Big Data and Distance

written by Jeremy Huggett | 15/11/2014

One of the features of the availability of increasing amounts of archaeological data online is that it frequently arrives without an accompanying awareness of context. Far from being a problem, this is often seen as an advantage in relation to 'big data' – indeed, Chris Anderson has claimed that context can be established later once statistical algorithms have found correlations in large datasets that might not otherwise be revealed.

The sheer quantity of data is argued to make quality less significant, so that the size of the datasets will offset any problems associated with errors and inaccuracies in the data to the extent that

> "It isn't just that 'more trumps some', but that, in fact, sometimes 'more trumps better'." (Mayer-Shönberger and Cukier 2013, 33).

Such proponents of 'big data' adopt a somewhat fetishistic belief in the power of systems to overcome the perceived limits of 'small data'. Fortunately the idea that 'big data' somehow carry an aura of truth, objectivity, and accuracy through size alone has been identified as a myth (for example, boyd and Crawford 2011).

In a recent post, Tim Hitchcock writes about big data and small data in the context of calls linking Big Data with a return to 'longue durée' history. He talks about 'macroscopes' (and see also *The Historian's Macroscope: Big Digital History* by Shawn Graham, Ian Milligan and Scott Weingart), visualisation tools that supposedly work at all scales, from the largest to the smallest, but Hitchcock observes that users of Big Data tend to miss the micro-scale altogether. This is reminiscent of the idea of 'distant reading' in literary computing which enables the identification of temporal and national devices, themes, tropes, and genres, whereas 'close reading' focuses on a very small canon, which, given the size of the potential set of texts, means that the reader has to abstract larger themes from a very much smaller sample (Moretti 2000). This issue of scale of focus equally applies within an archaeological context – how much data must be assembled in order to be able to address a research question with confidence? How many sites are necessary to be able to extrapolate conclusions to the larger corpus?

The idea of distance and data goes beyond just the question of scale. For example, there is also the issue of remoteness: the recognition that there are times when the detailed small-scale (human scale?) archaeological analysis is precisely what is required and yet the digital methods intervene and their mediation inserts distance from the object of study. Distance can also bring with it a sense of separation from the data – not necessarily in terms of the actual data to hand, but in relation to what those data purport to represent. Although the analyst is isolated from the object of record in a way that in some respects is no different to the relative isolation experienced through the medium of the printed volume, unlike the printed experience the individual is insulated by the quantity, and apparent quality, usability and flexibility of the digital data. Whether insinuating technological tools into the process of data collection or receiving volumes of 'primary' data transmitted from remote

digital archives, the increasingly arms-length relationship with those data introduces new dimensions to manipulating, understanding, and (re)communicating archaeological information.

While examples of archaeological 'big data' use are as yet rare, the foundations are being laid in the data infrastructures being constructed, the automated alignment of data from disparate sources, and the automatic extraction of data from published sources, for example. What will be the implications of 'big data' methodologies applied to archaeological data? How will they change our approach to archaeological synthesis, for example?

References

boyd, d. and Crawford, K. 2011 'Six Provocations for Big Data'

Mayer-Shönberger, V. and Cukier, K. 2013 *Big Data: a revolution that will transform how we live, work and think*. London: John Murray.

Moretti, F. 2000. 'Conjectures on World Literature', *New Left Review* 1, 54-68.