



This PDF is a simplified version of the original article published in Internet Archaeology under the terms of the Creative Commons Attribution 3.0 (CC BY) Unported licence. Enlarged images, models, visualisations etc which support this publication can be found in the original version online. All links also go to the online original.

Please cite this as: Aitchison, K. 2024 Big Data and Lots of Data are Not the Same Things: Small data sources in the social science of archaeologists, Internet Archaeology 65. <https://doi.org/10.11141/ia.65.3>

Big Data and Lots of Data are Not the Same Things: Small data sources in the social science of archaeologists

Kenneth Aitchison

'Big Data' is a misunderstood but technically definable concept; it is large volume, unstructured, heterogenous information, that is being generated at a high velocity - which normally means it is being constantly updated in real time. And, in many situations, this is very difficult to work with; an analogy might be to compare the curated content of Netflix, which is a (very) large but structured data set, with the unending chaos of YouTube content, which is Big Data. The first part of this paper considers big data, and how it is defined and used, with the second part considering how large amounts of small data can be used in sociology as applied to archaeology.

1. Background

While "Big data" has had a [Wikipedia entry](#) since 2010, the phrase reached a threshold of conventional acceptance when it first appeared in the Oxford English Dictionary in 2013. The phrase was added to the OED in the same batch of new vocabulary as some other technology-related terms: "crowdsourcing, e-reader, mouseover, redirect (the noun), and stream (the verb)", as well as "the noun and verb tweet (in the social networking sense)" (Gibson [2013](#)), and is defined as "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges" (OED [2023](#)).

Since entering widespread use, the defining properties or dimensions of big data have become accepted as the "3Vs" – volume, variety and velocity, a concept first introduced by Laney ([2001](#)). Volume refers to the amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing – now normally meaning in real time. The most important consideration of this is that all three properties are magnified – it does not just mean an increase in volume (more data), but this also relates to variety (more sources of data, which are both simultaneous and different) and velocity (the speed at which the data are acquired – not a static dataset, but one that is continually expanding).



Decades before this, there were discussions about data in historical research that led to the first published uses of the phrase 'big data' in 1980, in an article by the sociologist Charles Tilly –

“none of the big questions has actually yielded to the bludgeoning of the big-data people” (Tilly [1980](#), 8)

and separately

"the American reputation for Big Data and bigger research teams has been greatly exaggerated" (Tilly [1980](#), 21)

The word 'big' isn't necessary in Tilly's first sentence. The quote is not describing the magnitude of the data, but is emphasizing the concept – big questions and big data. Here, Tilly was responding to the historian Lawrence Stone's ([1979](#), 6) critique of the use of quantitative methods in historical research to generate lots of data and attempts to make it a 'science':

[their] "great enterprises are necessarily the result of team-work, rather like building the pyramids: squads of diligent assistants assemble data, encode it, programme it, and pass it through the maw of the computer, all under the autocratic direction of a team-leader. The results cannot be tested by any of the traditional methods since the evidence is buried in private computer-tapes, not exposed in published footnotes. In any case the data are often expressed in so mathematically recondite a form that they are unintelligible to the majority of historical profession. The only reassurance to the bemused laity is that the members of this priestly order disagree fiercely and publicly about the validity of each other's findings."

So Stone wasn't actually discussing big data, but was flagging up what would later be considered to be information overload.

Ultimately, the first documented use of the term "big data", in the sense that it is used in 2023, was in Cox and Ellsworth ([1997](#)) "Application-controlled demand paging for out-of-core visualization", which discussed big data as an issue from the perspective of seeking to manipulate and visualize data sets that could be too large to be stored and so simultaneously accessed, recognising that this was a particular problem in computational fluid dynamics.

2. Wicked Problems

Big data – the data - are wonderful. Big data – the concept - is difficult to work with. To paraphrase Raphael Mokades, Founder and CEO of Rare Recruitment, it is difficult because it needs machine learning, that means one third of the effort is in getting the data, one third is in cleaning the data, and one third is in training the system. You need to know what the wicked problem you are trying to address is, in order to make use of big data. And then identify the patterns within the big data that contribute to tackling that problem.

To think about the use and application of big data - is there a relationship between big data, wicked questions and wicked problems?



Wicked problems are very complicated and don't have an obvious answer. Wicked problems can be recognized by their complexity, inter-relatedness, ongoing nature, even by their lack of definition, that there is no stopping rule, and every problem is a symptom of another, this means they are incomplete and contradictory and have changing requirements that are often difficult to recognize.

Given the cascading, real-time nature of big data, this makes it possibly the only broad source of information that can be applied to wicked problems. And so, to reverse that argument is big data ever used to solve tame (i.e. non-wicked) problems?

If it's a tame problem, it doesn't need big data. It's not about whether it is a difficult problem or not – but tame problems can be clearly stated, have defined goals and once they are solved, they stay solved. So archaeology actually very rarely has wicked problems, and so very rarely is dealing with big data.

3. Small Data and the Sociology of Archaeologists

Here's the story of this researcher's journey into hard tame problems and lots of small data.

I graduated from the University of Edinburgh in 1992, with a decent understanding of the Iron Age in continental Europe and the later prehistory of south-west Asia, but absolutely no knowledge of how to get a job – or even if there were any jobs – in archaeology in Scotland. Immediately after graduation I was fortunate enough to be offered work by one of my former lecturers, doing archaeological excavation and reconstruction in Cyprus at the [Lemba Archaeological Research Centre](#), which was followed up by work as a research assistant at the University of Edinburgh and then fieldwork the next year in Syria at [Jerablus-Tahtani](#). So this had given me some experience, in recording and in holding a position of responsibility, supervising others. But it was seasonal, and was not financially rewarding.

In the winter of 1993-94, I was back in Scotland and had found work as a field archaeologist, digging for Kirkdale Archaeology at Stirling Castle. Where I was being paid, but I didn't know how long it would last, or what I might do next (beyond return to Syria, which indeed I did in the summers of 1994 and 1995). Ultimately, I ended up working for Kirkdale Archaeology on a few projects as I spent three good years oscillating between south-west Asia in the summers and Scotland in the winters (in terms of workplace temperatures – that was the wrong way round!).

I didn't have access to a source of information about who might possibly employ archaeologists in Scotland. So I did a little bit of research, first of all to find out where there were jobs in Scottish archaeology, and I phoned all of the potential employers in Scottish archaeology just to ask them the simplest question of – how many archaeologists work for you?



That led to the modest publication in a small article entitled 'Want a Job?' (Aitchison [1997](#)) in *Scottish Archaeological News*, the newsletter of the Council for Scottish Archaeology (now [Archaeology Scotland](#)). I didn't realise it at the time (despite having studied some undergraduate sociology) – what I was doing related to the sociology of professions, a well-defined academic discipline with the key publications having been Durkheim's *The Division of Labour in Society* ([1893](#)), Weber's *The Protestant Ethic and Spirit of Capitalism* ([1905](#)) and an essential article by Harold Wilensky ([1964](#)) called 'The Professionalization of Everyone?'

4. Profiling the Profession

That little article on work in Scottish archaeology generated some unexpected interest in England, and in 1998 I was invited by English Heritage (now [Historic England](#)), the [Council for British Archaeology](#) and the IFA – then the Institute of Field Archaeologists, now the [Chartered Institute for Archaeologists](#) – to do something on a much larger scale: to survey all the employers of archaeologists in the whole United Kingdom. This would pull together a picture of the scale and nature of the sector's entire workforce, and was something that had never previously been done comprehensively.

That project was called *Profiling the Profession*, and it resulted in a report published in 1999 (Aitchison [1999](#)), which highlighted information that employers, educators and individual archaeologists were interested in – how many people work in the sector? What do they do, where and for whom? And always the most interesting – how much do they get paid?

This had involved posting paper questionnaires with stamped addressed envelopes for responses to be returned in, and resulted in a snapshot picture of the sector on the census date of 16th March 1998.

As time passed, those data became increasingly outdated, but there was now a recognised demand for this information. So the exercise was repeated, approximately five years later to produce *Profiling the Profession 2002-03* for the Cultural Heritage National Training Organisation, working with Rachel Edwards (Aitchison and Edwards [2003](#)), and again after another five years - *Profiling the Profession 2007-08*, this time the client was the Institute of Field Archaeologists and the work was again done with Rachel Edwards (Aitchison and Edwards [2008](#)). These repeat exercises meant pulling together comparable (but not identical) data sets – as there were now different potential respondents, and slightly different questions, allowing for results that could be presented as both snapshots and longitudinal (time series) analyses.

The three substantial and data-heavy reports from these projects then formed the core of my PhD by research publications, *Demand and Supply in UK archaeological employment 1990-2010*, which I was awarded by the University of Edinburgh in 2011 (Aitchison [2011](#)).



5. Discovering the Archaeologists of Europe 2006-08

The 2007-08 UK report was produced as part of a larger, European project. Colleagues elsewhere in Europe had recognised the value of the UK work, and wanted to emulate this in their own countries. This was discussed at a series of annual meetings of the [European Association of Archaeologists](#), and a consortium was established to carry out research in twelve European countries, together with the European Association of Archaeologists, funded by the [European Commission](#) through the Leonardo da Vinci II funding stream.

Politically, the European Commission was interested in supporting work like this because direction had been set in the *Lisbon Agenda*. This was a strategy launched at the European Council meeting of European Union leaders in Lisbon in March 2000, which (over ambitiously) aimed at making the EU the world's most competitive economy by 2010 (European Council [2000](#)). The concept of transnational mobility was key, as delivered through the EU's *Action Plan for Skills and Mobility* (European Commission [2002](#)) – ensuring that European citizens could train, work and live in countries of the European Union other than those that they were born in.

The project established a shared methodology, ensuring comparable data sets were collected through individual partner-run projects in their own countries, resulting in both [national and transnational reports](#).

This was a remarkable step forward – there had not been any previous, controlled and quantified attempt to compare the scale and nature of archaeological practice in different European countries.

6. Discovering the Archaeologists of Europe 2012-14

Discovering the Archaeologists of Europe had managed to capture data almost exactly at the point of the onset of the global financial crisis of 2007-08 and so before the Great Recession that followed. This meant that it caught archaeology at an economic high point, when a lot of people were working in development-led archaeology.

The project had been a success, and after a few years the original participants (and many others) were keen to update the information – and to show the way that archaeology had been affected by the Great Recession. So a new consortium was assembled, this time under the coordination of the [York Archaeological Trust](#), and funding was secured.

The European Commission programmes that this was linked to were driven by post-recession recovery objectives, particularly as defined under the *New Skills for New*



Jobs initiative (European Commission [2008](#)) and then driven forward by the *Europe 2020* strategy (European Commission [2010](#)).

Ultimately, the project team found we had bitten off more than we could chew. The consortium was really big, it involved 21 partners, in 20 countries, 18 of which were in the EU. The European Commission funded it, under the Lifelong Learning Programme stream, and that programme had never funded a project with so many partners. It ran into difficulties; differing expectations and attitudes of separate partners' representatives led to methodologies diverging., meaning it was difficult to present comparable data in areas other than nine defined core areas. and reporting was inconsistent. This led to the European Commission initiating a formal audit mid-project – which supported the team to reach a successful conclusion on time and on budget.

As well as producing a [transnational report](#) comparing the situations in the 20 partner countries, individual [national reports](#) were also produced, including *Archaeology Labour Market Intelligence: Profiling the Profession 2012-13* (Aitchison and Rocks-Macqueen [2013](#)), the 4th in the series of UK reports.

7. Profiling the Profession Update

The organisations that had been partners in the Discovering the Archaeologists of Europe projects were keen to repeat the exercise – the European Association of Archaeologists established a Discovering the Archaeologists of Europe Community to facilitate ongoing discussions, but this was hampered by there being no easily accessible source of European Commission funding that could contribute to research like this.

Similarly, there was no easily accessible source of UK funding; previously, funding had been secured from Historic England and the other national heritage agencies as match funding to complement the European Commission funds received in 2007-08 and 2012-13. This all changed dramatically with the onset of the COVID-19 pandemic in 2020, and Historic England provided grant funding in June 2020 under the *Historic England Emergency Response Fund*.

The Project Summary that justified this application for 'emergency' funding identified the need for the project at being time-critical:

This year, the archaeological sector is facing two, once in a lifetime, events that could alter the sector beyond anything seen before - the Covid-19 pandemic and Brexit. Without knowing where the sector was before these events we will have no idea of what the effects will be, nor how to identify priorities for future.

This project will capture the time-sensitive critical data required to understand how these events will impact upon the sector. This will include capturing data on employment conditions, staff qualifications, diversity and training issues. Crucially – while the data on employment will be retrospective, describing the situation as it was, skills issues will be addressed in a forward-looking way – identifying what the sector feels it needs now and in the future.



8. Profiling the Profession 2020

This funding then allowed a new iteration of *Profiling the Profession* to run in 2020-21, gathering information from employer organisations as well as from individual archaeologists past, present and future – former archaeologists, people in work and archaeology students.

This reported (Aitchison *et al.* [2021](#)) with the data in reconfigurable and interrogatable tables alongside the authors' analysis, a step forward from the predecessor static PDF reports.

This project generated many data points - 400,000 data points for individuals and their jobs; 40,000 data points about organisations – and while they may be many, they are not big data - these are structured, interrogable small data.

Headlines from the research included figures for how many people were working in archaeology, who they worked for and how these numbers compared with other industries. Updated data about archaeologists – the industrial profile giving information about workers' age, gender, ethnicity, disability status, qualifications, salaries, and about their employers. A particularly noteworthy point emerged from the longitudinal time-series data - over time, the [gender balance in UK professional archaeology](#) had changed from being disproportionately male-dominated to match that of wider UK workforce.

9. Using Small Data

A further example of using an externally generated large, but 'small' data set to understand more about archaeology as a discipline – the UK's Higher Education Statistics Agency (HESA) now allows detailed extraction of data about subject studied, which allows for much more informative data to be extracted about the numbers of people studying archaeology; formerly (pre-2014 datasets), archaeological sciences could not be separated from forensic science.

Over the last five years, the numbers of archaeological students have been quite consistent, around 5,000 higher education students in total at any given time (extracted from [HESA: What do HE students study?](#)). This produced a calculated estimate of 2,165 graduates per annum from 2014-2020 (Aitchison and Dore [2022](#)). Comparing this figure with the data in Profiling the Profession 2020 on the total number of archaeologists in work, that means the number of graduates receiving degrees in archaeology is roughly the equal of one third of the total workforce, which (if all those graduates were to enter the archaeological workforce) would mean there would be a wholly new workforce every three years.

But that doesn't happen – and so that means we have data that can help investigate why those archaeology graduates do not work in professional archaeology.



10. Extrapolating from Small Data to Think about Lots of Money

We use data gathered on the number of people working in archaeology, together with data on salaries, cross-checked against published company reports to calculate the value of professional archaeology – in terms of the amount of money being spent on work in the sector.

The figures produced in the series of *State of the Archaeological Market* reports produced for [FAME](#) include estimates of the total amount of funding being received by commercial archaeology each year. The 2021 figures indicate that the value of commercial archaeology in the UK in 2021 was £247m (Aitchison and Rocks-Macqueen [2022](#), 16); the market in the USA is more than three times as big – Christopher Dore first identified in 2018 that the market for private-sector archaeological compliance work is worth more than a billion dollars (Dore [2018](#), 229). Around the world – more than four billion dollars is spent every year on archaeological practice (Aitchison and Dore [2023](#)). And such calculations, and their value to business and political planning, all rest upon exercises in gathering small data.

11. Conclusions

During the COVID-19 pandemic, confronted with the wicked problem of a new disease that was spreading rapidly with new variants emerging, and without certainty in treatment regimes, the United Kingdom's National Health Service was able to harness the power of big data to identify patterns with positive outcomes. They successfully linked and amalgamated diverse treatment trials data to reveal some unexpected patterns that would have remained hidden without the capabilities of big data analysis - an approach that led to a serendipitous discovery that risk of death declines when steroids are used in treatment of coronavirus (NHS Digital [2021](#)).

Archaeology doesn't deal with problems like this. Archaeology and heritage management rarely interact with big data, principally because big data are being generated at velocity (in real time). It was coincidental that the study of the UK archaeological workforce got a (substantial) small data boost as a consequence of COVID-19 – and gathering and using lots of small data is extremely valuable for archaeology.

Bibliography

Aitchison, K. 1997. 'Want a job?', *Scottish Archaeological News* 23, 4-5. <https://doi.org/10.6084/m9.figshare.20651808>



Aitchison, K. 1999. Profiling the Profession: a survey of archaeological jobs in the UK. York, London and Reading: Council for British Archaeology, English Heritage and Institute of Field Archaeologists. <https://profilingtheprofession.org.uk/wp-content/uploads/sites/5/Profiling-the-Profession-1997-82.pdf> <https://doi.org/10.5281/zenodo.7840412>

Aitchison, K. 2011 *Demand and Supply in UK archaeological employment 1990-2010*, Edinburgh Research Archive. <https://era.ed.ac.uk/handle/1842/33304>

Aitchison, K. and Dore, C. 2022 'We don't have a JOBS problem, we have a CAREERS problem', *The Archaeologist* **117**, 18-19. <https://doi.org/10.6084/m9.figshare.21385146.v3>

Aitchison, K. and Dore, C. 2023 'Billions of Dollars: Calculating the Size of the Heritage Compliance Sector', Conference paper presented at Society for American Archaeology 88th Annual Meeting, 1 April 2023.

Aitchison, K. and Edwards, R. 2003. Archaeology Labour Market Intelligence: Profiling the Profession 2002/03. Bradford: CHNTO. <https://profilingtheprofession.org.uk/wp-content/uploads/sites/5/Profiling-the-Profession-2002-3.pdf>

Aitchison, K. and Edwards, R. 2008 *Discovering the Archaeologists of Europe: United Kingdom. Archaeology Labour Market Intelligence: Profiling the Profession 2007-08*, Reading: Institute of Field Archaeologists. <https://profilingtheprofession.org.uk/wp-content/uploads/sites/5/Profiling-the-Profession-2007-8.pdf>

Aitchison, K., German, P. and Rocks-Macqueen, D. 2021 *Profiling the Profession 2020*. <https://profilingtheprofession.org.uk/> <https://doi.org/10.6084/m9.figshare.14333387>

Aitchison, K. and Rocks-Macqueen, D. 2013 *Archaeology Labour Market Intelligence: Profiling the Profession 2012-13*. <https://landward.eu/wp-content/uploads/2013/10/Archaeology-Labour-Market-Intelligence-Profiling-the-Profession-2012-13.pdf> <https://doi.org/10.5284/1024571>

Aitchison, K. and Rocks-Macqueen, D. 2022 *State of the Archaeological Market 2021*. <https://famearchaeology.co.uk/wp-content/uploads/2022/06/State-of-the-Archaeological-Market-2021-2.pdf> [Last accessed: 18 April 2023].

Cox, M. and Ellsworth, D. 1997 'Application-controlled demand paging for out-of-core visualization', *VIS '97: Proceedings of the 8th conference on Visualization '97*, Washington DC, IEEE Computer Society Press. 235–244. <https://doi.org/10.1109/VISUAL.1997.663888>

Dore, C. 2018 'Business challenges for the twenty-first century: the next forty years of private heritage management' in Francis P. McManamon (ed) *New Perspectives in Cultural Resource Management*, Abingdon and New York: Routledge. 229-239. <https://doi.org/10.4324/9781317327349-14>



Durkheim, E. 1893 *The Division of Labour in Society*, [English translation published 1933 New York: The Free Press].

European Commission 2002 *Action Plan for Skills and Mobility*. <https://eur-lex.europa.eu/EN/legal-content/summary/action-plan-for-skills-and-mobility.html> [Last accessed: 8 March 2023].

European Commission 2008 *New Skills for New Jobs. Anticipating and matching labour market and skills needs*, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0868:FIN:EN:PDF> [Last accessed: 17 April 2023].

European Commission 2010 *Europe 2020: a European strategy for smart, sustainable and inclusive growth*. <https://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf> [Last accessed: 17 April 2023].

European Council 2000 *Presidency Conclusions: Lisbon European Council 23 and 24 March 2000*. https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm [Last accessed: 8 March 2023].

Gibson, K. 2013 *A heads up for the June 2013 OED release*. <https://public.oed.com/blog/june-2013-update-a-heads-up-for-the-june-2013-oed-release/>

Laney, D. 2001 *3D data management: Controlling data volume, velocity and variety*, META group research note, February 6. <https://web.archive.org/web/20120304154148/http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [Last accessed: 28 February 2023].

Mokades, R. 2022 *Contribution to Institute of Business Ethics Webinar Ethics of AI and Recruitment*, 4th August 2022. <https://www.ibe.org.uk/resource/ethicsofusingaiinrecruitment-webinarrecording.html> [Last accessed: 18 April 2023].

NHS Digital 2021 *Lifesaving coronavirus treatment discovered by study using data from NHS Digital*. <https://digital.nhs.uk/news/2021/lifesaving-coronavirus-treatment-discovered-by-study-using-data-from-nhs-digital> [Last accessed: 18 April 2023].

Oxford English Dictionary 2023 'big data, n.', September 2023. <https://doi.org/10.1093/OED/5720738787>

Stone, L. 1979 'The Revival of Narrative: Reflections on a New Old History', *Past and Present* **85** (Nov. 1979), 3-24. <https://doi.org/10.1093/past/85.1.3>

Tilly, C. 1980 *The Old New Social History and the New Old Social History*, CRSO Working Paper 218, Ann Arbor: Centre for Research on Social Organization,



University of
Michigan. <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/50992/218.pdf?sequence=1&isAllowed=y> [Last accessed: 28 February 2023].

Weber. M. 1905 *The Protestant Ethic and Spirit of Capitalism*, English translation published 1930 London: Allen and Unwin.

Wilensky, H.L. 1964 'The Professionalization of Everyone?', *American Journal of Sociology* **70**(2), 137–58. <https://doi.org/10.1086/223790>