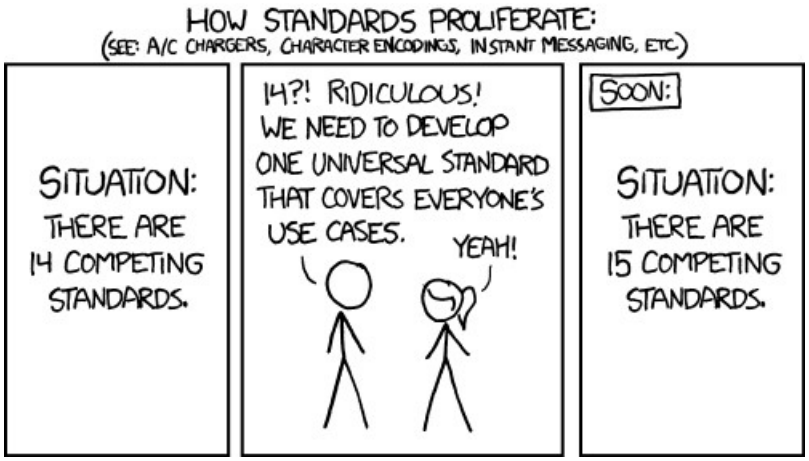
Data as Mutable Mobiles

written by Jeremy Huggett | 11/04/2023



(c) Randall Monroe https://xkcd.com/927/ (CC-BY-NC)

As archaeologists, we frequently celebrate the diversity and messiness of archaeological data: its literal fragmentary nature, inevitable incompleteness, variable recovery and capture, multiple temporalities, and so on. However, the tools and technologies that we have developed to use and reuse that data do their best to disguise or remove that messiness. Most of the tools and technologies that we employ in the recording, location, analysis, and reuse of our data generally try to reduce its complexity. Of course, there is nothing new in this – by definition, we always tend to simplify in order to make data analysable. However, those technical structures assume that data are static things, whereas they are in reality highly volatile as they move from their initial creation through to subsequent reuse, and as we select elements from the data to address the particular research question in hand. This data instability is something we often lose sight of.

The illusion of data as a fixed, stable resource is a commonplace – or, if not specifically acknowledged, we often treat data as if this were the case. In that sense, we subscribe to Latour's perspective of data as "immutable mobiles" (Latour 1986; see also Leonelli 2020, 6). Data travels, but it is not essentially changed by those travels. For instance, books are immutable mobiles, their immutability acquired through the printing of multiple identical copies, their mobility through their portability and the availability of copies etc. (Latour 1986, 11). The immutability of data is seen to give it its evidential status, while its mobility enables it to be taken and reused by others elsewhere. This perspective underlies, implicitly at least, much of our approach to digital archaeological data and underpins the data infrastructures that we have created over the years.

However, archaeological data may be mobile but it is not immutable. It changes as it travels and is altered as it encounters other data and is reused by researchers. It moves from one context to another and is decontextualised and recontextualised en route, possibly several times, before it is reused and then goes through the process again in relation to a different enquiry. The journeys that data undergo (e.g., Huggett 2022) foster a sense of distance being travelled, but a consequence of

this can be an arms-length relationship between the data and the user, encouraging a sense of remoteness through a reduction in both the context and the malleability and flexibility of its digital representation. Once data is incorporated within an infrastructure, though, it becomes institutionalised through its conformity with certain defined standards for format and representation. While the expectation is frequently expressed that such standards can mutate over time or be translated into new variants, the reality is that once standards are set, they tend to remain, and this is often taken as a sign of a successful standard. But such standards promote certain kinds of world views, certain forms of knowing, and also make some categories of data undocumentable and hence invisible (e.g., see Hacıgüzeller et al. 2022). They impose frictions on the data which limit them, constrain their subsequent use, and shape the meaning that can be derived from them.

So, where does this leave us when considering data reuse? The standard solution is to retain the messiness of the data rather than try and squeeze it into predefined structures and instead create a simple metadata summary to make the dataset easier to reuse, and of course this approach is widely applied across our digital archives (e.g., see Löwenborg 2018). But this apparent compromise corrals our data, silencing its variability and messiness as our attention transfers to the standardised metadata record. The metadata becomes increasingly used, not simply as a cataloguing or finding aid, but as data, and we are seeing any number of studies in which the 'what', 'where' and 'when' of classic archaeological metadata provides basic summary data that forms the basis of numerous distribution analyses and 'big data'-style studies. Metadata therefore shifts mode and 'becomes' data rather than how data is located or integrated: it effectively travels between being metadata and being data. Such metadata-derived data is another layer of abstraction further removed from the original, primary record but it has a different relationship with this original data because it is not just data itself but a means of finding data, so it carries the biases and worldviews of the infrastructure that created it and governs not just what can be found but also what can be known.

Metadata clearly has a role to play. However, we need to recognise that it is situated in a particular analytical and infrastructural context, however generalised it might appear to be. So we need to find ways to break out of this, to enable us to work with our messy data in a more flexible manner which embraces both its messiness and the unpredictability of our questions about it. What is needed is a different approach to the management and organisation of archaeological data, one which respects the messiness of the data by not enforcing it into a structure that may not be appropriate for all circumstances. Heilen and Manney (2023, 6) have recently argued that archaeological data need to be reconceptualised as living data that retain utility beyond the immediate purpose for which they were generated, although they see this as achieved through a more unified set of standards. I suggest that this living data requires a different approach, one that overcomes the institutionalisation of much of our data by postponing that standardisation. That deferral then makes its eventual application problem-centred and targeted on the specific research enquiry to hand.

This pushes against the widespread assumption that structured, standardised data are required from the outset, and rejects the drive for single ontologies that purport to capture the nuances of archaeological data and which are endlessly extended as new challenges are identified. Randall Monroe's well-known comic image (above) sees the proliferation of multiple standards as a problem – instead we should recognise that different research questions place different demands on data,

focus on different aspects of data, and so we should embrace the diversity in our data rather than supress it.

This post is part of a presentation given at CAA2023 Amsterdam in the session 'How do we ensure archaeological data are usable and reusable, and for whom? Putting the R in FAIR for archaeology's data', organised by Sara Perry and Holly Wright.

References

Hacıgüzeller, P., Taylor, J.S. and Perry, S. (2021) 'On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind', *Open Archaeology*, 7(1), pp. 1709–1730. https://doi.org/10.1515/opar-2020-0220.

Heilen, M. and Manney, S.A. (2023) 'Refining Archaeological Data Collection and Management', *Advances in Archaeological Practice*, 11(1), pp. 1–10. https://doi.org/10.1017/aap.2022.41.

Huggett, J. (2022) 'Data Legacies, Epistemic Anxieties, and Digital Imaginaries in Archaeology', *Digital*, 2(2), pp. 267–295. https://doi.org/10.3390/digital2020016.

Latour, B. (1986) 'Visualization and Cognition: Thinking with Eyes and Hands', *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6, pp. 1–40.

Leonelli, S. (2020) 'Learning from Data Journeys', in S. Leonelli and N. Tempini (eds.) *Data Journeys in the Sciences,* (Cham: Springer International Publishing), pp.1-24. https://doi.org/10.1007/978-3-030-37177-7.

Löwenborg, D. (2018) 'Knowledge production with data from archaeological excavations', in I. Huvila (ed.) *Archaeology and Archaeological Information in the Digital Society,* (Abingdon, Oxon: Routledge), pp. 38–53.