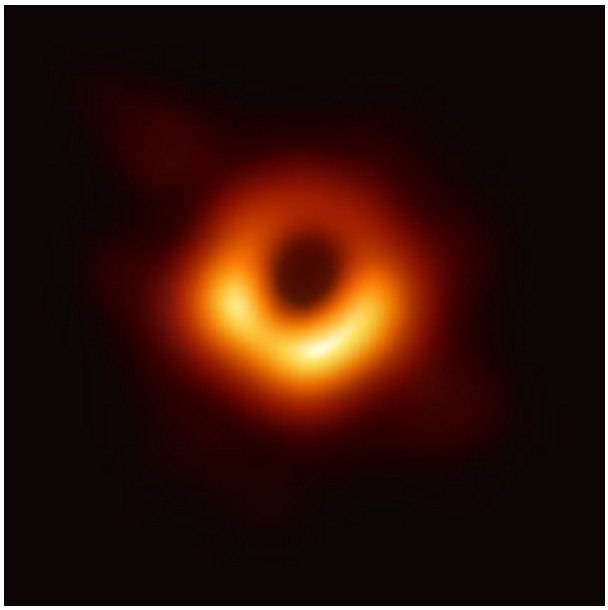
Nothing is Something

written by Jeremy Huggett | 31/03/2021



The black hole at the centre of Messier 97, via the Event Horizon Telescope (Wikimedia CC-BY)

Shannon Mattern has recently written about mapping nothing: from the 'here be dragons' on old maps marking the limits of knowledge and the promise of new discoveries, to the perception of the Amazon rainforest as an unpeopled wilderness until satellite imagery revealed pre-Columbian geoglyphs which had been largely invisible on the ground. In her wide-ranging essay, she makes the point that nothingness is always something: "A map of nothing demonstrates that an experiential *nothingness* depends upon a robust ecology of *somethingness* to enable its occurrence" (Mattern 2021). The question, of course, is what that something actually is.

Nothingness is something that has long been an issue in databases. Null is traditionally used to represent something missing. As null is not a value, it is technically and meaningfully distinct from zeros and empty strings which are values and hence indicators of something. Although this seems straightforward, the boundaries begin to blur when some guides to SQL, for instance, define null in terms of both missing and unknown values. After all, if something is missing, then we know we are missing it; if something is unknown, then we don't know whether or not it was ever something. Indeed, Codd, in his classic book on relational databases argued that null should also indicate *why* the data is missing, distinguishing between a null that is 'missing but applicable', and a null that is 'missing but inapplicable' (Codd 1990, 173), but this was never adopted. Consequently, nulls tend to have a bad reputation because of the ways they may variously be used (mostly in error) as representing 'nothing', 'unknown', 'value not yet entered', 'default value', etc. in part because of messy implementations in database management systems.

Aside from drilling into students the idea that you should never use zeros or empty strings to represent missing data, most archaeologists also have it instilled into them the concept that absence of evidence is never evidence of absence. However, that may not be entirely true: Eli Wallach (2019) has pointed out that, unlike most sciences, archaeology frequently draws inferences from absences (often when determining *terminus post* and *ante quem*, for instance) and Wallach provides several examples where archaeologists reason from an absence of evidence. Of course, the question remains, why is the evidence absent? Is it because it isn't there – for instance, the disappearance of an artefact type between two phases may represent a real absence and hence change – or is it because it wasn't recorded or because it wasn't recognised, in which case the absence is uncertain and more correctly a case of missing data, so drawing inferences in such a case would be unwise.

This kind of problem arose recently in a large data study which purported to show a relationship between social complexity and belief in powerful moralising gods (Whitehouse *et al*. 2019). Critics of the study pointed to, amongst other things, the way that unknown data had been translated into evidence for absence by re-coding missing data as known absences prior to analysis (Beheim *et al*. 2019, 2-3). This was justified by the original authors on the grounds that "Given the nature of the historical and archaeological record, if there was no evidence of moralizing gods we can treat them as being absent" (Beheim *et al*. 2019, S1.2). This would seem to be another case of drawing an inference from an absence, but it isn't quite so simple: the presence of moralizing high gods was coded present/absent in the data with 'absent' and 'not reported' being treated the same (Whitehouse *et al*. 2019). It also rather relies on a reasonably unambiguous recognition of absence whereas archaeological evidence for belief is highly interpretative (see, for example, Costopoulos 2019) so hardly unequivocal.

This underlines some of the archaeological challenges surrounding the handling of our fundamentally intransigent, incomplete, and partial data. We're increasingly accustomed to the idea that our data are situated, contingent, and theory-laden: that data are embedded with cultural, political, technological and spatial norms which vary through time and space and between creators. But what does this actually mean in practice? Very little, on the whole, since the way we record and manipulate our data strips all of this colour out in pursuit of what is perceived to be analytically useful. This goes beyond treating absence of evidence as evidence of absence, or combining unknown data and absent data, or treating nulls as unknown rather than missing, or any of the other more legitimate methods of data cleansing. There are multiple shades of grey behind what can superficially seem to be simple data. Why was that thing recognised to be a fragment of information in the first place and determined to be capable of being recorded? Why was it considered significant enough to be worthy of recording while other things were not? Why was the selected level of detail or terminology used chosen? Even recording something as 'unknown' lacks nuance – there are things we do not know, things we do not know we do not know, things we do not know we know, as well as things we think we know but do not know (Huggett 2020, 5). All these factors are affected by our disciplinary and professional biases, by what we perceive to be as significant or of value, by our black-boxing of techniques and decisions, and by institutional and commercial expectations, for example. Little or none of this is captured within our data, or within any accompanying metadata or paradata.

So data are never self-evident or straightforward, yet in a digital environment, their apparent manipulability can make them seem that way. While this may seem to be a primarily academic

argument, consider what may happen when data from multiple sources are combined (as in the example above), or when large datasets are incorporated within machine learning environments. What happens to the range of unknowns and the richness of knowns once incorporated in the algorithmic melange? In many respects – and often for perfectly good reasons – our data are frequently dumbed down to make them recordable and reusable, but the effects of this are rarely recognised in the throes of our data-based gratification when we treat our data "as simply 'means to ends' rather than as vital artefacts that also agentively construct and structure" (Smith 2018, 7).

Interestingly, there's to be a session at the 2021 CAA conference in June on absence of evidence as evidence of absence organised by Steve Stead, George Bruseker and Athanasios Velios which, among other things, will look at how to make documented absences interoperable and resusable. However, this rather presupposes that absence is documented in the first place and documented in a more refined manner than a simple (or not-so-simple!) present/absent basis. As Mattern says, nothing is not just a thing: it may be many things. The presences and absences in our data are always political and cultural, frequently deliberate and often accidental, and a key challenge is to acknowledge this and represent their many aspects.

References

Beheim, B., Atkinson, Q., Bulbulia, J., Gervais, W., Gray, R., Henrich, J., Lang, M., Monroe, M., Muthukrishna, M., Norenzayan, A., Purzycki, B., Shariff, A., Slingerland, E., Spicer, R., and Willard, A. (2019). "Corrected analyses show that moralizing gods precede complex societies but serious data concerns remain". *PsyArXiv preprint*. https://doi.org/10.31234/osf.io/jwa2n

Codd, E. (1990). *The Relational Model for Database Management. Version 2*. (Addison-Wesley, MA).

Costopoulos, A. (2019). "Moralizing gods update: Seshat still searching for something that isn't there". https://archeothoughts.wordpress.com/2019/12/02/moralizing-gods-update-seshat-still-searching-for-something-that-isnt-there/

Huggett, J. (2020). "Capturing the Silences in Digital Archaeological Knowledge". *Information*, 11(5), 278. https://doi.org/10.3390/info11050278

Mattern, S. (2021). "How to Map Nothing". *Places Journa*l, March 2021. https://placesjournal.org/article/how-to-map-nothing/

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

Wallach, E. (2019). "Inference from absence: the case of archaeology". *Palgrave Communications* 5, 94. https://doi.org/10.1057/s41599-019-0307-9

Whitehouse, H., François, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., Purcell, R., Ross, R., Larson, J., Baines, J., Haar, B. ter, Covey, A., and Turchin, P. (2019). "Complex societies precede moralizing gods throughout world history". *Nature*, 568(7751), 226–229. https://doi.org/10.1038/s41586-019-1043-4