Digital Data Relations

written by Jeremy Huggett | 28/06/2018



(adapted from original by Gerd Leonhard, CC-BY-SA 2.0)

We sometimes underestimate the impact of digital data on archaeology because we have become so accustomed to the capture, processing, and analysis of data using our digital tools. Of course, archaeology is by no means alone in this respect. For example, Sandra Rendgren, who writes about data visualisation, infographics and interactive media, recently pointed to the creation of a new genre of journalism that has arisen from the availability of digital data and the means to analyse them (2018a). But this growth in reliance on digital data should lead to a re-consideration of what we actually mean by data. Indeed, Sandra Rendgren suggests that the term 'data' can be likened to a transparent fluid – "always used but never much reflected upon" – because of its ubiquity and apparent lack of ambiguity (2018b).

It wouldn't be true to say there was a lack of reflection in archaeology – for example, in the current issue of *Advances in Archaeological Practice,* Ben Marwick and Suzanne Pilaar Birch point to a range of debates in the philosophy of science about the definition of data and conclude that what constitutes data "is dependent on who uses it, how, and for what purpose" (2018, 126). This is what, in the same issue, I describe as the slipperiness of data: the way in which it is negotiated and changes through time and through (re)use, "so that what was not considered to be data at one time may be of interest at a later stage and, conversely, what was once understood to be data may no longer be seen to have value" (Huggett 2018a, 98). But of course this is a long way from what data literally means.

Daniel Rosenburg provides a useful summary of the derivation of the term: from the plural of the Latin *datum*, 'that which is given' (2013, 18), and archaeologists frequently refer to 'raw data', which is often seen as data which are uncontaminated by methodological and theoretical biases (e.g. Carson 1997, 316), and hence effectively givens. Indeed, this is not unconnected to a supposedly 'scientific' approach in which constants, conversion factors, and the outputs of formal

procedures and hardware and software black boxes, can all be seen as givens. For example, Conal Boyce (2010, 192-195) distinguishes between 'givens' and 'tangibles' in chemistry experiments, where tangibles are non-givens and are often sensory in nature (he suggests they include audio, visual, tactile, olfactory and kinesthetic aspects). On the face of it, this might chime with the archaeological data experience – our sensory relationship with the trowel's edge, for instance – but Boyce goes beyond this and suggests that in this given/tangible paradigm the line between the two can be shifted more or less at will (2010, 197-196). In other words, in one case the line might be drawn such that most if not all elements are essentially treated as givens for simplicity (Boyce's example is an applet or computer simulation), whereas in another the black boxes are opened and the givens become tangible. It's an interesting model to consider, although Boyce warns of some of the risks of an over-emphasis on givens – the sanitisation of the subject, the over-reliance on the computer simulation, and a degrading of experience/knowledge (2010, 197-198). The idea that data might be perceived in terms of givens and tangibles would seem attractive, all the more so if the boundary between them is slippery and can be defined and redefined according to purpose and need.

However, it may be better to start from the position that all data are tangible rather than given. Indeed, as I pointed out some time ago (Huggett 2015, 17), both Chris Chippendale (2000) and Johanna Drucker (2011) have independently argued that 'capta' is a better term to use, both agreeing that data as 'capta' are captured or gathered, rather than given: it is "constructed as an interpretation of the phenomenal world, not inherent in it" (Drucker 2011 para 8), emphasising the interpretative, observer-dependent, nature of data/capta. Essentially, all data are cooked, rather than raw (see Huggett 2016).

So the idea of data may be fluid, but it is certainly not transparent (contra Rendgren). But this is not the end of it. Intriguingly, Sandra Rendgren suggests (2018b) that the introduction of digital technology alters our relationship with data:

> 'Data' began to be used for digitally encoded information, which can be stored and processed by computers. And as such, it begins to exist the moment it is recorded by the machine. And this is a crucial twist. All through the evolution of statistics through the 19th century, data was generated by humans, and the scientific methodology of measuring and recording data had been a constant topic of debate … The notion that data begins to exist when it is recorded by the machine completely obscures the role that human decisions play in its creation.

So to what extent have our perceptions of data changed within or been altered by a digital environment? Do we still hold to those 19th century perspectives of data as more or less unambiguous observations, that archaeological data is 'out there', waiting to be discovered? Can data still be perceived as a given, and if so, does making data digital change this or even reinforce this? Or if we accept the slippery fluidity of data, how does this affect data within a digital environment?

Kaufmann and Jeandesboz (2017, 316-319) recently defined a set of digital affordances, many of which directly relate to data and our use of (relationship with) it. These include the malleability and flexibility of digital devices, their storage capabilities, their searchability, their connectivity, their

computability, their interactive nature, and their creation and organisation of data. All of these – and more – in combination make for an unarguably attractive environment for data production, manipulation, and consumption. However, at the same time, it insulates us from the data though access to increasing quantities of data and their apparent quality, usability, and flexibility. For example, Gavin Smith has looked at how the consequences of the use of digital devices and data may be to obscure rather than reveal, and prioritises what he calls "data-based gratification" (2018, 2). Following boyd and Crawford (2012, 663), he points to the way that digital data sets can appear to come equipped with an aura of truth, objectivity and accuracy (2018, 3) and warns of the risk when we "learn to treat and utilise data in parochial and instrumental ways, as simply 'means to ends' … rather than as vital artefacts that also agentively construct and structure social experiences and environments" (2018, 7).

Smith (2018, 8-11) identifies three kinds of data-based relations that arise in a digital environment:

- ***Fetishisation***, where the significance of the data is inflated so as to assign them a higher level of insight than is warranted by virtue of their digital affordances;
- ***Habituation***, where the familiarity, proximity, accessibility, and apparent usability of digital data means that we overlook – or are unaware of – their underlying limitations, assumptions, inconsistencies (ignoring associated contextual metadata, for instance), and in the process perhaps revert to a more traditional perspective of the data as a set of givens rather than as capta;
- ***Seduction***, where we become enchanted by our access to digital infrastructures and data flows, using interfaces deliberately designed to encourage and ease our access whilst invisibly shaping it.

I've argued elsewhere that digital data change the terms of our engagement through their near-instant access, volume, and flexibility, and hence they have potentially transformative effects on the practice of archaeology now and in the future (Huggett 2018a, 101). The affordances of the digital intervene and mediate our production, access, and use of data, and in doing so, they have the potential, at least, to muddy our relationship with data in ways that may not be helpful to our archaeological practice. A digital environment which increasingly facilitates the aggregation of datasets into meta-analyses or large-scale synthetic analyses, based on the availability of large quantities of variable-quality data held in open repositories and used for purposes for which they were not originally intended, can inadvertently heighten the risks of fetishisation, habituation, and seduction of our digital data. This is an example of what I wrote about in an earlier post – that while as archaeologists we essentially agree on our subject of study, where we diverge is in how we arrive at an understanding of what we can say about it, our methodology for creating knowledge about the empirical remains of past worlds (Huggett 2018b), and our attitude and approach to digital data is a key aspect of this. At the very least, whether we are insinuating technological tools into the process of data collection or receiving volumes of 'primary' data transmitted from remote digital archives, our increasingly arms-length relationship with data introduces new dimensions to manipulating, understanding, and (re)communicating archaeological information which we need to be aware of (Huggett 2014).

References

Boyce, C. 2010 'On the boundary between laboratory 'givens' and laboratory 'tangibles',

Foundations of Chemistry 12 (3), 187-202. https://dx.doi.org/10.1007/s10698-010-9093-6

boyd, d. and Crawford, K. 2012 'Critical Questions for Big Data', *Information, Communication & Society* 15 (5), 662-679. https://dx.doi.org/10.1080/1369118X.2012.678878

Carson, C. 1997 'Laser Bones: copyright issues raised by the use of information technology in archaeology', *Harvard Journal of Law and Technology* 10 (2), 282-319.

Chippindale, C. (2000), 'Capta and data: On the true nature of archaeological information', *American Antiquity* 65 (4), 605–612. https://dx.doi.org/10.2307/2694418

Drucker, J. 2011 'Humanities Approaches to Graphical Display', *Digital Humanities Quarterly*, 5.1 http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Huggett, J. 2018a 'Reuse Remix Recycle: Repurposing Archaeological Digital Data', *Advances in Archaeological Practice* 6 (2), 93-104. https://dx.doi.org/10.1017/aap.2018.1 (also available at http://eprints.gla.ac.uk/161510/)

Huggett, J. 2018b 'A Romantic Digital Archaeology', *Introspective Digital Archaeology*, 12 May 2018. https://introspectivedigitalarchaeology.com/2018/05/12/a-romantic-digital-archaeology/

Huggett, J. 2016 'Deep-fried archaeological data', *Introspective Digital Archaeology,* 16 Oct. 2016. http://introspectivedigitalarchaeology.com/2016/10/16/deep-fried-archaeological-data/

Huggett, J. 2015 'Digital haystacks: open data and the transformation of archaeological knowledge', in Wilson, A. and Edwards, B. (eds.) *Open Source Archaeology: Ethics and Practice.* De Gruyter Open, pp. 6-29. https://dx.doi.org/10.1515/9783110440171-003 (also available at http://eprints.gla.ac.uk/114652/)

Huggett, J. 2014 'Big Data and Distance', *Introspective Digital Archaeology*, 15 Nov. 2014. https://introspectivedigitalarchaeology.com/2014/11/15/big-data-and-distance/

Kaufmann, M. and Jeandesboz, J. 2017 'Politics and "the digital": From singularity to specificity', *European Journal of Social Theory* 20 (3), 309-328. https://dx.doi.org/10.1177/1368431016677976

Marwick, B. and Pilaar Birch, S. 2018 'A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing', *Advances in Archaeological Practice* 6 (2), 125-143. https://dx.doi.org/10.1017/aap.2018.3 (also available at https://dx.doi.org/10.17605/OSF.IO/PY4HZ)

Rendgren, S. 2018a 'Data journalism for the people', idalab blog, 24 April 2018. https://idalab.de/blog/data-visualisation/data-journalism-for-the-people

Rendgren, S. 2018b 'What do we mean by "data"?', idalab blog, 20 June 2018. https://idalab.de/blog/data-science/what-do-we-mean-by-data

Rosenburg, D. 2013 'Data Before the Fact', in Gitelman, L. (ed.) *"Raw Data" is an Oxymoron*, Cambridge MA: MIT Press, pp. 15-40. (also available at

http://static1.1.sqspcdn.com/static/f/1133095/23310656/1376447540493/Rosenburg_RawData.pdf?token=dOxBykW5o6Wu4syltUyaoxKAnu0%3D)

Smith, G. 2018 ‘Data doxa: The affective consequences of data practices’, *Big Data & Society*, 5 (1), 1-15. https://dx.doi.org/10.1177/2053951717751551