Conference Abstracts

Jeremy Huggett, University of Glasgow
Datafication, Dataism, Data-centrism: Examining Big Data in Archaeology

We live in what some writers have described as the 'Age of Big Data', a time of data abundance, characterised as a revolution or a new gold rush, a new scientific paradigm. Archaeology is being shaped by Big Data to a greater or lesser degree, but compared with all the large-scale synthetic analyses undertaken within archaeology under the auspices of Big Data, there has been relatively limited enquiry into the phenomenon itself. In particular, what are the implications for archaeology of the interrelated concepts of datafication (an emphasis on quantification and automated data generation), dataism (a belief in the accuracy, completeness, and reliability of data), and data-centrism (an epistemological switch from theory-driven to data-driven research)? Big Data can be enthusiastically embraced for the transformative opportunities offered or seen as a socio-technical imaginary predicated on a belief in the pre-eminence of large datasets. Where does archaeology position itself?

Holly Wright, University of York
Economies of Scale: collaboration, big infrastructures and archaeological data

Now in its 23rd year, the Archaeology Data Service (ADS) is an OAIS compliant digital archive working at a national level in the UK, ensuring archaeologists have access to high quality and dependable digital resources, including openly licensed legacy data for reuse. The ADS acts as a metadata aggregator for archaeological data held by larger heritage agencies, smaller regional organisations and participates in international aggregation infrastructure projects such as ARIADNEplus, allowing users to access archaeological resources held in many countries from a single interface. Large-scale infrastructures can facilitate the building of long-term, complex relationships and active collaborations, not just technical solutions. This presentation will reflect on ARIADNEplus and the ADS, two large-scale online research infrastructures, including lessons learned, challenges and opportunities and how these types of infrastructures may help to create a more collaborative practice, now being actively explored within the SEADDDA Cost Action network.

Matt Peeples, Arizona State University
Large-scale data integration and archaeological synthesis: Ten years of collaborative work in the US Southwest
The Southwest Social Networks Project (now cyberSW) is a collaborative, multidisciplinary project focused on applying methods and models from network science to regional scale archaeological data from the U.S. Southwest and Mexican Northwest. The goals of this project include formally exploring the dynamics of interaction across the study area as well as developing new approaches toward incorporating archaeological cases in ongoing debates in the broader comparative social sciences. Over the last 10+ years, the project team has compiled a large and complex database with detailed information on tens of thousands of sites and millions of objects spanning a 1,000 year period. In this talk, I outline some of the challenges we have faced in collecting, integrating, analyzing, and interpreting these data and provide one example of how the size and scope of our data has allowed us to address new questions about the changing relationship between network positions and outcomes.

Gabriele Gattiglia, University of Pisa

*From datafication to Big Data: A possible approach for Data-driven Archaeology*

Digitisation has changed archaeology deeply and has increased exponentially the amount of data that could be processed, but it does not by itself involve datafication, which is the act of transforming objects, processes, etc. into a quantified format, so they can be tabulated and analysed. Datafication fits a Big Data approach and promises to go significantly beyond digitisation. To datafy archaeology would mean to produce a flow of data starting from the data produced by the archaeological practice, for instance, locations and relations between finds and sites. Through some example coming from the experience of MAPPA lab, it will be explained how this process would move archaeology towards data-driven research and Big Data. ArchAIDE, an H2020 funded project, for example, goes exactly in this direction with a tool for recognising archaeological potsherds which is able to create (and visualise) new data about findings and locations also generating new understanding.

Ben Marwick, University of Washington

*Big Data and Tool-driven Revolutions in Archaeological Science*

There is an argument in philosophy of science that revolutions in science are either ideas-driven or tool-driven. We explore this debate in light of recent efforts by many scientific disciplines to embrace methods to improve the reproducibility of their research. One of the most profound changes driven by this concern for reproducibility and transparency is from analysing data using tools dependent on pointing-and-clicking with a mouse in commercial software, to a new set of tools based on writing scripts in free and open source programming languages. These new tools are also distinctive because often the are the only way for researchers to work with big data. We first present bibliometric evidence for the status of archaeology as a science, followed by evidence for tool-driven change in the ecology as a paradigmatic case, and finally we review the evidence for archaeology in particular to test if the adoption of these new tools may be considered revolutionary. Our results show a positive citation effect for papers that use code. We consider to what extent the embrace of big data by archaeologists has driven this shift in tool use, and whether big data is making
archaeology more or less scientific. We discuss how computational approaches to improving reproducibility and transparency in archaeology are mediated and transformed by the use of code.

**Andreas Vlachidis, University College London**

**Reflections on excavating archaeological grey literature: and on the challenges in information extraction**

The largely unpublished reports generated by commercial or “rescue” archaeology, commonly known as “grey literature” contain a great deal of untapped information, highly relevant to the research and analysis of archaeological evidence. The presentation unfolds experiences and challenges in using Natural Language Processing techniques for “unlocking” and surfacing information from unstructured textual input, delivering structured outputs which enable new information access methods, based on linking worded representations to ontological definitions and formalisations for the purposes of information retrieval from heterogeneous data sources. The role of Named Entity Recognition, Relation Extraction, Negation Detection, and Word-Sense Disambiguation is presented in connection to a semantic annotation and automatic metadata generation endeavour, which spanned over ten years and two research projects, focusing on English, Dutch and Swedish grey literature.

**Daniel Pett, Fitzwilliam Museum**

**Do Museums have big data? Are they jumping on a trend?**

This talk will look at the use of the application of big data within the museum sector, where scale of data can be problematic for a very analogue based workforce. It will give some examples of projects that hold large datasets, techniques that have been applied and discuss their impact on the world.

**Anwen Cooper, University of Manchester**

**Chris Green, University of Oxford**

The 5-year ERC-funded English Landscape and Identities project, University of Oxford, set out in 2012 with the audacious idea of gathering together researchers from a multitude of backgrounds and seeing what they could do with an extravagant amount of legacy data (almost one million digital records in diverse formats together with substantial quantities of mapped information) generated across English archaeology over the last 150 years. Seven years on – the project completed, the final monograph submitted for publication – we reflect on the outcomes of this experiment with ‘big data’. In particular, we want to relay the positive outcomes of ‘thinking big’ – of working at significant spatial and temporal scales with ‘characterful data’, of being interpretatively bold, and of seeking to connect broad-brush computational analysis with the outcomes of intricate ‘bottom up’ research. We will discuss key project findings that developed specifically from working in this way, as well as contemplating the shifts in mindsets and in the emphasis of critical attention entailed.
Hector Orengo, Catalan Institute for Classical Archaeology
Arnau Garcia-Molsosa, University of Cambridge
Francesc C. Conesa, University of Cambridge
Cameron A. Petrie, University of Cambridge

Applications of machine and deep learning for the identification and analysis of archaeological sites at a continental scale using 40 years of multi-source satellite data.

Satellite data covering most of the Earth surface has been systematically collected during the last 40 years. Given the spatial, spectral and temporal resolution of multiband and other types of satellite-based data, these effectively constitute Big Data. In this presentation we will discuss how spatial statistics, machine and deep learning and other computational automated processes can be employed to query this enormous amount of multi-temporal and multi-source information in order to extract information of archaeological interest. In particular, the automated detection and classification of archaeological sites will be employed to illustrate current approaches. The paper will also briefly describe the conceptual and methodological workflows adopted to deal with massive satellite and aerial imagery datasets. These are based on the use of high-performance computing including parallel computing, cloud computing and virtual machines and the development of specialised algorithms and their derived code.

Andrew Bevan, University College London

Contemporary balancing acts between data, theory and method in archaeology

Recent floods of archaeological evidence have brought exciting new challenges and opportunities. Even so, such evidence still constitutes just one corner of a more enduring ternary relationship: between data, theory and method. These three components are widely understood as parts of almost any research cycle, but their specific meaning and relative priority in archaeology has nevertheless changed dramatically over the 150 years, in step with evolving or oscillating approaches to archaeological knowledge construction. Across all three components, we have also seen interesting shifts in scale, back and forth from micro-sampling to data-mining, from thick to thin descriptions, or from individual to systemic agencies, to evoke just a few aspects. This paper reflects on what current enthusiasm for large-scale data might mean in this wider context.