a larger initiative to analyze the linguistic history of South America, this study contributes valuable insights into the evolution and relationships within the Cariban language family, including its homeland location, initial expansion date, and migration patterns. It also addresses the challenge of incorporating archaeological data into general phylogenetic inference, especially when dealing with language families lacking written records over extended periods. Moreover, the classification based on lexical evidence motivates the search for shared innovations in phonology and morphology, facilitating the reconstruction of intermediate-level proto-languages through traditional and computer-assisted methods. The study further strengthens the foundation for research on language contact among South American native languages.

S11-5 Beyond broad strokes – the future of ancient DNA in the Americas

Kathrin Nägele[1]

[1]Max-Planck-Institute for Evolutionary Anthropology, Germany

Contact: kathrin_naegele@evampg.de

The growing interest in investigating the human history of the Americas through the archaeogenomic lens has led to an incredible increase in available ancient sequences from the Pleistocene to the colonial period. The by far most extensively studied region in the past two years has been the Caribbean, where ancient DNA Could show how the past populations in the Archaic and Ceramic Age of the Islands correspond to different genetic signatures, corroborating ideas about separate settlements based on the Linguistic models of Vescelius and Granberry. Still, genomic studies were unable to add more detail beyond those previously obtained from archaeological research. The low genetic variation – a consequence of repeated bottlenecks in the settlement processes of the Americas and the Insular Caribbean – impose strong limits on the identification of genetic diversity, leaving open many questions about the interactions between and within pre-colonial communities. Studies in the Pacific have shown how interpretations based on one line of evidence can lead to oversimplifications of past events, and how by integrating the sometimes conflicting lines of evidence, we can approximate the interactions in the past. It is evident, that for understanding the true shape of human history in the Americas, archaeogenomic results have to be integrated with archaeological and – where available – linguistic evidence.

This presentation discusses the past and future approaches in studying population dynamics in the Caribbean and beyond, and how novel methods in ancient DNA, paired with archaeological and linguistic expertise, can offer more comprehensive windows into the American past.