The Ontological Imperative When Researching in the Digital Age

Tom Liam Lynch\textsuperscript{a} and Hannah R. Gerber\textsuperscript{b}

\textsuperscript{a}School of Education, Pace University, NY, USA; \textsuperscript{b}Language, Literacy and Special Populations, Sam Houston State University, TX, USA

**ABSTRACT**

In this conceptual paper, the authors introduce the notion of an *ontological imperative* that researchers today must deeply consider when examining digital phenomena or using digital research tools. Making their argument within a mixed methods and multimethod framework, they draw heavily on the work of software theorists to argue that what it means to be digital is inherently unstable and that such instability makes both digital data and digital methods of research far less reliable than they first appear. One key claim is that failure to address the ontological imperative leads to epistemological instability wherein replicating studies with digital data and tools can be impossible. The authors respond by offering a 5-step methodological process to account for the ontology of the digital and, in doing so, strengthen researchers’ epistemological claims.

**KEYWORDS**

Digital methods; epistemology; methodology; ontology; software theory

---

**Opening**

*Digital* is a common word in today’s world. In its simplest form, it refers to processes and products that are mediated and shaped by computer software (Berry, 2011). On the one hand, the digital appears to be computational, objective, automated, electric, and ephemeral. And it is, in some ways. But computationality cannot be separated so easily from corporeality. Consider the etymology of the word. *Digital* entered the English language in 1450 by way of the Latin *digitālis*, which referred to “measuring a finger’s breadth” (Digital, 2016). Over time, other uses of the word developed like reference to whole numbers less than 10 (the number of fingers), humorous reference to the fingers themselves, and any keys on a musical instrument with a keyboard. Etymologically, the notion of the digital is far more human than it first appears. That is, whereas today one might commonly associate the digital with flashing and fast searching of information, streaming of entertainment, purchasing of goods, and interpersonal communication, the word’s roots fundamentally hearken materiality and fleshiness. The same is true in practice. For example, the digital requires computer processors. Such processors are produced by mining the earth for precious metals or ores that get refined into manufacturing materials such as copper, which is essential for electrical conduction in solenoids and relays. Or, although it is common to refer to the wonders of wireless connectivity, in fact, wirelessness relies on increasingly sophisticated wires (or cables) that human beings stretch across oceans and along continents (Lynch, 2016). In sum, the digital is not quite what it appears to be. Researchers of digital phenomena or who use digital tools must confront what we call an *ontological imperative*.

In what follows, we argue that a critical understanding of the ontology of the digital (what *digital* is) has direct methodological implications that can help the field avoid epistemological pitfalls associated with conducting research in the digital age. We draw heavily on the work of software theorists, exploring several examples of how software theorists conceptualize the digital. They reveal that linguistic instability and ideological leanings undergird what otherwise appear to be objective, neutral, and purely informational digital data and research tools. Second, we review the current state of mixed and multimethod research using Johnson and Onwuegbuzie (2004) as a guiding text. Third, we offer specific techniques for how researchers can address the ontological imperative in their methodological writings, particularly in their descriptions of data collection and resultant
epistemological ramifications. Finally, we conclude by offering suggestions for how researchers can bolster their understanding of the ontological imperative further, which, in turn, we believe will both strengthen their own research as well as the way the field views digital phenomena in the world.

The Ontological Imperative in Researching the “Digital”

Researchers (as well as the public) must consciously rupture common associations between the digital and qualities like objectivity, ephemerality, and neutrality. Rupturing such assumptions requires exploring the ways theorists have grappled with notions of the digital and computation, which theorists do in greatest depth under the umbrella of software theory. Software theory is a field of study devoted to examining how the nature of software “radically reshapes” (Berry, 2011, p. 10) human experience in the world. Scholars who contribute to software theory come from an array of fields, some not easily distinguished from software theory itself: software studies, code studies, platform studies, media theory, information science, and digital humanities—to name a few. We will use the terms software and digital interchangeably for present purposes and ease of reading, drawing attention to distinctions between the terms when necessary for clarity.

Software theorists emphasize the linguistic and ideological nature of software (Berry, 2011; Cox, 2013; Frabetti, 2015; Fuller, 2008; Galloway, 2012; Kitchin & Dodge, 2011; Lynch, 2015; Mackenzie, 2006; Manovich, 2001, 2013; Williamson, 2015). Software is linguistic insofar as what makes the digital world work is a complex assemblage of human and computational languages. Software is ideological insofar as its nature relies on reducing the complexity of experience to discrete categories and numbers for computation. As Berry (2011) writes in his definition of digital,

Something of the detail is always lost when moved from the phenomenal world to the discrete world off the computer. Digitalisation is therefore the simplification and standardisation of the external world so that it can be stored and manipulated within code. (p. 54)

Whoever determines what such simplification and standardization consists of cannot do so with pure neutrality. Assumptions and values about the world are bound to be at play, even if subtly so. We argue that researchers today must understand the linguistic and ideological nature of the digital and systematically address the ontological imperative.

Before proceeding, we wish to clarify what we mean by ontological imperative. We use the term ontological in the philosophical sense, referring to it as “a critical discipline rather than a positive philosophical science” that honors “empirical and statistical methods” but that locates such methods “in a more fundamental matrix of critical (ontological) thought” (Noonan, 2008, p. 580). Our interest is in naming the urgency for researchers to critically understand what the digital is—a critical question of being—as it relates to their methodologies and epistemological outputs. We use ontological with an awareness that in computer and information sciences, an ontology has a related but different meaning. Technically speaking, an ontology is a sophisticated mapping of information, categories, and relationships used to enable computers to automate insight and actions. Ontologies are what power recommendation engines and make artificial intelligence possible. Whereas philosophically an ontology might refer what it means for the digital to be, technical ontologies are concerned with what it means for the digital to do.

Understanding the digital as a linguistic and ideological construct might best be explored by considering it in terms of two kinds of authorship. First, software is itself authored. How is software written, by whom, and with what intensions? Second, software is used by human beings to author things in the world. How does software as an authoring tool enable and inhibit what can be known? In what follows, we examine the ontology of the digital with these two kinds of authorship in mind.

Software Itself is Authored

People write software. It does not just appear out of the ether. Theorists have argued for some time that all writing is inherently socially produced and its meaning always unstable (Spivak, 1976). Software is no exception. In her examination of the nature of software, Frabetti (2015) undertook a deconstruction of the notion of software, demonstrating that many of the binaries that are commonly used to define software are, in fact, inadequate. Whereas software is associated with quantification, it is also fundamentally reliant on alphabetic languages (i.e., code). Whereas software is associated with ephemerality, it requires physical hardware to operate. Whereas software might be associated with computational languages, the process of creating and maintaining software only occurs because of human communication via email, message boards, inky writing, and so forth.
Frabetti (2015) punctuated this latter insight with the counterintuitive declaration, “paper shapes software” (p. 111). Finally, Frabetti argued that despite associations between software and stability, the fact is that software is incredibly unstable, its nature is one of constant translation across multiple languages. For example, consider how often one’s laptop, tablet, applications, or phone needs to be updated. Those updates can be thought of as revisions to human and computational writing. Frabetti concluded her work by analyzing the ways that multiple layers of computational languages are combined to make software work, including the rapid firing of electrical pulses based on binary code.

The unstable linguistic nature of software is something media theorist Friedrich Kittler (1997) succinctly articulated when he compared software to a “postmodern Tower of Babel” (p. 148), comprising myriad different languages. He wrote that software:

... reaches from simple operation codes whose linguistic extension is still a hardware configuration, passing through an assembler whose extension is this very opcode, up to high-level programming languages whose extension is that very assembler ... We simply do not know what our writing does.

Kittler underscored some of the points that Frabetti (2015) made about software: it’s linguistic, it’s unstable, it’s in a constant state of translation, and no single author ever actually controls it.

Whereas some scholars focus on software as linguistic in the written sense, others emphasize that software can theoretically be considered a form of speech. Cox (2013) argued that software is less akin to written language than it is to spoken word. He argued that the execution of code makes it active in ways that is unlike written human language because code is language that acts. Cox further argued that court decisions protecting software as a form of free speech have added jurisprudential weight to theoretical claims that computer code is on par with one’s freedom to speak and write. The relationship between language and action is something that Berry (2011) explored at length as well when he unpacked how aesthetic coding competitions reward programmers who can create software that is written with aesthetic sensibility while executing programs in witty ways, blurring the lines between content and form, and written and spoken language.

The aforementioned review demonstrates briefly some reasons why researchers must take into account how the digital systems and artifacts they encounter are anything but neutral, objective, or stable. However, researchers must also problematize how digital research tools impact authorship, including their own ability to collect and analyze data.

Software is used to Author

If the nature of software itself is linguistic (i.e., it is unstable and socio-linguistically constructed), researchers must also consider that software is ideologically encoded (i.e., the values and assumptions of its authors are literally encoded in it). In his classic framing of software studies, Lev Manovich (2001) argued that there is a “logic of selection” (p. 123) in software that comprises “computer culture,” the result of which is that “authentic creation has been replaced by selection from a menu” (p. 124). For Manovich and others, users must be critically aware of the fact that every time they select an item from a drop down menu, fill out a delimited text box, or even click on a button (Pold, 2008), they are aligning their intentions with the “libraries of predefined objects” (Manovich, 2001, p. 124) of software developers. The kind of meaning that users can make is fundamentally limited by the way the application being used was designed. Any software application could have been designed otherwise, but for reasons usually beyond users’ grasp, it was not.

When one uses software to author meaning, one chooses to make what one wants to express comply with the limitations set out by software’s creators. The result is sometimes tragic, as was the case when NASA (i.e., National Aeronautics and Space Administration) scientists working on the ill-fated Columbia space shuttle mission let PowerPoint software dictate how to convey important information to administration (Adams, 2006, p. 407). Administrators never got the message and people died. Kitchin and Dodge (2011) argue that software is:

both contingent product of the world and a relational producer of the world ... written by programmers, individually and in teams, within diverse social, political, and economic contexts. The production of software unfolds—programming is performative and negotiated and code is mutable. Software possesses secondary agency that engenders it with high technicity. As such, software needs to be understood as an actant in the world—it augments, supplements, mediates, and regulates our lives and opens new possibilities—but not in a deterministic way. (p. 44)

Similar to Manovich’s warning that authorship via software becomes one of selection, Kitchin and Dodge add that what is authored can also take on its own life in the world, an insight Berry (2011) also noted with regard
to data streams that push and pull from users’ devices without their knowledge. Although authoring via software can be creative and original, Kitchin and Dodge (2011) add that one must also be skeptical: “... software is a powerful force for homogeneity, rather than the diversity that marks creativity” (p. 123). This last point merits underscoring for researchers. The same insights the authors make about aesthetic creativity applies to the creative work of research. Anytime in researchers’ workflow they choose a template, select a menu option, type into a textbox, or click a button, they are narrowing the possibilities for their understanding to comply with the predetermined limitations of software developers.

Research in the Digital Age

Research dealing with social phenomena and human activity is primarily conducted within three major empirical research traditions: quantitative research, qualitative research, and mixed and multimethods research. Quantitative research is typically identified with numeric data in controlled settings and is generated by surveys, tests, inventories, and checklists. Qualitative research deals with non-numeric data in naturalistic settings generated primarily through observation, interviews, and artifacts. Mixed methods research recognizes the power of combining both qualitative and quantitative approaches, a “class of research where the researcher mixes or combines qualitative and quantitative techniques, methods, approaches, concepts or language into a single study” (Johnson & Onwuegbuzie, 2004, p. 17), whereas multimethod research combines multiple qualitative elements or multiple quantitative elements. Although methodological purists have been quick to point out the incompatibility thesis (see Howe, 1988) that asserts that qualitative and quantitative approaches cannot be mixed due to their underlying philosophical and epistemological differences, it should be noted that all research is inherently mixed (see Onwuegbuzie, 2012), and, as such, should be considered plural, dynamic, and compatible when we design contemporary studies—particularly those that include online spaces and digital data generated online.

Mixed and multimethod research is rife with multiple philosophical stances, such as pragmatism in its various forms (Johnson & Onwuegbuzie, 2004; Onwuegbuzie, Johnson, & Collins, 2009), pragmatism-of-the-right philosophy (Putnam, 2002; Rescher, 2000), pragmatism-of-the-left philosophy (Maxcy, 2003; Rorty, 1991), dialectical pluralism (Biesta, 2010; Johnson, 2012; Johnson, Onwuegbuzie, De Waal, Stefurak, & Hildebrand, 2016) and critical dialectical pluralism (Onwuegbuzie & Frels, 2013). It has been argued that researchers should strive toward the radical middle in conducting research which should not be a passive and comfortable middle space wherein the status quo among quantitative and qualitative epistemologies is maintained, but rather a new theoretical and methodological space in which a socially just and productive coexistence among all research traditions is actively promoted, and in which mixed research is consciously local, dynamic, interactive, situated, contingent, fluid, strategic, and generative. (Onwuegbuzie, 2012, p. 192)

This is particularly relevant to researchers of digital spaces or who use digital methods. When adopting the radical middle stance, researchers are more likely to become critical of the data generated, understand that gaps occur within the data that they have collected, and, subsequently, learn to pay attention to the affordances and constraints of those particular data.

For instance, when researchers consider the ontology of the digital, it becomes clear just how unstable digital spaces are. What might be in an online space today, might be gone tomorrow. Anyone who has encountered a broken Uniform Resource Locator (URL) has experienced such ephemerality. Further, Terms of Service (ToS) and End User Licensing Agreements (EULA) are constantly updated and changed, thereby shifting a user’s expectations in accessing a site. For example, Figment.com (circa 2010), an online community publishing site for writers, was acquired by Random House in 2012. After the acquisition, Random House publicly indexed all of the writing reviews on their site, thereby moving the previously private reviews into a public space. In other instances, various online sites and online services may disappear overnight, causing both participants and researchers to search for similar places to inhabit. Anthropologist Bonnie Nardi’s research (cf. Pearce, 2010) of the Uru diaspora (Uru was an online massively multiplayer game published by Ubisoft) is a case in point that researchers must attend to the ebb and flow of online spaces that frequently open and close with little to no warning. Her research chronicled the flows of players as they moved to inhabit different games and online worlds when Uru closed its online game in 2008. She explored the tensions, disappointments, and stresses that come with being displaced from a digital space.

Research in digital spaces has increased in the last decade as scholars have begun to note the relationship between people’s online and offline daily lives (Hine, 2000; Leander, 2008). The two are not separate. Rather, meaning making moves and flows across and between spaces (Gerber, Abrams, Curwood, & Magnifico, 2017).
Researchers might think of “networked field sites” and pay close attention to the movement and flow of meaning making that happens in, around, across, and among users, sites, and resources (Gerber et al., 2017, p. xix). For example, Justine Secco, a senior director of communications at IAC (InterActive Corp), traveled from New York City to South Africa via London in 2013. During her London layover, she Tweeted out to her 170 followers, “Going to Africa. Hope I don’t get AIDS. Just kidding. I am white!” She then powered down her phone and settled in for the 11-hour flight to Johannesburg. Upon landing her phone was filled with texts and missed calls from family and friends indicating that she was the number one trend on Twitter due to her comments that had been deemed as racist by the Twittersphere. Upon landing in South Africa, she was fired from her job and subsequently sunk into a deep depression (Ronson, 2015). As this brief example suggests, our lives are inextricably intertwined with what we do both online and offline; actions and activities in both spaces inform each other.

Digital spaces are porous (Burnett & Merchant, 2014) and, as such, require researchers to consider the methods by which they trace user movement. For example, a student does not simply remain in one classroom all day and glean all of their knowledge about a particular topic within those four walls. A sociocultural view of learning (Hutchins, 1995; Lave & Wenger, 1991; The New London Group, 1996; Vygotsky, 1978) suggests that the student is informed by countless interactions and intersections with a variety of peoples, experiences, resources, and circumstances. The same can be said for online activities. People move among and between websites and online social groups through hyperlinks, retweets, reblogs, likes, and shares. These traces of learning are known as “networking residues” (Grimes & Fields, 2012, p. 43). Due to the complexity of studying people in online spaces, and understanding that singular data points will not paint a rich and nuanced picture as to how users move across and between spaces, mixed and multimethod research approaches are logical for better understanding the meaning-making activities that happen between and across spaces.

On Mixed and Multimethods Research in the Digital Age

Although research in the digital age is increasing, and a growing number of researchers are seeing the importance of researching interactions in online spaces (Black, 2008; Curwood, 2014; Gatson, 2011; Lammers, 2016), mixed methods and multimethod research is not the primary method used to examine digital and online communication practices. Snelson (2016) conducted a systematic literature review of social media research and found that between 2007 and 2014, only 259 studies had been conducted around social media use across all disciplines, with the main site of inquiry being Facebook. Out of the 259 studies, only 55 of the studies used some form of mixed methods or multimethod approaches (approximately 21%); additionally a large majority of the researchers of those studies did not self-identify their approaches as mixed or multimethod in the study’s description. Furthermore, Snelson noted that only a smaller subset of those mixed methods research studies relied upon the use of social media metadata in their mixing of data sources. (Note: Snelson does not give an exact number of studies that use social media metadata, but it is less than 11 studies total, because it is within this bracket that she lists the example of network analysis). This is in line with the findings of Onwuegbuzie, Gerber, and Abrams (2017) who conducted a prevalence rate study of mixed methods research within the field of communication. These researchers analyzed two of the highest ranked communications journals from 2009 to 2014 (Human Communication Research and Journal of Computer Mediated Communication) and found that less than 10% of the empirical articles in these two aforementioned journals used mixed or multimethod research approaches (4.4% and 9.7%, respectively). In both cases, Onwuegbuzie et al. (2017) and Snelson (2016), noted that the authors of the sampled studies did not identify the approaches that they used in their research as mixed or multimethod within their study’s description.

Because a plethora of data exists in digital spaces and, as previously indicated, an individual’s meaning making flows and moves across and between porous digital spaces, it is crucial to consider employing multiple methods to capture the richest and most nuanced picture of the meaning-making activities that are occurring with users. This means that in order to capture what is happening in digital spaces, researchers must think beyond qualitative and quantitative binaries and towards mixed and multimethod approaches. As Kafai and Fields (2013) indicated in their rationale for mixing methods to study young gamers:

> Some researchers have begun to acknowledge that the dividing lines between quantitative and qualitative methodologies create a false dichotomy because each of the perspective[s] contributes to our understanding of what, when, and why players engage in these worlds. Furthermore, the complexity of virtual worlds indicates that not any one data source alone but the triangulation of many may do better justice in understanding player practices, purposes, and psychology. (p. 265)

Further, Gerber et al. (2017) suggest that multidimensional data collection will help researchers paint a more nuanced picture. Multidimensional data collection stems from capturing multimodal data sources from across
Various domains in an effort to paint the richest and most nuanced picture of the activities that are occurring in and across online spaces. In addition, many data sources used in research that includes online spaces incorporate data that are both numeric and non-numeric, such as the various social media metadata associated with a singular re-Tweeted meme. The numeric data (e.g., number of retweets, number of geolocations) can be examined alongside the qualitative comments and images, thereby providing the researcher a broader picture than just the multimodal Tweet alone, or than just than the numeric data associated with the Tweet (see Gerber & Lynch, 2017). Onwuegbuzie et al. (2017) refer to data that have both numeric and non-numeric value as *multi-data* (p. 29), suggesting that researchers could combine these multidata with a variety of traditional and non-traditional data collection methods (e.g., interviews, surveys) to obtain a better understanding of “multi-sited, hyperlinked, and hypermobile practices” of a given social phenomenon with “increased flexibility and reflexivity” (Gerber et al., 2017, p. 169). Researchers must be creative yet purposeful when selecting multiple methods to understand their topics of inquiry, and particularly so in the digital age when information flows shift and change on a daily basis.

**Digital Methods**

Although we can argue that mixed and multimethod research is a more logical approach for better understanding social interactions and meaning-making experiences in and across digital spaces, we would be remiss to ignore the impact that social media analytics, which uses big data sets and hails from the field of digital methods (Rogers, 2013), can have on educational research. Digital methods are methods that exist natively within digital spaces, and are only made available through the advent of code, such as agent-based modeling, web scraping, social media analytics, network analysis, and geovisualization techniques. For the purposes of our research, when we discuss digital methods, particularly those dealing with social media, we also refer to the techniques that require a researcher to access back end metadata (data about data) from digital spaces. As explained in the next section, these metadata are ontologically complicated by what data the companies make available and/or what data that web-based tools can access.

**The Ontological Imperative in Digital Methods**

We argue that common understandings of the digital often ignore the complexity of what the digital actually is. Borrowing concepts from software theory, we suggest that the very nature of the digital is inseparable from software, and that software is an unstable linguistic construct that embodies the assumptions and ideologies of those who produce it. Today, the work of researchers across disciplines—and across qualitative, quantitative, mixed, and multimethod approaches—is mediated and shaped by software. On the one hand, this makes new kinds of work possible with unprecedented efficiency. For example, sociologists Burgess and Matamores-Fernández (2016) showcased the possibilities of engaging in issue mapping (a sociological research method that looks at various social issues, including the main actors within the issue) within social media spaces. They mapped the #gamergate controversy across multiple social media platforms, namely Tumblr, YouTube, and Twitter (including social network analysis and the mining of back end user data). Their research highlighted the methodological challenges and complexities inherent in attempting to accurately trace cultural dynamics across platforms. Such challenges are above the surface, however. Hidden from view, the nature of the digital means that the very data that researchers seek and their tools for examining data are themselves unstable (linguistically) and slanted (ideologically). So, how can researchers respond? Writing about digital humanities and social science research, Rieder and Röhle (2012) offer some guidance:

> The heuristic function of digital research methods in the humanities is mostly focused on the finding of patterns, dynamics, and relationships in data. By rendering certain aspects, properties, or relations visible, these tools offer us particular perspectives on the phenomena in which we are interested. They suggest specific ways to view and interpret the data at hand. Though their results may be visually impressive and intuitively convincing, the methodological and epistemological status of their output is still somewhat unclear. (p. 70; italics in original)

Limited not only to the humanities, these authors observe that access to bigger data sets and faster ways to make visual sense of those data might make it easier for researchers to mistake “intuitively convincing” representations of data for knowledge. They go on to identify five specific challenges for social science researchers: the lure of objectivity, the power of visual evidence, black-boxing, institutional perturbations, and the quest for universalism. For the sake of space, we will limit our discussion to considerations of objectivity, visual evidence, and black-boxing.
The Lure of Objectivity

Rieder and Röhle (2012) argued that enthusiasm for the use of software in research can be traced to a historical and popular belief in “mechanical optimism”: the impression that machine processing endows results with a higher epistemological status (p. 72). As we previously suggested, digital methods, software, and computers are often associated with greater objectivity than more explicit human-centered approaches to research. Such an association is understandable, but incomplete and inaccurate. “While the ‘plodding’ capacities of machines can be usefully integrated into many kinds of research,” the authors go on to state, “they should not be taken to guarantee a higher epistemological status of results” (p. 73). As the software theorists above make clear, software gives the appearance of being inherently objective, but it actually masks subjective influence in ways that the average user cannot easily observe: the code itself, the interface design, the structure of information, and the activity of algorithms.

The Power of Visual Evidence

One of the offshoots of the growth of digital data has been the spread of data visualization. Data visualizations are not new (Tufte, 2006), but their proliferation in digital, dynamic, and interactive forms is a more recent phenomena (Yau, 2011). Rieder and Röhle (2012) offer two valuable insights with regard to visualizations and images in digital methods of research. First, they argue visualizations are inherently reductive. That is, visualizations distill the complexity of data into interpretations that fit within the parameters of visual forms like screen size, color, resolution, and so on. And yet, visualizations “can carry so much argumentative weight despite their reductive character” (p. 74; italics in original). Second, these authors noted that, historically, a visualization or image was used as an “object of investigation” (p. 74). In part, it is our collective familiarity with visualizations as prepared modes of communication that lead to visualizations’ rhetorical power. However, with the rise of digital data collection and analysis, visualizations become “an epistemic device” (Rieder & Röhle, 2012, p. 74)—they become ways to make meaning of phenomena rather than to represent meaning that had already been made. We will return to the significance of these insights later in this article. Briefly, we wish to note that when it comes to digital phenomena, visualizations appear to make static and knowable elements that are always in flux.

Black-Boxing

Black-boxing refers to the fact that most software is opaque to users in two main ways (Rieder & Röhle, 2012). First, much of the code that comprises software (i.e., source code) is often unavailable to researchers, which means that different assumptions or prejudices of its developers go unexamined. Second, researchers often lack the unique skills to examine source code or algorithms, which might be referred to as code literacy. Researchers would need to have the technical knowledge to know what they were examining. Earlier, we mentioned that software often comprises many computational and human languages—including advanced mathematics. The combination of inaccessibility and linguistic opacity makes what it means to be digital always somewhat illusive. That is, although users experience particular effects of software, that which is experienced is illusory and a far cry from the complexity of how software itself operates. Recall Kittler’s prior reference comparing software to the unstable Tower of Babel. If it is true that no individual programmer can know the full scope of what his or her “writing does,” it is certainly the case that no researcher using software can fully account for all the ways software (and the people who create it) might shape the meaning made through its use.

Software is an unstable and ideological linguistic construct. It narrows the way its users can make meaning through it. It is a black box that researchers are hard pressed to understand. But software is also ubiquitous. So, what are researchers to do? Rieder and Röhle (2012) proposed the beginnings of a solution. They concluded that when working with digital methods,

questions of methodology should play a much larger role in the documentation of research projects. In order for the results to be challenged and critically assessed, there needs to be a high degree of transparency regarding assumptions, choices, tools, and so forth. (pp. 80-81)

Despite challenges, researchers who engage with digital phenomena and who use digital research tools must attend to the ontological imperative (rigorously examining the nature of the software involved in their research
and respond methodologically with an increased commitment to thorough transparency that determines the what, the where, the how, and the type of data that they were able and not able to collect. In what follows, we offer concrete suggestions for researchers to consider.

**Addressing the Ontological Imperative as Part of Methodology**

In earlier sections, we framed the ontological imperative in terms of software theory and mixed and multi-method research. Specifically, we argued that researchers take epistemological risks when they do not account methodologically for the ontology of the digital. Fortunately, researchers can do more to learn about and make transparent how the hidden aspects of the digital impact their work. In what follows, we offer five steps with guiding questions to support researchers in work that includes digital data and tools.

When addressing the ontological imperative, it is important to do so as it relates to both (a) digital *data sources and services* (i.e., Twitter metadata) and (b) digital tools that researchers use to access, analyze, and render such data (i.e., TCAT tool—Twitter Capture Archiving Tool). To guide researchers in this work, we offer five questions to consider when explicating the ontological imperative in one’s methodological discussion:

1. **What digital tools, systems, and services are at play in my study? Who created them and why?**
2. **What data do these digital tools, systems, and services render?**
3. **What hidden limitations might there be to the data rendered via these digital tools, systems, and services?**
4. **What are the epistemological implications of this ontological analysis?**
5. **What are the axiological implications of this ontological analysis?**

We briefly consider each of the five questions in sequence to illustrate what accounting for the ontological imperative might look like.

**What Digital Tools, Systems, and Services are at Play in My Study? Who Created Them and Why?**

Consider an example of a researcher collecting Twitter data about the death of David Bowie. As part of a workshop offered by the Queensland University of Technology’s Digital Media Research Centre’s Digital Method Summer School, Hannah (second author) received a dataset that was collected with a web-based tool—the Digital Methods Initiative (DMI) Twitter Capture Archiving Tool (TCAT). The dataset included a pull of tweets that referred to “David Bowie” after the artist’s death in January 2016. There are two organizations involved in this example: DMI and Twitter. Firstly, DMI provided a wonderful service when creating the web-based tool for collecting social media data. They lowered a barrier to entry for social science researchers. Their tool enables researchers to simply enter values (such as hashtags) into textbox fields in order to collect data. Their motive for doing so is not driven by profit but rather for the purposes of creating knowledge. Secondly, Twitter is the purveyor of data in this case. Twitter makes its users’ data available via a specific technical process (using an application programming interface or API), which DMI understands and with which DMI integrates. It is reasonable to assume that Twitter makes business decisions about what data to make available at least in part based on its business interests and not academic or epistemological potential. The Digital Methods Initiative TCAT tool allows researchers to choose from several export options, including relatively familiar file formats like .tsv or .txt.

**What Data do These Digital Tools, Systems, and Services Render?**

Using the web-based tool, Hannah received 35 fields of data: id, time, created_at, from_user_name, text, filter_level, possibly_sensitive, withheld_copyright, withheld_scope, truncated, retweet_count, favorite_count, lang, to_user_name, in_reply_to_status_id, source, location, lat, lng, from_user_id, from_user_realname, from_user_verified, from_user_description, from_user_url, from_user_profile_image_url, from_user_UTC_offset, from_user_timezone, from_user_lang, from_user_tweet_count, from_user_followercount, from_user_friendscount, from_user_followers_count, from_user_friends_count, from_user_listed, from_user_withheld_scope, and from_user_created_at. The data were made available via a .tsv file, which was then opened in Excel.
What Hidden Limitations Might There be to the Data Rendered via These Digital Tools, Systems, and Services?

The data presented above are incomplete. At the time of writing, Twitter identified in its developer documentation at least 150 fields of data that are technically available (see https://dev.twitter.com/overview/api). Yet Hannah received only 35. As Rieder and Röhle (2012) suggested in our earlier example, the data exist in a black-box that many social science researchers would be hard pressed to decrypt. To access the data themselves, a researcher would have to take several steps. First, a researcher has to navigate to Twitter’s developer website. Second, a researcher has to know where to look to see what fields of data are available. Third, a researcher has to be comfortable enough with a programming language (e.g., R, Python) to make an API call to Twitter to request data—keeping in mind that the process of doing so changes at least slightly perennially. Fourth, a researcher has to set up an application with Twitter that provides passwords and keys to securely request and receive the data. And finally, the researcher has to know how to make sense of the data or at least to export them into another software program where the researcher can analyze the data (e.g., Excel). It is also worth remembering that even Twitter’s developer documentation does not list all the information on its users that the company collects. Rather, Twitter makes a strategic sampling of data available for developers and researchers to access in a manner that serves its business interests, which might relate to data that are commercially desirable or just technically easier to share.

Incompleteness of data is just one limitation that merits examination. Another is the fluidity of both the data themselves and the way they are accessed. As mentioned previously, Terms of Service (ToS) and End User Licensing Agreements (EULA) are constantly in flux. This means that any dataset identified and acquired—as well as the process for doing so—can change at any time and in multiple ways. First, executives and developers at Twitter can decide to make more or less data available whenever they wish. A reader of this article who decides to recreate our David Bowie dataset above will likely receive at least slightly different data due to the different limitations that Twitter might impose. Twitter can also choose to tighten or loosen restrictions on how it permits developers to collect data, limiting attributes like institutional affiliation, purpose, geography, or maximum size of datasets allowed to be collected. Second, DMI might choose to update its web-based tool (TCAT), which could change the types of data gathered, the number of companies from whom data can be collected, and the mechanisms for exporting or analyzing data. Importantly, even a researcher who makes a direct API call using a programming language does not escape this second type of fluidity because programming languages themselves change. Further, programming languages like R or Python rely heavily on packages and libraries to perform some of their more advanced and useful functionalities. Packages and libraries are pre-written banks of code that have been created to make certain processes faster and more broadly accessible. A novice programmer does not have to learn to code algorithms for text analysis. He or she just has to install the right package from the community. But software is unstable and like the programming languages themselves, packages and libraries change with time. Further, they too are black boxes to non-experts with predetermined assumptions and activities encoded into them yet mostly hidden from view.

What are the Epistemological Implications of this Ontological Analysis?

Were a social science researcher to account for the epistemological implications of the aforementioned ontological analysis, there are several aspects that he or she would want to address. First, results from any analysis are temporal—they exist in a moment. Not only are the lives and activities of human beings using social media in flux, but so too is the way data are collected and made available. This has a direct impact on the replicability of research. Whereas replicability is often viewed by researchers as a key part of methodological rigor and epistemological validity, when dealing with social media data replicability might in fact be impossible: the data themselves might not change, but all the channels for retrieving and making sense of such data can. To counteract such impossibility, a researcher must take care to describe the precise conditions of how data are gathered: what services by which companies are used, the companies’ current policies surrounding such data collection, the total number of data fields available, significant restrictions to such policies, significant changes to the tools, and so on. In addition, a researcher might want to anticipate what aspects of their data collection and analysis are especially susceptible to future replication. For example, one might note whether a company’s process for collecting data changed during the writing of an article. Or, a researcher might note explicitly that he or she used a specific library in Python, the details of which are opaque but might have particular impacts on the research results.
In addition, Rieder and Röhle (2012) warn that researchers should be particularly aware of the way any resultant data analysis uses visualizations. Even in the case that a researcher describes in detail the affordances and limitations of the digital data collected, any accompanying visualization of data should be noted as also being temporal. Visualizations (i.e., graphs, charts) have great rhetorical power, which is one reason they can be so compelling in research studies. They make something known at a single glance. However, as we have discussed, visualizations can also convey a sense of epistemological stability that is uncharacteristic of digital phenomena. Any visualization of digital data collection or analysis should contain a disclaimer addressing its temporality. This might be undertaken in the methodology section of one’s study or in the way a visualization is labeled in a research report. Or, if a researcher is replicating a study, they should add a methodological note in the context section, which outlines the differences occurring in the methods that they employed compared to the methods that they were attempting to replicate.

What Are the Axiological Implications of this Ontological Analysis?

Axiology indicates that one must pay particular ethical and moral attention to one’s research processes (Denzin & Lincoln, 2011). This is particularly the case with digital data that have been generated within online spaces that are fraught with ontological considerations of permanence, replicability, and completeness. Additionally, publicness of digital data is an area of contention in research that includes online spaces. Some researchers claim that the data are already public by the very nature of the Terms of Service and the API (Berry, 2011). Other researchers claim that care and attention need to be given to any data that have been generated in digital spaces due to the nature of how digital data can always be traced back to a user, thereby divulging a user’s identity and thwarting any anonymity a researcher might have thought that they procured (Gerber et al., 2017; Zimmer, 2010). Additionally, because of the continual updating of Terms of Service, API changes, and commercially invested shifts of taking digital data from password protected private forums to open access public spaces (such as the previous Figment.com example), researchers should continually think through how their research might harm an individual if data were made identifiable by de-identifying and anonymizing participants and information about these data (such as in geolocation, @mentions, etc.) as they work through aggregating data into big data sets. Comparable care should be taken to protect participants’ identities when unpacking aggregated data from big data sets. It is still possible to drill down into data and discover an individual’s identity and, therefore, researchers must take care in how they work through these data and examine whether informed consent and other measurements for protection should be considered.

Conclusion

As digital data and tools become increasingly ubiquitous in social science research, researchers must strengthen their critical understanding of how the very nature of the digital impacts the questions they ask, the data they analyze, and the meaning they make. We have argued that addressing this ontological imperative is inherently multi-methodological because although much of the data collected and the analytical techniques employed might appear empirical, objective, and quantitative, in fact, the ontology of software demands that more critical and qualitative attention be paid to the hidden subjectivities, biases, and ideologies undergirding researchers’ methodologies. We even go as far as to suggest that replicability—a hallmark of rigorous research—is sometimes impossible to achieve when researching digital phenomena or using digital tools. All this points toward a fulfillment of what Onwuegbuzie (2017) referred to as the “in the moment” nature of research in the digital age. Although we agree that “methodological innovation” (p. xvii) is necessary, we add that the most important innovation is not necessarily creating new methods, per se. The crux of the issue is not exactly the methods being used, but rather the need to systematically account for the ontological imperative and, in so doing, strengthen our collective work.

Researchers who wish to better acquaint themselves with the ontology of the digital can do so in several ways. First, we recommend using the references below as a reading list. Form a software theory reading group, which, in its simplest form, might be a small collection of colleagues who will agree to read, meet, and discuss the implications of software theory on research methodology. Second, we suggest that organizers of research conferences create space for drawing attention to the ontological imperative. For instance, a call for proposals might refer to this article and work by software theorists. Third, the next time readers are collecting or analyzing digital data or using digital tools, we recommend a moment of pause to ask the five suggested questions. Fourth, in reviewing work for journals and other peer-reviewed publications, we suggest that reviewers apply the five questions to the author’s work and if the work is lacking in providing clarity around the nature of the digital,
reviewers should suggest that the author attend to the ontological imperative. Raising one’s awareness in even this small way will, we believe, have radical effects on how researchers view the epistemological and methodological complexity of living in increasingly digital times.

The etymology of digital is a fitting way to both begin the current article as well as to conclude. We shared how the word’s origins offer a sobering conceptual insight for researchers: although we think of the digital today as invisible, empirical, and objective, its etymology is very human. The origins of the word reveal a corporeality to its use that is lost in its computationality today. We conclude by noting another layer of the etymology. We made the case that the ontological imperative must be taken seriously by researchers, that the ontology of the digital informs the methodology of the digital and consequently informs the epistemology of the digital. In addition, it should not escape notice in the word’s etymology that the very qualitative nature of the hand and fingers were used to think quantitatively, even to make music possible. In short, the word’s origins allude to mixed and multimethod paradigms. It is our conviction that researchers of digital phenomena and those who use digital tools should embrace mixed and multimethod approaches—if not because 21st century innovations are erasing stark distinctions between qualitative and quantitative research, then because it is simply what digital has meant for nearly 700 years.

References


