Digital heritage research re-theorised:

Ontologies and epistemologies in a world of big data

Dr Chiara Bonacchi, Lecturer in Heritage, Division of History and Politics, University of Stirling

Marta Krzyzanska, PhD Candidate, Department of Anthropology and Archaeology, University of Cambridge and Research Assistant, Division of History and Politics, University of Stirling

Abstract:

This article provides the first theoretical treatment of the ontologies and epistemologies of digital heritage research at the time of the interconnected and social web, based on extensive empirical and analytical investigation. We draw on observations and concepts developed while conducting the first study of public experiences of the past that utilised big data – over 1.4 million Facebook posts, comments and replies – to revisit or generate new theory from the ground up. We expect that this will help scholars from a range of fields in the humanities, social and computing sciences who are interested in undertaking digital heritage research to understand the deeper implications of their work, the complexities and limitations of the knowledge they create, and its value in exposing the processes via which heritage is made and assessed.

Keywords: digital heritage, ontology, epistemology, activism, big data.

1. Introduction

1.1. Aims

This article proposes a theorisation of digital heritage research at the time of the collaborative web (Ruppert, Lowe and Savage 2013; Housley et al. 2014), arguing that, in order to adequately engage with digital heritage research from a conceptual point of view, it is vital to understand technicity - 'the technology considered in its efficacy or <u>operating functioning</u>' (Hoel and van der Tuin 2013, 187; emphasis by the authors). Furthermore, as today's web is increasingly interconnected and generating deluges of data objects, we cannot prescind from examining digital heritage it in the context of big data. For these reasons, we will discuss theoretical aspects of digital heritage research through a practice-based approach informed by the first study to use big data with the aim of exploring public perceptions and experiences of the past in the present (Bonacchi, Altaweel and Krzyzanska 2018; this work was undertaken by the two authors together with Mark Altaweel).

The field of digital heritage started to take shape in the 1990s but has expanded more substantially in the last ten years, focussing especially on the roles and impact of digital technologies in museum and gallery contexts (for foundational work on the topic, see Fahy 1995; Anderson 1997; Arvanitis 2002; Galani and Chalmers 2002; Parry 2007a, 2007b, 2010; Cameron and Kenderdine 2010). Since then, published material in this area has been concerned primarily with digital engagement with the past or the digitisation of analogue resources (see all the citations given above and, more recently, Kidd 2011; Geismar 2012; Ridge 2014; Were 2015; King, Stark and Cooke 2016; Díaz-Andreu 2017; Jones et al. 2017), whereas literature dealing specifically with digitally-born and digitally-enabled research remains very limited. Moreover, the few existing studies of this kind tend to take one of two directions. A first group reflects on overarching problems - such as the implications of the

2

social web for heritage scholars - without being directly grounded in empirical work (e.g., Bonacchi and Moshenska 2015; Perry and Beale 2015; Richardson 2018). Contrasting with this, in a second group of publications, digital tools are leveraged to conduct digital heritage or humanities research, without discussing the underlying ontological and epistemological issues at stake (Marwick 2013; Wevers and Verhoef 2017; Zuanni 2017). Here we argue that, whilst each of these two approaches is legitimate, neither of them is sufficient on its own to explore the complexities of advancing heritage theory at the time of the data deluge. We stress the need to delve into practice, in order to uncover the ethical, ontological and methodological shifts driven by web environments that are now for a large part controlled by private black boxes such as Facebook or Twitter.

In a world of big data, technicity is simultaneously part of both the methodology that is used and the subject being investigated (Rieder et al. 2015). How can we then even begin to grasp the impact of social media platforms for digital cultural engagement without knowing how networking sites are collecting information? How can we critique these practices if we do not have an in depth and first-hand understanding of how they function? How can we decolonise collections and disrupt dominant ways of organising and representing information, if we are not able to track the creation and circulation of data by mining, collecting and critiquing it? Interestingly, John Dewey's work on experiential learning and the importance of acquiring and consolidating knowledge 'by doing' has become one of the pillars of heritage interpretation scholarship, but it does not seem to be applied when it comes to digital heritage research (Dewey 1938; Silberman 2007). The same paradox can be highlighted for concepts such as material thinking (Carter 2004), co-production (Bonacchi 2018a) and digital DIY (Gauntlett 2015), which are popular but have not practically contributed to advance digital heritage researchers' direct engagement with technicity in any meaningful way.

This gap between theory and practice, emerged with the postmodernist turn of the 1990s, has been recently re-discussed in a range of fields - from heritage, to contemporary and public archaeology – with the aim to resolve it. Matsuda (2016), for example, notes the dichotomy between the economic and critical paradigms of public archaeology, while Olsen and Pétursdóttir (2017) have written on the topic of *Theory Adrift. The Matter of Archaeological Theorizing*, denouncing the issue of theory that is completely detached from the substance of practice and only exists in its formal essence. Rieder et al. (2015) also address a similar problem, noting the contrast between highly rhetorical and detail-devoid theory in research about big data, in contrast to a-theoretical and practice-centred big data research.

1.2. Case study

In this article, we will address the issue outlined above by focussing on a case study concerned with uses of the ancient past – from the Iron Age to the Early Medieval period – in Brexit-related micro-activism performed on Facebook (Bonacchi, Altaweel and Krzyzanska 2018). Following Marichal (2013), political micro-activism is defined as a kind of activism that takes place on social media and expresses people's political identities and 'hoped for' futures. Brexit is the process of deciding upon and implementing changes in the membership of the UK in the European Union (Clarke, Goodwin and Whiteley 2017; Gardner 2017; Bonacchi, Altaweel and Krzyzanska 2018; Fuchs 2018).[1]

In March and April 2017, the authors extracted all the posts, comments and replies published since 2013 on public Facebook pages that contained the word *Brexit* in their title.[2] Utilising R Free and Open Source Software (R), we obtained a *corpus* of over 1.4 million posts, comments and replies (henceforth referred to as *documents*) derived from 364 Facebook

pages. We then applied a set of 569 keywords (see Ancient Identities 2017 for a complete list of the keywords that were used) characterising people, places and ideas from 800BCE to 800CE, to locate a subset of 2529 documents where the Iron Age, Roman or Early Medieval pasts were mentioned (Table 1). The keywords were identified through desk research, initially informed by a workshop organised in November 2016 as part of the *Ancient Identities in Modern Britain* project (Ancient Identities 2016), funded by the Arts and Humanities Research Council, UK.

Table 1. Messages extracted from Facebook pages containing the term 'Brexit' in their title or description.

We categorised the metadata available for the Facebook pages we had extracted, to understand who set them up (whether it was an individual or a group; if the group had already been established offline and the shared values and qualities of its members – e.g., same local community, ethnicity, party affiliation, etc.), and their motivations for doing so (where they intending to provide a neutral space to share news and to debate, or did they wish to support a particular side?). With R and Python programming languages, we then analysed the sub-set of 2529 documents via term frequencies (counts of the recurrence of specific words inside a *corpus* of texts), term associations (useful to understand what terms are found more frequently next to the most recurring terms in a *corpus*), and topic modelling (a probabilistic method to identify the themes that are present in a large *corpus* of texts; Bonacchi, Altaweel and Krzyzanska 2018). Applying these methods, together with close reading, was useful to reveal how certain periods featured, and the narratives they were woven into.

Results showed that a parallel between the Roman Empire and the European Union appeared prominently in discussions about Brexit on public Facebook pages. This parallel is the centre of a number of oppositions - for example between civilisation and barbarism, cultural diversity and homogeneity, freedom and domination – around which three main myths, of origin, resistance and collapse, are constructed. These myths draw on much of the same repertoires of images and ideas from the ancient past, which are however utilised to express and back sometimes divergent positions. Findings also cast light on the long-term legacy of heritage-related scholarly thinking, filtered through formal education and mass media communication over the past few decades. In doing so, the research opens up questions about the impact of heritage experts and expertise on people's opinions regarding issues that are core to our societies (see also Bonacchi 2018b).

Whilst the research context and findings were presented in full in a previous publication (Bonacchi, Altaweel and Krzyzanska 2018), here we will reflect instead on the concepts and practices we developed to reach those results. Each of the following sections will revisit the process of designing and completing the study, and, by problematising technicity, it will build new theory blocks on the ontologies and epistemologies of digital heritage research from the ground up.

2. Digital heritage ontologies

2.1. Digital heritage as process and as (big) data

What is digital heritage and what does it mean to research it? In our study, we framed heritage as the processes and outcomes of engaging with elements of the past – material and immaterial – and attributing social and cultural meanings to them in the present (Smith 2006;

Harrison 2013). Here we define the sub-field of digital heritage as examining interactions of this kind that are enabled by the Internet and the outcomes of such processes (the footprints – including data - that are produced). We will consider both of these aspects in turn, starting from the former.

In her work on Internet ethnographies, Christine Hine (2015) explains how human activities can unfold across online and offline fields, arguing that, for this reason, ethnographic investigation undertaken online should be conceived as potentially interlinked with offline ethnography, rather than separate from it. Expanding on this idea, we define a context of heritage production to be any activity, occurring online or offline, as part of which human or non-human agents engage with the past more or less incidentally (e.g. political activism, re-enactment, art making, neighbourhood surveillance, etc.). The genesis – and, hence, the study - of digital heritage naturally takes place within a given context of heritage production, with the understanding that this can cut across web and analogue spaces, each of which becomes a field of ethnographic investigation.

Facebook was the field of investigation on which we focussed our research, and the entry point to begin mapping contexts of heritage production where ideas, materials and places from 800BCE to 800CE had been called upon. In order to derive a list of contexts that would be as close as possible to the ways in which their generators framed them, we extracted all the Facebook pages whose title included words describing the general periods that we were dealing with - Iron Age, Roman and Early Medieval. We then qualitatively analysed the description of each of the pages that were retrieved and defined categories for heritage contexts (**Table 2**). Subsequently, we decided to explore one context in particular - political micro-activism. We realised, however, that, in order to do so, we would need to concentrate

on a specific topic, and chose Brexit both for its contemporary relevance and because this is a neologism that has exactly the same meaning internationally. We anticipated that the absence of semantic ambiguity would allow us to more easily retrieve Facebook pages rich in content about Brexit, and this was proved to be the case.

Table 2. Examples of contexts of heritage production where the Roman past is called upon, based on an analysis of public Facebook pages.

In order to analyse digital heritage as processes, we must turn to consider digital heritage as outcomes or (big) data. When we approach online fields to examine uses of the past in the present, it is likely that our questions will only be answered if we are able to navigate 'big' digital heritage data, especially if the research is undertaken on and via social media. Big data has been variously defined by commentators, who however mostly agree that it possesses qualities including being very large in volume, exhaustive in scope, relational, of high velocity, variety, flexibility, and fine granularity (Kitchin 2013, Kitchin 2014). In a more recent article, Kitchin and McArdle (2016) revisit this set of characteristics through substantial and comparative empirical work, which led them to confirm that a big dataset will feature most but not necessarily all of these categories. Being able to navigate big data in digital heritage research is important to develop a good understanding of the context of heritage production that is investigated. In our case, it was vital to extract all the 1.4 million documents posted on public Facebook pages that contained the word *Brexit* in the title. Only 2529 of these documents actually mentioned the ancient past, but the relative scale and 'weight' of the corpus became clear only in relation to the total dataset (**Table 1**).

As noticed by Pink et al. (2018) data has materiality, it accumulates and fills the space of servers, it breaks and leaks. It is mundane and produced everyday (Pink et al. 2017), even from research undertakings, and it can become heritage depending on how we more or less consciously curate it over time. Although it is beyond the scope of this article to reflect on the idea of 'future heritages' and 'heritage futures' (Harrison 2016; Harrison et al. 2016), it has been important to touch upon this point, in order to highlight the artificiality of a marked subdivision between heritage as processes and as outcomes, since the two are constantly morphing into one another.

2.2. Agency

A large part of digital heritage is generated directly by individuals, whether working for relevant institutions or interacting with heritage resources independently. Another part, however, is produced by machines, as in the case of CCTV cameras in museums, sensors recording movement within the historic environment or, indeed, 'bots' that operate online. The latter have received great attention recently and especially after rumours that private companies such as Cambridge Analytica and states like Russia had been using them to influence democratic processes including the US Presidential Elections and the Brexit referendum. It has been suggested that Russian state-funded media (e.g., RT or Sputnik), together with social media 'trolls' and bots, have worked to spread misinformation amongst Western populations and to impact their voting preferences (Narayanan et al. 2017; Tabansky 2017). This non-human agency may have drawn on the past as well, as our study seems to evidence.

In Bonacchi, Altaweel and Krzyzanska (2018), we noted that the number of messages posted multiple times on the same or on different Facebook pages was not high. However, it is

important to underline that there was a minority of cases where multi-posting occurred and we cannot completely exclude that these messages were produced with the purpose of influencing Brexit discussions, possibly via fake profiles. An example – albeit on the pro-remain side - is that of an article about the Anglo-Saxons, published by the historical fiction writer Martin Wall (2016) in the BBC History Magazine, the link to which appeared 15 times accompanied by a message that was posted by the same user on different Facebook pages.[3] The post reiterates the origin myth that presents the British and Celts as indigenous people:

> 'Great Britain is a wrong name. In around the year 500 the Anglo-Saxons (+Jutes) from North-West Europe conquered in a rude way the land of the British and of the Kelts [sic] in Scotland and Wales and later Ireland. A large part of the real British left the island and went to Bretagne. Although large parts of the real Brits who had not enough tradition in warfare and the Kelts had become Roman Catholics while the county was occupied by the Romans, no soldiers from the Romans defended them and the Roman army had already left the island also. All in all there is hardly any logical argument to call England "Britain", let be "Great Britain" ! What is also completely unlogical is that the English are racists now, unless you consider that the victory over the Brits and the Kelts and later over all the English colonies was always mainly supported by the feeling that all others than the English were inferior humans. Is also this habit a background for the racism, the discrimination and xenofobia behind the Brexit????'

While it may be that someone simply decided to manually post the same message multiple times to reach a wider audience, it is also possible that the message was circulated automatically by software. Furthermore, both RT and Sputnik featured amongst the media websites linked in the messages containing references to the Iron Age, Roman and Early

Medieval periods, even though the total number of such links was rather low. The past may have been intentionally leveraged to suit various political agendas via human and non-human agencies. Whilst the latter are not entirely concealed, they are certainly rendered more opaque by social media fields, despite increasing efforts to develop methods for the identification of bots and fake accounts (Paavola et. al 2016; Stukal et al. 2017).

So far, we have discussed the agency behind individual interactions with the past through the posting of single Facebook posts, comments and replies. However, as we have collected these documents via Facebook pages, we must also consider the kind of group agency (if any) that lies behind these entities. Being public, a Facebook page may (or not) serve as a public sphere – a space where topics can be discussed by collectives (Habermas 1997; McKee 2004). In our work, we showed that the majority of public Facebook pages dedicated to Brexit were not set up with the aim of hosting exchanges of opinions, but, rather, to present and re-iterate views that individuals already possessed (Bonacchi, Altaweel and Krzyzanska 2018, 179-180). This finding is consistent with those of studies more generally concerned with the role played by Facebook in the Brexit referendum and showing the formation of two distinct echo chambers, one in favour of leaving the European Union and the other of remaining (Del Vicario et al. 2017). The fact that Facebook did not function as a public sphere during, before and soon after the referendum that decided the membership of the UK in the European Union, however, does not mean that this should always be the case. Since platforms and their functioning are in constant flux, it is in fact vital to continue questioning the nature of interventions that are made via the Facebook field as well as through other online ones.

The digital heritage emerging from political micro-activism revolving around Brexit and performed through Facebook, is – counterintuitively from what Facebook, as a social media platform, might lead to think – primarily the expression of entrenched beliefs constructed over time. Hence, when we examine how the past features in 'discussions' about Brexit, we tend to gather insights about aggregated individualism – that is of distinct identities that manifest themselves in spatial proximity but tend to remain independent from each other.

3. New and old epistemologies

3.1. Data structures and heritage utilisers

In order to discuss the epistemologies of digital heritage in a world of big data, it is necessary to first introduce the ways in which data structures are configured and how they can be accessed and meaningfully interpreted by heritage utilisers. In explaining the genesis and management of big data, Zwitter (2014) identifies three relevant groups of stakeholders: generators, collectors and utilisers; collectors have substantial epistemological weight, because they decide what material should be gathered and in what form. As argued by Rieder et al. (2015, 30), big data is never raw but 'pre-cooked' and the collector - hardly ever a researcher – decides on its structure, which is extremely variable and mixed, both in the present and over time. Facebook, for example, has a much more complex structure than several other social media (Rieder et al. 2015), with multiple levels (posts, comments, replies) under each of a number of parallel entities defined as groups, public pages, and individual profiles. These levels can be extracted to different extents depending on the entities they belong to and in full only if they have been shared on Facebook pages, which are set up as public spaces for interaction. Their access and availability cannot be restricted by their administrators, if not by blocking specific users. When mining all the levels of a public

Facebook page, we gather textual data – including links to images and videos - that is structured, but differently from other parallel entities on the Facebook field.

Data architectures might change further as a result of the approach we choose to undertake text mining. A first approach - and the one we opted for - is to 'enter' the database of a given web platform through its Application Programming Interface (API). APIs are designed by collectors to regulate the kinds of information that utilisers may extract (usually a subset of the data available), and they can be accessed using a number of different programming languages. Gathering data in this way, however, is subject to the restrictions that social media and other web-based companies decide to introduce at any given point, with considerable impact on interoperability – the possibility of utilising data collected over time to perform a same kind of analysis (Lipworth et al. 2017, 493).

Following the Cambridge Analytica scandal, Facebook started to require utilisers to register and verify an application before mining a public page. Without an application of this kind, utilisers can extract Facebook posts, but not comments and replies. These changes are in step with similar ones introduced by Instagram in 2016, which effectively meant that 'only those seeking to commercialise users' data could access' such data via the API (Huffer and Graham 2017: 2. Methods). In addition to granting larger businesses an advantage over heritage researchers, this situation raises concerns about representation. It makes it more difficult to analyse the discussions unfolding on public Facebook pages in their totality, leading to the likely exclusion of less official voices, who, based on our experience of researching the *Heritage of Brexit*, tend to prefer responding to posts published by others. The perspectives of individuals and groups who have public *personae* and are the creators of authorised

discourses (Smith 2006) may thus be over-represented and they acquire greater visibility compared to those of private citizens.

These challenges may become even more problematic if, in order to extract heritage-relevant web data, we do not utilise APIs, but scrape the content of Facebook pages as it appears when searched via a browser. Data is retrieved from the html code of a website and requires familiarity with html structures and with the 'selectors' through which parts of these structures are mined. This can be slower and more convoluted than going via the API route, especially for heritage researchers trained in the humanities and social sciences. On the other hand, workflows developed in this way are likely to remain current for relatively longer, because changes to the html code of web platforms have a direct impact on the 'look and feel' of these sites, which social media try to minimise.

The last option is available only in some cases – for instance, for Twitter – and consists of purchasing data from the collecting company. This path however is costly and beyond the possibilities of most heritage utilisers, who operate on relatively small budgets, regardless of whether they are based in higher education, the cultural or creative sectors. Such approach also brings along an additional set of questions about the transparency and reliability of the results. There is no way of checking whether, in actual fact, the data provided by the company matches given and desirable parameters, potentially leaving commercial enterprises wider scope and agency to influence research agendas and outcomes. It is thus urgent that heritage researchers improve their programming skills to have greater possibilities to choose the most appropriate and ethical methods for data collection and management. Another reason why digital upskilling is of vital importance is that digital heritage data needs considerable manipulation before it can be analysed, because of its multi-layered and mutable

nature. In turn, this requires automating a high number of tasks and workflows, while cloud computing competencies might be necessary to set up server instances with adequate combinations of memory or storage. Furthermore, coding is central to the design and management of non-relational databases that can host data with differential levels of detail (Rieder et al. 2015).

In our case, these motivations are exemplified by the substantial evolution of the workflows used to extract content from Brexit-centred Facebook pages, and by the adjustment of data storage and management strategies resulting from the increase in information volume and complexity, as the collection of documents progressed. When we started mining higher numbers of Facebook pages, we also encountered issues related to the fluid velocity of data change. As text was being continuously created and deleted, in the time elapsed between the retrieval of the ID of a Facebook page and the extraction of associated content, the whole page or certain documents could disappear. Our R script would then fail to recognise the ID of a deleted page, and return an error that stopped the loop setup iterating over multiple pages. This technical issue highlights not only that we are gathering a snapshot of the existing data at a particular moment in time, but also that each fragment of the dataset has been extracted from a different time slice. To an extent, this is also true for smaller sets of data, however the effect is hugely amplified when the length of the data collection process is stretched from minutes or hours into days and weeks, aided by the design of bespoke methods to deal with error identification and handling in an automated way (Krzyzanska, Bonacchi and Altaweel 2018).

3.2. Data-intensive heritage ethnographies

The epistemological implications of what we have observed are far-reaching. The ending part of the previous section explained the impossibility to map the exact timescale of a data collection exercise (and of the production of specific digital heritages). Furthermore, as previously noted, agency is rendered more opaque on web platforms, to the point that it becomes extremely difficult to distinguish between human and non-human contributions in any clear-cut way. As we cannot define populations and samples nor claim representativeness, the research we perform is intrinsically ethnographic - rather than quantitative - even if we apply data-intensive techniques.

In our case study, for example, the analysis led to identify simple myths where the ancient past was used to support pro-leave or pro-remain positions, but we could have not quantified their recurrence in any meaningful way even if we had wished to. Our exploration concluded that such myths exist in contemporary society – with a sense that their diffusion is international in scope – and they are leveraged to relate to Brexit. Data-intensive ethnographies lose the representational element of traditional quantitative research and some (albeit not all) of the possibilities for contextualisation that classically define ethnographic practice. We started with a qualitative investigation of the metadata of 364 public Facebook pages containing the word Brexit in their title, to understand how they had been created – the social actors and motivations behind their emergence. We then mapped the content of all these Facebook pages through topic modelling and compared the themes that had emerged to those appearing in the subset of data containing period-specific terms. We then applied term frequencies and associations to orientate the scope of close reading. To research heritage in a world of big data entails embracing this new methodology and epistemology, which come to detail and refine the idea of a quali-quantitative approach described by Latour, Jensen and Venturini as:

[…] a way to define what is a longer lasting social order *without* making the assumption that there exist two levels [of individuals and structures]. […] Instead of having to choose and thus to jump from individuals to wholes, from micro to macro, [we want to] occupy all sorts of other positions, constantly rearranging the way profiles are interconnected and overlapping. (Latour, Jensen and Venturini 2012: 591).

Leveraging both close and distant 'reading' contributes to enhance the extent to which we can create ethical epistemologies of heritage. Close reading is key to respond to recent changes made by companies to their APIs, which do not permit the extraction of discussion threads in their totality, as we have seen before. It can also be useful to protect voices that might be more vulnerable or simply more private and less official. Excluding those voices would mean deleting their interpretations of the past and their entanglements with personal memories and family histories (Merriman 1991; Bonacchi 2018b). Even so, some biases relating to the kinds of subjects whose views are represented through digital heritage research will remain. So far, literature has in fact demonstrated that people with certain socio-demographic characteristics, personality traits and abilities use specific web platforms more and differently than others (Hargittai 2007; Ryan and Xenos 2011; Alhabash and Ma 2017).

However, as long as these considerations are accounted for and we understand digital heritage research as qualitative and immersive exploration, the knowledge we create can be deemed legitimate. Ceasing to generate it would deepen the rift between the very detailed understanding that commercial parties and those with greater economic resources have (or can obtain) of people's interactions with the past and the knowledge that the wider population and public institutions may gain thanks to the work of academic researchers.

As showed in the *Heritage of Brexit* (Bonacchi, Altaweel and Krzyzanska 2018), the past is intertwined with the present and makes up part of all those symbols and images that people draw upon in order to relate to contemporary problems, and talk about their political affiliations and inclinations. Data on how these processes unfold is a powerful tool that skilled but malicious data utilisers may covertly weaponise against data generators. As heritage researchers, we must ensure that we have the ability to provide counter-narratives, through robust and sound digital heritage activism that exposes all those processes through which the past is appropriated for instrumental ends including political ones (Harrison 2018). We have a duty to study the ways in which heritage is made and assessed online. In doing so, it is important to consider the ethics of data usage to minimise chances that, while endeavouring to be critical in the public interest, we are instead endangering citizens or abusing their rights and prerogatives. A thorough discussion of the ethics of digital heritage research is however left to a future publication, since it would require substantial additional space to acknowledge and adequately reflect on the literature developed in this area until now.

4. Conclusions: paradigm shifts in heritage studies

This article points towards a paradigm shift in heritage studies, which we will now summarise in the light of the reflections and findings that were discussed. The notion of *paradigm* was introduced by Kuhn in 1962, in a volume entitled *The Structure of Scientific Revolutions,* to describe a set of theories, priorities and rules that is shared by researchers. Paradigm changes occur suddenly, orientated by an achievement that is 'sufficiently unprecedented to attract an enduring group of adherents away from competing modes of

scientific activity', and simultaneously 'sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve' (Kuhn 1962, 22).

The existing and prevailing paradigm of heritage studies is strongly rooted in post-modernism (Kristiansen 2014), critical theory, and – generally - deeply qualitative approaches suitable to research designs that are immersive and locomotive. As we have seen, however, this paradigm fails when we engage with digital heritage in a world of big data, and – following Kuhn – necessitates to be rapidly substituted by an alternative one that we here propose could be centred on three pillars. The first pillar concerns research priorities, which should be concentrated on the investigation of digital heritage ontologies as the processes and outcomes of interacting with elements of the past – in its material and immaterial traits – via online fields in possible conjunction and synergy with offline ones, and with the understanding that heritage outcomes are both clues guiding our exploration of interactions that have already occurred and artefacts and heritage in the making. These ontologies have opaque agency, constantly mutable temporality and globalised spatiality. Digital heritages are made by aggregated individualism that escapes clear-cut definitions according to the dimensions of time and geography. This leads to the second pillar – the failure of quantitative approaches whilst the very large nature of big data is at the core of the failure of qualitative ones. We then forge the idea of data-intensive ethnographies, or the use of data-intensive techniques to navigate big data and exploring it, revealing spectra of diversity that are present in our contemporary world but which we cannot fully measure or quantify. We detect them and reflect upon them with a renewed critical apparatus that is developed from the very observation and analysis of technicity.

The third pillar consists in the re-shaping of the role of heritage scholars. Not only focused on one body of theory, on one sub-discipline, on theory alone, they are engaged fully with

operational aspects of digital technologies, to develop a new and *ad hoc* body of theory from the ground up, working through empirical engagements. It is a kind of researcher that embraces simultaneously the tradition of Enlightenment and that of Romanticism, to create new world views by showing to data collectors their idiosyncrasies, the paradoxes and abuses but also the possible innovative and socially activist applications of digital heritage processes. Our hope for the future is that a discussion of technicity will no more be perceived as somehow foreign to heritage, but a natural element of the digital heritage research field.

Notes

1. On 23 June 2016, a referendum was held to decide on this matter. It resulted in the victory of the camp advocating that the UK should cease being member of the European Union.

2. We collected all the messages published since 2013, when the referendum was announced, and which were still available at the time.

3. The message was published on the public Facebook pages: Brexitbible, UK Brexit, Brexit.Europe, brexitballs, BrexitUK, Brexit Apologies, GB Brexit, LSEBrexitVote, MemesVsBrexit, Economists for brexit, BrexitWarwick, BrexitUpdates and BrexitBargain.

References

Alhabash, S. and M. Ma. 2017. A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? *Social Media + Society* 3(1): 205630511769154. DOI: 10.1177/2056305117691544.

Ancient Identities. 2016. *Welcome to Ancient Identities Today.* Available at: http://ancientidentities.org (accessed 15 June 2018).

Ancient Identities. 2017. *Keywords*. Available at: https://docs.google.com/spreadsheets/d/e/2PACX-1vSQB3A8Bfa5CtDg6Weh35gVLVbYOAwrIG9HEDYjMMri5xr_d3fEvvCa34FYGUJEMnwFivO6i3tXcn96/pub?output=csv (accessed 7 June 2018).

Anderson, M. 1997. *The wired museum: emerging technology and changing paradigms*, edited by K. Jones-Garmil, 11–32. Washington, DC: American Association of Museums.

Arvanitis, K. 2002. "Digital, virtual, cyber or network museum? A search for definitions as an act of interpretation". Paper presented to *Museum, communication and new technologies,* 1st international museology conference, Department of Cultural Technology & Communication, University of the Aegean, Mytilene, Greece (May 31-June 2, 2002).

Bonacchi C. 2018a. Co-*producing knowledge online*. Bristol: AHRC Connected Communities Foundation Series, Arts and Humanities Research Council. Available at:

https://connected-communities.org/wp-content/uploads/2018/07/Knowledge_Online_SP.pdf
(accessed 14 December 2018).

Bonacchi, C. 2018b. Public archaeology cannot just 'fly at dusk': the reality and complexities of generating public impact. *Antiquity* 92(366): 1659-1661. DOI: 10.15184/aqy.2018.231.

Bonacchi, C. and G. Moshenska. 2015. "Critical Reflections on Digital Public Archaeology". *Internet Archaeology* 40: 10.11141/ia.40.7.1.

Bonacchi, C., M. Altaweel, and M. Krzyzanska. 2018. "The heritage of Brexit: Roles of the past in the construction of political identities through social media". *Journal of Social Archaeology* 18(2): 174-192. DOI: 10.1177/1469605318759713.

Cameron, F., and S. Kenderdine. 2010. *Theorizing digital cultural heritage: a critical discourse*. Cambridge, Massachusetts: MIT Press.

Carter, P. 2004. *Material Thinking: The Theory and Practice of Creative Research*. Carlton, Vic: Melbourne University Publishing.

Clarke, H.D., M.J. Goodwin, and P. Whiteley. 2017. *Brexit: why Britain voted to leave the European Union*. Cambridge and New York: Cambridge University Press.

Del Vicario, M., F. Zollo, G. Caldarelli, A. Scala, and W. Quattrociocchi. 2017." Mapping social dynamics on Facebook: The Brexit debate". *Social Networks* 50: 6-16. DOI: 10.1016/j.socnet.2017.02.002.

Dewey, J. 1938. *Experience and Education*. New York: Simon & Schuster.

Díaz-Andreu, M. 2017. "Introduction to the themed section 'digital heritage and the public'". *International Journal of Heritage Studies* 23(5): 404-407. DOI: 10.1080/13527258.2017.1286780.

Fahy, A. 1995. "New technologies for museum communication". In *Museum, media, message*, edited by E. Hooper-Greenhill, 82–96. London: Routledge.

Fuchs, C. 2018. *Nationalism 2.0. The Making of Brexit on Social Media*. London: Pluto Press.

Galani, A., and M. Chalmers. 2002. "Can you see me?: exploring co-visiting between physical and virtual visitors". In *Museums and the Web 2002: selected papers*, edited by D. Bearman, and J. Trant, 31–40. Archives & Museum Informatics.

Gardner, A. 2017. "Brexit, boundaries and imperial identities: A comparative view". *Journal of Social Archaeology* 17(1): 3-26. DOI: 10.1177/1469605316686875.

Gauntlett, D. 2015. *Making media studies: the creativity turn in media and communications studies*. New York: Peter Lang.

Geismar, H. 2012. "Museum + Digital =?" In *Digital Anthropology*, edited by H.A. Horst and D. Miller, 266-287. London & New York: BERG.

Habermas, J. 1997. "The public sphere". In R *Contemporary Political Philosophy: An Anthology*, edited by E. Goodin and P. Pettit, 105-108. Oxford: Blackwell Publishers.

Hargittai, E. 2007. Whose Space? Differences Among Users and Non-Users of Social Network Sites. *Journal of Computer-Mediated Communication* 13(1): 276-297. DOI: 10.1111/j.1083-6101.2007.00396.x.

Harrison, R. 2013. *Heritage: critical approaches*. Milton Park, Abingdon and New York: Routledge.

Harrison, R. 2016. "Archaeologies of Emergent Presents and Futures". *Historical Archaeology* 50 (3): 165-180.

Harrison, R. 2018. Critical heritage studies beyond epistemic popularism. *Antiquity* 92(365): e9. DOI: 10.15184/aqy.2018.223.

Harrison, R., N., Bartolini, C. DeSilvey, C. Holtorf, A. Lyons, S. Macdonald, S. May, J. Morgan, and S. Penrose. 2016. "Heritage Futures". *Archaeology International* 19: 68-72. DOI: 10.5334/ai.1912.

Hine, C. 2015. *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury Academic.

Hoel, A.S., and I. Van der Tuin. 2013. "The Ontological Force of Technicity: Reading

Cassirerand Simondon Diffractively". *Philosophy & Technology* 26: 187-202.

Housley, W., R. Procter, A. Edwards, P. Burnap, M. Williams, L. Sloan, O. Rana, J. Morgan, A. Voss, and A. Greenhill. 2014. "Big and broad social data and the sociological imagination: A collaborative response". Big Data & Society 1(2): 2053951714545135. DOI: 10.1177/2053951714545135.

Huffer, D. and S. Graham (2017) The Insta-Dead: The rhetoric of the human remains trade on Instagram. *Internet Archaeology* 45. DOI: 10.11141/ia.45.5.

Jones, S., S. Jeffrey, M. Maxwell, A. Hale, and C. Jones. 2017. "3D heritage visualisation and the negotiation of authenticity: the ACCORD project". *International Journal of Heritage Studies* 24(4): 333-35. DOI: 10.1080/13527258.2017.1378905.

Kidd, J. 2011. "Enacting engagement online: framing social media use for the museum". Information Technology & People 24(1): 64-77. DOI: 10.1108/09593841111109422.

King, L., J.F. Stark, and P. Cooke. 2016. "Experiencing the Digital World: The Cultural Value of Digital Engagement with Heritage". *Heritage & Society* 9(1): 76-101. DOI: 10.1080/2159032X.2016.1246156.

Kitchin, R. 2013. "Big data and human geography: Opportunities, challenges and risks". *Dialogues in Human Geography* 3(3): 262-267. DOI: 10.1177/2043820613513388.

Kitchin, R. 2014. The Data Revolution: Big Data, Open Data, Data Infrastructures & Their

Consequences. Thousand Oaks, CA: Sage.

Kitchin, R., and G. McArdle. 2016. "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets". *Big Data & Society* 3(1): 205395171663113. DOI: 10.1177/2053951716631130.

Kristiansen, K. 2014. "Towards a New Paradigm? The Third Science Revolution and its Possible Consequences in Archaeology". *Current Swedish Archaeology* 22: 11-34.

Krzyzanska, M., C. Bonacchi, and M. Altaweel. 2018. *Heritage-of-Brexit*. figshare. Code. https://doi.org/10.6084/m9.figshare.5966818.v1.

Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, B., P. Jensen, and T. Venturini. 2012. "'The whole is always smaller than its parts' – A digital test of Gabriel Tarde's monads". *British Journal of Sociology* 63(4): 590–615. DOI: 10.1111/j.1468-4446.2012.01428.x.

Lipworth, W., P.H. Mason, I. Kerridge, and J.P.A. Ioannidis. 2017. "Ethics and Epistemology in Big Data Research". *Journal of Bioethical Inquiry* 14(4): 489-500. DOI: 10.1007/s11673-017-9771-3.

Marichal, J. 2013. "Political Facebook groups: Micro-activism and the digital front stage". First Monday 18(12). DOI: 10.5210/fm.v18i12.4653.

Marwick, B. 2013. Discovery of Emergent Issues and Controversies in Anthropology Using Text Mining, Topic Modeling, and Social Network Analysis of Microblog Content. In Y. Zhao and Y. Cen (Eds.) Data Mining Applications with R, Amsterdam: Academic Press, Elsevier, pp. 63-93. DOI: 10.1016/B978-0-12-411511-8.00003-7.

Matsuda, A. 2016. "A Consideration of Public Archaeology Theories". *Public Archaeology* 15(1): 40-49. DOI: 10.1080/14655187.2016.1209377.

Merriman, N. 1991. *Beyond the Glass Case: the Past, the Heritage and the public in Britain*. Leicester: Leicester University Press.

McKee, A. 2004. *The public sphere: an introduction*. Cambridge and New York: Cambridge University Press.

Narayanan, V., P.N. Howard, B. Kollanyi, and M. Elswah. 2017. "Russian Involvement and Junk News during Brexit". Data Memo 2017.10. Oxford, UK: Project on Computational Propaganda. Available at: http://comprop.oii.ox.ac.uk/research/working-papers/russia-and-brexit/ (accessed 14 June 2018).

Olsen, B., and Þ. Pétursdóttir. 2017. "Theory adrift: The matter of archaeological theorizing". *Journal of Social Archaeology* 18(1): 97-117. DOI: 10.1177/1469605317737426.

Paavola, J., T. Helo, H. Jalonen, M. Sartonen, A-M. Huhtinen. 2016. "Understanding the Trolling Phenomenon: The Automated Detection of Bots and Cyborgs in the Social Media". *Journal of Information Warfare* 15(14): 100-111.

Parry, R. 2007a. *Recoding the museum: digital heritage and the technologies of change*. London and New York: Routledge.

Parry, R. 2007b. "Digital heritage and the rise of theory in museum computing". *Museum Management and Curatorship* 20(4): 333-348. DOI: 10.1080/09647770500802004.

Parry, R. 2010. *Museums in a digital age*. London and New York: Routledge.

Perry, S., and N. Beale. 2015. "The Social Web and Archaeology's Restructuring: Impact, Exploitation, Disciplinary Change". *Open Archaeology* 1(1): 153-165. DOI: 10.1515/opar-2015-0009.

Pink, S., S. Sumartojo, D. Lupton, and C. Heyes La Bond. 2017. "Mundane data: The routines, contingencies and accomplishments of digital living". *Big Data & Society* 4(1): 205395171770092. DOI: 10.1177/2053951717700924.

Pink, S., M. Ruckenstein, R. Willim, and M. Duque. 2018. "Broken data: Conceptualising data in an emerging world". *Big Data & Society* 5(1): 205395171775322. DOI: 10.1177/2053951717753228.

Richardson, L-J. 2018. "Ethical Challenges in Digital Public Archaeology". *Journal of Computer Applications in Archaeology* 1(1): 64-73. DOI: 10.5334/jcaa.13.

Ridge, M. (Ed.) 2014. *Crowdsourcing Our Cultural Heritage*. Farnham, Surrey, England: Ashgate Publishing Limited.

Rieder, B., R. Abdulla, T. Poell, R. Woltering, and L. Zack. 2015. "Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring 'We are all Khaled Said'". *Big Data & Society* 2(2): 205395171561498. DOI: 10.1177/2053951715614980.

Ruppert, E., J. Lowe, and M. Savage. 2013. "Reassembling Social Science Methods: The Challenge of Digital Devices". *Theory, Culture & Society* 30(4): 22-46. DOI: 10.1177/0263276413484941.

Ryan, T. and S. Xenos. 2011. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior* 27(5): 1658-1664. DOI: 10.1016/j.chb.2011.02.004.

Silberman, M. 2007. "Introducing *The Handbook of experiential Learning*". In *The Handbook of Experiential Learning*, edited by M. Silberman, 1-10. Hoboken: John Wiley & Sons, Inc.

Smith, L. 2006. *Uses of Heritage.* London and New York: Routledge.

Stukal, D., S. Sanovich, R. Bonneau, and J.A. Tucker. 2017. "Detecting Bots on Russian Political Twitter". *Big Data* 5: 310–324. DOI: 10.1089/big.2017.0038.

Tabansky, L. 2017. "Cybered influence operations: towards a scientific research agenda". Oslo: The Norwegian Atlantic Committee. Available at: http://www.atlanterhavskomiteen.no/files/dnak/Documents/SPL%202-17.pdf (accessed 14 June 2018).

Wall, M. 2016. "10 things you (probably) didn't know about the Anglo-Saxons". *History Extra*. Available at: http://www.historyextra.com/article/alfred-great/10-facts-anglo- saxons-history (accessed 7 February 2018).

Were, G. 2015. "Digital heritage in a Melanesian context: authenticity, integrity and ancestrality from the other side of the digital divide". *International Journal of Heritage Studies* 21(2): 153-165. DOI: 10.1080/13527258.2013.842607.

Wevers, M., and J. Verhoef. 2017. "Coca-Cola: An Icon of the American Way of Life. An Iterative Text Mining Workflow for Analyzing Advertisements in Dutch Twentieth-Century Newspapers". *Digital Humanities Quarterly* 11(4). http://www.digitalhumanities.org/dhq/vol/11/4/000338/000338.html.

Zuanni, C. 2017. Unintended Collaborations: interpreting archaeology on social media. *Internet Archaeology* 46. DOI: 10.11141/ia.46.2.

Zwitter, A. 2014. Big Data ethics. *Big Data & Society* 1(2): 205395171455925. DOI: