The data deluge

Andrew Bevan

Postprint of Bevan, A. 2015. The data deluge, *Antiquity* 89.348: 1473–1484
doi:10.15184/aqy.2015.102

Introduction

Archaeology has wandered into exciting but daunting territory. It faces floods of new evidence about the human past that are largely digital, frequently spatial, increasingly open and often remotely-sensed. The resulting terrain is littered, both with data that are wholly new and data that were long known about but previously considered junk. This paper offers an overview of this diluvian information landscape and aims to foster debate about its wider disciplinary impact. In particular, I would argue that its consequences: a) go well beyond the raw challenges of digital data archiving or manipulation and should reconfigure our analytical agendas; b) can legitimately be read for both utopian and dystopian disciplinary futures; and c) re-expose some enduring tensions between archaeological empiricism, comparison and theory-building.

Overground, underground, systematic and serendipitous

It has become popular to speak of a 'big data' phenomenon in the social and natural sciences, as well as to a lesser extent in the humanities. Narrowly defined, the term 'big data' refers to those structured or unstructured digital datasets that are so colossal (often hundreds of terabytes or more) that they present unusual contemporary challenges with regard to everything from basic storage to read/write/search functions, analysis and visualisation. In truth, only a few kinds of archaeological information even approach this level of challenge: certain kinds of remote-sensing or ancient DNA datasets undoubtedly do, and in the future so too might certain kinds of automatically gathered evidence about public engagement (for example, online user statistics and geo-social networking related to heritage experiences). While most archaeological evidence may not typically be quite so vast, it is nonetheless at least as rich and challenging in other ways and the quantities are rising sharply (see Figure 1).

One major source of new information is remote sensing and there are at least six key developments that, together, have ushered in a remote-sensing revolution over the last decade.

- First, since the advent of high spatial resolution (i.e. ≤1m) commercial satellites and the freeing up of global positioning systems (GPS) at the end of the 1990s, archaeologists have increasingly benefited from a combination of viable base maps and location-aware fieldwork (e.g. where surveyors use devices that record where they are working in the landscape to a horizontal spatial accuracy of at least ±5–10m). This has largely removed the awkward positional uncertainties and paper mapping of previous decades. Earth viewers (e.g. Google Earth), on-board handheld GPS base maps and smartphone applications have become part of routine practice.
- Second, also worth noting is the sharply increased global geographic coverage of this imagery. That is partly been driven by the gradual accumulation of satellite-borne datasets collected for non-archaeological purposes (QuickBird, Ikonos, WorldView, Landsat, ASTER), but also by enlightened re-dissemination of historical imagery (e.g. declassified

CORONA or wartime aerial photos; Cowley *et al.* 2010; Casana & Cothren 2013). Terrestrial geophysics can now also declare itself a truly landscape-scale set of methods. Towed or pushed multi-sensor systems can cover many hectares of ground quickly, especially across the world's ploughed and open agricultural landscapes (Keay *et al.* 2009; White *et al*. 2013; Neubauer *et al.* 2014).

- Third, several of our airborne and ground-based techniques offer actively sensed data (where readings are taken based on a purpose-generated laser, radar pulse, electrical current, etc,. rather than passively from an existing source such as electromagnetic reflectance from sunlight) and new kinds of 3D depth information. For example, the advent of ground-penetrating radar (GPR), plus theoretical and software advances allowing GPR profiles to be interpolated as 2D plan views or 3D image cubes, has made this the most volumetric and geo-archaeologically sophisticated of remote-sensing approaches (Sala & Linford 2012; Conyers *et al*. 2013; Herrmann 2013). At a larger scale, both synthetic aperture radar (SAR) and airborne laser-scanning (LiDAR) now offer excellent cloud-, (and in the latter case) woodland- or rainforest-penetrating ways to collect topographic information at both coarse and fine spatial scales (Devereux *et al*. 2008; Chase *et al.* 2011). This opens up some of the hitherto most poorly known parts of the world to rapid survey even if ground-truthing (e.g. for site date and function) will remain a key step.
- Fourth, we are about to become far more obsessed with deep spectral resolution. As imaging, and especially 'hyper'-spectral imagery (sometimes also called imaging spectroscopy, where the electromagnetic spectrum is sampled more or less continuously rather than with gaps), becomes increasingly captured by unmanned aerial vehicles and plane-mounted systems. We are starting to be able to recover the kinds of spectral signatures for archaeological phenomena that have long been anticipated but largely unachievable due to the coarseness of much multispectral satellite coverage to date.
- A fifth aspect is that we can now undertake meaningful time series analysis, either by systematically collecting new diurnal, seasonal or inter-annual data (e.g. Poirier *et al*. 2013; Casana *et al*. 2014) or by fusing existing evidence in new ways (e.g. Menze & Ur 2012). As archives of historical aerial photographs and declassified Cold War imagery (see above) also come online, such analysis can be extended backwards to the early years of the twentieth century (albeit only for certain temporal windows, where serendipitous coverage permits, and typically only for qualitative visual comparison).
- Sixth, the last 10–15 years have seen plummeting costs in, and an overall democratisation of, remote sensing, at least for certain applications. As noted above, archaeological findspots can now often be located to within 5–10m by anyone with a handheld GPS or by viewing the locations on Google Earth. This enables less technically minded archaeologists and members of the public to contribute, and allows less well-resourced projects to develop modest remote-sensing programmes. The cost of high-resolution satellite imagery has dropped rapidly as the number of commercial platforms has diversified. Furthermore, structure-from-motion/multi-view stereo (SfM

and more broadly 'computer vision') methods provide an ultra-low-cost method for 3D data capture, merely by taking a sufficient number of overlapping images with a standard digital camera (Ducke *et al*. 2011; Green *et al*. 2014). Many of these low cost remote-sensing methods are also increasingly scale-agnostic in their application, with new opportunities to document and analyse not just sites and landscapes, but also artefacts (e.g. Bevan *et al*. 2014). At this 'artefact-scale', a further major advance has been the use of portable X-ray fluorescence (pXRF; Milić 2014) to achieve a rough impression of surface material composition. While it is worth sounding a note of caution about what SfM and pXRF can and cannot achieve (as well as stressing the continuing need for specialist input), both methods alleviate a whole series of traditional operational bottlenecks (e.g. needing to bring the object to the equipment), thereby allowing for much larger sampling programmes.

The above review offers a quick sense of the sheer pace of change in remote-sensing applications, both *in situ* and in the storeroom, but it would be a mistake to assume that such technologies were exclusively responsible for our current floods of archaeological information. For example, while the enormous threat posed to archaeology by modern construction is a story many decades old in certain Western countries (e.g. Webley *et al.* 2012), what has changed more recently is access to 'grey literature' reports from developer-led archaeological mitigation (Fulford & Holbrook 2011). This is in step with, but less widely known than, the massively expanded access provided by journal digitisation. As methods for digitising, searching, summarising and mining grey reports have advanced (Haselgrove & Moore 2007; Jeffrey *et al*. 2009; Vlachidis & Tudhope 2011), the sheer scope of fresh evidence that they represent has also become clear. In addition, while the gold standard of regional-scale digital inventories in the UK has long been the 'Sites and Monuments Record' (aka Historic Environment Record)—and these inventories continue to be important—we can now also point to a range of other synthetic georeferenced datasets for large areas, covering specific categories of archaeobotanical, metallurgical, zooarchaeological and landscape evidence (e.g. Bland 2005; Shennan & Conolly 2007; Conolly *et al*. 2012; Cooper & Green 2015). Increasingly, other old legacy datasets lurking in hard copy can be 'crowdsourced' via a huge group of interested volunteers (who each assist with individually small but collectively significant data collection tasks online, see below) and georeferenced via semi-automatic place-name look-up ('geocoding'). Further categories of existing data are finding new and unanticipated onward uses: perhaps the most important (albeit still contentious) example is the treatment of tens of thousands of radiocarbon dates 'as data', that is, as a proxy for aggregate quantities of human activity and therefore potentially as evidence for highs and lows in human population across time and space (e.g. Shennan *et al*. 2013). Likewise, there are also vast quantities of genetic data (ancient and modern, from humans, plants, animals and vestigial in 'dirt') as well as multiple long-term climate proxies (lake cores, speleothems, tree-rings) that are increasingly being shared digitally and tuned to address questions of mainstream archaeological interest.

A final point to note, but in many ways perhaps the most important, is the fact that almost all of the above categories of information are increasingly released into a public commons under liberal data licenses that encourage rather than prohibit 'downstream' re-use (e.g. Creative Commons' CC0, CC-BY). They are also increasingly the subject of peer-reviewed 'data papers' in publications such as *Internet Archaeology*, *Journal of*

Open Archaeology Data or *Open Context*, in which the strengths, weaknesses, biographies and re-use potential of this information is discussed (see also Huggett 2014; Kansa *et al*. 2014). This widening access and broader contextualisation is occurring not only 'top-down' with respect to the outputs from large-scale data collection exercises, but also 'bottom-up' by unlocking the potential of many smaller scale datasets that previously circulated in rather eccentric, narrow and conditional networks of colleague-to-colleague gifting (see below).

Mining and mashing; cases and controls

One obvious result of the data deluge is that, at least in certain parts of the world, we cannot in all good conscience claim "we don't yet have enough evidence" or that we should "wait till the evidence is in". The question then becomes how best to mine, mix and otherwise analyse a potential embarrassment of riches. It is not possible to do justice here to the full range of techniques that can be applied to such datasets, but some general points are worth making. First, there should be no shame in espousing empirical, often inductive, classification methods that continue to unpick the large-scale 'systematics' of this growing archaeological evidence. Even if this heralds, for some, an unwelcome return to culture-historical topics such as population-scale effects, continental-scale comparison and deep artefact traditions (Dunnell 1971; Roberts & Vander Linden 2011; Shennan 2011; Bevan 2014). In fact, there remains a clear role here for archaeology to contribute to more general debates about how best to collate, split or connect complex spatio-temporal datasets (Womble 1951; Barbujani *et al.* 1989; Huson & Bryant 2006).

Second, however dense it becomes, archaeological evidence will always remain patchy, with levels of uncertainty and variable expert opinions that are hugely challenging. In certain instances, the diluvian arrival of new data is a good reason to favour Bayesian inferential approaches that provide a formal framework both for expressing the strengths and weaknesses of existing knowledge, and then for incorporating new updates to it (Buck *et al.* 1996; Ortman *et al*. 2007; Fernández-López de Pablo & Barton 2015). Moreover, in addition to standard Bayesian methods, there is now also a range of further related approaches to data uncertainty (Beaumont 2010; Crema 2012; Bevan *et al*. 2013; Ducke 2014).

Third, we need to be clearer about analytical objectives. Is this growing body of evidence to be harnessed for 'pure prediction' or for wider behavioural, processual and interpretative insight? In certain heritage management situations, hyper-automated pure prediction may be both necessary and sufficient, for example to anticipate the extent of archaeological mitigation needed due to a proposed commercial development, to predict likely conservations risks to museum objects or to target coarse-grained public engagement strategies in cost-efficient ways. However, in the vast majority of archaeological situations, we *do* wish to interpret what the evidence is saying about past behaviours or processes. For example, we may wish to use archaeological settlement and field-system evidence not simply to anticipate where further examples of such archaeology might exist for protection purposes, but also to understand changing social logics influencing where people live in the landscape and how they organise it. It may also be of interest to know how these logics vary both regionally in the same time period and within the same region over different time periods. In such instances we need to: a) develop methods that are more transparent and grapple better with change detection, comparative distinctions and cause-effect wherever possible (bearing in mind the likelihood of equi- and multi-final outcomes, but not taking these as an excuse for never attempting formal analysis); b) recognise that both bottom-up agencies and patterning on a larger scale can be interesting; and c) balance more data-led treatments of our evidence with a continuing attention to behavioural theory and close, contextual archaeology.

More generally, faced with bewilderingly diverse but patchy sets of new archaeological information, we should be careful to avoid two kinds of fallacy: a) to presume that analytical confusion in the present necessarily implies the existence of behavioural complexity in the past; and b) to assume that real behavioural complexity in the lived past necessarily requires similar complexity in our present-day modelling. Put a different way, there is a risk that we explain away poorly constructed analyses and awkwardly complex analytical results by claiming they are due to a past reality that was behaviourally complex and irreducibly entangled. Even when we *do* think that a past situation was not straightforward, the fitting of complex models to complex data (just because we now can) will not necessarily lead to a more sophisticated understanding (e.g. Box 1976). Simplicity of approach remains a great virtue.

Finally, if we wish to make sense of the ever increasing, but still very uneven evidence, then we should follow the lead of subjects such as epidemiology or ecology and think far more in terms of cases-and-controls. This is one of the most robust approaches developed so far for evidence-based inference in situations where fresh, randomised experiments are impossible (Woodman & Woodward 2002). Typically, 'cases' refer to those observations that positively exhibit a property of interest (e.g. pottery of a certain type found at different sites) while 'controls' refer to those observations that do not (e.g. other types of pottery at the same sites). The idea can be further generalised and mapped as spatial surface via the notion of a continuous 'relative risk' of one kind of observation or measurement *vs* another (e.g. Kelsall & Diggle 1995; Hazelton & Davies 2009). An example from the spatial analysis of metal detecting data is the comparison of 2D kernel densities of one kind of metal find (e.g. Iron Age coins as the cases) with another (e.g. coins of other types as the controls) or with metal detector finds overall (Bevan 2012). More broadly, such compare-and-contrast approaches to archaeological data in high volumes are invariably much stronger strategies than single variable discussions, as recent work in multi-method terrestrial geophysics also makes clear (Kvamme *et al*. 2006). The rationale in all of these situations is what we might call 'adversarial', a term here borrowed very loosely from journalism (e.g. Hanitzsch 2007: 373) to place emphasis on the need to create and compare countervailing voices or datasets when faced with perceived biases in investigation. Hence the adversarial archaeological researcher seeks to mitigate investigative biases by juxtaposing one subset of evidence with the rest of a parent sample, with another distinct subset that has similar recovery conditions, or with other evidence known to be subject to quite different recovery conditions. Some archaeologists have long adopted this comparative, contrastive approach as a matter of course (see Smith & Peregrine 2012), but there are good reasons to make it more practically and theoretically explicit. In certain instances, this shift in emphasis from an obsession with one category of evidence to a strategic juxtaposition of several, may present unexpected challenges to current record-keeping practice. For example, if we are interested in modelling the locational preferences exhibited by Iron Age settlements recorded in a sites and monuments record, it may well be useful to recruit not just evidence about these Iron Age cases, but also about settlement locations in *any other* archaeological period in the study region, or alternatively all known archaeological findspots in the SMR. The latter two categories of 'control' give us a sharper way to handle biases in our record, rather than simply assuming the existing data is representative. However, asking for *all* SMR records because you are interested in one category of evidence is not, in many operational contexts worldwide, something that is currently easy to do in terms of permissions, or currently supported by adequate financial resourcing at the level of the record-keeping body itself.

Citizens and scientists

As briefly mentioned above, archaeologists are also crossing a major watershed regarding how they disseminate information (Lake 2012), with a dramatically increased emphasis on open-access and third-party licensing in various flavours (traditional publications, software source code, raw data). A related opportunity is provided by our growing ability to collect and discuss archaeological evidence in more public-facing and participatory ways (e.g. Kansa *et al.* 2011; Bonacchi 2012). Archaeological 'citizen science' has a long history offline involving volunteer archaeological and historical societies (especially, but not exclusively, in the UK), academic community collaborations (e.g. most recently, The Survey of Hillforts n.d.) and its fair share of contentious encounters (e.g. over metal detecting). But in many ways both the pros (predominantly) and cons (more rarely in my view) are now being turbo-charged by online coordination and modern digital documentation. For example, perhaps the earliest and most enduring example of massive online crowd-sourced documentation in archaeology is the UK-based Megalithic Portal (n.d. in operation since 2001). Over the last decade, vocational archaeologists have greeted it with an interesting mixture of enthusiasm (e.g. as a huge, community-driven database) and dismissiveness (e.g. about perceived poor data quality or about its inclusion of fringe archaeologies). Overall I would suggest that many or most would now consider it to be an extremely successful venture. Regardless and put simply, the archaeological record is a massive, spatially scattered, constantly threatened, and rapidly dwindling resource, and as a subject we have traditionally only received the support of rather small amounts of money. So there are compelling reasons to share responsibility for data collection, and where possible analysis and interpretation, beyond a group of traditional specialists (see also Bevan *et al.* 2014). Overall, 'volunteer sensing' is both a trend we cannot avoid (given metal detectors and earth viewers) and one we should be enthusiastic about (in certain circumstances and if conducted responsibly). There are many recent examples of effective collaboration, including the UK Portable Antiquities Scheme (n.d.), community-based geophysics projects (Geophysics and the Landscape of Hertfordshire n.d.) and SfM-led 3D modelling of archaeological features (MicroPasts n.d.; HeritageTogether n.d.; ACCORD n.d.).

A more pernicious feature of the current growth phase in digital archaeological information is the risk that it will make us less, not more, equal as researchers. For example, it is the historically wealthy, computer-savvy, ploughzone-rich and *Metallschock*-ed parts of the world that are arguably best placed to collect high-quality remote-sensing datasets. Even when other parts of the world have useful coverages of this kind (e.g. because they have been under surveillance in military conflicts), it is often Western institutions that retain best access to such information, and it is often researchers who have grown-up (educationally speaking) in data-, resource- and technique-rich Western archaeological settings who will be academically best placed to exploit it. Likewise, the subsidising of open access arguably raises as many problems as it solves, opening up research to a wider group but often involving premiums paid to traditional publishers by wealthy institutions for super-charged academic exposure. Hence, while remaining optimistic overall about the impact of the data deluge, there are good grounds for continuing concern about how such flows might recapitulate wider economic, political and knowledge asymmetries.

Final thoughts

In a sense, the metaphor of a single deluge of archaeological data implies too clear-cut a before-and-after moment, when in fact we will have to get used to chronic annual flooding. Regardless, the operating conditions for archaeological research are being transformed. For example, even if it will always remain crucial for archaeologists to frame arguments clearly, narrate them in an attractive way, contextualise closely and build theory, the need for more specific computational and numerical skills is also likely to grow further in importance alongside the growing complexity of our evidence and the scope of our explanatory ambitions (for a recent perspective on the latter, see Kintigh *et al*. 2014). More generally, the way we understand 'reproducible research' is likely to change (Stodden *et al*. 2014; rOpenSci n.d.) and require us to revisit which stages of the research process are given due credit and transparency as academic outcomes. Should it remain just the summary results of a research project that are so treated, or should data papers, collection protocols and analytical software also be peer-reviewed, published and given due intellectual credit? The answer to the latter questions is almost certainly yes, but the mechanisms by which this is done remain very new and not yet widely known. More generally, there is a pressing need to demonstrate that our information glut does indeed lead to better (rather than just more) archaeology and part of that challenge is related to the trade-off between archaeological empiricism and broader knowledge construction (to force a slightly unfair distinction). Matthew Johnson's concern (2011) that archaeologists are prone to disciplinary episodes of hyper-empiricism is fair: the risk is that we will now enter a rather undirected phase of gathering ever greater amounts of evidence, whilst assuming that it otherwise largely speaks for itself. In many instances, a more adversarial analytical approach is called for, in which we harness increased flows of archaeological information to set up explicit forms of comparison-and-contrast. Ultimately, whilst we should avoid undermining the kinds of detailed, hard-won specialist knowledge that remain key to a proper understanding of the archaeological record, this comparative effort will also need to cross some persistent sub-disciplinary barricades, and work 'ultra-longitudinally' across different chronological periods; periods that are usually treated separately in the same study region, 'trans-geographically' across traditional regional or modern political borders, and 'counter-artefactually' across constellations rather than single classes of material culture.

References

ACCORD n.d. Available at: https://accordproject.wordpress.com/ (accessed 01 July 2015).
BARBUJANI, G., N.L. ODEN & R.R. SOKAL. 1989. Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology* 38: 376–89. http://dx.doi.org/10.2307/2992403
BEAUMONT, M. 2010. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics* 41: 379–405. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144621
BEVAN, A. 2012. Spatial methods for analysing large-scale artefact inventories. *Antiquity* 86: 492–506. http://dx.doi.org/10.1017/S0003598X0006289X
– 2014. Mediterranean containerization. *Current Anthropology* 55: 387–418.

BEVAN, A., J. CONOLLY, C. HENNIG, A. JOHNSTON, A. QUERCIA, L. SPENCER & J. VROOM. 2013. Measuring chronological uncertainty in intensive survey finds. A case study from Antikythera, Greece. *Archaeometry* 55: 312–28. http://dx.doi.org/10.1111/j.1475-4754.2012.00674.x

BEVAN, A. X.J. LI, M. MARTINÓN-TORRES, S. GREEN, Y. XIA, K. ZHAO, ZH ZHAO, SH. MA, W. CAO & T. REHREN. 2014. Computer vision, archaeological classification and China's Terracotta Warriors. *Journal of Archaeological Science* 49: 249–54. http://dx.doi.org/10.1016/j.jas.2014.05.014

BEVAN, A., D. PETT, C. BONACCHI, A. KEINAN-SCHOONBAERT, D. LOMBRAÑA GONZÁLEZ, R. SPARKS, J. WEXLER & N. WILKIN. 2014. Citizen archaeologists. Online collaborative research about the human past. *Human Computation* 1(2): 183–97. http://dx.doi.org/10.15346/hc.v1i2.9

BONACCHI, C. (ed.). 2012. *Archaeology and digital communication. Towards strategies of public engagement*. London: Archetype Publications.

BLAND, R. 2005. Rescuing our neglected heritage: the evolution of the Government's policy on portable antiquities and treasure. *Cultural Trends* 14: 257–96. http://dx.doi.org/10.1080/09548960600573526

BOX, G.E.P. 1976. Science and statistics. *Journal of the American Statistical Association* 71: 791–99. http://dx.doi.org/10.1080/01621459.1976.10480949

BUCK, C.E., W. CAVANAGH & C.D. LITTON. 1996. *Bayesian approach to interpreting archaeological data*. Chichester: Wiley.

CASANA, J. & J. COTHREN. 2013. The CORONA Atlas Project: orthorectification of CORONA satellite imagery and regional-scale archaeological exploration in the Near East, in D.C. Comer and M.J. Harrower (eds.) *Mapping Archaeological Landscapes from Space* (Briefs in Archaeology 5): 33-43. New York: Springer.

CASANA, J., J. KANTNER, A. WIEWEL & J. COTHREN. 2014. Archaeological aerial thermography: a case study at the Chaco-era Blue J community, New Mexico. *Journal of Archaeological Science* 45: 207–19. http://dx.doi.org/10.1016/j.jas.2014.02.015

CHASE, A.F., D.Z. CHASE, J.F. WEISHAMPEL, J.B. DRAKE, R.L. SHRESTHA, K.C. SLATTON, J.J. AWE & W.E. CARTER. 2011. Airborne LiDAR, archaeology, and the ancient Maya landscape at Caracol, Belize. *Journal of Archaeological Science* 38: 387–98. http://dx.doi.org/10.1016/j.jas.2010.09.018

CONOLLY, J., S. COLLEDGE, K. MANNING, K. DOBNEY, B. STOPP & S.J. SHENNAN. 2012. The origin and spread of stock-keeping in the Near East and Europe [dataset]. Archaeology Data Service (collection 1141). http://dx.doi.org/10.5284/1016121

CONYERS, L.B., J.M. DANIELS, J.A. HAWS & M.M. BENEDETTI. 2013. An Upper Palaeolithic landscape analysis of coastal Portugal using ground-penetrating radar. *Archaeological Prospection* 20: 45–51. http://dx.doi.org/10.1002/arp.1439

COOPER, A. & C. GREEN. 2015. Embracing the complexities of 'Big Data' in archaeology: the case of the English Landscape and Identities project. *Journal of Archaeological Method and Theory*. http://dx.doi.org/10.1007/s10816-015-9240-4 [online first]

COWLEY, D., R. STANDRING & M. ABICHT (ed.). 2010. *Landscapes through a lens. Aerial photographs and historic environment*. Oxford: Oxbow.

CREMA, E.R. 2012. Modelling temporal uncertainty in archaeological analysis. *Journal of Archaeological Method and Theory* 19: 440–61. http://dx.doi.org/10.1007/s10816-011-9122-3

DEVEREUX, B.J., G.S. AMABLE & P. CROW. 2008. Visualisation of LiDAR terrain models for archaeological feature detection. *Antiquity* 82: 470–79. http://dx.doi.org/10.1017/S0003598X00096952

DUCKE, B. 2014. An integrative approach to archaeological landscape evaluation: locational preferences, site preservation and uncertainty mapping, in E. Meylemans, J. Poesen & I. In 't Ven (ed.) *The archaeology of erosion, the erosion of archaeology. Proceedings of the Brussels conference, April 28–30, 2008* (Relicta Monografieën 9 Archeologie): 13–22. Brussels: Flanders Heritage Agency.
DUCKE, B., D. SCORE & J. REEVES. 2011. Multiview 3D reconstruction of the archaeological site at Weymouth from image series. *Computers and Graphics* 35: 375–82. http://dx.doi.org/10.1016/j.cag.2011.01.006
DUNNELL, R.C. 1971. *Systematics in prehistory*. New York: The Free Press.
FERNANDEZ-LOPEZ DE PABLO, J. & C.M. BARTON. 2015. Bayesian estimation dating of lithic surface collections. *Journal of Archaeological Method and Theory* 22: 559–83. http://dx.doi.org/10.1007/s10816-013-9198-z
FULFORD, M. & N. HOLBROOK. 2011. Assessing the contribution of commercial archaeology to the study of the Roman period in England, 1990–2004. *The Antiquaries Journal* 91: 323–45. http://dx.doi.org/10.1017/S0003581511000138
Geophysics and the Landscape of Hertfordshire n.d. Available at: https://hertsgeosurvey.wordpress.com/ (accessed 1 July 2015).
GREEN, S., A. BEVAN & M. SHAPLAND. 2014. A comparative assessment of structure from motion methods for archaeological research. *Journal of Archaeological Science* 46: 173–81. http://dx.doi.org/10.1016/j.jas.2014.02.030
HANITZSCH, T. 2007. Deconstructing journalism culture: toward a universal theory. *Communication Theory* 17: 367–85. http://dx.doi.org/10.1111/j.1468-2885.2007.00303.x
HASELGROVE, C. & T. MOORE (ed.). 2007. *The later Iron Age in Britain and beyond*. Oxford: Oxbow.
HAZELTON, M.L. & T.M. DAVIES. 2009. Inference based on kernel estimates of the relative risk function in geographical epidemiology. *Biometrical Journal* 51: 98–109. http://dx.doi.org/10.1002/bimj.200810495
HeritageTogether n.d. Available at: http://heritagetogether.org/ (accessed 1 July 2015).
HERRMANN, J.T. 2013. Three-dimensional mapping of archaeological and sedimentary deposits with ground-penetrating radar at Saruq al-Hadid, Dubai, United Arab Emirates. *Archaeological Prospection* 20: 189–203. http://dx.doi.org/10.1002/arp.1456
HUGGETT, J. 2014. Promise and paradox: Accessing open data in archaeology, in C. Mills, M. Pidd & E. Ward (ed.) *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities*. Sheffield: HRI Online Publications.
HUSON, D.H. & D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–67. http://dx.doi.org/10.1093/molbev/msj030
JEFFREY, S., J. RICHARDS, F. CIRAVEGNA, S. WALLER, S. CHAPMAN & Z. ZHANG. 2009. The Archaeotools project: faceted classification and natural language processing in an archaeological context. *Philosophical Transactions of the Royal Society A* 367: 2507–19. http://dx.doi.org/10.1098/rsta.2009.0038
JOHNSON, M. 2011. On the nature of empiricism in archaeology. *Journal of the Royal Anthropological Institute* 17: 764–87. http://dx.doi.org/10.1111/j.1467-9655.2011.01718.x
KANSA, E.C., S.W. KANSA & E. WATRALL. 2011. *Archaeology 2.0: new approaches to communication and collaboration*. Los Angeles (CA): Cotsen Institute of Archaeology Press.
KANSA, E.C., S.W. KANSA & B. ARBUCKLE. 2014. Publishing and pushing: mixing models for

communicating research data in archaeology. *International Journal of Digital Curation* 9(1): 57–70. http://dx.doi.org/10.2218/ijdc.v9i1.301
KEAY, S., G. EARL, S. HAY, S. KAY, J. OGDEN & K.D. STRUTT. 2009. The role of integrated geophysical survey methods in the assessment of archaeological landscapes: the case of Portus. *Archaeological Prospection* 16: 154–66. http://dx.doi.org/10.1002/arp.358
KELSALL, J.E. & P.J. DIGGLE. 1995. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine* 14: 2335–43. http://dx.doi.org/10.1002/sim.4780142106
KINTIGH, K.W., J.H. ALTSCHUL, M.C. BEAUDRY, R.D. DRENNAN, A.P. KINZIG, T.A. KOHLER, W.F. LIMP, H.D.G. MASCHNER, W.K. MICHENER, T.R. PAUKETAT, P. PEREGRINE, J.A. SABLOFF, T.J. WILKINSON, H.T. WRIGHT & M.A. ZEDER. 2014. Grand challenges for archaeology. *American Antiquity* 79: 5–24. http://dx.doi.org/10.7183/0002-7316.79.1.5
KVAMME, K., E. ERNENWEIN, M. HARGRAVE, T. SEVER, D. HARMON & F. LIMP. 2006. New approaches to the use and integration of multi-sensor remote sensing for historic resource identification and evaluation. SERDP Report. Available at: http://cast.uark.edu/home/research/geophysics/multi-sensor-data-fusion-for-historic-resource-identification/data-fusion-final-report (accessed 1 July 2015).
The Megalithic Portal n.d. Available at: http://www.megalithic.co.uk/ (accessed 1 July 2015).
MENZE, B.H. & J.A. UR. 2012. Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proceedings of the National Academy of Sciences* 109: E778–E787. http://dx.doi.org/10.1073/pnas.1115472109
MicroPasts n.d. Available at: http://crowdsourced.micropasts.org/ (accessed 29 April 2015).
MILIĆ, M. 2014. pXRF characterisation of obsidian from central Anatolia, the Aegean and central Europe. *Journal of Archaeological Science* 41: 285–96. http://dx.doi.org/10.1016/j.jas.2013.08.002
NEUBAUER, W., C. GUGL, M. SCHOLZ, G. VERHOEVEN, I. TRINKS, K. LÖCKER, M. DONEUS, T. SAEY & M. VAN MEIRVENNE. 2014. The discovery of the school of gladiators at Carnuntum, Austria. *Antiquity* 88: 173–90. http://dx.doi.org/10.1017/S0003598X00050298
ORTMAN, S.G., M. VARIEN & T.L. GRIPP. 2007. Empirical Bayesian methods for archaeological survey data: an example from the Mesa Verde region. *American Antiquity* 72: 241–72. http://dx.doi.org/10.2307/40035813
POIRIER, N., F. HAUTEFEUILLE & C. CALASTRENC. 2013. Low altitude thermal survey by means of an automated unmanned aerial vehicle for the detection of archaeological buried structures. *Archaeological Prospection* 20: 303–307. http://dx.doi.org/10.1002/arp.1454
ROBERTS, B.W. & M. VANDER LINDEN. 2011. Investigating archaeological cultures: material culture, variability, and transmission, in B.W. Roberts & M. Vander Linden (ed.) *Investigating archaeological cultures: material culture, variability, and transmissions*: 1–21. New York: Springer.
rOpenSci n.d. Available at: https://ropensci.org/ (accessed 1 July 2015).
SALA, J. & N. LINFORD. 2012. Processing stepped frequency continuous wave GPR systems to obtain maximum value from archaeological data sets, in *Ground-Penetrating Radar (GPR), 2010 13th International Conference* (IEEE Proceedings). http://dx.doi.org/10.1109/ICGPR.2010.5550093.
SHENNAN, S. 2011. Descent with modification and the archaeological record. *Philosophical Transactions of the Royal Society B* 366: 1070–79. http://dx.doi.org/10.1098/rstb.2010.0380

SHENNAN, S.J. & J. CONOLLY. 2007. The origin and spread of Neolithic plant economies in the Near East and Europe [dataset]. Archaeology Data Service (collection 452). http://dx.doi.org/10.5284/1000093

SHENNAN, S., S.S. DOWNEY, A. TIMPSON, K. EDINBOROUGH, S. COLLEDGE, T. KERIG, K. MANNING & M.G. THOMAS. 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications* 4: 2486. http://dx.doi.org/10.1038/ncomms3486

SMITH, M.E. & P. PEREGRINE. 2012. Approaches to comparative analysis in archaeology, in M.E. Smith (ed.) *The comparative archaeology of complex societies*: 4–20. Cambridge: Cambridge University Press.

The Survey of Hillforts n.d. Available at: http://www.arch.ox.ac.uk/hillforts-atlas-survey.html (accessed 29 April 2015)

STODDEN, V., F. LEISCH & R.D. PENG (ed.). 2014. *Implementing reproducible research*. Boca Raton: Chapman & Hall/CRC.

UK Portable Antiquities Scheme n.d. Available at: https://finds.org.uk/ (accessed 29 April 2015).

VLACHIDIS, A. & D. TUDHOPE. 2011. Semantic annotation for indexing archaeological context: a prototype development and evaluation. *Metadata and Semantic Research Communications in Computer and Information Science* 240: 363–74. http://dx.doi.org/10.1007/978-3-642-24731-6_37

WEBLEY, L., M. VANDER LINDEN, R. BRADLEY & C. HASELGROVE (ed.). 2012. *Development-led archaeology in north-west Europe*. Oxford: Oxbow.

WHITE, R.H., C. GAFFNEY & V.L. GAFFNEY. 2013. *Wroxeter, the Cornovii and the urban process. Final report on the Wroxeter Hinterland Project 1994–1997. Volume 2 characterizing the city* (Journal of Roman archaeology supplementary series 68). Oxford: Archaeopress.

WOMBLE, W.H. 1951. Differential systematics. *Science* 114: 315–22. http://dx.doi.org/10.1126/science.114.2961.315

WOODMAN, P.E. & M. WOODWARD. 2002. The use and abuse of statistical methods in archaeological site location modelling, in D. Wheatley, G. Earl & S. Poppy (ed.) *Contemporary themes in archaeological computing*: 22–27. Oxford: Oxbow.

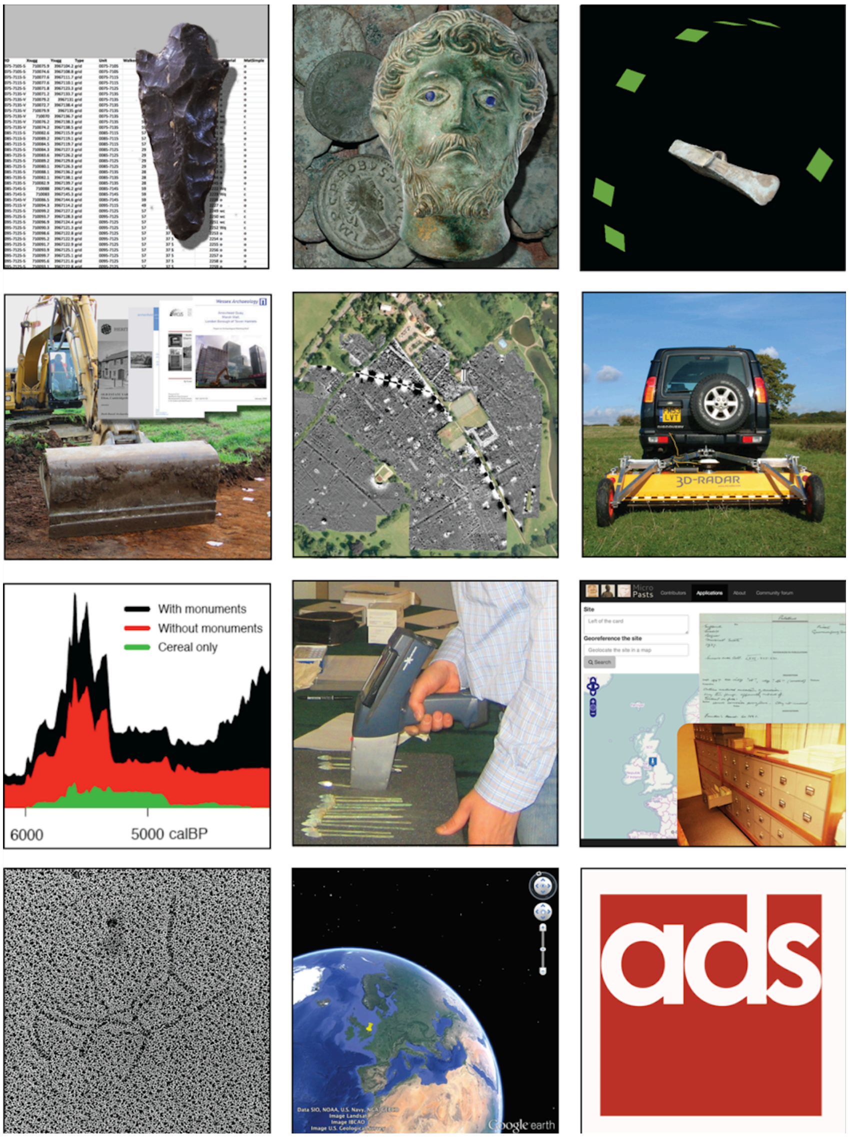

Figure 1. Archaeological evidence in flood, top left to bottom right: people's incidental spreadsheets as digital natives; finds from the UK Portable Antiquities Scheme; 3D modelling via structure-from-motion; digital 'grey literature' from developer-led archaeology; large-scale community-led geophysics; vehicle-towed GPR; radiocarbon dates 'as data'; crowd-sourced archival transcription; ancient and modern DNA; georeferencing via earth viewers; open data; and digital repositories.