

Review

Data Legacies, Epistemic Anxieties, and Digital Imaginaries in Archaeology

Jeremy Huggett 

Archaeology, School of Humanities, University of Glasgow, Glasgow G12 8QQ, UK;
jeremy.huggett@glasgow.ac.uk

Abstract: Archaeology operates in an increasingly data-mediated world in which data drive knowledge and actions about people and things. Famously, data has been characterized as “the new oil”, underpinning modern economies and at the root of many technological transformations in society at large, even assuming a near-religious power over thought and action. As the call for this Special Issue recognizes, archaeological research is socially and historically situated and consequently influenced by these same broader developments. In archaeology, as in the wider world, data is the foundation for knowledge, but its capacity is rarely reflected upon. This paper offers just such a reflection: a meditation on the nature of archaeological digital data and the challenges for its (re)use. It asks what we understand by data: its etymology and comprehension, its exceptionality and mutability, its constructs and infrastructures, and its origins and consequences. The concept of the archaeological data imaginary is introduced to better understand approaches to the collection and use of archaeological data, and a case study examines how knowledge is mediated and remediated through the data embedded in grey literature. Appreciating the volatility and unpredictability of digital data is key in understanding its potential for use and reuse in the creation of archaeological knowledge.

Keywords: data; practice; data journey; data friction; imaginary; grey literature; digital; archaeology



Citation: Huggett, J. Data Legacies, Epistemic Anxieties, and Digital Imaginaries in Archaeology. *Digital* **2022**, *2*, 267–295. <https://doi.org/10.3390/digital2020016>

Academic Editors: Markos Katsianis, Tuna Kalayci and Apostolos Sarris

Received: 13 April 2022

Accepted: 16 May 2022

Published: 19 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

“Data is, we imagine, an immaterial thing; or at least ethereal, made of light and electricity, processed at superhuman speed, transmitted in real time. The everyday world we move in seems dense and slow by comparison. The landscape is slower again; thick, heavy and persistent.” [1]

The mercurial nature of digital data described above can make it seem unpredictable, and yet at the same time, data can appear to be self-evident and unproblematic. From an archaeological perspective, data can be both material and immaterial, objective but biased, precise but inaccurate, tangible but ephemeral, detailed but generalized. It can be big or small, fast or slow, quantitative or qualitative, comprehensive yet selective. It is open to misrepresentation, misconception, misunderstanding, misreading, and misinterpretation. It can be distorted, altered, wrangled, and reshaped into something entirely new, used for different purposes and according to different agendas [2]. Despite its many perils and imperfections, data is nevertheless the basis for archaeological interpretation and for determining what is and is not known about the past through its material remains. Indeed, archaeologists are all too accustomed to the lacunae in their data, whether viewed colourfully as a “vast fiendish jigsaw invented by the devil as an instrument of tantalising torment” missing an unknown number of pieces [3] (p. 5) or more prosaically as a discipline defined by the challenges of working with gaps and absences in its primary data [4] (p. 204). Such characteristics are often considered to be key distinguishing features of archaeological data, but this has not been broadly reflected in considerations of digital data beyond technical discussions of data handling and arguments supporting the importance of digital data archiving. This paper seeks to redress this imbalance by

considering the nature, characteristics, and implications of digital data for the construction of archaeological knowledge.

That data is so fundamental can be seen in the way that many of the “revolutions” or paradigm shifts experienced within archaeology are related to data (e.g., [5]), whether they are new ways of approaching data, the creation of new kinds of data, a greater abundance of data, or faster processing of data in larger quantities than before. Some revolutions are based on technological developments and adoptions, for example, the Radiocarbon revolution of the 1950s and the Quantitative revolution associated with the New Archaeology, while more recently, Kristiansen [6,7] has identified a new scientific revolution associated with the combination of DNA and isotope studies with big data. If anything, there have been more “turns” than revolutions in archaeology: for example, Sørensen [8] lists the linguistic turn, the spatial turn, the practice turn, the material turn, the affective turn, and the ontological turn. Many of these have their origins outside archaeology, derived from scientific and technological advances or theoretical developments in other disciplines, but they nevertheless entail changes in approaches to data—how it is gathered, analyzed, processed, and perceived. As reflections of changes in wider society, not just in archaeology, the list may be extended further with the Information Age and the digital turn (the incorporation of computers and other digital devices in practice). We may even be experiencing the Second Digital turn, defined by Carpo as a nonhuman, post-scientific method combining big data and artificial intelligence: “Just as the digital revolution of the 1990s (new machines, same old science) begot a new way of making, today’s computational revolution (same machines, but a brand new science) is begetting a new way of thinking.” [9] (p. 7). A data revolution is characterized by Kitchin in terms of “the profound datafication” of our lives [10] (p. 3) and the extent to which data has come to govern our knowledge and experience. These informational, digital, and data turns or revolutions are experienced within archaeology as much as elsewhere, presenting both challenges and opportunities for the ways in which archaeological data is handled, reproduced, manipulated, analyzed, archived, circulated, and reused (e.g., [11–13]), but the implications of these are as yet barely recognized.

2. Introducing Imaginaries

Terms such as “revolution” or “turn” carry connotations of abrupt dislocations or changes in direction. These do not necessarily reflect the experience of technological, methodological, or theoretical changes within archaeology which often take place over extended periods of time and may be only identified in retrospect. Furthermore, many of these developments are broadly contemporary with others, rather than being succeeded or replaced, and there may be a degree of interdependence involved. An alternative which might better characterize the situation is the concept of the *imaginary*, which is commonly applied in anthropology and science and technology studies and seems appropriate in the context of an inquiry into digital data. Strauss [14] outlines a complex range of origins for the term, derived from Cornelius Castoriadis’s concept of the imaginary as the shared unifying core conceptions of a group [14] (p. 324), Jacques Lacan’s view of the imaginary as an illusion or fantasy created in response to a psychological need [14] (p. 326), and the imaginaries of Benedict Anderson and Charles Taylor consisting of communities of people sharing concerns and practices [14] (pp. 329–330; see also [15,16] for a more extensive analysis of the use of the term). As Strauss suggests, the imaginary goes beyond simply shared facts or knowledge and covers a spectrum between indisputable knowledge and ignorance, including “explicit knowledge of imagined facts, implicit cultural beliefs, and dissociated, repressed, and fantasized knowledge” [14] (p. 339). So, for example, Marcus [17] sees the possibilities of a *technoscientific imaginary* associated with scientific practice and potentially ranging across reflective, visionary, and innovative scientific thought and the imagination of science fiction writers [17] (p. 4) before narrowing it down to the social and culturally embedded imaginaries of scientists “tied more closely to their current positionings, practices, and ambiguous locations in which the varied kinds of science they do are possible at all” [17] (p. 4). Similarly, Jasanoff and Kim [18] (pp. 122–123) define a

sociotechnical imaginary as a system of meaning which enables collective interpretations of social reality and a shared sense of identity and attachment to a group, which are both achieved through and support scientific and technological advances. They are “collective, durable, capable of being performed; yet they are also temporally situated and culturally particular.” [19] (p. 19). The idea of the imaginary “emphasises that political agendas are driven by culturally specific belief and value systems that produce different forms of techno-political order” and invites “a close reading of the various expectations and concerns, the diverse norms, mores and ideologies that help these imaginaries to form, stabilise, proliferate, and endure” [20] (p. 91).

Like Jasanoff, Ruppert sees imaginaries as something enacted rather than determined, which can lead to unintended or unexpected consequences. She suggests that some of the most powerful sociotechnical imaginaries concern digital technologies and big data [21] (p. 14). These imaginaries “not only shape what is thinkable but also the practices through which actors perform them. . . they are taken up in practices through which new paradigms or ways of thinking are propagated.” [21] (p. 19). In Bucher’s study of Facebook, for instance, she introduces the concept of the *algorithmic imaginary*, which “does not merely describe the mental models that people construct about algorithms but also the productive and affective power that these imaginings have.” [22] (p. 41). She highlights the way in which algorithms are experienced and encountered as part of the “‘force relations’ that give people a ‘reason to react’” and the consequent feedback loop through which the algorithm is moulded in turn [22] (p. 41).

The fact that these algorithms operate on data, together with the very ubiquity of data, means that *data imaginaries* have also been characterized by several authors. For example, Beer defines the data imaginary as

“ . . . part of how people imagine data and its existence, as well as how it is imagined to fit with norms, expectations, social processes, transformations and ordering. The data imaginary is about how data are imagined in the social world and how they intervene in the connections between people and between people and organisations, nation states, media and their material environment.” [23] (p. 18)

Beer emphasizes how the data imaginary, like Bucher’s algorithmic imaginary, operates in a feedback loop, with the imaginary shaped by data practices and those practices in turn shaped by the visions of the imaginary, leading to an intensification of data analytics, for example [23] (p. 18). In her discussion of the use of data in response to the coronavirus pandemic, Leonelli also highlights the importance of the vision entailed in data imaginaries. This is a vision of the technical, human, institutional, and data resources that can be developed and combined to address a problem without assuming those resources necessarily exist yet:

“Hence the choice of the term ‘imaginaries’: rather than just collections of ideas, these are ways in which data science is routinely *imagined* and *performed* by researchers, policymakers, and various publics and stakeholders. They typically do not amount to a coherent plan or a systematic philosophy of data use; they are also not necessarily stable and can rapidly adapt to changing research conditions.” [24] (p. 5, emphasis in original)

Elsewhere, *data ideology* is employed as a term in much the same way as data imaginary: for example, Poirier et al. define data ideology as “constituted through the complex cultural and institutional forces that shape particular, yet always collective, ideas and values about data sharing and data infrastructure design.” [25] (p. 214). Similarly, Dourish and Gómez Cruz suggest their *data narratives*—narratives in, from, and of data—broadly parallel data ideologies [26] (p. 6). Alternatively, Gray uses the notion of *data worlds* as a means of exploring

“ . . . the *performative* capacities of data infrastructures: what they do and could do differently, and how they are done and could be done differently... to consider

how data infrastructures may be involved in not just the *representation* but also the *articulation* of collective life, while at the same time being the products of social and institutional work themselves.” [27] (p. 6, emphasis in original)

Data imaginaries, data ideologies, or data worlds all provide ways of conceptualizing the assumptions, practices, and consequences surrounding data. For example, Jasanoff argues that the value of studying imaginaries is the way in which it “allows at its best a deep meditation on the basis of a technological society’s particular forms of sightedness and blindness, and the trade-offs that inevitably accompany attempts to build a shared normative order.” [28] (p. 339). Nor are they exclusive categories: multiple imaginaries or worlds may coexist and interact. For example, Leonelli explores five data imaginaries related to COVID-19 studies, linked to surveillance, predictive modelling, causal explanation, logistical decision-making, and socio-environmental needs [24] (p. 4). Similarly, Gray characterizes three overlapping aspects of data worlds: the world-making capacities of data infrastructures, the collective identities of those making and using data, and the transnational scope of data infrastructures [27] (pp. 12–13). This underlines that data imaginaries may be interrelated rather than a single entity since they may represent different fields of practice, different possible futures, different sets of trade-offs, and different effects or consequences. As Beer [23] (p. 19) suggests, the data imaginary is a component of the broader sociotechnical imaginary. In the same way, a data imaginary may itself consist of a subset of imaginaries, each focused on a different intervention or characteristic. For example, Poirier et al. identify four data ideologies at work within cultural anthropology: what they call “all or nothingism” (related to the sharing of data); “ethico-political sensitivities” (a “data-clutching” ideology holding that all data is sensitive and requires protection); “institutionally reinforced individualism” (where credit, reward, and advancement emphasize the individual and discourage collaboration); and “interpretive exceptionalism” (an emphasis on the unique nature of anthropological data) [25] (pp. 220–223). Underlying all four ideologies or imaginaries is the question of sharing and reusing data.

3. The Archaeological Data Imaginary

Understood in this light, the archaeological data imaginary can be conceived as consisting of several interlinked imaginaries. These include imaginaries connected to open research, to research infrastructures, to big data and algorithmic analysis, and to the nature of archaeological data itself. Separately and jointly, they carry implications for approaches to archaeological data.

3.1. The Open Data Imaginary

Archaeology has a long-standing relationship with the concept of open data, although how it is operationalized has varied considerably. In terms of digital data, archaeology is in certain respects well-served with free access to data provided by a range of national institutions and other organizations (e.g., see contributions to [29]). However, different degrees of openness may be encountered, ranging from open to view-only, through a limited ability to download data, to full download access limited only by the requirement to cite the source [30] (pp. 7–8). Although there may be a general perception, within the subject at least, that archaeology is relatively open in terms of a willingness to share data, the reality often falls some way short as rights over data are frequently retained. This is embedded in some professional codes through the right of primacy: for example, the Chartered Institute for Archaeologists (UK) specifies that a member has up to ten years within which to publish data during which time they retain near-exclusive rights [31] (para. 4.4). Similarly, the European Association of Archaeologists specifies a ten-year period within which an archaeologist may have prior rights, and while they are expected to make results as accessible as possible during this period, it is only after ten years that records should be freely available to others [32] (para. 2.7). Consequently, archaeologists traditionally treat their datasets as proprietary to a degree, given their investment in collecting them in the first place [33] (p. 9). To some extent, such a position was easily defensible given traditional

practices which made data-sharing cumbersome and awkward at best, but the growth in digital data removes this physical limitation. Marwick et al. observe that this data-ownership mindset should become obsolete in favour of data-stewardship, recognizing that data are collected and shared on behalf of the broader community [33] (p. 9), a view expressed more robustly by Kansa [34] (p. 507), for example.

The challenge to archaeological practice cannot be underestimated. For example, a permissive definition of “openness” is summarized in the Open Definition: “Knowledge is open if anyone is free to access, use, modify, and share it—subject, at most, to measures that preserve provenance and openness” [35]. A more recent characterization of openness is encapsulated within the FAIR principles: data that is findable, accessible, interoperable, and reusable (e.g., [36,37]). In the Open Definition, the means by which openness is achieved is not specified, whereas the FAIR principles outline how openness might be accomplished (for instance, emphasizing machine-readable metadata and standardized protocols, vocabularies, etc.) while maintaining that the barrier to entry that these methods represent is set deliberately low [36] (p. 4). The emphasis of the FAIR principles is on more rigorous management and stewardship of data [36] (p. 7) and is described as the culmination of years of agreements and actions by publishers, repositories, funding agencies, and others [37] (p. 28), which may lie behind its greater appearance of institutionalization.

As in other disciplines, the FAIR acronym has become something of a buzzword in discussions of open archaeological data, providing a model against which data provision can be measured and assessed (for example, see contributions to [29,38]), although its generic nature may not be altogether helpful [39]. It also embeds a series of expectations surrounding the standardization of data and metadata, in part required in order to allow for the automated discovery and integration of datasets. This goes beyond standardized digital communications protocols and persistent identifiers to include shared languages for knowledge representation and common vocabularies, for instance. These ontologies, high-level classes of information, and associated standards are not widely discussed in archaeology [40], and recent studies demonstrate that the structures created with such tools do not necessarily capture the nuances of archaeological data (e.g., [41]). Consequently, the kinds of cultural changes required to develop open approaches in archaeology and beyond via stronger data mandates from funders, incentivization of open data practices, and appropriate funding for data infrastructures (e.g., [33,37,42]) are not the whole story: the open-data imaginary risks fetishizing data sharing without fully considering the implications of what is shared, how, and for whom.

3.2. The Data Infrastructure Imaginary

A key feature of development in archaeology over the past twenty years or so has been the development of digital infrastructures designed to support data management, data access, data sharing, and data preservation (e.g., [43,44]). In addition to making data findable, accessible, interoperable and reusable (i.e., FAIR), these digital infrastructures provide data integration features, using structured data to support interdisciplinary research (e.g., [45,46]). They can be described as “scholarly eco-systems”, supporting scholarly development and the use of research resources, tools, and methods and the outputs they enable [47] (p. 3). In this way, archaeological practice is increasingly embedded in digital infrastructures in much the same way as communication infrastructures have become fundamental to supporting activities in the modern world and largely disappear into the background. It is only when such infrastructures break down or behave unexpectedly that they become foregrounded and the degree of dependence on them becomes apparent. For example, Historic Environment Scotland’s archaeological digital collections, including Canmore (the National Record of the Historic Environment) and PastMap, the National Collection of Aerial Photography (NCAP), Britain from Above (historic aerial imagery), and the Buildings at Risk Register, were taken offline for a month from December 2021 to January 2022 in order to resolve the Apache Log4j critical vulnerability in their internet services. Such a significant downtime, even over a holiday period, underlined the extent to

which these services were relied upon for desk-based assessments in commercial archaeology as well as for general research, as evidenced in social media queries and responses around that time. It is when such tools are demonstrated to be central and indispensable components of regular practice that they can be revealed as imaginaries in terms of what they enable (or disable) in practice (e.g., [48] (p. 2)).

Such infrastructural data imaginaries organize and promote certain practices over others and therefore act as technological gatekeepers through the ways in which they provide and control access to information. For example, Day [49] (p. 4) argues that such sociotechnical infrastructures mediate their access by hiding the assumptions that underlie them. A simple way of demonstrating this in practice is by comparing search tools used by national and transnational archaeological infrastructures. Canmore (<https://canmore.org.uk/> accessed on 22 March 2022), maintained by Historic Environment Scotland, provides access to records for ancient and historic monuments and allied resources across Scotland. Searching for the site classification “motte and bailey” (a form of medieval timber and earthwork castle), a term offered by the Canmore search interface, returns 43 records for Scotland. Turning to the Archaeological Data Service’s UK-wide ArchSearch faceted browser (<https://archaeologydataservice.ac.uk/archsearch/browser.xhtml> accessed on 22 March 2022) and performing the same search, restricting it to the ADS snapshot of Canmore and to the geographic area of Scotland, returns 345 records. The same search applied in the ARIADNE trans-national data infrastructure (<https://ariadne-infrastructure.eu/portal/> accessed on 22 March 2022) returns 45 site records (excluding fieldwork reports). The two additional sites returned by the ARIADNE portal are actually in England, and a consequence of border changes between England and Scotland as they also appear in Historic England’s Heritage Gateway records but not in Canmore. The difference between 43 and 345 records is more problematic, however. On inspection it appears that ArchSearch is using a fuzzy search, returning sites that are mottes (without baileys) alongside the motte and baileys that were specifically requested. In contrast, Canmore uses an exact search, as expected. Indeed, adding a search for mottes to a search for motte and baileys in Canmore returns a total of 346 records. This is not to suggest that either of these well-established data tools is wrong as such. However, at first sight, the two generate very different results from the same data because the underlying search implementation is different and in neither case can this be seen or controlled by the end user. In short, they control access to digital data in non-transparent ways, underlining their function as an imaginary and emphasizing that the choices made by those who control such infrastructures determine what data, in what form, and which other resources are made available in the future [50] (p. 15). As Okune et al. observe, it is critical to reflect on who is included and excluded in the design and use of these infrastructures, and they argue for what they call “inclusive knowledge infrastructures”, which are “mindful of the diversity of human needs, identities, abilities, experiences, and forms of knowing” [51] (p. 13).

3.3. The Big Data Imaginary

Archaeology has seen significant growth in the quantity and availability of digital data in recent years. For example, the Archaeology Data Service’s Annual Report for 2020–2021 indicates that its archive holds over 3.6 million files in 308 file formats, totalling 25.28 TB of data [52] (p. 7). These are spread across 4087 collections including databases of artefacts, burials, coins, building surveys, fieldwork archives and reports, scientific datasets, and sites and monuments records [52] (p. 11). Over 1.4 million metadata records are catalogued and searchable via ArchSearch, and their unpublished fieldwork library currently contains over 64,000 reports. Elsewhere, Historic Environment Scotland’s Canmore passed 1 million digital objects in 2020 [53], while the Portable Antiquities Scheme for England and Wales has just over 1 million records relating to more than 1.5 million objects [54]. In the USA, the Digital Archaeological Record (tDAR) was reported in 2019 to contain over 395,000 documents, 21,000 images, and 1700 databases documenting around 1100 archaeological projects [55] (p. 42). Turning from national collections to single archaeological

projects, for example, a network analysis of connections between sites in southwestern USA employed a database of over 4.3 million ceramic artefacts from more than 700 sites and more than 4800 obsidian artefacts from 140 sites [56] (p. 5785). Similarly, a project examining changing distributions of settlements, cemeteries, field systems, and finds across England from c.1500BC to AD1086 employed a database of over 900,000 records collated from a wide variety of different sources, including regional Historic Environment Records and the Portable Antiquities Scheme [57] (p. 245), constituting over 3 GB of data and over 100 GB of GIS data [58] (p. 298).

These examples certainly represent large datasets in archaeological terms, and the availability and accessibility of increasing quantities of archaeological data through digital infrastructures has made a significant difference to research and desk-based assessments, as well as to archaeological practice more generally. As a result, there has been some debate within archaeology in reaction to the growth in use of large if not “big” datasets in archaeological analysis (e.g., [12,13,59,60]). Another example of concepts and associated technologies borrowed from beyond the immediate discipline, “big data” has become more than a buzzword. As in the wider world, it can be seen as representing a mythology in which data are framed as raw materials to be mined alongside a corresponding shift in theory and methodology (e.g., [12] (p. S13)). For example, Holdaway et al. warned that one attraction of big data is to make archaeology appear more scientific and more suited to predictive and probabilistic inference [59] (p. 874). It is also closely associated with the resurgence of artificial intelligence and the expanding use of neural networks and deep learning technologies, which are dependent upon large datasets and can be seen in part as a rebranding of big data (e.g., [61]). The narrative of big data claimed that comprehensive coverage of a topic was possible due to the quantities of data available making sampling unnecessary; in turn, such quantities meant that biases were removed and errors compensated for in the data, and analyses focused on the search for correlations within the data, (in)famously characterized by Anderson’s declaration of the “end of theory” [62] (see [12] (p. S13)).

Whether archaeological data truly constitutes “big data” is open to debate (e.g., [12,13]), but Ruppert argues that the force of the big data imaginary is in the adoption of new mindsets and paradigms, and the way in which data practices are reconfigured:

“... they not only shape what is thinkable but also the practices through which actors perform them. So, while some commentators declare big data as hype, these pronouncements underestimate the material and political effects of imaginaries as they are taken up in practices through which new ways of thinking are propagated.” [63] (p. 635; see also [12] (pp. S12–S14))

While there have been warnings in archaeology about the risk of gathering larger and larger quantities of data in the assumption that masses of data enable it to largely speak for itself (e.g., [8,12,64]), the technologies themselves are embedded with a big-data data-driven ethos and consequently influence approaches to data and analysis. For example, a common feature of digital imaginaries is the perception that the technology provides a neutral means to an end, but this overlooks the potential drawbacks and theoretical/methodological implications that lie beyond. In archaeology, for instance, as long as geographical information systems were primarily used as a means of mapping data, they were seen to present few problems and many advantages. However, the analytical tools provided by GIS presented greater challenges in terms of embedded theory and methods and required a more critical approach (e.g., [65,66]). In the same way, the idea of big data as being little more than a capacity to search, aggregate, and cross-reference large datasets (e.g., [67] (p. 663)) may imply that big data practice is simply a case of doing the same thing as before, only with more. However, the imaginary of big data reveals that the underlying changes are more radical and more broadly affect the attitudes and methodologies applied to data (e.g., [61,67]), which offers profound challenges to established methods as well as potential new opportunities for archaeological practice [60].

3.4. *The Data/Capta Imaginary*

How data is perceived in archaeology has fluctuated and changed for many years. In part, these changes mirror distinctions drawn between the sciences and humanities, and since archaeology sits on the cusp of both, it might reasonably be expected that archaeological approaches to data would be complicated as a result. At various times, archaeology has taken a more scientific or a more humanistic position towards data, and currently both approaches (at least) to data are in use. For instance, Sørensen warns of challenges to data that he identifies with the “scientific turn”: the “unhelpful return to the ethos of letting ‘data speak for itself’ . . . because—as the popular legend goes—‘facts do not lie’ and thus become associated with ‘truth’ . . . quite often leading to a liberal, even careless, use of scientific data . . . ” [8] (p. 102).

One challenge is that the nature of what is meant by data in archaeology can appear at times to be self-evident and is infrequently discussed. The idea of “data” is often subsumed within higher-level discussions of the “archaeological record” and associated issues surrounding the deposition, survival, and preservation of material evidence, while remaining largely silent on the meaning of “data” itself (e.g., [68,69]). The terms “facts” and “data” are often used interchangeably (e.g., [70,71]), even as in “data: *see* facts” [70] (p. 184). However, the etymology of the terms is different: the origin of “data” is in the Latin verb, *dare*, to give, while “fact” is derived from the Latin *facere*, that which was done, occurred, or exists [72–74]. In the 17th and early 18th centuries, data referred to truths, or to claims that were accepted for the sake of argument—matters of faith or belief, even—not discovered through experiment or investigation. During the 18th century, Rosenberg identifies a shift in which data “came to evoke a particular sort of representational entity upon which one could operate through systems of calculation, classification, and communication, while holding the question of referential truth in abeyance.” [73] (p. 566). “Data” and “facts” could increasingly be substituted for each other [74] (p. 390). By the end of the 18th century, references to data had expanded from primarily mathematics and theology to medicine, economics, natural history, and geography, and its meaning had shifted to facts in evidence determined by experiment, experience, or collection, with data becoming the result of an investigation rather than its premise [72] (pp. 32–33). As Rosenberg describes it, data “went from being reflexively associated with those things that are outside of any possible process of discovery to being the very paradigm of what one seeks through experiment and observation.” [72] (p. 36). Data subtly shifted from being given, from being beyond argument, to being something that is captured or extracted through observation and computation.

Although Rosenberg’s characterization has received some criticism (e.g., [75]), the meaning of data has changed over time, and the shift he identifies from data as given to data as captured finds an echo in Chippindale’s contention that archaeological data

“ . . . are not data at all, for they are practically never given to us by the archaeological record. They are actually *capta*, things that we have ventured forth in search of and captured with all that the idea of capture implies; hunting is a dangerous and uncertain business in the rugged country of archaeological material.” [76] (p. 605, emphasis in original)

Although there was no subsequent published discussion following Chippindale’s Forum piece in *American Antiquity*, the idea of data as *capta* was later developed by Drucker, who argued that humanistic enquiries should acknowledge “the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, *taken*, not simply given as a natural representation of pre-existing fact.” [77] (para. 3, emphasis in original). She concluded that “*Data are capta*, taken not given, constructed as an interpretation of the phenomenal world, not inherent in it.” [77] (para. 8, emphasis in original). The fact that the term “data” is still used in preference to “*capta*” might suggest that archaeologists—and humanists more broadly—recognize data as *capta* but continue to use “data” to avoid confusion (c.f. [10] (p. 5), [75] (para. 7)). In reality,

this serves to disguise different attitudes to data and to theory in general: for instance, realist, positivist, or modernist (data as “given”) versus constructivist, phenomenologist, or postmodernist (data as “capta”) approaches (e.g., [77] (para. 3), [78] (pp. 111–114)).

The importance of the concept of “capta” in relation to data lies in its interpretative character. Hence, Drucker describes it as

“... situated, observer co-dependent, and partial. Its variables are, in theory, infinite, but they are always present in some degree or measure by virtue of the performative and participatory character of interpretative information. Interpretation depends upon and is an expression of an individual reading in a particular set of circumstances and never presumes to completeness or observer independence.” [77] (para. 29)

This resonates with approaches to data in archaeology. For instance, in their discussion of excavation strategy, Andrews et al. argued that “Obviously all observations involve interpretation, simply because observations become comprehensible when they confront the pre-expectations which are held by the observer” [79] (p. 526). Similarly, Wylie argued that “what archaeologists recognize as data, and what they infer to be its evidentiary significance, are necessarily functions of the “pre-understandings” they bring to bear” [4] (p. 204). This sets up what Wylie identifies as an “epistemic anxiety”: that archaeological data are of value only if they can be used as evidence, but those same data only exist because they conform to the pre-expectations and conventions of archaeological practice, in what she describes as a “vicious circularity”. Leonelli sees this as a paradox, namely that

“... despite their epistemic value as ‘given’, data are clearly made. They are the results of complex processes of interaction between researchers and the world, which typically happen with the help of interfaces such as observational techniques, registration and measurement devices, and the rescaling and manipulation of objects of inquiry for the purposes of making them amenable to investigation.” [80] (p. 813)

Likewise, Buccellati identifies an archaeological paradox in which data only come into existence through observation rather than being given [81] (p. 6).

To put it simply, therefore, two distinct approaches to data in archaeology can be broadly identified (following Kitchin [82] (pp. 17–22)). On the one hand, data are seen to be empirical and unarguable, while on the other, data are seen to be interpretative, negotiable. The first perspective sees data as relatively straightforward raw pieces of information about the world, so that where instruments are used to capture the data, what they capture is representative of pre-existing (given) characteristics and they do so in a neutral, objective manner independently of any philosophical thought. The role of the archaeologist is to use the most representative means of capturing the data and, as far as possible, creating as complete a record as is feasible, at the same time minimizing any errors or biases in the data. These data provide the groundwork for subsequent interpretation and understanding which is therefore more reliable and consistent. The second perspective sees things rather differently. Data do not exist independent of the archaeologist: they are brought into being through archaeological practice. The research agenda, available techniques and technologies, physical and temporal constraints, conventions, learned behaviors, etc. all make a difference to what data are created, giving rise to multiple possible datasets. For instance, the same measurement can be captured by different instruments in different ways, giving rise to different resolutions in the outcomes. The archaeologist still seeks to work in a consistent, rigorous, professional, and objective manner but is limited by choices, restrictions, requirements, and resources. This means that data are situated contextually, culturally, socially, politically, technologically, and spatially.

In reality, however, the boundaries between these two positions are blurred, giving rise to the paradox that Leonelli identifies: that data are treated as pre-existing givens whilst at the same time being a consequence of human action. In archaeology, therefore, data are seen as primarily interpretive rather than being purely empirical [79] (p. 526),

while at the same time, essentially descriptive data are treated as objective empirical data. In the process, an interpretive archaeology is transformed into a (supposedly) objective, more scientific archaeology, and the value-laden, theory-laden, process-laden selection and organization of data is largely set aside. Chippindale associates this reversion of *capta* to data with the advantages of digital data:

“As the manipulation of data becomes easier, so it also becomes easier to treat the data as given things rather than to enquire after just what these given things are, just where they come from, just what uncertainties, assumptions, classifications, and concepts their created existence depends upon.” [76] (p. 609)

This is a consequence of Sørensen’s scientific turn—or, more specifically, a computational turn in which data are processed as symbols and numbers (c.f. [83] (pp. 27–30). “Data” becomes employed “to describe the values on which computation is effectuated independent of any question of representational truth” [74] (p. 390), as is frequently seen in “big data”-style analyses. In the process, data are treated as a neutral and objective reflection of the archaeological record (c.f. [84] (pp. 71–72)).

3.5. *The Distant Data Imaginary*

A particular characteristic of digital data is a distancing effect: the heady combination of the availability, accessibility, and quantities of data making data more available as it paradoxically becomes more remote. As a consequence, “increasing access to increasing amounts of data has to be set against greater distance from that data and a growing disconnect between the data and knowledge about that data” [30] (p. 13). Ellul, for example, observed how proximity to technology can replace knowledge of the primary elements from which things are made, occasioning “profound mental and psychic transformations which cannot yet be assessed” [85] (p. 325), and the deskilling, alienation, and distancing effects of technology has been a common critique in recent years. For instance, Ingold argues that “... the project of technology has been to capture the skills of craftsmen or artisans, and reconfigure their practice as the application of rational principles whose specification has no regard for human experience and sensibility.” [86] (p. 61). For example, the interface in a digital environment mediates the range of negotiations between investigator and data. Drucker observes that “*No single innovation has transformed communication as radically in the last half century as the GUI.* In a very real, practical sense we carry on most of our personal and professional business through interfaces. Knowing how interface structures our relation to knowledge and behaviour is essential.” [87] (p. vi, emphasis in original). This interface operates both at the physical level (the screen, the keyboard, the mouse) and at a more metaphysical level (the operating system, application software) and “... can easily resemble a device of capture, as a journey is plotted out through which mechanisms are triggered and set off that reduce autonomy into zones of non-knowledge.” [88] (p. 173). This can promote a “push button” approach in which complex routines can be easily applied with little understanding of their background, operation, or consequences (e.g., [89] (sect. 5), [90] (p. 131)).

This distancing can be visualized in several ways. For example, consider archaeological field data, increasingly commonplace in digital archives. It is often considered that such data starts with the material evidence in the ground. For instance, Buccellati argues for the primordially atomic nature of the archaeological record in which “we do not fragment an observed whole, nor do we impose an analytical fragmentation” [81] (p. 234). However, from a “*capta*” perspective, such data are not the start of the process but a consequence of multiple decisions, predeterminations, and perspectives which precede the moment of capture and impose constraints on the observations recorded as data. For example, Carver’s Field Research Procedure (e.g., [91] (pp. 35–37)) (see Figure 1) characterizes the sequence of reconnaissance, evaluation, and project design that come before and define and structure the nature of the investigation itself. Each of these prior stages themselves entail data creation and data analysis. For instance, evaluation may include desktop assessment, the creation of a deposit model, and the creation of a research agenda that

will then influence the project design which in turn establishes the strategies for data acquisition in the field. These field data are consequently enmeshed in a web of other data, decisions, and determinations, and cannot be seen as truly primordial. The field data are themselves influenced by a combination of cultural and taphonomic processes, and the observer articulates their knowledge based on experience, research objectives, professional expectations, customary actions, commercial and temporal constraints, and so on, in order to identify, select, and categorize the data (e.g., [92,93]). Once recorded, those data records are then drawn together through a process of interpretation and analysis into a data-structures report or its equivalent, which in turn provides the basis for a site report. This may subsequently be incorporated into a synthetic volume—a period or regional-based study, for instance. It is easy to see how the report is much more remote from the original site data than, say, the context record. However, consider what happens when other interventions occur: for instance, the synthetic work is incorporated into other studies, or the grey literature report is referred to in another project or another synthetic work. These are still further removed from the original site data, often building on interpretations of those data rather than the data themselves. Additionally, of course, it is becoming increasingly common to deal with data at second or third hand (or more) rather than going back to the primary data, which themselves are only a proxy for the original material evidence which is long gone.

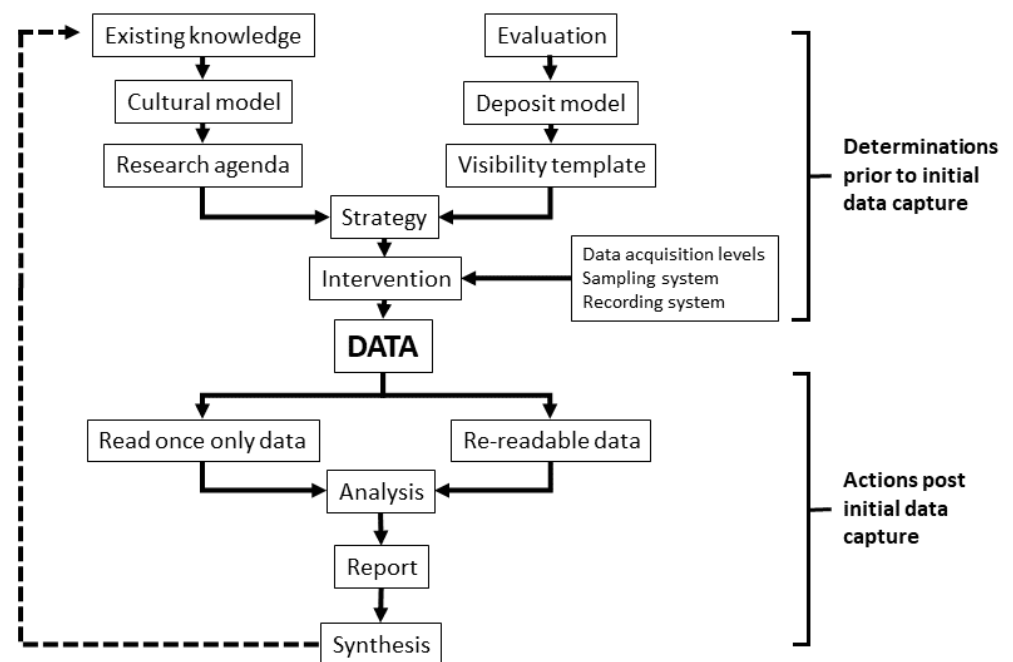


Figure 1. Data relations within Carver's Field Research Procedure (adapted from [94] (p. 136)).

Another way of visualizing this distancing is to consider the relationship between archaeological practice and the data it generates (see Figure 2). This model illustrates the transitioning of archaeology from a craft practice into a more automated, computerized form [95] (pp. 418–419). The model tries to show that the different categories or elements of practice overlap with each other, and that practice may operate at one or more stages depending on factors such as the resources available, experiences and preferences of the practitioners, and professional expectations. From a digital perspective, there are also chains of dependency—the automation of a particular task requires there to be systematization and standardization, for instance. Additionally, there are questions of the changing relationships between practitioner and the digital—where the balance of agency between the two is different across the model. However, there are also implications for the perception and handling of data in this model. For example, at the craft level, there is close physical proximity to the data—direct personal observations made in the encounter with the material

evidence. Standardization introduces a degree of distance from the data through a closer definition of what is of interest and how it should be captured, and hence it can be perceived to be in tension with craft practice. Systematization sees the introduction of tools that increase distance again—for example, using photogrammetry rather than manual drawing to capture sections and surfaces. Finally, automation introduces still further distance, with technological devices performing the encounter with the archaeology and capturing data about it.

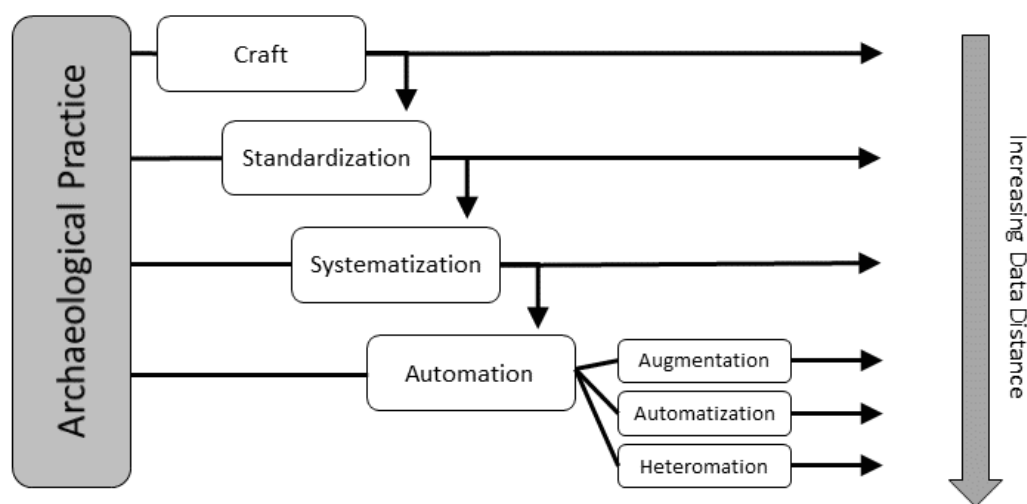


Figure 2. Archaeological practice and distance from data (adapted from [95] (p. 418)).

The intervention of technology can therefore introduce distance into the data capture process. For instance, there is a growing trend for the use of digital photogrammetry as a replacement for traditional section drawing and planning, in which imagery and 3D models produced through digital photogrammetry are used in the field to draw the contexts and to record the textual and graphical information describing the context in the database (e.g., [96–98]). The Aide Mémoire Project [99] has begun to look at the implications of digital versus hand-drawing for archaeological understanding, and their initial survey results speak to a sense of distance introduced by the digital application. For example, respondents commented that “I find it’s much easier in my experience to simply switch off and fall into the mechanical click-click of photogrammetry” and “Drawing a context is like reading a page in a book, photographing (and subsequent tracing) is like photocopying that page.” [99] (p. 12). Another respondent said, “If drawing starts to be seen as more akin to transcription than to translation as a result of digital tools, I think we will lose perspective on our own processes of knowledge generation”; and finally, “We’re never capturing reality; at least hand drawings are unfailingly honest about that.” [99] (p. 14). Such perspectives capture a clear sense of an arms-length relationship through technology which fosters distance or remoteness from the data.

4. The Many Characters of Data

These instances of data imaginaries combine in different ways and in different situations to contribute to the archaeological data imaginary. In the process, they underline the inherent diversity and complexity of archaeological data in different contexts ranging from the empirical/interpretative nature of data, attitudes and approaches to data, proximity to data, and ultimately the construction of infrastructures that enable subsequent access to and reuse of data. Huvila [100] characterized the idiosyncratic nature of archaeological information (or data in this context), arguing that

“The heterogeneity and fragmentary nature of archaeological information, destructive nature of archaeological work, coexistence of multiple epistemologies and standards of information work and representation of information and long

temporal time span of the archaeological subject matter and archaeology itself all impede effective and efficient management of archaeological information.” [100] (p. 158)

The survey of digital archaeological repositories reported by Geser et al. [39] underlines some of the challenges associated with the management of such diverse data.

However, it is possible to see archaeological data as more exceptional in character than it really is. For example, the uniqueness of archaeological data derived from the “excavation is destruction” trope is frequently employed as a justification for the preservation of remains or intervention in the field. It has also played a crucial role in calls on funding agencies to support digital archival infrastructures given the added fragility of digital data and the need to actively curate it (e.g., [101,102]). Yet, although great play is made of the distinctiveness of archaeological data and the unrepeatability of its data collection process, it is not alone in this regard, and other disciplines might make equal claim. For instance, a musicologist could point to the uniqueness of a live performance, or an ethnographer to the interview event occurring at a single place and time. Similarly, the range of epistemologies and data standards in archaeology is not an exceptional feature amongst humanities disciplines, nor is their frequent conflict with the generic digital systems and infrastructures which are built upon formalized data descriptions to support data sharing, for example. The long temporal scope of archaeological data is also not in itself an exclusive characteristic, nor is the fragmentary nature of archaeological data and its inherent biases, as parallels for both can be found in other disciplines.

What does make archaeological data unique, however, is the way in which it is used as a proxy for past human behavior, or, as Huvila puts it, the fact that archaeological data is seldom directly informative about matters of archaeological interest [100] (p. 153). Data recovered by archaeologists stands as proxy for the complexities of human social and economic activities in the past (e.g., [103] (p. 12)). For example, quantitative measures may stand for qualitative aspects, such as GIS distance measures as proxy for a sense of place or knowledge about a landscape, or friction calculations as proxy for accessibility within a landscape. Elsewhere, material remains act as proxies: the presence/absence of artefacts as indicators of trading networks, the substitution of one characteristic artefact with another associated with population replacement, the style and decoration on artefacts used as signifiers of social or group identity, and so on. Proxies are also extensively used within more scientific approaches in archaeology, with, for example, environmental change through time assessed through proxies including pollen data, tree rings, and sediment analysis, for instance. Multiple proxies are frequently used in conjunction with each other, such as in GIS studies, network analyses, and, indeed, in “big data” analyses with their search for correlations. However, the interpretation of the outcomes of such proxy data investigations is highly influenced by the nature of those same data, to the extent that it may never be possible to dispense with “controlled conjecture” [104] (p. 5) because of the fragmentary, partial, and temporal aspects of the data, which means it is both representation and sample. Collins [105], for example, identified what he called a “sequence of contingencies” which intervene between past human action and the archaeologist’s perception of that action, and which emphasize the incompleteness of archaeological data (Figure 3). The consequence of the range of cultural, taphonomic, and archaeological factors mean that much is unrepresented, unrecognized, and unknown, and hence archaeological data is “haunted by absences” [69] (p. 178), “shadowy” [4] (p. 204), and characterized by ignorance and silences [106]. As Chippendale describes it, “Archaeology is plagued in many an instance with poorly defined variables (usually thought of as ‘data’) drawn from ill-understood populations, and with uncertain articulations between the entities whose logical relations we seek to understand.” [76] (p. 611).

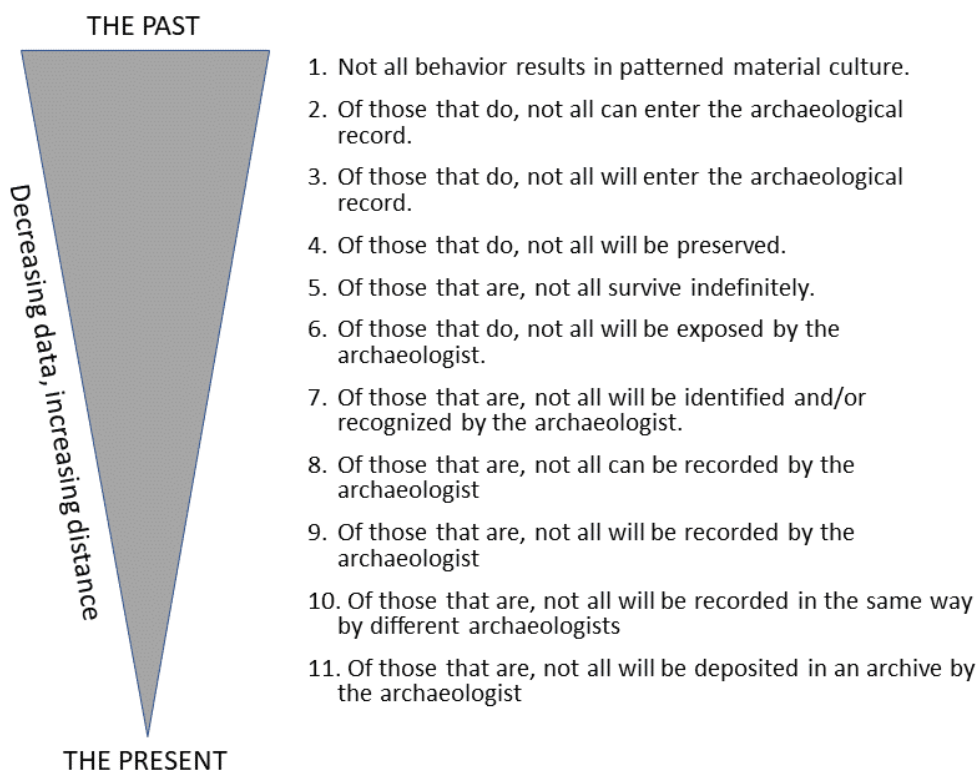


Figure 3. An extended version of Collins’s “sequence of contingencies” [105] (stages 1–7 from Collins) and its relationship to data.

The sequence of contingencies that constitute archaeological data is compounded with the recognition that in different situations and at different times, the character of data itself may take on a range of different forms and potentials, sometimes simultaneously and even contradictorily. For example, Kitchin identifies a range of *technical* categories of data which influence their subsequent use [10] (pp. 6–11). For instance, data may be *quantitative* (numeric) and within this may be nominal, ordinal, interval, or ratio data, or *qualitative* (non-numeric textual or image data). Data may also be *structured* (fitting a data model and typical of a relational database), *semi-structured* (loosely structured without a data model but capable of a reasonably consistent set of categories as seen in XML tagging of text), or *unstructured* (highly variable data lacking a data model and defying categorization). Data may also be captured *directly* through observation or survey, for instance, or *indirectly* as a by-product of another function or purpose (for example, the archaeological use of satellite imagery, LiDAR, or seismic data, originally captured for other forms of prospection). Data may be *primary* (pertaining to the direct observation of the material archaeological evidence and/or data created by the researcher for a specific program of work), *secondary* (made available and capable of reuse by others for replication or reanalysis), or *tertiary* (derived from the analysis and reworking of data by others, typically in the form of summary or amalgamated data). There may be *indexical* data (such as context numbers and small-finds numbers, which enable the unique identification of data objects and their subsequent linking and processing), *attribute* data (aspects of the phenomenon being examined), and *metadata* (data describing the data). Metadata itself falls into several categories: *descriptive* (e.g., ownership, authorship, subject, description, location), *structural* (e.g., organization and coverage of the data), *administrative* (how the dataset was created, file formats, permissions, etc.), and *process* metadata, or *paradata* (decisions and processes surrounding the collection and processing of the data).

Other categorizations of data are also possible. For instance, Marwick and Birch identify four different types of archaeological data, although they emphasize that their list is not comprehensive [42] (p. 126) (see also [50] (pp. 23–24)). *Observational* data

includes site and artefact descriptions and measurements, *computational* data includes the outputs of simulations and computer models, *experimental* data includes results from laboratory procedures and field experiments, and *records* are largely concerned with textual documentation. Elsewhere, for example, Purdam and Elliot [107] (pp. 28–29) define an eight-part typology of data for social scientists based on the way data are generated, which overlaps in certain respects with the categories identified by Marwick and Birch. Then, of course, there is the classic division into *raw* and *derived* data. Leaving aside the observation that the concept of “raw” data is an oxymoron and that all data are cooked in some way [108], [109] (p. 184), this seeks to distinguish between data in its most basic, unprocessed form (typically quantitative values resulting from direct measurement) and data that has been manipulated, cleaned, and processed into what might otherwise be called information (e.g., [30] (pp. 15–17)). As Longino [110] (p. 397) suggests, the term “raw” attempts to get back to an origin point where data is least processed. Although it might be expected that the unprocessed data would be where the true value of the data lies, it is often the subsequent processing of the data that is seen to add value, even to the extent that the processed data becomes perceived as the more important product and the “raw” data is set aside in its favor.

Furthermore, digital data can be categorized according to its *format*, for instance: XML files or word processor formats containing textual and semi- or unstructured data; various database formats, comma-separated values, or delimited text files containing structured data; shapefiles, geography markup language, or georeferenced TIF files containing spatial data; and JPEG or uncompressed TIF files containing photographic or image data (e.g., [111]). Each format carries its own implications, requirements, and constraints. For example, data tables in the form of comma-separated values files will require separate documentation to describe the relationships between tables that were originally incorporated within the proprietary database formats. Other filetypes comprise multiple files—for example, a shapefile consists of three mandatory and up to seven optional datafiles, all of which must be present for the data to be usable without problems. Some file types may also contain information about their character and origin: for instance, GIS geoprocessing tools may update associated metadata with details of the processing applied to the data, while EXIF metadata is captured in digital photographs containing information about shutter speed, focal length, exposure, time, date, and so on.

What these characterizations highlight is that digital data only exists as a relatively simple conceptual entity at a very high level of abstraction. Drill down into the character of data and it becomes apparent that, far from being “a thing”, it is at once many things, consisting of a rich and complex set of characteristics which are combined and recombined in many ways during its lifetime. Manovich, for example, highlights the variability of digital media which can exist in different, potentially infinite versions, and the way that it can be transcoded, or translated, into different formats [83] (pp. 36–47). So rather than a static unchanging entity, data is unstable, capable of translation from one mode to another, from one format to another, from one category to another, for one purpose or another. So, for example, qualitative data is frequently transformed into quantitative data for analytical purposes using a numeric code to replace a descriptive attribute in a process which entails abstraction and a reduction in the information carried. Resolving data differences arising through variable data recovery and multiple recording methods is a commonplace task in archaeological meta-analyses [106] (pp. 6–7), together with the host of data-cleansing techniques which transform data in often poorly documented ways. Data in proprietary file formats is translated into open formats for archival purposes, which preserves the data but may lose relationships or other characteristics in the conversion, requiring accompanying documentation to fill the gap. Semi-structured or unstructured data may be incorporated within a structured data model, in the process smoothing out variability or omitting aspects considered to be unimportant for the study at hand, or which simply do not fit within the schema (e.g., [41]). The provision of directly captured data, typically from instruments, may be set aside in favor of processed datasets (e.g., [112] (Section 3.1)) although not in all

cases (e.g., [113] (Section 3.1)). The distinctions between primary, secondary, and tertiary data highlight the transitioning of data from what was originally collected (so-called “raw” data) and what is subsequently processed and interpreted (e.g., [50] (p. 27)), incorporating a range of data transformations in the process. Add to this volatile mix the sequence of contingencies classically associated with archaeological data, together with the fact that the same data entities may be evidence for multiple phenomena [50] (p. 28), along with the fact that what is perceived to be data will depend on the context of use and the preconceptions of the users, and the result is a highly complex, highly unstable, and highly unpredictable product.

5. Data Travels and Frictions

This mode- and medium-switching capacity of digital data can be seen as a strength as well as a weakness. As Leonelli has argued, what makes data so powerful as sources of evidence is their mutability: “The multiple ways in which they are transformed and modified to fit different uses as they travel across space, time and social situations.” [114] (p. 6). Their mobility and capacity for adaptation is what makes it possible for data to be reused, reanalyzed and recycled in different contexts. Data “acquire or shed components, merge with other data, shift shape and labels, change vehicles and companions, and such transformations prove essential to their usability by different audiences and purposes” [114] (p. 6). This contrasts with Latour’s perspective of data as defined by their mobility, while their power is derived from their immutability or unchanging nature. To Latour, data are “immutable mobiles” [115] which circulate without substantive change in themselves. Instead, the potentiality of data—or the idea of data as prospective evidence [116] (p. 77)—means that its value as data at any point in time or place is not certain or fixed, that what actually counts as data can change at any stage ([117] (p. 16)), and it may only become clear at some unspecified future point, all of which makes the initial capture of data, its subsequent curation, and continued accessibility crucial interventions for future knowledge creation. Indeed, Leonelli proposes that digital repositories do not actually preserve data but

“... objects which may or may not be used as data (or data sources) ... as soon as the effort is made to use such objects as data or acquire data from them (for example, through measurement), they are at least minimally modified to fit the ever-evolving physical environments and research cultures within which they are valued and interpreted.” [114] (p. 7)

The metaphor of *data journeys* has been developed in relation to the movement of data from their site of production to other sites of subsequent use within and beyond the same field of research [116] (p. 39) (see summary in [118] (pp. 3–4)). An archaeological analysis may typically employ data created by the analysts themselves, data created by other archaeologists, and data that travels into archaeology from other specializations and disciplines, for instance (e.g., [71,119]). Leonelli defines three stages through which data travel—de-contextualization and re-formatting, re-contextualization, and reuse [120] (pp. 4–5)—each of which entails packaging and repackaging of the data ([116] (pp. 24–31)). De-contextualization is seen to consist of metadata creation which enables the data to move beyond its original context and to be integrated with other data, while re-formatting ensures that the data is compatible with other datasets [120] (p. 4). Although this may simply imply shifting file formats, data conversion and standardization may be complicated by the risk of data loss or misinterpretation. Re-contextualization enables the data to be used within a new, quite possibly unanticipated, context and requires knowledge about the origins of the data, the procedures used in its collection, and any subsequent modifications, in order to evaluate the data and determine whether they are appropriate for the present use [120] (pp. 4–5). This appraisal is facilitated by the metadata created at the previous stage. Finally, reuse sees the data brought to bear in its new-found context, in new formats, in conjunction with other new datasets. This journey implies a reversal of the classic data–information–knowledge–wisdom model in that a would-be re-user of data approaches a dataset through

an understanding of the knowledge–information–data disarticulation of the original data creator [11] (p. 100) that is supposedly provided through the accompanying metadata.

This data journey model is not without its problems when applied to the archaeological situation and in certain respects has higher expectations of data than is often the case. For instance, most archaeological metadata associated with datasets in support of their reuse typically focus on data discovery rather than how the data originated. Although paradata (process metadata or data provenance) may be seen as incorporating documentation about equipment, expertise, protocols, and procedures (e.g., [121] (p. 45)), if present at all, it currently tends to focus primarily on the means by which the data have been processed rather than the preunderstandings and constraints behind their original capture and processing [106] (pp. 2–3). This introduces a remoteness from the data through a detachment from and consequent loss of the implicit, tacit knowledge concerning prior decisions and limitations. These may also be lost during the de-contextualization as the data and metadata are classified and characterized, which is not a neutral process. Accordingly, a perception may arise that the data is more objective, more reliable, and more complete than may be the case.

A useful metaphor related to the idea of data journeys is the concept of *data friction*, originally characterized by Edwards [122] in the context of climatology and subsequently applied to other fields. Data friction

“... describes what happens at the interfaces between data ‘surfaces’: the points where data move between people, substrates, organizations, or machines—from one lab to another, from one discipline to another, from a sensor to a computer, or from one data format ... to another ... Every movement of data across an interface comes at some cost in time, energy, and human attention.” [123] (p. 669)

Edwards also characterized *computational friction*: “the struggle involved in transforming data into information and knowledge” [119] (p. 20), and related to data friction. Consequently, friction can be seen to be encountered across the range of different technologies, standards, and practices associated with the creation, processing, curation, and re-use of data, often drawn from different contexts, during its travels.

Although Edwards’s definition of data friction typically casts it as associated with the movement of data, friction may equally arise within the immediate context of data creation and use, before it travels into different settings. In archaeology, for instance, data friction arises through human agency in terms of both the original deposition and eventual materialization of what might become data, as well as through the range of non-human agencies that determine what might survive as potential data. For example, Edgeworth describes the affordances and constraints experienced during the act of excavation which influence the archaeological encounter [93,124] and which may be seen as resulting in friction in the recognition and capture of data in the field. Collins’s “sequence of contingencies” (Figure 3) can therefore be viewed through a different lens as data friction in action, long before recorded data moves beyond its creators for use by others. Common data practices during archaeological data capture can also be seen as attempts to reduce friction. For example, standardization across pro-forma recording systems and the use of recording templates seeks to normalize data from within a specific site to ensure that defined characteristics are captured, and subsequently eases the movement of data from different sites into centralized organizational recording systems and potentially thereafter into large-scale curatorial infrastructures. Data friction is also experienced in the translation of datasets, whether salvaging old data or repurposing data (e.g., [125,126]). For instance, Huvila has observed the conditioning of archaeological information work by paper-based norms [127] (p. 235), which gives rise to friction through its digitization. Similarly, Harrison describes the ways in which steps to clean and standardize an old dataset masks inconsistencies and gaps in the data in order to increase interoperability and access. She argues that transparency about methods of data cleaning and manipulation builds confidence in the data [126] (p. 89), and thereby helps to reduce data friction. Interfaces to data also introduce friction in both initial recording and subsequent reprocessing and in the integration of data with other data,

through the ways in which they constrain and structure as much as they facilitate access, framing interactions and representations.

Data friction may also be characterized in socio-political terms, arising between people, organizations, and infrastructures in the form of conflicts, frustrations with processes, variations and disagreements over practices, and so on (e.g., [123] (p. 669), [128] (p. 425)). For example, Kansa [129] describes a range of issues affecting archaeological infrastructures, such as the ethical, legal, and practical challenges of data, ownership, power inequities, and the consequences of dependencies on both commercial infrastructures and open-source software libraries, all of which might also be seen as contributing to data friction. Many of the archaeological responses to digital agency may also be linked to friction, including the subversion of recording systems, failures to streamline recording methods, and delays introduced in primary recording practice, for instance [95] (pp. 422–423).

Friction is generally presented as a negative, problematic characteristic to be resolved, with benefits accruing from its resolution. However, as Borgman [50] (pp. 68, 80) points out, ontologies, taxonomies, thesauri, metadata structures, and the like are developed to facilitate interoperability and hence reduce friction, but at the same time, these systems may create friction in terms of incompatibilities between schema, conflicting standards, variability in data sharing arrangements, and emphasize boundaries between different communities and domains. The mental and labor costs involved in the creation, handling and management of metadata is itself a form of friction—*metadata friction* ([122] (p. xvii); [123] (p. 673))—and the same can equally be extended to the development and maintenance of thesauri, wordlists, ontologies, etc., which may disincentivize their use in overcoming frictions (for example, see [130] on the development and implementation of the CIDOC CRM ontology). Similarly, the kinds of challenges identified by Kansa [126] in relation to the archaeological use of commercial data infrastructures and open-source software underlines this conflicting aspect: they may overcome friction and introduce value-for-money in terms of what they enable, but at the same time they may reintroduce friction through the constraints and risks which they bring.

From other perspectives, however, friction can be seen in positive terms rather than something that is necessarily required to be overcome. For example, Caraher's characterization of a "slow archaeology" [131,132] emphasizing the humanity of archaeological data over their complexity is supported by others who highlight slow approaches to field methodologies (e.g., [133,134]), for instance, or to the design of interfaces which encourage slower, more reflective engagements with data (e.g., [135]). Friction in these terms is seen to slow down processes with beneficial results, allowing time and energy to reflect and engage with data, effectively as a response to the data imaginary. Slow approaches to data help to reveal the forgotten or overlooked aspects of data, to expand knowledge of the data, to reflect on the value of data and the practices that generated them, to engage with the collaborative aspects of data, to participate in the thoughtful reworking of data and its evolution and transformation, and to create narratives about the data [13] (pp. 103–105). In this light, slow data practice can be seen as a form of positive data friction which draws attention to the dependencies, shortcomings, and uncertainties of data on their journeys through the range of socio-technical assemblages and infrastructures [136] (p. 276).

6. Case Study: Journeys and Frictions in Grey Literature

Archaeological data have typically journeyed from their original identification through to their recording, processing, reporting, and ultimate incorporation within a larger infrastructure awaiting their re-discovery. In light of the data frictions experienced en route, are data still capable of being re-used in full knowledge of their origination, their strategies of recovery, the procedures applied, the constraints experienced, and any modifications applied? Edwards [122] uses the term *infrastructural inversion* to describe this process of inspecting the original data on which subsequent analyses have been based so that confidence in the data can be gained through an understanding of the character of the data, including any inconsistencies and variability: "data aren't data until you have turned the

infrastructure upside down to find out how it works" [122] (p. 20). For instance, one of the transformational bodies of data increasingly used within UK archaeology is the collection of unpublished ("grey literature") archaeological site reports which have become important as a reference source for new archaeological investigations including pre-development assessments (the origin of many of the grey literature reports themselves). They also provide a resource for regional and national synthetic studies, and increasingly for automated data mining to extract information about periods of sites, locations of sites, types of evidence, and so on (e.g., [137]). Despite this, archaeological grey literature itself has not been closely evaluated as a resource for the creation of new archaeological knowledge. Such reports on their own would typically be considered to represent secondary archaeological data, inserting a degree of distance between the descriptive discussion in the report and the original observations in the field. However, access to the site records behind the report should reduce this distance by enabling the discussion presented in the report to be evaluated to a greater extent, for instance by tracking where uncertainties in the primary record have been transformed into assertions in the report.

The Archaeology Data Service (ADS), based in the UK, currently holds over 64,000 unpublished archaeological reports in its digital Library (<https://archaeologydataservice.ac.uk/library/> accessed on 21 March 2022), covering a wide range of archaeological interventions including Schemes of Investigations for archaeological work, desk-based assessments, watching briefs, evaluations, excavations, geophysical surveys, environmental assessments, and artefact reports, for example. It should be emphasized that this collection of grey literature held by the ADS is in a state of near-constant flux, with over 1400 registered organizations and individuals [138] adding data and reports to the collection via the OASIS information system, which was developed to report investigations to regional and national Historic Environment Records and for subsequent deposit with the ADS. The collection therefore continues to grow rapidly in scale and scope with hundreds of records added or edited each week [138]. This means that a considerable investment in time and effort is required by the ADS to ensure that the various generations of reports and associated datasets are interlinked and cross-referenced, a process which may not necessarily be carried out at the time of deposit (Green, pers. comm.).

One means of assessing the nature of the data journeys and frictions experienced by such reports is to evaluate the extent to which the data embedded within them ("grey data") is capable of being re-used in full knowledge of its origination, the strategies of recovery, and any constraints which may have applied in advance, during, and after its collection and analysis. The grey literature report essentially packages the data collected through interventions, and its data journey can be unpacked by reviewing the original data in order to understand the travels and the frictions the data has experienced in its repackaging as a report. For this investigation, a small sample of unpublished reports that were accompanied by site archives was extracted, consisting of 15 cases produced in 2021 relating to two watching briefs, seven evaluations, and six excavations (see Appendix A for details). Eight archaeological commercial organizations were represented in this sample, three appearing more than once. Data friction was experienced from the outset since it transpired that the tool used to extract the sample did not retrieve records from the full collection of 64,000 reports, or from the set of 34,288 grey literature reports indicated as a separate resource in the ADS ArchSearch browser, but from an unspecified sample somewhere in between because of a variety of complex issues currently being resolved (Green, pers. comm.). Consequently, the sample is more random than anticipated yet at the same time limited by a range of less-than-random data issues dependent on a range of factors, all of which are presently hidden behind the interface. For the purposes of the present investigation, this is not an issue, but clearly a larger-scale study could not have complete confidence in the data as currently constituted. This conveniently illustrates the challenges associated with handling data from multiple origins, multiple authors, multiple depositors, and deposited at different times, and the frictions associated with their data journeys.

The Guide to Good Practice for archiving excavation and fieldwork data produced by the Archaeology Data Service [139] recognizes that different scales of archaeological intervention require different levels of digital archives. An *index-level archive*, derived from a desk-based assessment and/or evaluation or watching brief, primarily consists of narrative reports (including the initial project specification and project design documents), together with tabulated lists of contexts, artefacts, etc. [139] (pp. 16–19). This index-level definition provided the basic template for submissions via OASIS [139] (p. 19). Beyond this, the *assessment-level archive* was defined for where work proceeded beyond an initial evaluation [139] (pp. 19–20). This includes an updated project design, the report itself, specialist databases, the site matrix (as a text or CAD file, or scanned original), the context database, and the location of all interventions (as coordinate lists, CAD file, or scanned trench plan). The *research-level archive* is reserved for significant excavations and consists of a more extensive and complex archive [139] (p. 20). On this basis, it might be expected that the archives associated with the reports in the current selected sample would be situated at the index or assessment level, depending on whether the intervention was a watching brief, evaluation, or excavation.

At face value, few of the cases conformed to the definition of an index-level archive, since all were missing one or more elements. Insofar as can be judged by eye, all reports in the sample incorporated digitally created location maps, site plans, and feature/section drawings, alongside the report text. Details of contexts, artefacts, etc., were generally provided in summary tables in the reports but were omitted in one instance. All the site archives accompanying the reports contained photographic image data, primarily of excavated features, although in 8 of the 15 cases (including several excavations), these photographs were the sole category of data included in the archive other than the report itself. The remaining seven cases, consisting of watching briefs, evaluations, and excavations, included scanned original context sheets and catalogues/registers of drawings, artefacts, etc., alongside the images. Only one (an evaluation) included a full-context database as a set of comma-separated files with an accompanying entity relationship diagram. Five of the seven cases with scanned context sheets or a database also included CAD data (generally site plans rather than section drawings, etc.), while a sixth provided zipped GIS shapefiles rather than CAD data. Since all the cases appeared to have used CAD/GIS to create the various maps and plans in the reports, their absence from almost two-thirds of the site archives in the sample is notable. This may be because more data remains to be added to the archives in the future, and in relation to site location maps, it may be a result of the use of copyright Ordnance Survey map data, which was not licensed for reuse. However, in their current condition, data which was evidently available to the authors of the reports in this sample was not included in the accompanying archive. In six cases, the Scheme of Investigation (setting out the scope and strategy for the work) was either included as an appendix to the report or provided separately in the site archive. Prior reports (for example, desk-based assessments, surveys, etc.) were often referenced in the report text but in most cases, these have not (yet) been deposited in the wider ADS Library, although this may change with time. Although in general the reports themselves conformed to templates specific to each originating organization, the same is not true of the site archives: there was no clear pattern of inclusion/exclusion relating to authorship, scale of investigation, or depositing organization.

These results broadly confirm observations elsewhere. For example, in 2017, Richards [140] reviewed the state of archiving at the assessment level and concluded that few projects met those minimum standards, with stratigraphic matrices rarely preserved and specialist spreadsheets of finds and animal bones not regarded as part of the core archive. Instead, project archives more often consisted of a text report and a collection of photographs of trenches and features [140] (pp. 229–230), as is the case with the examples examined here. Richards suspected that more digital data was likely available but was not included, partly as a consequence of cost, and concluded that the definition and enforcement of a minimum standard for an adequate archive was one of the greatest challenges

facing archaeological curators and repositories [140] (p. 230). The present sample would seem to support this argument: as presently constituted, there remains variability amongst the site archives for similar investigations of a similar scale, and similar variability even within deposits made by the same organization, although as noted earlier, this may change as records are added and updated.

In many respects, therefore, archaeological grey literature and its associated site archives form another data imaginary, consisting of the expectations and realities of data, alongside the social practices, professional standards, and institutional requirements associated with its creation, archiving, and subsequent potential for reuse. Although the case study is small, there is considerable variability in terms of the journeys and frictions experienced by the data, most of which are largely tacit at best. Even when the data providers are the same people or organizations, a variety of significant frictions are evident, and the journeys represented by the individual cases vary to a considerable degree. For example, the study highlights evidence for hidden decisions taken surrounding the inclusion of data, without revealing the reasons behind those decisions. It also demonstrates how aspects of the data might be obscured behind opaque interfaces or omitted from the archive altogether, restricting understanding of the specific data context and increasing the risk of misuse or misrepresentation. The presence of grey literature reports without site archives, and the variability of those archives where present, makes it very difficult to trace the data journeys involved in those archaeological interventions. In a very real sense, the ultimate data friction is when that data is absent. Where reports of earlier evaluations, desk-based assessments, surveys, and evaluations were also absent, a full understanding of the contextual and situated nature of the data becomes even more difficult. Although the standardized format of the reports means that they include aims and objectives of the work, it is difficult to trace how these may have changed over the duration of the archaeological inquiries where those earlier reports are missing from the archive. Uncertainties about the nature of the data journeys or the data frictions experienced means that the range of entanglements present in the data may remain largely obscure, and the analytical selectivity contained within the reports is unable to be as easily evaluated as might be desirable. The inability at present to have clear knowledge about the data journeys and the presence of often tacit data frictions makes it difficult to be confident about the subsequent use of the grey data in grey literature and its incorporation into the kind of larger amalgamated datasets typically used in more remote, even automated, analyses.

7. Discussion

Markham describes data as “the material result of a series of choices made at critical junctures” [141] (p. 514). She emphasizes that data is a consequence of lived experience: the practices, skills, conversations, and categorizations which combine in the effort to capture the essence or meaning of the object of study. In the same way, archaeological data are a consequence of the lived experiences of excavators, surveyors, supervisors, finds specialists, illustrators, site directors, heritage managers, academics, archivists, and a host of others who engage with and influence archaeological data through its data journeys. However, archaeological data go beyond this in the sense that the lived experience is not limited to the present or the future, but also includes past encounters: archaeological data are the detritus of past lives as well as the analytical objects of investigators past, present, and future. The “sequence of contingencies” of archaeological data underlines that engagements with data take place within situated contexts in place and time, which influence what data survive, are recognized, and are considered important or relevant to capture, and what data are set aside, obscured, or otherwise forgotten (e.g., [106] (pp. 7–9)).

The archaeological data imaginary is interwoven with encounters such as these, which influence how archaeology manages the issues surrounding open data, the design and construction of digital infrastructures, the treatment and handling of data in analytical contexts, and broader philosophical and conceptual approaches to data. What is also clear is that many of these incorporate internal contradictions: solutions which seek to resolve

perceived problems or action desired outcomes, but which introduce new sets of challenges and often unforeseen consequences.

So, for instance, we might recognize the importance of not treating archaeological data as givens, as unproblematic or as self-evident, not expecting the data to speak for themselves, even if that makes the task more difficult and more time consuming. However, how can the creative or performative nature of data be represented? Additionally, how might this be operationalized? Documentation is the traditional means by which the range of syntactic, semantic, and coverage anomalies in the data ought to be resolved so that a subsequent (re)user can fully understand the decisions and methods that the data have undergone. However, documentation and its digital surrogate, metadata, place significant demands on time and energy. Acquiring the historical chain of events and actions back to the origins of the data is a non-trivial task, and attempting to reduce the labor demands of doing so by introducing automatic or at least semi-automatic methods is a largely unsolved problem. It also overlooks that much of this information is primarily tacit, rarely articulated, and in some cases impossible to be consciously represented (e.g., [106] (pp. 7–9)). There is a long tradition of archaeologists using unstructured recording methods during fieldwork or laboratory work, such as the use of diaries and notebooks, which record their thoughts, ideas, theories, mistakes, decisions, and actions, often alongside the more formalized proforma methods. However, such methods often sit uncomfortably within a commercial environment, and where they exist, such reflexive records tend to be reserved for the physical rather than the digital archive. Some digital recording systems have been designed with less structured recording in mind: for instance, data records in the Federated Archaeological Information System (FAIMS) are associated with annotation fields which are said to mimic the handwritten marginalia in proforma sheets [142] (p. 49). However, such approaches tend to take a limited perspective by focusing on the more easily recognized and codified forms of tacit knowledge [106] (pp. 9–11). At a simpler level, developing recording systems which maintain a versioning approach would enable revisions to the data to be automatically recorded so that new interpretations, alterations to descriptions, modifications to extents, and the application of processing tools would be logged as part of the data. This would still lack the benefit of marginalia representing the accompanying discussions and debates—the conversations at the edge of the trench [134] (p. 146).

Similar issues arise at the infrastructural level, where the potential distance from data is perhaps at its greatest and consequently the need for data about the data is crucial. Digital archives operate differently to traditional material archives, not only in the sense that they are not subject to the same physical constraints or costs, but in several uniquely digital respects. For instance, Blom identifies a fundamental shift between traditional archives, where content is distinct from infrastructure, and digital archives: where “once the archive is based on networked data circulation, its emphatic form dissolves into the coding and protocol layer, into electronic circuits or data flow” [143] (p. 12). This dematerialization is what enables the collapse of time and space that supports remote access, but at the same time it constrains access through the systems and algorithms used to manage and reveal data (as illustrated above). Furthermore, the digital content is frequently recontextualized as part of its data journey, which means that the archive is dynamic and constantly changing [144] (pp. 16–17), as revealed in the case study above. This recontextualization draws on the data for different purposes and presents significant challenges for would-be producers of metadata: “There is no unique minimal and sufficient set of metadata for any given dataset, since sufficiency depends on the use(s) to which the data are put.” [145] (p. 335). Perhaps inevitably, therefore, infrastructural metadata largely focusses on the support of the FAIR principles and is subject to its limitations. The handling of archived digital data is also unique in comparison to traditional archives: digital data is preserved by a combination of refreshing (to reduce the risk of data loss) and migration (to avoid technological obsolescence). Consequently, digital preservation is concerned with ensuring the authenticity of copies of the original data, whereas in a traditional archive, the emphasis is placed on preservation of the physical media (e.g., [146] (p. 46)). The original digital

data only exist as long as they remain accessible through their original technology, but ultimately it is copies of data which persist through multiple migrations and translations so that the original lives on only in reproductions. As digital archivists emphasize (e.g., [147]), this means that digital data does not lend itself to benign neglect, and yet the costs and efforts entailed in archiving the quantities of archaeological data can make it difficult to do more than store the data in the hope that a demonstrable demand for its reuse makes it feasible to properly document and archive them. In the end, this is what lies behind the idea that the best way to preserve archaeological data is to reuse it, and yet the ability to reuse it is critically dependent on the documentation efforts required to make it accessible. Moreover, the ability simply to access data is not sufficient in itself. Since data are prepared in specific contexts with specific questions in mind and using specific sets of criteria and tools, they may not easily be re-tasked for different purposes. This makes it important to be able to track back to the origins of the data in order to properly understand the full data journey through to the present rather than simply reusing high-level pre-cooked data.

8. Conclusions

What is currently lacking in archaeology is a detailed understanding of the journeys and frictions experienced by data. Considerable efforts have been expended in debating the nature of the archaeological record (e.g., [69,70]), but substantially less attention has been paid to the basis of that record: the character of the data themselves. The concept of data journeys could be used as a framework from which to evaluate the (re)use of data, developing the small case study here into a large-scale analysis of grey literature and associated site archives over time and incorporating questionnaires and interviews with data depositors. Similarly, data friction could provide a mechanism with which to investigate the range of issues surrounding the creation, manipulation, archiving, and subsequent reuse of data. Such studies would usefully sit alongside calls for similar ethnographic-style studies to examine the construction of the ontologies used in archaeology [40] and to investigate the emphasis on structured data and structuring mechanisms [41], for instance.

Operating within the archaeology data imaginary clearly presents a host of challenges. By no means all of these are unique to the discipline, and many can be seen as features of any disciplinary and interdisciplinary data relations. However, while archaeology can undoubtedly learn from the lessons of others, and adopt and adapt methodologies from elsewhere, many of the solutions and approaches need to arise out of the recognition of the particularities of the archaeological imaginary and respond to the specifically archaeological encounters with data. Thinking about archaeological data provides the basis for rethinking our engagement with it: from its initial recognition and capture through to its subsequent incorporation into analyses at all levels. In the end, an appropriately critical approach to our data can only make the archaeological knowledge we create with it more robust.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The grey literature reports and archives used in the case study are openly available from the Archaeology Data Service (<https://archaeologydataservice.ac.uk/> accessed on 22 March 2022). Appendix A lists the reports and archives used with associated DOIs.

Acknowledgments: I would like to thank Markos Katsianis, Tuna Kalayci, and Apostolos Sarris for their invitation to contribute to this Special Issue and for their support in seeing it to its completion. Aspects of this paper were initially trialled in presentations to the Museum of London Archaeology Research Seminar series in December 2021 and to the University of Glasgow Archaeology Research Seminar series in January 2022, and I thank Sara Perry and Gareth Beale, respectively, for chairing and facilitating the subsequent discussions. I am also grateful to Katie Green of the Archaeology Data Service for her advice and assistance regarding their grey literature collections. I thank the anonymous

reviewers for their constructive and helpful feedback. As ever, any errors or misconceptions are my own.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Reports used in the grey literature case study sample were extracted from the Archaeology Data Service archive using the search tool (<https://archaeologydataservice.ac.uk/archives/view/greylit/query.cfm> accessed on 21 March 2022) and filtered to return reports for the year 2021 which also had site archives. Watching briefs, evaluations, and excavations were selected from the resulting set, setting aside the historic building surveys, geophysical surveys, artefact, and environmental reports that were also returned. The final sample therefore consisted of the following cases:

Table A1. Watching briefs.

Site	Organization	Project Archive
Jubilee Car Park Extension, Bishops Waltham, Hampshire: Archaeological Watching Brief Report	Cotswold Archaeology	https://doi.org/10.5284/1084812
Archaeological Watching Brief at 66 High Street, Witney, Oxfordshire OX28 6HJ	John Moore Heritage Services	https://doi.org/10.5284/1090402

Table A2. Evaluations.

Site	Organization	Project Archive
Archaeological Evaluation Trenching on Land Adjacent to 28 Hill Cottage Gardens, Southampton	Archaeological Research Services	https://doi.org/10.5284/1086865
Archaeological Evaluation on Land off Oak Avenue, Scawby, Lincolnshire	Archaeological Research Services	https://doi.org/10.5284/1088100
Court Place Gardens, Iffley, Oxford: Archaeological Evaluation	Cotswold Archaeology	https://doi.org/10.5284/1085015
Land off Loperwood Lane, Calmore, Hampshire: Archaeological Evaluation	Cotswold Archaeology	https://doi.org/10.5284/1085004
Land off Braybrooke Road, Desborough, Northamptonshire	Oxford Archaeology	https://doi.org/10.5284/1056884
Land off Clunch Pit Lane, Reach, Cambridgeshire	Oxford Archaeology	https://doi.org/10.5284/1090486
Partridge Hill Farm Doncaster: Archaeological Evaluation	Wessex Archaeology	https://doi.org/10.5284/1090497

Table A3. Excavations.

Site	Organization	Project Archive
16–22 Coppergate, York-The Roman Burials AY12	York Archaeological Trust	https://doi.org/10.5284/1084816
Savile House Music Practice Room, New College, Oxford	Oxford Archaeology	https://doi.org/10.5284/1088123
Grange Farm, Main Road, Cannington, Somerset: Archaeological Excavation	Cotswold Archaeology	https://doi.org/10.5284/1084992
Thomas Turner and Co.'s 'Suffolk Works': The History and Archaeology of a Sheffield Steel and Cutlery Works	Archaeological Research and Consultancy at the University of Sheffield	https://doi.org/10.5284/1090410
Land off North Road, South Molton, Devon	AC Archaeology	https://doi.org/10.5284/1090500
Proposed Recreation Ground, Greet Road, Winchcombe, Gloucestershire: Archaeological Excavation	Cotswold Archaeology	https://doi.org/10.5284/1085008

References

1. Whitelaw, M. Landscape, Slow Data and Self-Revelation. Available online: <https://teemingvoid.blogspot.com/2009/05/landscape-slow-data-and-self-revelation.html> (accessed on 21 March 2022).
2. Huggett, J. Data as Flux. Available online: <https://introspectivedigitalarchaeology.com/2021/05/20/data-as-flux/> (accessed on 21 March 2022).
3. Bahn, P.G. *The Bluffer's Guide to Archaeology*; Oval Books: London, UK, 1999; ISBN 978-1-902825-47-2.
4. Wylie, A. How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways. *Sci. Technol. Hum. Values* **2017**, *42*, 203–225. [[CrossRef](#)]
5. Huvila, I. Be Informed of Your Information. *Curr. Swed. Archaeol.* **2014**, *22*, 47–51. [[CrossRef](#)]
6. Kristiansen, K. Towards a New Paradigm? The Third Science Revolution and Its Possible Consequences in Archaeology. *Curr. Swed. Archaeol.* **2014**, *22*, 11–34. [[CrossRef](#)]
7. Kristiansen, K. The Nature of Archaeological Knowledge and Its Ontological Turns. *Nor. Archaeol. Rev.* **2017**, *50*, 120–123. [[CrossRef](#)]
8. Sørensen, T.F. The Two Cultures and a World Apart: Archaeology and Science at a New Crossroads. *Nor. Archaeol. Rev.* **2017**, *50*, 101–115. [[CrossRef](#)]
9. Carpo, M. *The Second Digital Turn: Design Beyond Intelligence; Writing Architecture*; MIT Press: Cambridge, MA, USA; London, UK, 2017; ISBN 978-0-262-53402-4.
10. Kitchin, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, 2nd ed.; Sage Publications Ltd.: Thousand Oaks, CA, USA, 2021; ISBN 978-1-5297-3375-4.
11. Huggett, J. Reuse Remix Recycle: Repurposing Archaeological Digital Data. *Adv. Archaeol. Pract.* **2018**, *6*, 93–104. [[CrossRef](#)]
12. Huggett, J. Is Big Digital Data Different? Towards a New Archaeological Paradigm. *J. Field Archaeol.* **2020**, *45*, S8–S17. [[CrossRef](#)]
13. Huggett, J. Is Less More? Slow Data and Datafication in Archaeology. In *Critical Archaeology in the Digital Age, Proceedings of the 12th IEMA Visiting Scholar's Conference, Buffalo, NY, USA, 6–7 April 2019*; Cotsen Digital Archaeology Series; Garstki, K., Ed.; UCLA Cotsen Institute of Archaeology Press: Los Angeles, CA, USA, 2022; pp. 97–110, ISBN 978-1-950446-30-8.
14. Strauss, C. The Imaginary. *Anthropol. Theory* **2006**, *6*, 322–344. [[CrossRef](#)]
15. McNeil, M.; Arribas-Ayllon, M.; Haran, J.; Mackenzie, A.; Tutton, R. Conceptualizing Imaginaries of Science, Technology, and Society. In *The Handbook of Science and Technology Studies*; Felt, U., Fouché, R., Miller, C.A., Smith-Doerr, L., Eds.; The MIT Press: Cambridge, MA, USA, 2017; pp. 435–463, ISBN 978-0-262-03568-2.
16. Willim, R. Imperfect Imaginaries: Digitisation, Mundanisation, and the Ungraspable. In *Digitisation: Theories and Concepts for Empirical Cultural Research*; Koch, G., Ed.; Routledge, Taylor & Francis Group: London, UK; New York, NY, USA, 2017; pp. 53–77, ISBN 978-1-315-62773-1.
17. Marcus, G.E. Introduction. In *Technoscientific Imaginaries: Conversations, Profiles, and Memoirs*; Marcus, G.E., Ed.; University of Chicago Press: Chicago, IL, USA; London, UK, 1995; pp. 1–9, ISBN 978-0-226-50444-5.
18. Jasanoff, S.; Kim, S.-H. Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva* **2009**, *47*, 119. [[CrossRef](#)]
19. Jasanoff, S. Future Imperfect: Science, Technology, and the Imaginations of Modernity. In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*; Jasanoff, S., Kim, S.-H., Eds.; University of Chicago Press: Chicago, IL, USA, 2015; pp. 1–33, ISBN 978-0-226-27652-6.
20. Rieder, G. Tracing Big Data Imaginaries through Public Policy. The Case of the European Commission. In *The Politics and Policies of Big Data: Big Data, Big Brother?* Saetnan, A.R., Schneider, I., Green, N., Eds.; Routledge: London, UK, 2018; pp. 89–109.
21. Ruppert, E. *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data*; Erasmus School of Social and Behavioural Sciences; Erasmus University Rotterdam: Rotterdam, The Netherlands, 2018; ISBN 978-90-75289-25-1.
22. Bucher, T. The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms. *Inf. Commun. Soc.* **2017**, *20*, 30–44. [[CrossRef](#)]
23. Beer, D. *The Data Gaze: Capitalism, Power and Perception*, 1st ed.; Society and Space; SAGE Publications: Thousand Oaks, CA, USA, 2018; ISBN 978-1-5264-3692-4.
24. Leonelli, S. Data Science in Times of Pan(Dem)Ic. *Harv. Data Sci. Rev.* **2021**, *3*. [[CrossRef](#)]
25. Poirier, L.; Fortun, K.; Costelloe-Kuehn, B.; Fortun, M. Metadata, Digital Infrastructure, and the Data Ideologies of Cultural Anthropology. In *Anthropological Data in the Digital Age: New Possibilities—New Challenges*; Crowder, J.W., Fortun, M., Besara, R., Poirier, L., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 209–237, ISBN 978-3-030-24925-0.
26. Dourish, P.; Gómez Cruz, E. Datafication and Data Fiction: Narrating Data and Narrating with Data. *Big Data Soc.* **2018**, *5*, 205395171878408. [[CrossRef](#)]
27. Gray, J. Three Aspects of Data Worlds. *Krisis J. Contemp. Philos.* **2018**, *1*, 4–17. Available online: <https://archive.krisis.eu/three-aspects-of-data-worlds/> (accessed on 22 March 2022).
28. Jasanoff, S. Imagined and Invented Worlds. In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*; Jasanoff, S., Kim, S.-H., Eds.; University of Chicago Press: Chicago, IL, USA, 2015; pp. 321–341, ISBN 978-0-226-27652-6.
29. Jakobsson, U.; Novák, D.; Richards, J.D.; Štular, B.; Wright, H. Digital Archiving in Archaeology: The State of the Art. *Internet Archaeol.* **2021**, *58*. Available online: <https://intarch.ac.uk/journal/issue58/index.html> (accessed on 22 March 2022).

30. Huggett, J. Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge. In *Open Source Archaeology: Ethics and Practice*; Wilson, A.T., Edwards, B., Eds.; De Gruyter Open: Warsaw, Poland, 2015; pp. 6–29, ISBN 978-3-11-044017-1.
31. Chartered Institute for Archaeologists Code of Conduct: Professional Ethics in Archaeology. Available online: <https://www.archaeologists.net/codes/cifa> (accessed on 15 February 2022).
32. European Association of Archaeologists European Association of Archaeologists Code of Practice. Available online: https://www.e-a-a.org/EAA/About/EAA_Codes/EAA/Navigation_About/EAA_Codes.aspx (accessed on 15 February 2022).
33. Marwick, B.; d’Alpoim Guedes, J.; Barton, C.M.; Bates, L.A.; Baxter, M.; Bevan, A.; Bollwerk, E.A.; Bocinsky, R.K.; Brughmans, T.; Carter, A.K.; et al. Open Science in Archaeology. *SAA Archaeol. Rec.* **2017**, *17*, 8–14. [[CrossRef](#)]
34. Kansa, E. Openness and Archaeology’s Information Ecosystem. *World Archaeol.* **2012**, *44*, 498–520. [[CrossRef](#)]
35. Open Knowledge Foundation Open Definition 2.1—Open Definition—Defining Open in Open Data, Open Content and Open Knowledge. Available online: <https://opendefinition.org/od/2.1/en/> (accessed on 16 February 2022).
36. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
37. Stall, S.; Yarmey, L.; Cutcher-Gershenfeld, J.; Hanson, B.; Lehnert, K.; Nosek, B.; Parsons, M.; Robinson, E.; Wyborn, L. Make All Scientific Data FAIR. *Nature* **2019**, *570*, 27–29. [[CrossRef](#)]
38. Niccolucci, F.; Richards, J. (Eds.) *The ARIADNE Impact*; Archaeolingua Foundation: Budapest, Hungary, 2019; ISBN 978-615-5766-31-2.
39. Geser, G.; Richards, J.D.; Massara, F.; Wright, H. Data Management Policies and Practices of Digital Archaeological Repositories. *Internet Archaeol.* **2022**, *59*. [[CrossRef](#)]
40. Huggett, J. Lost in Information? Ways of Knowing and Modes of Representation in e-Archaeology. *World Archaeol.* **2012**, *44*, 538–552. [[CrossRef](#)]
41. Hacıgüzeller, P.; Taylor, J.S.; Perry, S. On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind. *Open Archaeol.* **2021**, *7*, 1709–1730. [[CrossRef](#)]
42. Marwick, B.; Birch, S.E.P. A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. *Adv. Archaeol. Pract.* **2018**, *6*, 125–143. [[CrossRef](#)]
43. Richards, J.D.; Niven, K.; Jeffrey, S. Preserving Our Digital Heritage: Information Systems for Data Management and Preservation. In *Visual Heritage in the Digital Age*; Ch’ng, E., Gaffney, V., Chapman, H., Eds.; Springer Series on Cultural Computing; Springer: London, UK, 2013; pp. 311–326, ISBN 978-1-4471-5534-8.
44. Huvila, I. Ecology of Archaeological Information Work. In *Archaeology and Archaeological Information in the Digital Society*; Huvila, I., Ed.; Routledge: Abingdon, UK; New York, NY, USA, 2018; pp. 122–142, ISBN 978-0-415-78843-4.
45. Kintigh, K.W.; Altschul, J.H.; Kinzig, A.P.; Limp, W.F.; Michener, W.K.; Sabloff, J.A.; Hackett, E.J.; Kohler, T.A.; Ludäscher, B.; Lynch, C.A. Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. *Adv. Archaeol. Pract.* **2015**, *3*, 1–15. [[CrossRef](#)]
46. Wright, H.; Richards, J.D. Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures. *J. Field Archaeol.* **2018**, *43*, S60–S67. [[CrossRef](#)]
47. Benardou, A.; Champion, E.; Dallas, C.; Hughes, L.M. Introduction: A Critique of Digital Practices and Research Infrastructures. In *Cultural Heritage Infrastructures in Digital Humanities*; Benardou, A., Champion, E., Dallas, C., Hughes, L.M., Eds.; Routledge: Abingdon, UK, 2017; pp. 1–14, ISBN 978-1-315-57527-8.
48. Taylor, C. *Modern Social Imaginaries*; Duke University Press: Durham, UK, 2004; ISBN 978-0-8223-3255-8.
49. Day, R.E. *Indexing It All: The Subject in the Age of Documentation, Information, and Data*; The MIT Press: Cambridge, MA, USA, 2014; ISBN 978-0-262-02821-9.
50. Borgman, C.L. *Big Data, Little Data, No Data: Scholarship in the Networked World*; The MIT Press: Cambridge, MA, USA, 2015; ISBN 978-0-262-02856-1.
51. Okune, A.; Hillyer, R.; Albornoz, D.; Posada, A.; Chan, L. Whose Infrastructure? Towards Inclusive and Collaborative Knowledge Infrastructures in Open Science. In *Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure, Proceedings of the 22nd International Conference on Electronic Publishing, Toronto, ON, Canada, 28 June 2018*; Chan, L., Mounier, P., Eds.; OpenEdition Press: Marseille, France, 2018. [[CrossRef](#)]
52. Archaeology Data Service Annual Report 1st August 2020–31 July 2021. Available online: <https://archaeologydataservice.ac.uk/about/annualReports.xhtml> (accessed on 21 February 2022).
53. Safdar, I. Canmore Hits One Million! Available online: <https://blog.historicenvironment.scot/2020/02/canmore-hits-one-million/> (accessed on 21 February 2022).
54. British Museum Portable Antiquities Scheme Database. Available online: <http://finds.org.uk/database> (accessed on 21 February 2022).
55. Witze, A. Disappearing Digital Data. *Am. Archaeol.* **2019**, *23*, 40–45.
56. Mills, B.J.; Clark, J.J.; Peeples, M.A.; Haas, W.R.; Roberts, J.M.; Hill, J.B.; Huntley, D.L.; Borck, L.; Breiger, R.L.; Clauset, A.; et al. Transformation of Social Networks in the Late Pre-Hispanic US Southwest. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5785–5790. [[CrossRef](#)]

57. Green, C.; Gosden, C.; Cooper, A.; Franconi, T.; ten Harkel, L.; Kamash, Z.; Lowerre, A. Understanding the Spatial Patterning of English Archaeology: Modelling Mass Data, 1500 BC to AD 1086. *Archaeol. J.* **2017**, *174*, 244–280. [[CrossRef](#)]
58. Cooper, A.; Green, C. Embracing the Complexities of ‘Big Data’ in Archaeology: The Case of the English Landscape and Identities Project. *J. Archaeol. Method Theory* **2016**, *23*, 271–304. [[CrossRef](#)]
59. Holdaway, S.J.; Emmitt, J.; Phillipps, R.; Masoud-Ansari, S. A Minimalist Approach to Archaeological Data Management Design. *J. Archaeol. Method Theory* **2019**, *26*, 873–893. [[CrossRef](#)]
60. Gattiglia, G. From Digitization to Datafication. A New Challenge Is Approaching Archaeology. In *Il Telescopio Inverso: Big Data e Distant Reading nelle Discipline Umanistiche, Proceedings of the AIUCD 2017 Conference, 26–28 January 2017, Rome, Italy*; Ciotti, F., Crupi, G., Eds.; Associazione per l’Informatica Umanistica e la Cultura Digitale: Florence, Italy, 2017; pp. 29–33. [[CrossRef](#)]
61. Elish, M.C.; Boyd, D. Situating Methods in the Magic of Big Data and AI. *Commun. Monogr.* **2018**, *85*, 57–80. [[CrossRef](#)]
62. Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Available online: <https://www.wired.com/2008/06/pb-theory/> (accessed on 21 February 2022).
63. Ruppert, E. Different Data Futures: An Experiment in Citizen Data. *Stat. J. IAOS* **2019**, *35*, 633–641. [[CrossRef](#)]
64. Bevan, A. The Data Deluge. *Antiquity* **2015**, *89*, 1473–1484. [[CrossRef](#)]
65. Brouwer Burg, M. It Must Be Right, GIS Told Me so! Questioning the Infallibility of GIS as a Methodological Tool. *J. Archaeol. Sci.* **2017**, *84*, 115–120. [[CrossRef](#)]
66. Hacıgüzeller, P. GIS, Critique, Representation and Beyond. *J. Soc. Archaeol.* **2012**, *12*, 245–263. [[CrossRef](#)]
67. Boyd, D.; Crawford, K. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Inf. Commun. Soc.* **2012**, *15*, 662–679. [[CrossRef](#)]
68. Collis, J. The Nature of Archaeological Evidence. In *Companion Encyclopedia of Archaeology*; Barker, G., Ed.; Routledge: London, UK; New York, NY, USA, 1999; pp. 81–127, ISBN 978-0-415-06448-4.
69. Lucas, G. *Understanding the Archaeological Record*; Cambridge University Press: Cambridge, UK, 2012; ISBN 978-1-107-01026-0.
70. Lucas, G. *Writing the Past: Knowledge and Literary Production in Archaeology*; Routledge: London, UK; New York, NY, USA, 2019; ISBN 978-0-367-00105-6.
71. Wylie, A. Archaeological Facts in Transit: The ‘Eminent Mounds’ of Central North America. In *How Well do Facts Travel? The Dissemination of Reliable Knowledge*; Howlett, P., Morgan, M.S., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2011; pp. 301–322, ISBN 978-0-521-19654-3.
72. Rosenberg, D. Data before the Fact. In *“Raw Data” Is an Oxymoron*; Gitelman, L., Ed.; MIT Press: Cambridge, MA, USA; London, UK, 2013; pp. 15–40, ISBN 978-0-262-51828-4.
73. Rosenberg, D. Data as Word. *Hist. Stud. Nat. Sci.* **2018**, *48*, 557–567. [[CrossRef](#)]
74. Rosenberg, D. Data. In *Information: A Historical Companion*; Blair, A., Duguid, P., Goeing, A.-S., Grafton, A., Eds.; Princeton University Press: Princeton, NJ, USA, 2021; pp. 387–391, ISBN 978-0-691-17954-4.
75. Lavin, M. Why Digital Humanists Should Emphasize Situated Data over Capta. *Digit. Humanit. Q.* **2021**, *15*, 13.
76. Chippindale, C. Capta and Data: On the True Nature of Archaeological Information. *Am. Antiq.* **2000**, *65*, 605–612. [[CrossRef](#)]
77. Drucker, J. Humanities Approaches to Graphical Display. *Digit. Humanit. Q.* **2011**, *5*. Available online: <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (accessed on 22 March 2022).
78. Lanigan, R.L. Capta versus Data: Method and Evidence in Communicology. *Hum. Stud.* **1994**, *17*, 109–130. [[CrossRef](#)]
79. Andrews, G.; Barrett, J.C.; Lewis, J.S.C. Interpretation Not Record: The Practice of Archaeology. *Antiquity* **2000**, *74*, 525–530. [[CrossRef](#)]
80. Leonelli, S. What Counts as Scientific Data? A Relational Framework. *Philos. Sci.* **2015**, *82*, 810–821. [[CrossRef](#)] [[PubMed](#)]
81. Buccellati, G. *A Critique of Archaeological Reason: Structural, Digital, and Philosophical Aspects of the Excavated Record*; Cambridge University Press: Cambridge, UK, 2017; ISBN 978-1-107-11029-8.
82. Kitchin, R. *Data Lives: How Data Are Made and Shape Our World*; Bristol University Press: Bristol, UK, 2021; ISBN 978-1-5292-1514-4.
83. Manovich, L. *The Language of New Media*; MIT Press: Cambridge, MA, USA, 2002; ISBN 978-0-262-63255-3.
84. Svensson, J.; Guillén, O.P. What Is Data and What Can It Be Used For? Key Questions in the Age of Burgeoning Data-Essentialism. *J. Digit. Soc. Res.* **2020**, *2*, 65–83. [[CrossRef](#)]
85. Ellul, J. *The Technological Society*; Vintage Books, Random House: New York, NY, USA, 1964; ISBN 978-0-394-70390-9.
86. Ingold, T. *Being Alive: Essays on Movement, Knowledge and Description*; Routledge: London, UK; New York, NY, USA, 2011; ISBN 978-0-415-57683-3.
87. Drucker, J. *Graphesis: Visual Forms of Knowledge Production*; Harvard University Press: Cambridge, MA, USA, 2014; ISBN 978-0-674-72493-8.
88. Dieter, M. Dark Patterns: Interface Design, Augmentation and Crisis. In *Postdigital Aesthetics: Art, Computation and Design*; Berry, D.M., Dieter, M., Eds.; Palgrave Macmillan: London, UK, 2015; pp. 163–178, ISBN 978-1-349-49378-4.
89. Huggett, J. The Past in Bits: Towards an Archaeology of Information Technology? *Internet Archaeol.* **2004**, *15*. [[CrossRef](#)]
90. Lock, G.; Pouncett, J. Spatial Thinking in Archaeology: Is GIS the Answer? *J. Archaeol. Sci.* **2017**, *84*, 129–135. [[CrossRef](#)]
91. Carver, M.O.H. *Archaeological Investigation*; Routledge: London, UK; New York, NY, USA, 2009; ISBN 978-0-415-48918-8.
92. Edgeworth, M. *Acts of Discovery: An Ethnography of Archaeological Practice*; BAR International Series; Archaeopress: Oxford, UK, 2003; ISBN 978-1-84171-504-9.
93. Edgeworth, M. Follow the Cut, Follow the Rhythm, Follow the Material. *Nor. Archaeol. Rev.* **2012**, *45*, 76–92. [[CrossRef](#)]

94. Carver, M.O.H. Field Archaeology. In *Companion Encyclopedia of Archaeology*; Barker, G., Ed.; Routledge: London, UK; New York, NY, USA, 1999; pp. 128–181, ISBN 978-0-415-06448-4.
95. Huggett, J. Algorithmic Agency and Autonomy in Archaeological Practice. *Open Archaeol.* **2021**, *7*, 417–434. [[CrossRef](#)]
96. Dell’Unto, N.; Landeschi, G.; Apel, J.; Poggi, G. 4D Recording at the Trowel’s Edge: Using Three-Dimensional Simulation Platforms to Support Field Interpretation. *J. Archaeol. Sci. Rep.* **2017**, *12*, 632–645. [[CrossRef](#)]
97. Taylor, J.; Issavi, J.; Berggren, Å.; Lukas, D.; Mazzucato, C.; Tung, B.; Dell’Unto, N. “The Rise of the Machine”: The Impact of Digital Tablet Recording in the Field at Çatalhöyük. *Internet Archaeol.* **2018**, *47*. [[CrossRef](#)]
98. Berggren, Å.; Gutehall, A. Going from Analogue to Digital: A Study of Documentation Methods during an Excavation of the Neolithic Flint Mines at Pilbladet, Sweden. *Curr. Swed. Archaeol.* **2018**, *26*, 119–158. [[CrossRef](#)]
99. Morgan, C.; Petrie, H.; Wright, H.; Taylor, J.S. Drawing and Knowledge Construction in Archaeology: The Aide Mémoire Project. *J. Field Archaeol.* **2021**, *46*, 614–628. [[CrossRef](#)]
100. Huvila, I. Management of Archaeological Information and Knowledge in Digital Environment. In *Knowledge Management, Arts, and Humanities; Knowledge Management and Organizational Learning*; Handzic, M., Carlucci, D., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 147–169, ISBN 978-3-030-10921-9.
101. Richards, J.D. Digital Preservation and Access. *Eur. J. Archaeol.* **2002**, *5*, 343–366. [[CrossRef](#)]
102. Richards, J.D. Managing Digital Preservation and Access: The Archaeology Data Service. In *Managing Archaeological Resources: Global Context, National Programs, Local Actions; One World Archaeology*; McManamon, F.P., Stout, A., Barnes, J.A., Eds.; Left Coast Press: Walnut Creek, CA, USA, 2008; pp. 173–194, ISBN 978-1-315-42493-4.
103. Wylie, A. Archaeological Cables and Tacking—The Implications of Practice for Bernstein’s “Options beyond Objectivism and Relativism”. *Philos. Soc. Sci. Soc.* **1989**, *19*, 1–18. [[CrossRef](#)]
104. Verboven, K. Introduction: Finding a New Approach to Ancient Proxy Data. In *Complexity Economics: Building a New Approach to Ancient Economic History*; Verboven, K., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 1–18, ISBN 978-3-030-47898-8.
105. Collins, M.B. Sources of Bias in Processual Data: An Appraisal. In *Sampling in Archaeology*; Mueller, J.W., Ed.; University of Arizona Press: Tucson, AZ, USA, 1975; pp. 26–32, ISBN 978-0-8165-0482-4.
106. Huggett, J. Capturing the Silences in Digital Archaeological Knowledge. *Information* **2020**, *11*, 278. [[CrossRef](#)]
107. Purdam, K.; Elliot, M. The Changing Social Science Data Landscape. In *Innovations in Digital Research Methods*; Halfpenny, P., Procter, R., Eds.; SAGE Publications Ltd: London, UK, 2015; pp. 25–58, ISBN 978-1-4462-0309-5.
108. Gitelman, L. (Ed.) “Raw Data” Is an Oxymoron; Infrastructures Series; MIT Press: Cambridge, MA, USA; London, UK, 2013; ISBN 978-0-262-51828-4.
109. Bowker, G.C. *Memory Practices in the Sciences*; MIT Press: Cambridge, MA, USA, 2005; ISBN 978-0-262-02589-8.
110. Longino, H.E. Afterword: Data in Transit. In *Data Journeys in the Sciences*; Leonelli, S., Tempini, N., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 391–399, ISBN 978-3-030-37176-0.
111. Archaeology Data Service Guidelines for Depositors. Available online: <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml> (accessed on 25 March 2022).
112. Schmidt, A.; Ernenwein, E. Guide to Good Practice: Geophysical Data in Archaeology. Available online: https://guides.archaeologydataservice.ac.uk/g2gp/Geophysics_Toc (accessed on 25 March 2022).
113. Payne, A. Laser Scanning for Archaeology: A Guide to Good Practice. Available online: https://guides.archaeologydataservice.ac.uk/g2gp/LaserScan_Toc (accessed on 25 March 2022).
114. Leonelli, S. Learning from Data Journeys. In *Data Journeys in the Sciences*; Leonelli, S., Tempini, N., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–24, ISBN 978-3-030-37176-0.
115. Latour, B. Visualization and Cognition: Thinking with Eyes and Hands. *Knowl. Soc. Stud. Sociol. Cult. Past Present* **1986**, *6*, 1–40.
116. Leonelli, S. *Data-Centric Biology: A Philosophical Study*; University of Chicago Press: Chicago, IL, USA, 2016; ISBN 978-0-226-41647-2.
117. Leonelli, S. What Distinguishes Data from Models? *Eur. J. Philos. Sci.* **2019**, *9*, 22. [[CrossRef](#)]
118. Bates, J.; Lin, Y.-W.; Goodale, P. Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows. *Big Data Soc.* **2016**, *3*, 205395171665450. [[CrossRef](#)]
119. Wylie, A. Radiocarbon Dating in Archaeology: Triangulation and Traceability. In *Data Journeys in the Sciences*; Leonelli, S., Tempini, N., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 285–301, ISBN 978-3-030-37177-7.
120. Leonelli, S. What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big Data Soc.* **2014**, *1*, 2053951714534395. [[CrossRef](#)]
121. Kansa, S.W.; Atici, L.; Kansa, E.C.; Meadow, R.H. Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data. *Adv. Archaeol. Pract.* **2020**, *8*, 40–52. [[CrossRef](#)]
122. Edwards, P.N. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*; MIT Press: Cambridge, MA, USA, 2010; ISBN 978-0-262-01392-5.
123. Edwards, P.N.; Mayernik, M.S.; Batcheller, A.L.; Bowker, G.C.; Borgman, C.L. Science Friction: Data, Metadata, and Collaboration. *Soc. Stud. Sci.* **2011**, *41*, 667–690. [[CrossRef](#)]
124. Edgeworth, M. The Clearing: Archaeology’s Way of Opening the World. In *Reclaiming Archaeology: Beyond the Tropes of Modernity*; Ruibal, A.G., Ed.; Routledge: London, UK, 2013; pp. 33–43.

125. Richards-Rissetto, H.; Landau, K. Digitally-Mediated Practices of Geospatial Archaeological Data: Transformation, Integration, & Interpretation. *J. Comput. Appl. Archaeol.* **2019**, *2*, 120–135. [CrossRef]
126. Harrison, L.K. Closing the Loop on the Digital Data Lifecycle: Reviving a Salvage Archaeology Dataset. In *Critical Archaeology in the Digital Age, Proceedings of the 12th IEMA Visiting Scholar's Conference, Buffalo, NY, USA, 6–7 April 2019*; Cotsen Digital Archaeology Series; Garstki, K., Ed.; UCLA Cotsen Institute of Archaeology Press: Los Angeles, CA, USA, 2022; pp. 79–96, ISBN 978-1-950446-30-8.
127. Huvila, I. Putting to (Information) Work: A Stengersian Perspective on How Information Technologies and People Influence Information Practices. *Inf. Soc.* **2018**, *34*, 229–243. [CrossRef]
128. Bates, J. The Politics of Data Friction. *J. Doc.* **2018**, *74*, 412–429. [CrossRef]
129. Kansa, E.C. On Infrastructure, Accountability, and Governance in Digital Archaeology. In *Critical Archaeology in the Digital Age, Proceedings of the 12th IEMA Visiting Scholar's Conference, Buffalo, NY, USA, 6–7 April 2019*; Cotsen Digital Archaeology Series; Garstki, K., Ed.; UCLA Cotsen Institute of Archaeology Press: Los Angeles, CA, USA, 2022; pp. 141–152, ISBN 978-1-950446-30-8.
130. Bruseker, G.; Carboni, N.; Guillem, A. Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. In *Heritage and Archaeology in the Digital Age; Quantitative Methods in the Humanities and Social Sciences*; Vincent, M.L., López-Menchero Bendicho, V.M., Ioannides, M., Levy, T.E., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 93–131, ISBN 978-3-319-65369-3.
131. Caraher, W. Slow Archaeology: Technology, Efficiency, and Archaeological Work. In *Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology*; Averett, E.W., Gordon, J.M., Counts, D.B., Eds.; The Digital Press, The University of North Dakota: Grand Forks, ND, USA, 2016; pp. 421–441.
132. Caraher, W. Slow Archaeology, Punk Archaeology, and the 'Archaeology of Care'. *Eur. J. Archaeol.* **2019**, *22*, 372–385. [CrossRef]
133. Perry, S. Why Are Heritage Interpreters Voiceless at the Trowel's Edge? A Plea for Rewriting the Archaeological Workflow. *Adv. Archaeol. Pract.* **2018**, *6*, 212–227. [CrossRef]
134. Morgan, C.; Wright, H. Pencils and Pixels: Drawing and Digital Media in Archaeological Field Recording. *J. Field Archaeol.* **2018**, *43*, 136–151. [CrossRef]
135. Opitz, R.S.; Johnson, T.D. Interpretation at the Controller's Edge: Designing Graphical User Interfaces for the Digital Publication of the Excavations at Gabii (Italy). *Open Archaeol.* **2016**, *2*, 1–17. [CrossRef]
136. Poirier, L. Data, Knowledge Practices, and Naturecultural Worlds: Vehicle Emissions in the Anthropocene. In *The Palgrave Handbook of the Anthropology of Technology*; Bruun, M.H., Wahlberg, A., Douglas-Jones, R., Hoeyer, K., Kristensen, D.B., Winthereik, B.R., Eds.; Palgrave Macmillan: Singapore, 2022; pp. 273–290, ISBN 978-981-16-7083-1.
137. Richards, J.; Jeffrey, S.; Waller, S.; Ciravegna, F.; Chapman, S.; Zhang, Z. The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0: New Approaches to Communication and Collaboration*; Kansa, E.C., Kansa, S.W., Watrall, E., Eds.; Cotsen Institute of Archaeology Press: Los Angeles, CA, USA, 2011; pp. 27–56.
138. Evans, T. OASIS V Is Here. Available online: <https://archaeologydataservice.ac.uk/blog/oasis/?p=742> (accessed on 7 April 2022).
139. Richards, J.D.; Robinson, D. (Eds.) *Digital Archives from Excavation and Fieldwork: A Guide to Good Practice*; AHDS Guides to Good Practice; Oxbow: Oxford, UK, 2000; ISBN 978-1-900188-73-9.
140. Richards, J.D. Twenty Years Preserving Data: A View from the United Kingdom. *Adv. Archaeol. Pract.* **2017**, *5*, 227–237. [CrossRef]
141. Markham, A.N. Troubling the Concept of Data in Qualitative Digital Research. In *The Sage Handbook of Qualitative Data Collection*; Flick, U., Ed.; Sage Reference: Los Angeles, CA, USA, 2018; pp. 511–523, ISBN 978-1-4739-5213-3.
142. Ballsun-Stanton, B.; Ross, S.A.; Sobotkova, A.; Crook, P. FAIMS Mobile: Flexible, Open-Source Software for Field Research. *SoftwareX* **2018**, *7*, 47–52. [CrossRef]
143. Blom, I. Introduction: Rethinking Social Memory: Archives, Technology, and the Social. In *Memory in Motion; Archives, Technology and the Social*; Blom, I., Lundemo, T., Røssaak, E., Eds.; Amsterdam University Press: Amsterdam, The Netherlands, 2017; pp. 11–38, ISBN 978-94-6298-214-7.
144. Dekker, A. Introduction: What It Means to Be Lost and Living (in) Archives. In *Lost and Living (in) Archives: Collectively Shaping New Memories; Making Public*; Dekker, A., Ed.; Valiz: Amsterdam, The Netherlands, 2017; pp. 11–25, ISBN 978-94-92095-26-8.
145. Michener, W.K.; Brunt, J.W.; Helly, J.J.; Kirchner, T.B.; Stafford, S.G. Nongeospatial Metadata for the Ecological Sciences. *Ecol. Appl.* **1997**, *7*, 330–342. [CrossRef]
146. Duranti, L. The Impact of Digital Technology on Archival Science. *Arch. Sci.* **2001**, *1*, 39–55. [CrossRef]
147. DeRidder, J.L. Benign Neglect: Developing Life Rafts for Digital Content. *Inf. Technol. Libr.* **2011**, *30*, 71–74. [CrossRef]