Facsimile narratives: Researching the past in the age of digital reproduction

Mateusz Fafinski *
Faculté des lettres, University of Lausanne, Switzerland

Abstract

Taking a cue from the reflections and contributions made by manuscript, archival and historical studies, this paper proposes a new approach to the methodology of digital facsimiles. It asks how we research the past in the age of digital facsimiles and what are our biggest challenges and opportunities. By broadening the definition of what digital facsimiles are and re-evaluating their relationship with databases seeing databases as primarily aggregations of digital facsimiles the paper argues that the key methodological challenge lies in acknowledging the narrative within them. This can only be achieved if we incorporate the existing humanities methodologies in any data-related pursuits. The article demonstrates how notions of objective thresholds in data aggregations are in reality failures to notice and acknowledge the inherent narratives within the digital facsimiles that constitute them. At the same time, while deeply connected, digital facsimiles have to be recognised as ontologically separate from their exemplars. This influences greatly the way we research the past and a failure to recognize both the inherent narrativity and the cognitive gap will always lead to questionable results.

Correspondence:
Mateusz Fafinski, Faculté des lettres, Section des sciences du langage et de l'information
Université de Lausanne
Quartier Chamberonne, bât. Anthropole, bureau 3093, Switzerland.
E-mail:
matfafinski@web.de

1. What Are Facsimile Narratives About? An Introduction

Taking a cue from the reflections and contributions made by manuscript, archival and historical studies, this paper proposes a new approach to the methodology of digital facsimiles. It asks how we research the past in the age of digital facsimiles and what are our biggest challenges and opportunities. By broadening the definition of what digital facsimiles are and re-evaluating their relationship with databases – seeing databases as primarily aggregations of digital facsimiles – the paper argues that the key methodological challenge lies in acknowledging the narrative within them. This can only be achieved if we incorporate the existing humanities methodologies in any data-related pursuits. The article demonstrates how notions of objective thresholds in data aggregations are in reality failures to notice and acknowledge the inherent narratives within the digital facsimiles that constitute them. At the same time, while deeply connected, digital facsimiles have to be recognised as ontologically separate from their exemplars. This influences greatly the way we research the past and a failure to recognize both the inherent narrativity and the cognitive gap will always lead to questionable results.

In 1982, Stanisław Lem, Polish futurologist and science-fiction writer, described 'histocannons'—massive computational machines akin to historiographical artillery, able to retroactively 'predict' the past of distant societies, based on the principles of data science (Lem, 2007 [1982], pp. 54–56).

Digital Scholarship in the Humanities © The Author(s) 2021. Published by Oxford University Press on behalf of EADH.
doi:10.1093/llc/fqab017

Those 'histocannons', due to their concentrated power, were able to use predictive modelling to chart the precise course of history (and, due to enormous space distance, the present) of distant planets. Their home? 'The Institute of Historiographical Machines'. Their results? Mangled. And without cultural context and proper humanist-led interpretation often incomprehensible and misleading.

With the 'Institute' and the 'histocannons', Lem gave a vision of digital historiography radically different from the data-positivist Asimov (2004 [1951]) and accurately, if ironically, predicted some of the developments and pitfalls in digital humanities. Asimov wanted a world in which data will deliver history and thus also the future. But his concept was deeply flawed and it was Lem's, admittedly pessimistic vision that is close to our lived digital reality. This vision stressed two main aspects: the enormous possibilities brought by the computational power at the disposal of digital humanists, and the downsides of the belief in absolute objectivity of predictive models. Forty years later Lem's 'histocannons' almost hit the target in their pessimism: aggregating data is not the objective silver bullet Asimov wanted it to be, and the past cannot be researched on data-positivist terms. Just like Lem's novel this article wants to challenge the entrenched assumptions. The point of departure is the critique of two latent concepts of digital humanities scholarship: the concept of the inherent neutrality of databases, and the positivistic model of digital humanities inquiry. It might be assumed that the concept of neutrality of databases does not need to be disproved anymore as it has been repeatedly challenged (boyd, 2016). And yet in practice we see that this disproving did not stop databases being treated like cheap and ready ammunition for 'histocannons' producing often questionable results (e.g. Whitehouse et al., 2019; Safra et al., 2020). This state of things can be explained by a lack of acknowledgement that databases are embedded in a narrative.

In order to explore the continued existence, allure, and popularity of this positivistic model I draw attention to a more complex understanding of digital facsimiles (digital attempts at representations of heritage media) and databases (the aggregations of digital facsimiles). In particular, databases are hereby redefined as collections of forms of knowledge that are, without exception, embedded in a narrative. I also try to show that any epistemology of the digital that will be created has to be rooted in the existing epistemological tradition to be effective.

I posit that when we deal with historical information, large aggregations of digital facsimiles are always collected into a narrative, and are neither inherently objective nor do they represent a more neutral form of historical knowledge assembling. While such collections might be used to create models that enhance our understanding, those models will not work without an acknowledgement of the narrative.

The idea of the database as a form of narrative has been explored by Folsom (2007), but when it comes to the study of the past, the impact of this exploration has not yet been fully realized and researched. While historiography treats narratives as its bread and butter, digital historiography needs more methodological footing in this area. This 'narrative turn', conceptualizing databases as guided by (or within) embedded narratives, helps to understand how data in itself cannot reveal information about the past and is not neutral. Moreover, the inquiry into that problem has been conducted from two separate, but related, directions—data science and digital humanities, and those two research strands rarely compare their notes on methodology.

At the same time, I postulate and address the three paradoxes of digital facsimiles: how they exist in the public sphere but are researched in the private one; how they make sources more accessible but actually create an additional degree of separation between the viewer and the source; and how aggregation, while not a more objective way of studying the past, can still potentially bring us new, otherwise inaccessible, information. I also analyze the way in which digital facsimiles are being consumed by both specialists and non-specialists. Thanks to such a combined methodological approach, I show that it is not the embedded narrative that poses a problem but rather the lack of acknowledging its existence. The narrative is unavoidable, but its acknowledgement is both manageable and necessary. This means that methods of digital research actually result in an increasing number of subjective layers. Thus, approaches claiming that the results of digital method-driven inquiries are more objective than the 'traditional' scholarship are to be rejected.[1] The recent explosion in digital humanities projects creates a particular form of information noise in

which the inherent narratives are even more obscured. The question of our relationship with digital facsimiles and the narratives that bind them together in databases is then particularly relevant. In such conditions it is difficult to critically engage with the results of digital scholarship, and that state of affairs therefore has to be remedied.

2. Digital Facsimiles and Their Habitat

History is, in many respects, a data science 'par excellence'. The very principle of using sources in order to write narratives about the past (an endeavor often recognized to be less or more quixotic, depending on the methodological stance of its practitioners) resembles the basic tenets of data science. When digital methods emerged, the methodological situation became more complicated. If the old understanding was that 'methodologies are usually byproducts of a philosophical view' (Popper, 1960, p. 54), when it comes to digital historiography, the methodologies we use are mostly a byproduct of our praxis. In other words: we make do with what we have. This situation is partly a result of what can be termed as the inheritance model of digital humanities inquiry. The inheritance model means that the tools of digital humanists used in the early stages of the field (and partly to this day) have mostly been created for purposes very different than the study of the past. There are, of course, exceptions to the tools 'inherited' from other disciplines.[2] This situation has led to the creation of a mixed methodology, which tries to use tools not necessarily made to perform a particular task.[3] Therefore, very often our methodology has been the product of an on-going experimentation. The best example, and maybe the most striking, is the birth-project of digital humanities, the *Index Thomisticus*, the complete lemmatization of Thomas Aquinas' work, which started in 1949 on IBM punch cards (Vanhoutte, 2013; Nyhan et al., 2015).

Further, because there exists a significant strand in continental scholarship that has dealt with the problem of multidisciplinarity in history by relegating certain disciplines to the status of *Hilfswissenschaften*—auxiliary sciences; a similar fate awaited scholars of digital historiography as it has been categorized as a method, and thus of subservient value, even by some champions of its application.[4] There is, of course, another extreme: a strand opposed to seeing digital historiography as in any way subservient. There is a research about the incoming 'power of big data to illuminate the shadows of history' (Guldi and Armitage, 2014, p. 117), as if just the properly managed data itself could be a gateway to new discoveries about the past.[5]

Both of these approaches—'digital humanities as *Hilfswissenschaften*' and 'digital humanities as a blank-slate independent discipline'—are problematic. In this complicated methodological situation, we need a theory of 'possible history' to create a viable digital historiography (Koselleck, 1985, p. 155). It is important to employ meaningful explanations when visualizing patterns or mapping our data, or in the general use of digital methods. 'Meaningful explanations' mean multilateral and explicit narratives that acknowledge the biases of our data, the facsimile character of what is visible on the screen, and provide an explicit narration of the conceivable queries by the end users of our tools. Which queries are possible within a database or even within a single digital facsimile is dependent on the narratives written by their creators. A query as a method of use changes the way we look for information and how the search process is conducted (Fiormonte et al, 2015, pp. 191–95; Antonijevic and Cahoy, 2018). Assembling our scholarship is just as dependent on data structures as was the case with 'traditional' methods, but now a further layer—the narrative of our queries—is added. Those layers obscure the inner workings of digital methodologies and thus have to be documented—their narrative has to be spelled out as well. In addition, a query not as a tool but as a method is a problematic concept, prompting fundamental questions about the originality of research and the task of a historian (Hitchcock, 2008, pp. 81–90). Prescott called it the 'danger of googleization' (2008, p. 14).

This need for 'meaningful explanations' in light of its challenges can then only be fulfilled if we look into the process of knowledge extraction in digital historiography. Almost all study of the past, today involves an element of the digital; thus, when a scholar claims to be starting a 'traditional' project, they are bound to include at least some digital tools in its execution. It

has been a protracted revolution, but one that has changed significantly the way we write and disseminate history. In the center stands the digital facsimile—the key unit of 'possible history' in the digital age.

What are digital facsimiles? The problem with this term is that in the last thirty years, it has vastly evolved and therefore contains multitudes of definitions, approaches, methodologies, and production and display tools. At its core is an understanding of 'reproduction': a digital copy, mostly used to denote a digital image of a manuscript (Ciula, 2009). But digital facsimiles are also understood as already combining visual and textual layers: '[they] are both texts and electronic resources' (Donaldson, 1997, p. 178). Some scholars like Tarte and Endres propose to understand them as *avatars*, artefacts crossing into the digital realm (Endres, 2014). Digital facsimiles are most often seen as digital reproductions (Burns, 2014, p. 158) of sources (primary and secondary): images made to be displayed on computer screens; for example, photographed manuscript pages, 3D models of artefacts, and scanned books.

But this definition of the digital facsimile as a 'reproduction' can be expanded and modified. First of all, the key is recognizing the subjectivity of facsimiles. They are, in fact, not exact copies, not reproductions, but representations of a source, by necessity altered (Dahlström, 2019). When a digital facsimile of a manuscript is created, color corrections are made and, moreover, a 3D object is moved into 2D. Some viewing tools, like Mirador, even allow a limited agency for the end-user in those corrections by allowing individual manipulation of the images on the fly. When a textual digital facsimile is made—like a transcription or transliteration—the encoding of text becomes the basic interpretative layer. A good example of this practice is the *Open Arabic Periodical Editions* project (Grallert, 2020). This ontological difference has far-ranging consequences. As representations, digital facsimiles are geared towards interactivity and (obviously) do not reproduce the source in a material similar to the original (like 'analog' facsimiles).[6] This interactivity might be extremely basic, and digital facsimiles might constantly fail to achieve their interactive goal, but it remains a goal nonetheless. The ability to manipulate them is also one of the key aspects of how they can influence the perception of their physical counterparts through recontextualization or virtual rejoining of previously dispersed fragments (Tarte, 2011).

Digital facsimiles are then every form of a digital representation of a source, not only those that intend to graphically reproduce those sources and they are ontologically separate. The form difference between, for example, a photographed manuscript page and a digitally stored text of its transcription, occurs only on the side of the human consuming those digital facsimiles.

Therefore, every point of historical data transformed into a digital form is a digital facsimile—an attempt to produce a representation of a source, primary or secondary: a digital simulacrum. Digital facsimiles never exist in isolation—they are always accompanied by some form of metadata.[7] This metadata can be extremely simple and thus (wrongly) assumed to be non-existent: a filename or just an address in the storage space or on the web; or it can be elaborate, providing context, periodization, or linked with other facsimiles, becoming more semantic.

One more aspect of digital facsimiles has to be addressed: loss. This is a crucial feature of the conversion from analog to digital (Treharne, 2013, p. 466). 'Digital facsimile, as any other surrogate, is a representation of some of the facts of the document, but not all of them' (Pierazzo, 2016, p. 94). Too often we focus on what the digital can bring, forgetting about all the elements that we lose. The process of digitization always involves the choice of what will be omitted. If we take as an example a medieval book, only selected layers can be transmuted from a plenitext—that is, words, images, spaces, folios, quires, binding, and all other material aspects of a book or fragment (Treharne, 2013, p. 470)—into the digital realm. These specific 'poetics of loss' will weigh heavily on our considerations here. One is inclined to see this loss as a wound—an irreparable void that gapes in the digital realm. This wound has far-reaching implications for the narratives that are built on the basis of these digital facsimiles; to borrow a literary theory term, they are 'narrative wounds' that skew the way the narrative functions (Fleishman, 2018).

This only strengthens the observation that digitization is not the same as preservation. Digitization can foster preservation attempts by re-focussing attention

on the originals and thus helping with their restoration. But it is somewhat ironic that optical media storage strategies are actually rather short-lived (NIST, 2007). In a study of rescuing data resources, over twenty factors inhibiting recovery of data were identified and, together with factors known to be dangerous for traditional media like humidity, most of them focussed on the problematic sustainability of digital forms (Ross and Gow, 1999, pp. iv–vi). Digital facsimiles are representations, not reproductions (and even as reproductions they would not preserve their originals, for that reproduction would only show the source at a particular moment in time).[8] This can only be fully understood when we view digital facsimiles as entities separate from the sources they aim to represent. The initial optimism assuming the merger of the original with the copy thanks to digital reproduction (Davis, 1995) seems not to be applicable anymore, although belief in this theory can account for the still present danger of the positivist trap and the underlying assumption of equating the source and its digital facsimile. The very creation of a digital facsimile is a complex process of remediation (Kichuk, 2007). The 'aura' of Benjamin remains of course an important aesthetic category when discussing this divide (Benjamin, 2012b)—what parts of the original work's conditions, if any, are 'preserved' in its digital form? (Bratu Hansen, 2008).

Digital facsimiles can also be aggregated, thus forming databases: from the point of view of a researcher of the past, these databases are the collections of digital facsimiles created according to at least one aggregating principle. These principles can be ontological (based on the characteristic of the digital facsimiles themselves) or epistemological (based on a question or problem). A different way of dividing digital facsimiles would be according to whether they have been automatically aggregated or curated. Automatic aggregation requires some form of metadata standardization—either applied to the digital facsimiles or extracted from it, if they are machine-readable. Curated aggregation is usually carried out as part of a research project. Both of those types can also exist in mixed forms where automatically aggregated collections are reaggregated through curating and vice versa.

And here especially historians cannot forget that their sources possess an aesthetic dimension. Digital facsimiles as an aesthetic category are not without their political and social repercussions. Their very existence is possible thanks to the age of digital reproduction, which, although it shares many features with the age of mechanical reproduction (Benjamin, 2012a), is distinct in one major feature: its place in the public or private sphere. While not exclusively, most digital facsimiles are researched in the private sphere. Their aggregations are available in the public sphere 'par excellence': the Internet. Nevertheless, access may be limited for research purposes, often behind paywalls, or only available for the chosen few curators, by accidental or deliberate lack of proper documentation.[9] In other words, while aggregations of digital facsimiles exist and can be viewed in the public sphere, they are usually worked on in the private sphere, in front of our personal computers.[10]

This presents the first paradox: that of consumption. Digital facsimiles allow a completely different form of heritage consumption, which although akin to the experience of browsing a library is nevertheless a distinct phenomenon which changes the nature of our experience (Prescott, 2020). Digital facsimiles are not only used in research, but are also 'hyperconsumed', that is, consumed for non-functional purposes by members of the public, students, and researchers, as pure objects of desire (Lipovetsky and Charles, 2005). Their digital presence is then detached not only from their material context but also from a research framework; they become 'browsing objects' – heritage entities consumed in a form of phenomenological parentheses. In those moments of hyperconsumption, the interpretative layer embedded in them by their creators is even more important, for they are not to be consciously theorized by their viewers; in other words, only the methodological framework anticipated by the creator of the digital facsimile and the curiosity of the viewer guide the user in that moment. Even for purely scholarly applications, the form in which a source is handled and the tools used have great consequences for the outcomes of our research (Treharne, 2011). In the digital world the division between hyperconsumption and consumption is never a sharp one and the coexistence of those two means that everyone is a 'professional' and everyone is

an 'amateur'. The paradox of consumption means that while the 'connoisseur' might be gone (Prescott, 2020, pp. 39–41), we are surrounded by digital splinters that at any given moment can be hyperconsumed. This process is tightly bound to our everyday experience, social media platforms like Instagram or Twitter create a plane of interaction where this hyperconsumption can constitute a way of connecting with the materiality of facsimilized objects. At the same time the narrative behind the creation of those digital facsimiles becomes the guiding factor of this hyperconsumption (Green, 2018). Medieval manuscripts for example appear then in a form of isolated snippets, echoing the nineteenth century practice of cutting out illuminations or singular pages. The decision what to cut is then part of an inexplicit narrative. This phenomenon is not limited to social media interactions. Some digital repositories, like the *Bibliothèque virtuelle des manuscrits médiévaux*, for some of the manuscripts host digital facsimiles of only selected decorations or pages. The narratives that underline those choices are not only curatorial but also economic.

The politics of digital facsimiles manifest themselves in many ways. When talking about digital representations of sources, the question of unseen labour is a prominent one (Whearty, 2018). The imbalance of credit and remuneration between the creator of the digital facsimile, be it a scan or a data entry, and its curator seems to mirror the broader imbalances characteristic of late capitalism. This raises the issues of erasure and the human cost of creating digital facsimiles. Even though an overwhelming majority of digital facsimiles and their aggregations are free to view, using and reworking them is often not.[11] Big data create its own particular normativity based on the creation of engagement rules along the lines accepted by a somewhat artificially created majority (Harper, 2017). When it comes to digital facsimiles, a similar process is at play: the choice of sources which form a base for digital facsimiles reflects a particular aesthetic as well as the epistemological programmes of their curators.[12] Moreover, even those programs are severely dependent on and restricted by the funding of digital projects. Digital initiatives are overwhelmingly reliant on third-party funding initiatives with their own agendas. Thus, aesthetics and epistemology become at the end of the day determined by a late capitalist dynamic of monetization. This aesthetic dimension feeds into the narrative of the database.

3. The (Pseudo)Positivist Trap

The use of digital methods can perpetuate the myth of objectivity, entrenched in the positivist understanding of history. Spurred by the legend of big data neutrality, such a view is grounded in the belief that 'although data are big, it can be explained in terms of a relatively simple model with a small number of hidden factors and their interaction' (Alpaydin , 2016, p. ix). It assumes that even though we cannot grasp the vast amounts of data, they are, after passing a certain threshold of magnitude which is notoriously difficult to define (Blanke and Prescott, 2016, p. 186), 'objective'. What this 'objectivity' brought by technology means is rarely understood, and can be only properly grasped by looking at the past (Christin, 2016). 'Objectivity' has been shown to actually be riddled with biases and dangerous assumptions (boyd, 2016). But 'objectivity' is also a myth. There is no 'raw data' that comes before the 'fact' in data science (Rosenberg, 2013) and, as famously quipped by Bowker (2008, p. 184) 'raw data are an oxymoron'.

An additional problem with big data in history is that some forms of historical information (e.g. manuscript images, quintessential digital facsimiles) are inherently more difficult to handle in computing environments than even immense amounts of statistical information (Blanke and Prescott, 2016). They are rarely shallow data, and they do not aggregate well.

In historiography, the existence of 'objective narratives' has been long debunked (Barthes, 1975; White, 1984), but a lack of methodological consideration in terms of digital methods means that although in theory a subjective narrative is accepted, in practice it is still latent. The introduction of digital facsimiles in large, even massive numbers in the past two decades, and the easy way in which those lend themselves to seemingly impartial analysis, creates what we can call a simulacrum of objectivity.

I call it a 'positivist trap', for it relies on a simplistic understanding of positivist approaches (thus it might even be referred to as being 'pseudopositivist'). This trap of practical positivism creates a form of facsimile empiricism—the sources in their digitized equivalent

are not only treated in the same way as their physical counterparts (which is a perception-like problem) but are given an even greater degree of perceived objectivism. 'Objectivity' (which is in reality an assumption) thus enables an uncritiqued bias replication. Arguments have been made that contrary to the pervasive assumption about the diminishing importance of individual interpretations in the presence of big data, actually the opposite is true—big data foster strong individual interpretations (Jauréguiberry, 2017). The same is true in the face of digital facsimiles (even though they operate on a different magnitude level)—their hypermodern character actually makes them more subjective as objects of inquiry. This modern pseudopositivist approach might be seen as an attempt to escape the post-modern trap of a relativistic approach to 'truth' in the past (Jenkins, 1995) through hypermodernity. This assumption just perpetuates the positivist understanding of the 'historical fact' as a brick used to reconstruct past reality; but data instances are not even historiographical 'facts', for they lack the extraneous interpretation and suprasource information necessary for the creation of a historical synthesis.[13]

Digital facsimiles thus present a second paradox—while they make the source material more accessible, they simultaneously create an additional degree of separation between the source and us. This creates a gap reminiscent of the *écart* of Simondon (2007, p. 274). This gap is not a prolongation of our distance from the source—for it makes it actually closer to us—but a distortion of our cognitive field. In this gap exists the digital gaze. It makes us behave as if we were interacting with a product of the past, while we are manipulating and looking at a simulacrum. This might lead to the revival of the pseudopositivist trap, akin to when a historiographical fact (an interpretation) becomes conflated with the source itself. Acknowledging the *écart* does not mean accepting a nihilistic and relativistic mode of writing history. It does not mean that every conclusion that we reach with the help of digital facsimiles is relative and thus, ultimately, pointless. On the contrary, it makes writing history an even more crucial pursuit by highlighting the mechanisms that underpin it in the digital age.

One could make an argument that similar mechanisms can be shown for a so well 'domesticated' phenomenon as an edition of a source. But the difference is fundamental. A scholarly edition does not mimic the source as a digital facsimile does. An edition is clearly a separate object ontologically (an edition is an 'approximation' of how historians assume the original text to have been intended by the addition of emendations and variant readings). Editions have their own problems with narratives and their nature as an intermediary can be missed or forgotten but is very rarely, if ever, hidden. This does not make an edition a neutral object. It is neither the source nor its facsimile and often presents discussion (a choice of a variant) as attestation. Paradoxically it was the arrival of digital facsimiles that allowed us to critically assess the old printed editions at a larger scale and see those problems. Before, the choices of the editors, while documented in the footnotes, remained transparent only to a degree. The text encoded in an edition was not authentic but it seemed so for many users because it often was the first (or even the only) point of contact with a source. Broadly available digital facsimiles have made it possible for us to see the ontological difference between the source and the edition more clearly. Seeing this we can only concur that 'digital simulacra', are then explicitly not just new editions (Treharne 2013, p. 465).

That being said, digital facsimiles can form parts of editions, even obscuring them visually (e.g. superimposing an edition on a digital facsimile, giving the possibility to seamlessly interact with both). There are multiple forms and best practices of using digital facsimiles as editions (Prescott, 2008, pp. 8–12). And digital editions provide us with the clearest sightings of the 'editorisches Narrativ'—'editorial narration' (Ralle, 2016, p. 152), a tacit acknowledgement that the creator of a digital edition is guiding the end user according to a particular narrative. The way a digital edition is defined is also different. First and foremost, it has to be defined by its possibilities—not through can but through could; in such a definition it is the embedded interpretative layer that truly shapes the character of a digital edition. The proliferation of digital editions, offering a practical framework for diplomatic and side-by-side single manuscript editions, is changing the very concept of the approximation-driven critical edition. This will, and already does, change the way we interact with historical sources.

This already complex methodological position becomes even more complicated when digital facsimiles become aggregated, forming databases. Contrary to the mantra of big data, digital facsimiles will not validate a positivist approach to history—data are unable to reveal to us the truth about the past by itself; similarly, there is no critical mass of digital facsimiles that when aggregated enable us to 'see' history.

The process of digitization brings with it then also digitized narratives. But those are not only representations of already existing (e.g. historiographical) narratives but also separate narratives in their own right. Thus, digital facsimiles are separate ontologically from their exemplars, bringing with them their own problems and narratives. But at the same time, they are also anchored in the already existing narratives surrounding their exemplars. This double narrative entanglement of digital facsimiles makes them particularly complex methodologically.

4. Facing the Narrative

If, then, data are not history (and a map is not the territory), how to explain the third paradox, that an aggregated collection of digital facsimiles actually allows us to gain new knowledge? This is because of what those collections of data can do: open new interpretative possibilities and inform our narratives better. But data, like record, are not evidence, and therefore can only potentially serve as such (Furner, 2004). The narrative is the hinge around which the whole methodology of digital facsimiles revolves. It is through the acknowledgment of the narrative that historians can simultaneously avoid the pseudopositivist trap and enhance their research when using digital facsimiles. Digital facsimiles and their aggregations can, when properly used, even save us from committing errors (Prescott, 2008, pp. 15–22). Cautiously, one can remark that even though digital facsimiles are not historiographic 'golden tickets', they do offer a knowledge production platform. When working with digital facsimiles, we are researching a created narrative, but this narrative highlights patterns that can be otherwise difficult to notice or to ascertain. Knowledge production in databases happens then not because there is a critical mass of data that causes it, but because of the connection between data points highlighted by the narrative. It is the story, not the data points that does it. Understanding aggregation as an epistemological method is consistent with what has long been acknowledged in historical practice. 'There is always more at stake in historical knowledge than what is contained in the sources' (Koselleck, 1985, p. 153). Databases, through the narratives embedded within, become 'systems of meaning production' (White, 1984, p. 21). This realization forces us to reassess the relationship that we perceive between embedded narratives and digital facsimiles. If we are to use digital facsimiles as tools, we need to have a consistent methodology for them.

Lev Manovich posited that the narrative and the database form 'natural enemies' and threaten each other's very existence (Manovich, 2001). In response or rather in sync with this theoretical reflection, some scholars have noted that a database is actually a quasi-literary genre on its own (Folsom, 2007), and that in practice databases and narratives can also function as complementary objects with the power to supplement each other (Hayles, 2007). From a digital historiography perspective though, the database as an aggregation of digital facsimiles (or extracts of those, which we could call digital regesta) actually *contains* a narrative; its curation, selection, even the initial digitization that has led to its creation, are already acts of historiographical choice that bind digital facsimiles into a narrative. Their creators should then be digital activist archivists, practitioners critically aware of the narrative within. Archivists have already faced and tackled many problems connected with the narrative entanglement of aggregation and use of records but their theoretical contribution remains largely unreflected upon (Caswell, 2016). They have also noted how '[t]he describer selects certain layers for inclusion, and decides which of those to foreground. In this process, there is analysis, listing, reproduction, and so on, but its primary medium is narrative.' (Duff and Harris, 2002, p. 276).

When we notice this narrative, we see that databases are not 'innocent' collections by virtue of their form, for their arrangement itself is an element, an expression of the underlying narrative. It is a dynamic not dissimilar to how the metadata of archives is the product of 'archival representation' (Yakel, 2003). Hermeneutically, those narratives are not hindering our perceptions of databases, but the lack of

'acknowledgement' of their existence does. In that respect they function very much like the prejudices in Gadamer's philosophy, their existence is necessary for our understanding of the concepts in which they are embedded (Gadamer 2004, pp. 277–304). But the real theoretical turn in such understanding of the concept of prejudice has been brought by Habermas' critique of Gadamer, in which he pointed out that there is a distinction between prejudices that are reflected and not reflected upon (Habermas, 1982, p. 283). The Gadamer/Habermas debate over the understanding of hermeneutics (Mendelson, 1979) highlights the growing importance of acknowledgement as a practice. Acknowledgement then becomes a methodological necessity.

Thus, every database is a narrative, and this fact has to be readily recognized. By extension, databases are susceptible to all the biases of a narrative account in addition to their own, facsimile-related problems. In trying to understand the relationship between digital facsimiles and their aggregations (databases), it is possible to see digital facsimiles—the digital simulacra of sources—as the narremes of databases.[14] The databases then form the 'systems of narrative' of those facsimiles (Barthes, 1975).

Embedded narratives can be explicit or implicit (or strong and soft). The choices can be made visible from the start. In strong, 'curated' databases, either the historian-curator has made the choice or the documentation of the narrative has been supplied in a form that makes it explicit—in other words, a narration of the narrative has been supplied. The choices can also be hidden, making the narratives implicit, for either their contents are automatically generated on the basis of a selection algorithm that is not made visible, or the necessary documentation (narration) is not supplied or non-existent.

As Hayden White remarked commenting on Lévi-Strauss, there is an almost indefinite plurality of chronologies to represent the passage of time (White, 1984, p. 11)—similarly there is an almost indefinite plurality of schemas to aggregate digital facsimiles. Those various schemas are the strength not the weakness of digital methods—they allow for novel ways of ordering both data and problems. However, the lack of narrative transparency in those schemas makes aggregations problematic. We are not used to making the narratives visible, for on the one hand we often fall victim to a pseudopositivist trap (digital facsimiles and their aggregations as 'more objective' on their own and thus not requiring the explication of a narrative), and on the other our reflection has been hindered by the fact that the tools we use were not made to be used by historians in the first place. These tools have mostly been repurposed from other disciplines, and their methodologies have been inherited. They were not made to be concerned with implicit narratives, unstable pasts, or plural histories. But the results created with their help exist in the historiographical field and are interpreted accordingly. The growing understanding among digital practitioners that particular elements of our tools possess a semantic significance and the ability to make arguments (Andrews and van Zundert, 2018) shows that there is already an understanding that the *form* of the tools influences the narrative they are presenting. In other words, an aggregation principle of digital facsimiles influences the narrative that will shape the final 'product'.

One method of escaping the problems encountered when working with the inherited tools has been to structure data in the most granular way feasible. The TEI standard is a good example of such a strategy. It was a viable stop-gap solution (especially in light of the efficiency in computing and the access that it brought), but it came not without peril. Granular structuring can lead to overstructuring, in which the implicit (not immediately evident) narratives of such digital facsimiles become increasingly heavy. When, for example, we proceed to map digital facsimiles structured in such a rigid way, we do not map our source, we map its implicit interpretation or its metadata. Both remain a form of a narrative. Again, this is not a bad practice *per se*, but unacknowledging it creates false assumptions about the past.

To a certain extent, in accordance with Gadamer's hermeneutics, narratives are simply inevitable phenomena—that is, narratives will be embedded no matter how their creators try to avoid them. Gadamer recognizes that prejudice is a part of every interpretation. In the realm of digital facsimiles, one can say that the narrative is the necessary part of their creation. Their very existence is an act of interpretation, and this act carries its creator's viewpoint (Gadamer, 2004, pp. 267–72). There is no prescriptive method of reaching narratives—the hermeneutics of

digital facsimiles is an interpretation of an interpretation. This is for us, as digital practitioners, the crucial point: narratives in our work are *unavoidable*.

Source criticism must thus be enhanced in the methodological toolbox of any historian by digital facsimile criticism. Today, historians become curators of digital facsimiles, even if they do not identify as digital humanists. There is, epistemologically, no way of avoiding an engagement with the curation of digital facsimiles. The emergence of information networks forced the functions on both sides of the invention-construction/operation divide of technical objects to converge, thus leading to the function of a technician, capable of understanding and manipulating objects on both sides of this divide (Simondon 2012, pp. 9–10). Similarly, historians must extend their digital capabilities. Methodologically and practically speaking, this is achievable in its totality only in narrow fragments of the field. Thus, the emergence of a historian-curator (of digital facsimiles) makes multidisciplinarity and collaboration not an added value anymore but a necessity—just as the emergence of world history as a system has made the need for theory in history obvious (Koselleck, 1985, p. 103). This cooperation, due to the constraints of the sheer volume of concepts that need to be grasped, will have a 'piecemeal' character, but nevertheless will have to be a constant factor.[15]

Working on digital facsimiles and their aggregations starkly actualizes the argument of Benedetto Croce: 'The practical need, underlying every historical judgment, gives to all history the character of »contemporary history«, because however extremely remote chronologically are the events in it, history in reality refers to the present needs and situations on which those facts exert their influence.' (Croce, 1937, p. 4, transl. MF). But our work becomes 'contemporary history' not only because of the repercussions of its results on our present situation, but also because what we work on is a creation of our time. It is not only the contemporary concern then that informs our work (Roberts, 1995, p. 95), but the very contemporary creation of our material. In other words: when working on the past with digital tools, we must incorporate digital methods, but also consider the political and practical consequences of our choice of digital facsimiles and their relationship with the sources. Only then can we grasp the relevance of digital facsimiles as elements of 'contemporary history'. Their cognition as historical objects as well as the processes within them (in this case the embedded narratives) is only possible through a cognition of the constellations in which they exist (Adorno, 1973, pp. 165–6). We cannot make the narratives in the databases explicit without acknowledging their external relationships.

The recognition of aggregations of digital facsimiles as narratives makes it possible to identify the plurality of discourses within them. They are not 'pasts' but 'histories', and therefore not more (or less) objective than other types of historical narratives. And even digital facsimiles have been around for long enough to start developing their own history—another layer to their interpretative framework (Balbi and Magaudda, 2018, pp. 154–213). But the way they are being used both by their curators and by their end users requires a new philosophy of digital history that centres their facsimile character, their inherent bias, and their use by query.

5. Conclusions

The role but also the praxis of history is changing rapidly. The omnipresence of digital facsimiles means that the need to understand and potentially adopt digital methodologies has to extend outside the digital humanities proper. The use of digital facsimiles of secondary literature is already a widespread practice, and it has fundamentally changed the way we interact with it. The same is now true of the sources. For the overwhelming majority of students, the first encounter with the traces of the past is through a digital facsimile.

Thus, the question of digital methodology in history is no longer (if it ever was) confined to the field of digital humanities. It touches everyone who works with the past. For digital practitioners, who are both the users and creators of tools, it means that it is now necessary to make the narratives of digital tools explicit by providing them with proper documentation and explanation. This will help to create a more level playing field in the most common method of digital facsimile consumption: user queries. This should also mean finally regarding digital tools and platforms as equal to monographs and peer-reviewed articles as forms of academic output in the humanities.

This work of making narratives explicit is crucial for the next stage of developing digital methodologies. It can help historians reposition digital facsimiles and databases as the complex historiographical tools that they really are, thus making them less likely to be used in simplistic positivist interpretations of the past. The fear that the results of digital humanities undertakings will be discounted as 'unscientific' if the pretense of objectivity is not be upheld, has led to inexplicit narratives. The resulting work can misrepresent the past and contribute to the exclusion of marginalized groups in digital humanities. This danger of misrepresentation looms over every study that ignores facsimile narratives.

The creation of digital facsimiles and their aggregations (databases) involves countless choices which result in loss, and which are undertaken according to a certain framework. These choices and losses bind together into narratives of the past—thus, our digital facsimiles and our databases are the part of historiography, which is no longer seen as a purely objective undertaking.

The 'rediscovery of the narrative' in digital form should make us more aware, but it should also lead us to a different praxis—one where we focus on creating tools and methodologies that consider the nature of our data. Close reading together with data analysis will be the methodological goal of those tools (Blanke and Prescott, 2016, p. 191). They will also strive towards a more equal representation of the workers involved in their creation. The semantic overlap of terminology between data science and archaeology can be used to our advantage. Just as sites are excavated, so data are 'mined'. And just as an excavation should not begin without a clear research idea in mind, the creation of digital tools, or digital facsimiles, should not. A lack of initial theoretical reflection will not make them neutral, but will make it more difficult to recognize inherent biases.

The chief argument of this article is epistemological. How is knowledge about the past acquired, but also embedded and interpreted through digital facsimiles and their aggregations (databases)? The answer proposed here is that all three of these processes involve narrative creation. Data in itself is not knowledge. Even the greatest concentration of data (i.e. passing the elusive threshold to 'big data', if it even can be passed) will not on its own bring us closer to getting to know the past. Concentrating data does not write history. The epistemology of digital humanities is therefore an epistemology of the digital narrative. At its basic level, this narrative is present in every form of a digital facsimile and database—it is impossible to create them without embedding them in elements of our own narrative. Making this narrative explicit (and accepting its existence) is therefore not only a matter of ethics (making the position known) but also of epistemology. Only then can we move to the next stage: the interpretation of those subjective narratives in order to gain knowledge about the past. The knowledge production process in digital humanities cannot be decoupled from this subjective layer.

Digital humanities as a form of humanist practice do not exist in isolation from other branches of the humanities. The basic methodological tenets of its functioning need therefore to be derived from the humanities' aesthetics, ethics, and epistemology in order for digital humanities to stand on its own. By acknowledging the narrative turn, we might avoid the pseudopositivist pitfalls of data science and at the same time gain new knowledge about the past. To a certain extent, what is proposed here is that dealing with data is at its core an undertaking deeply rooted in the humanities, and is usually performed by weaving data into a narrative. This narrative heritage is a tenet that still has to be properly acknowledged both inside the humanities and (maybe even more) outside them.

And finally then—back to Lem. In *Wizja Lokalna*, where almost all processes are permeated with computerized microchips, 'historians are still people [. . .] for humanities cannot be automated [. . .] and just cannot be fully given over to the logical circuits' (Lem, 2007 [1982], p. 143). The 'histocannons', fortunately, cannot be left to run on their own.

Acknowledgements

The author would like to thank Sarah Schlüssel, Elaine Treharne, and the anonymous reviewers for their helpful and generous comments on this article. This article would not have been possible without the support the author has received as the Text Technologies Fellow at Stanford University in 2018/19.

References

Adorno, T.W. (1973). *Negative Dialektik: Jargon der Eigentlichkeit*. Frankfurt am Main: Suhrkamp.

Alpaydin, E. (2016). *Machine Learning: The New AI, MIT Press Essential Knowledge Series*. Cambridge, MA: MIT Press.

Andrews, T.L. and van Zundert, J.J. (2018). What are you trying to say? The interface as an integral element of argument. In Bleier, R., Bürgermeister, M., Klug, H.W., Neuber, F., Schneider, G. (eds.), *Scholarly Digital Editions as Interfaces*. Norderstedt: BoD, pp. 3–34.

Antonijevic, S. and Cahoy, E.S. (2018). Researcher as Bricoleur: Contextualizing humanists' digital workflows. DHQ 12.

Asimov, I. (2004). *Foundation, The Foundation Series*. New York: Bantam Books.

Balbi, G. and Magaudda, P. (2018). *A History of Digital Media: An Intermedia and Global Perspective*. London: Routledge.

Barthes, R. (1975). An Introduction to the Structural Analysis of Narrative. *New Literary History*, **6**: 237–72.

Benjamin, W. (2012a). Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit, in: *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit: drei Studien zur Kunstsoziologie*, Edition Suhrkamp. Suhrkamp, Frankfurt am Main, pp. 7–44.

Benjamin, W. (2012b). Kleine Geschichte der Photographie. In: *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit: drei Studien zur Kunstsoziologie*, Edition Suhrkamp. Frankfurt am Main: Suhrkamp, pp. 45–64.

Blanke, T. (2018). Predicting the Past. DHQ 12.

Blanke, T. and Prescott, A. (2016). Dealing with Big Data. In Griffin, G., Hayler, M. (eds.), *Research Methods for Reading Digital Data in the Digital Humanities*. Edinburgh: Edinburgh University Press, pp. 184–205.

Bowker, G.C. (2008). *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

boyd, d. (2016). Undoing the neutrality of big data. *Florida Law Review, 67:* 226–32.

Burns, J.E. (2014). Digital facsimiles and the modern viewer: medieval manuscripts and archival practice in the age of new media. *Art Documentation: Journal of the Art Libraries Society of North America*, **33**: 148–67.

Bratu Hansen, M. (2008). Benjamin's aura. *Critical Inquiry*, **34**: 336–75.

Caswell, M.L. (2016). 'The Archive' Is Not an Archives: On Acknowledging the Intellectual Contributions of Archival Studies. Reconstruction 16.

Christin, A. (2016). From Daguerreotypes to Algorithms: Machines, Expertise, and Three Forms of Objectivity. *SIGCAS Computers and Society*, **46**: 27–32.

Ciula, A. (2009). The palaeographical method under the light of a digital approach. In Rehbein, M., Sahle, P., Schassan, T. (eds.), *Codicology and Palaeography in the Digital Age*. Norderstedt: BoD—Books on Demand, pp. 219–37.

Croce, B. (1937). La storia come pensiero e come azione. *La Critica*, **35**, 1–35.

Da, N.Z. (2019). The computational case against computational literary studies. *Critical Inquiry*, **45**: 601–39.

Dahlström, M. (2019). Copies and facsimiles. *International Journal of Digital Humanities*, **1**: 195–208.

Davis, D. (1995). The work of art in the age of digital reproduction (an evolving thesis: 1991–1995). *Leonardo*, **28**: 381–6.

Donaldson, P. (1997). Shakespeare and electronic textuality. In Sutherland, K. (ed.), *Electronic Text: Investigations in Method and Theory*. Oxford: Clarendon Press.

Duff, W.M. and Harris, V. (2002). Stories and names: Archival description as narrating records and constructing meanings. *Archival Science*, **2**: 263–85.

Endres, W. (2014). 'More than Meets the Eye: Going 3D with an Early Medieval Manuscript', in Mills C., Pidd M. and Ward E., *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. The Digital Humanities Institute, Sheffield. https://www.dhi.ac.uk/openbook/chapter/dhc2012-endres

Fiormonte, D., Numerico, T., Tomasi, F. (2015). *The Digital Humanist: A Critical Inquiry*. Brooklyn: Punctum Books.

Fleishman, I. (2018). *An Aesthetics of Injury: The Narrative Wound from Baudelaire to Tarantino*. Evanston: Northwestern University Press.

Folsom E. (2007). Database as genre: the epic transformation of archives. *PMLA*, **122**(5): 1571–79.

Frischer, B. (2011). Art and science in the age of digital reproduction: from mimetic representation to interactive virtual reality. *Virtual Archaeology Review*, **2**: 19–32.

Furner, J. (2004). Conceptual analysis: a method for understanding information as evidence, and evidence as information. *Archival Science*, **4**, 233–65.

Gadamer, H.-G. (2004). *Truth and Method*, revised edn. London: Bloomsbury Academic.

Grallert, T. (2020) *Open Arabic Periodical Editions*. https://openarabicpe.github.io/ (accessed 03 February 2020).

Green, J. (2018). Digital Manuscripts as Sites of Touch: Using Social Media for 'Hands-On' Engagement with Medieval Manuscript Materiality. *Archive Journal*. http://www.archivejournal.net/?p=7795 (accessed 30 November 2020).

Guldi, J. and Armitage, D. (2014). *The History Manifesto*. Cambridge: Cambridge University Press.

Habermas, J. (1982). *Zur Logik der Sozialwissenschaften*. Frankfurt am Main: Suhrkamp.

Harper, T. (2017). The big data public and its problems: big data and the structural transformation of the public sphere. *New Media & Society*, **19**: 1424–39.

Harvey, C., and Press, J. (1995). *Databases in Historical Research: Theory, Methods and Applications*. London: Macmillan Education UK.

Hayles, N.K., (2007). Narrative and database: natural symbionts. *PMLA*, **122**: 1603–8.

Hitchcock, T. (2008). Digital searching and the re-formulation of historical knowledge. In Greengrass, M., Hughes, L.M. (eds.), *The Virtual Representation of the Past, Digital Research in the Arts and Humanities*. Aldershot: Ashgate, pp. 81–90.

Howard, R. (1995). Apollo 13. Universal Pictures.

Jauréguiberry, F. (2017). L'individu hypermoderne face aux big data. *Socsoc*, **49**: 33–58.

Jenkins, K. (1995). *On 'what is history?' From Carr and Elton to Rorty and White*. London: Routledge.

Kichuk, D. (2007). Metamorphosis: remediation in Early English Books Online (EEBO). *Literary and Linguistic Computing*, **22**: 291–303.

Koselleck, R. (1985). *Futures Past: On the Semantics of Historical Time*. Cambridge, MA: MIT Press.

Lem, S. (2007). Wizja lokalna. Wydawnictwo Literackie Kraków-Wrocław.

Lipovetsky, G. and Charles, S. (2005). *Hypermodern Times*. Cambridge, UK; Malden, MA: Polity Press.

Manovich, L. (2001). *The Language of New Media, Leonardo*. Cambridge, MA: MIT Press.

Mendelson Jack. The Habermas-Gadamer Debate. *New German Critique*. 1979. (18) 44. 10.2307/487850.

NIST/Library of Congress (LC) Optical Disc Longevity Study. (2007). https://www.loc.gov/preservation/resources/rt/NIST_LC_OpticalDiscLongevity.pdf (accessed 12 October 2019).

Nyhan, J., Flinn, A., Welsh, A. (2015). Oral History and the Hidden Histories project: towards histories of computing in the humanities. *Literary Linguistic Computing*, **30**: 71–85.

Pierazzo, E. (2016). Digital Scholarly Editing: Theories, Models and Methods. London: Routledge.

Piotrowski, M., (2018). Digital Humanities: An Explication. In Burghardt, M. and Müller-Birn, C. (Hrsg.), *INF-DH-2018*. Bonn: Gesellschaft für Informatik e.V.

Popper, K.R. (1960). *The Poverty of Historicism*. Basic Books, New York.

Porter, D. (2018). Zombie Manuscripts: Digital Facsimiles in the Uncanny Valley. *Presented at the International Congress on Medieval Studies, Kalamazoo*. http://www.dotporterdigital.org/zombie-manuscripts-digital-facsimiles-in-the-uncanny-valley/ (accessed 11 November 2019).

Prescott, A. (2008). The imaging of historical documents. In Greengrass, M., Hughes, L.M. (eds.), *The Virtual Representation of the Past, Digital Research in the Arts and Humanities*. Aldershot: Ashgate, pp. 7–22.

Prescott, A. (2020). Ways of Seeing Manuscripts: Exploring Parker 2.0. In: Albritton B., Henley G., Treharne E. (eds.), *Medieval Manuscripts in the Digital Age*. London: Routledge, pp. 37–54.

Ralle, I.H. (2016). Maschinenlesbar—menschenlesbar. Über die grundlegende Ausrichtung der Edition. Editio. Internationales Jahrbuch für Editionswissenschaft. International Yearbook of Scholarly Editing. *Revue Internationale des Sciences de l'Edition Critique*, **30**: 144–56.

Roberts, D.D. (1995). *Nothing But History: Reconstruction and Extremity After Metaphysics*. Berkeley: University of California Press.

Roman, S. and Palmer, E. (2019). The growth and decline of the western roman empire: quantifying the dynamics of army size, territory, and coinage. *Cliodynamics*, **10**: 76–98 . 10.21237/C7clio10243683

Rosenberg, D. (2013). Data before the fact. In Gitelman, L. (ed.), *'Raw Data' Is an Oxymoron, Infrastructures Series*. Cambridge, MA; London, England: The MIT Press, pp. 15–40.

Ross, S. and Gow, A. (1999). Digital archaeology: rescuing neglected and damaged data resources: a JISC/NPO study with the Electronic Libraries (eLib) Programme on the preservation of electronic materials, Electronic libraries programme studies. London: Library Information Technology Centre.

Safra, L., Chevallier, C., Grèzes, J., Baumard, N. (2020). Tracking historical changes in trustworthiness using machine learning analyses of facial cues in paintings. *Nature Communications*, **11**: 4728. https://doi.org/10.1038/s41467-020-18566-7

Simondon, G. (2007). L'individuation psychique et collective: A la lumière des notions de Forme, Information, Potentiel et Métastabilité. Paris: Editions Aubier.

Simondon, G. (2012). Technical mentality. In Murray, A., De Boever, A., Roffe, J. (eds.), Boever, A.D. (Tran.)., *Gilbert Simondon: Being and Technology*. Edinburgh University Press, pp. 1–16.

Tarte, S. (2011). 'Interpreting Ancient Documents: Of Avatars, Uncertainty and Knowledge Creation', *ESF Exploratory Workshop on Digital Palaeography*. University of Würzburg, 22 July 2011.

Topolski, J. (1984). Metodologia historii. Państwowe Wydawnictwo Naukowe, Warszawa.

Treharne, E. (2013). Fleshing out the text: The transcendent manuscript in the digital age. *Postmedieval*, **4**: 465–78.

Treharne, E. (2011). The Good, the Bad, the Ugly: Old English Manuscripts and their Physical Description. In *The Genesis of Books, Studies in the Early Middle Ages*. Turnhout: Brepols Publishers, pp. 261–83. https://doi.org/10.1484/M.SEM-EB.1.100485

Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change*. Chicago: The University of Chicago Press.

Vanhoutte, E. (2013). The gates of hell: history and definition of digital humanities computing. In Terras, M.M., Nyhan, J., Vanhoutte, E. (eds.), *Defining Digital Humanities: A Reader*. Farnham, Surrey, England; Burlington, VT: Ashgate Publishing Limited,.

Whearty, B. (2018). 'Invisible in 'The Archive': Librarians, Archivists, and The Caswell Test', *English, General Literature, and Rhetoric Faculty Scholarship*. https://orb.binghamton.edu/english_fac/4 (accessed 15 November 2020).

White, H. (1984). The question of narrative in contemporary historical theory. *History and Theory*, **23**: 1–33.

Whitehouse, H., François, P., Savage, P.E., et al. (2019). Complex societies precede moralizing gods throughout world history. *Nature*, **568**: 226–9.

Wittmann, H. (1975). Théorie des narrèmes et algorithmes narratifs. *Poetics*, **4**: 19–28.

Yakel, E. (2003). Archival representation. *Archival Science*, **3**: 1–25.

Notes

1 A lot of the criticism of the so-called 'computational linguistics' branch of digital humanities comes from the perception of their methods to be either positivistic or essentially re-hashing the arguments made already by 'traditional' scholarship (Da, 2019). A lot of this criticism is valid when it comes to the use of statistical models, sample sizes, and the certain flirtation with a positivistic feedback loop. On the other hand, while computational methods do indeed allow us to see the forest for the trees (Underwood, 2019), their future does not lie in becoming a 'hard science' branch of humanities, but in being part of the narrative, developing own models and creating a more independent methodological and theoretical reflection. This can only be done through more precise definition of what they are (Piotrowski, 2018).

2 A good and poignant example is the EPPT (Edition Production and Presentation Technology), which has spurred its own methodological discussion (Prescott, 2008). In a fitting comment on the sustainability of digital tools, it does not exist anymore.

3 I am reminded here about the scene in the film Apollo 13 when the astronauts have to fit a squared peg into a round vent using only the supplies that they have on board of their small spaceship. If they do not succeed, they will suffocate. Resembling the astronauts in this analogy, digital humanists need to make do with tools not designed to study the past (Howard, 1995).

4 This strand has also not been completely absent in the English-speaking world (Harvey and Press, 1995).

5 Maybe best exemplified by the research published in journals like *Cliodynamics*. A recent article posited that the optimal time for the division of the Roman Empire was mid-reign of Marcus Aurelius, exactly in 171 CE (Roman and Palmer, 2019, p. 17). None of the authors are Roman historians but the article purports to illuminate the past just on the basis of systems modelling. It is a prime example of a positivist trap of 'pure data' that, in this case, ignores a large portion of historiographic discussion on the subject.

6 Interactivity and the interactive feedback are especially important for facsimiles existing in virtual reality, but I argue that even "classic" digital facsimiles are geared towards interaction first and foremost. For the interactive turn in digital reproduction, see (Frischer, 2011).

7 One can, when considering digitized manuscripts, specifically define metadata as an integral part of the digital facsimile according to the formula '(digital images +

structural metadata + additional data)', which understanding is perfectly valid and amounts to the same principle: the necessity of metadata existence in the digital facsimile mix (Porter, 2018).

8 Attempts to present some aspects of digital representations, namely the virtual reality ones, as distinct from the mimetic (by introducing the category of the interactive) rely on a false assumption that the act of manipulating a digital object is not a mimetic action in itself (Frischer, 2011).

9 A great case study is the main German manuscripts database, http://www.manuscripta-mediaevalia.de/. Apart from a truly clumsy interface, the user will be soon confronted with a fascinating but not immediately obvious database narration choice: the structure of the database is guided by old printed manuscript catalogues. In a narrative turn those are available deep in the underbelly of the website in a digitized form, in which every page is a separate PDF document, creating a disjointed polyphony of metadata that even though it guided the very creation of this databank, exists separated from it. The replacement portal in the works (to be released in beta in 2021), handschriftenportal.de, promises greater interoperability and to bring this disjoined information together.

10 Notwithstanding experiments of shared workspaces (like the *Digging Deeper* project at Stanford University), the bulk of digital work actually happens in the private sphere, even when its results are sometimes almost immediately visible in the public one.

11 Although this is not always the case, as the example of Voss Manuscripts from Leiden shows. There, online reproductions require a subscription while the manuscripts from which they were made can, in principle, be viewed for free. A curious case when the access to the simulacrum is more expensive than the access to the exemplum. Perhaps not so surprising in the era of digital capitalism. https://primarysources.brillonline.com/browse/vossiani-latini.

12 This phenomenon is not dangerous in itself, but the obscuring of its existence is. For an interesting example of the discussion on how those choices influence the perception of sources in the public sphere, one can look at the discussions inside the #medievaltwitter community on posting snippets of medieval manuscripts which conform to a particular concept of what can be considered 'pretty'. Again, it is not the choice which is problematic, but the lack of its argumentation and making it explicit.

13 On historical and historiographical facts as atoms of historical methods (Topolski, 1984, pp. 185–192). Topolski's fascination with the interaction between cybernetics and methodology of history was due partly to the inclusion of inherent contradictions as a paramount part of cybernetic systems (where he saw parallels with the Marxist dialectics). Nevertheless, his observations on the mutual influence of methodological toolboxes of digital sciences and historiography, made for the first time over 50 years ago, remain groundbreaking and largely unknown in the West.

14 Narremes are an attempt at defining the basic building blocks of narratives. On narremes and their role as basic units of narrative structure (Wittmann, 1975).

15 On 'piecemeal' cooperation of disciplines Popper, 1960, pp. 79–80). Popper's critique of holism does not hold in light of what the digital methods bring—the sum of aggregated data does bring more than the separate parts of it when put through an interpretative lens and when the embedded narrative is acknowledged and engaged with—but his intuitive understanding of disciplines cooperating on an *ad hoc* basis in order to solve problems encountered seems to be well applicable. Popper's position is of course understandable in his historical context, but, in contrast to the more Marxist inclined philosophers of history, he had not considered the possibility of the arrival of digital methods and thus the arrival of a different form of prediction: inside the datasets (Blanke, 2018). Reading Popper is then for digital practitioners (as for any humanists in general) an exercise that has to be done with extreme caution. Probably the biggest lesson a digital humanist can take from Popper though is that totality does not bring truth—big data will not remove biases from our digital facsimiles just by the virtue of its scale.