Journal of Archaeological Science xxx (2017) 1–21

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Geospatial Big Data and archaeology: Prospects and problems too great to ignore☆

Mark D. McCoy

Department of Anthropology, Southern Methodist University, P.O. Box 750336, Dallas, TX, 75275-0336, USA

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form
27 May 2017
Accepted 1 June 2017
Available online xxx

Keywords:
Geospatial
Big Data
Spatial technology
Cyberinfrastructure
Data science

ABSTRACT

As spatial technology has evolved and become integrated in to archaeology, we face a new set of challenges posed by the sheer size and complexity of data we use and produce. In this paper I discuss the prospects and problems of Geospatial Big Data (GBD) – broadly defined as data sets with locational information that exceed the capacity of widely available hardware, software, and/or human resources. While the datasets we create today remain within available resources, we nonetheless face the same challenges as many other fields that use and create GBD, especially in apprehensions over data quality and privacy. After reviewing the kinds of archaeological geospatial data currently available I discuss the near future of GBD in writing culture histories, making decisions, and visualizing the past. I use a case study from New Zealand to argue for the value of taking a data quantity-in-use approach to GBD and requiring applications of GBD in archaeology be regularly accompanied by a Standalone Quality Report.

1. Introduction

Archaeology has long recognized that spatial location is a core variable in our field (Spaulding, 1960). Today, we create, use, and share geospatial archaeological data on an unprecedented scale. In a recent paper, Bevan outlined many of the challenges we face with "floods of new evidence about the past that are largely digital, *frequently spatial*, increasingly open and often remotely sensed" (Bevan, 2015:1473, emphasis added). As our locational datasets grow, and become more accessible, so does apprehension about data quality, privacy (especially the protection of the locations of archaeological sites), and how best to manage large and growing geospatial data. At the same time, we have amassed such large databases that, on some topics, it would be disingenuous to claim we do not yet have enough data (Bevan, 2015:1477).

There is a growing literature in archaeology aimed at bringing attention to how we can best use technology (Kintigh, 2006; Snow et al., 2006) to achieve our larger disciplinary goals (e.g., Kintigh et al., 2014). The need for larger and more integrated geospatial data and analyses cross-cuts virtually all of our goals and aspirations as a science (Table 1). These require us to produce data and results that are scientific (testable, replicable), authentic (a faithful representation of the archaeological record and the human past), and ethical (protects cultural resources). To that end, I am guided in this paper by three questions: 1) What kinds of geospatial data are available today? 2) How will larger and more accessible geospatial databases shape the near future of archaeology? And, using a case study from New Zealand, I examine the question, 3) What can we do now about apprehensions regarding data quality, privacy, and the growing size of archaeological geospatial datasets?

These questions – what data is available, what will be the impacts of larger and more accessible data, and what can we do mitigate our concerns about data – exemplify current debates about Big Data in general, and Geospatial Big Data specifically. Geospatial Big Data (GBD) can be broadly defined as data sets that include locational information and exceed the capacity of widely available hardware, software, and/or human resources. Before we go further, it is important to note that as of today, nearly all archaeological datasets fall short of being defined as GBD since the volume of data we work with rarely outstrips the capacity of available resources; with the exception of remotely sensed data (satellite imagery, lidar). But, while the volume of archaeological geospatial datasets are currently manageable, there are at least two good reasons we should begin to think about our geospatial datasets as GBD. First, due to the fragmentary nature of archaeological material evidence we are compelled to work with a broad variety of sources of data, to code complex contextual information in to a

☆ The special issue was handled by Meghan C.L. Howey and Marieka Brouwer Burg.

E-mail address: mdmccoy@smu.edu.

http://dx.doi.org/10.1016/j.jas.2017.06.003
0305-4403/

Table 1
Grand Challenges for Archaeology and the Need for Larger Geospatial Data and Analyses. Kintigh et al. (2014:5) summary of the "most important scientific challenges" for archaeology highlight a number of areas where the need for larger geospatial datasets and analyses is paramount. The purpose of this paper is identify how we are currently building, using, and sharing geospatial data; what larger geospatial datasets will mean for the near future; and suggest ways we may overcome apprehension over the adequacy of large geospatial datasets – otherwise known as Geospatial Big Data – that will be necessary to meet our disciplinary goals.

General Topic	Examples of the Need for Larger Geospatial Datasets and Analyses
Emergence, Communities, and Complexity	"Archaeological data on cities range from small architectural details and short-lived cities to broad patterns of heterogeneous urban textures covering many square kilometers and presenting a historical depth of millennia. Consequently, characterizing long-term urban fabrics and animating associated behaviors via computational modeling *requires enormous data archives and substantial computational infrastructure*" (Kintigh et al., 2014:10, emphasis added). "Conflict is notoriously difficult to identify and quantify through archaeological remains … *more systematic and large-scale analyses are certainly necessary*." (Kintigh et al., 2014:10, emphasis added). "Inequality can be systematically inferred through studies of landscape, monuments, residences, and mortuary remains … *Quantitative dynamic modeling* to emplace general models of sociopolitical change in specific prehistoric and historical settings … will be critical to our success." (Kintigh et al., 2014:9, emphasis added).
Resilience, Persistence, Transformation, and Collapse	"The archaeological record is replete with examples of the rise and fall of communities of all scales … With *recent advances in the quantity and quality of archaeological and historical studies, we can uncover robust patterns in societal collapses over time and space*." (Kintigh et al., 2014:11, emphasis added).
Movement, Mobility, and Migration	"Typically, archaeologists have explored human mobility through a case-study approach based on archaeological and ancillary data from small-scale research projects. However, we also see the need for *regional- and continental-scale studies that match the scale of the problem to the scale of particular interactions*." (Kintigh et al., 2014:13, emphasis added).
Cognition, Behavior, and Identity	"…. how did humanity arise?... a *massive body of emerging data* are critical to resolving this question…" (Kintigh et al., 2014:15, emphasis added). "Tracking and evaluating localized arrangements and reconfigurations … *necessitates extensive investments in digital spatial datasets* that incorporate LiDAR, geophysical, and other three-dimensional data that allow virtual exploration and analysis." (Kintigh et al., 2014:15, emphasis added).
Human Environment Interaction	"How do humans perceive and react to changes in climate and the natural environment over short- and long-terms?... The challenge is to move from case or regional studies to larger scale comparative research, and to learn how to make generalizable statements about how people make choices that draw on universal biases in cognition … [this] will require *making data from relatively small field projects widely accessible and increasing current technological capabilities* to allow for studies of human-environment interaction to increase in scope and complexity" (Kintigh et al., 2014:18–19, emphasis added).

digital format, and to interpolate trends across time and space using sparse data. These types of problems (variety, veracity, visualization) mirror issues raised by Big Data (see also Huggett, 2016). Second, from the perspective of data science our data are probably best classified as 'embryonic' Geospatial Big Data in that they are likely to grow extremely large in volume in the future. We have the opportunity now to shape our growing geospatial datasets before it becomes necessary to come up with specialized solutions for common tasks. It is also important to note that the problem of best practices regarding geospatial data is well-known to the subfield of geospatial archaeology, as well as archaeology that engages with computer and data science. As the science and technology dealing with GBD evolves, the hyper-technical side of archaeology is more important than ever. But, since GBD is already influencing how we write culture history, visualize our research, and participate in public discourse about science and heritage, I felt it is timely to review and comment on this topic for a broad audience in as non-technical terms as is reasonable.

2. Geospatial big data and archaeology

Today, we refer to any information, "of or relating to the relative position of things on the earth's surface" as *geospatial* data (Collins English Dictionary). Geospatial Big Data (GBD) is geospatial data that exceeds the capacity of widely available resources (i.e., hardware, software, human resources) and requires specialized effort to work with. Applied research in GBD tends to be driven by the perceived economic benefit of mining data to reveal spatial relationships that make businesses more cost efficient, enhance insight in to customer's behavior, and help industry make better

Please cite this article in press as: McCoy, M.D., Geospatial Big Data and archaeology: Prospects and problems too great to ignore, Journal of Archaeological Science (2017), http://dx.doi.org/10.1016/j.jas.2017.06.003