UNIVERSITY OF COPENHAGEN

# Digital Humanities and networked digital media

Finnemann, Niels Ole

Article – Theme section

# Digital humanities and networked digital media

## Niels Ole Finnemann

*This article discusses digital humanities and the growing diversity of digital media, digital materials and digital methods.*

*The first section describes the humanities computing tradition formed around the interpretation of computation as a rule-based process connected to a concept of digital materials centred on the digitisation of non-digital, finite works, corpora and oeuvres.*

*The second section discusses "the big tent" of contemporary digital humanities. It is argued that there can be no unifying interpretation of digital humanities above the level of studying digital materials with the help of software-supported methods. This is so, in part, because of the complexity of the world and, in part, because digital media remain open to the projection of new epistemologies onto the functional architecture of these media.*

*The third section discusses the heterogeneous character of digital materials and proposes that the study of digital materials should be established as a field in its own right.*

## Humanities computing – computing the humanities?

In recent years, digital humanities have become a dominant framework for discussions of digital methods within the humanities. The use of software-supported methods, however, has a long history. Roberto Busa started to develop a tool for performing text searches within a digitised corpus of Thomas Aquinas's works, *Index Thomisticus*, in the 1940s. [1] Another project was the attempt to develop a "love-letter generator" in 1952 (Wardrip-Fruin, 2011). A wider range of perspectives was introduced by Roman Jakobson (1960), who regarded binary distinctions as constitutive in language and literature (e.g., the metaphor-metonymy opposition), and Noam Chomsky, whose publications on generative grammar influenced computer programming theory (e.g., Chomsky, 1957). Likewise, the psychologists Miller, Galanter & Pribram (1960) widened the interpretation of digital processes from computational to information-processing or even artificial intelligence. These contributions also initiated a move from static structuralism to dynamic perspectives within the humanities.

In Susan Hockey's *History of Humanities Computing*, only Busa is included as pioneering what she denotes as an "interdisciplinary academic area of activity" (Hockey, 2004: Part 1.1). According to Hockey, initiatives in the 1960s were rudimentary. IBM sponsored a conference on *Literary Data Processing* in 1964. *Computers and the Humanities* began publication in 1966, but series of conferences did not appear more regularly until the 1970s. In 1978, The Association for Computers and the Humanities (ACH) was formed as an outcome of conferences held in the UK and the US, alternately. A core result was the establishment of the Text Encoding Initiative (TEI), which developed *Guidelines for Electronic Text Encoding and Interchange*, the first version being published in 1994.

Hockey's delimitations indicate the defining characteristics. The contributions of Jakobson and Chomsky were interpretations of real, linguistic phenomena and processes, while humanities computing defines itself around the digitisation of non-digital originals: "concerned with the applications of computing to research and teaching within subjects that are loosely defined as 'the Humanities', or in British English 'the Arts.'" (Hockey, 2004: 1.1 Introduction). Computation is praised for the "the rigor and systematic unambiguous procedural methodologies characteristic of the sciences". The idea is to use these characteristics "to address problems within the humanities that had hitherto been most often treated in a serendipitous fashion" (Hockey, 2004: 1.1 Introduction). The interpretation of computation is the primary and explicit constituent, while the area of application is a broad range of objects studied within the humanities. Underneath, the idea of a finite, non-digital original work serves as a more implicit constituent.

Similarly, for Unsworth (2002, Abstract):

> Humanities Computing is a practice of representation, a form of modelling [...] a way of reasoning and a set of ontological commitments, and its representational practice is shaped

by the need for efficient computation on the one hand, and for human communication on the other.

One might wonder what was meant by "ontological commitments" and a "need for efficient computation" and why these ideas occupied such a prominent place in the ontology of humanities computing. The answer is that they put "humanities computing, or rather the computing humanist, in the position of having to do two things that mostly, in the humanities, we don't do: provide unambiguous expressions of ideas, and provide them according to stated rules." (Unsworth, 2002, VI). However, formal expressions do not meet these expectations. The interpretation of the role of formalism was probably written using a word processor, allowing the author to articulate any sort of ambiguity on the level of semantics. The word processor is based on formalisms. Ambiguities can be articulated because formalisms do not define the content and meanings conveyed. Formalisms are not *per se* bearers of any logic; they are syntactical, modular and modifiable in function and meaning. This does not undermine the need for formalisations, but it does mean that formalisms are syntactical derivatives of human thought processes and not their basic form.

Thus, within the humanities computing tradition, "computing" includes only a subset of computational processes. It does not include a large part of what the computing humanist and most people actually do with a computer. Computing as a general notion of what can be done with computers cannot be restricted to include only unambiguous rule-based procedures such as those that Turing (1936) described with his concept of the universal computer. Today, most computational processes are enacted on the basis of what Turing denoted as a choice machine (Turing, 1936, p. 232), leaving the next step or sequence of steps to be specified by a human operator.

The ontological interpretation of computation within humanities computing serves to privilege the formalist methodology as intrinsically superior to other methodologies. It does not ignore the humanities at large, but it minimises the obligation to have the results acknowledged by colleagues studying non-digital original materials with other methods. The question of whether formalisms may serve as an epistemological foundation re-enters the discourse within humanities computing when confronted with the question of the relationship between human language and formal languages. The answer given is that we cannot fully articulate our knowledge of the world in a simple, coherent formal language. In Unsworth's wording:

> Much of this map-making will be social work, consensus-building, compromise [...] Consensus-based ontologies (in history, music, archaeology, architecture, literature, etc.) will be necessary, in a computational medium, if we hope to be able to travel across the borders of particular collections, institutions, languages, nations, in order to exchange ideas. (Unsworth, 2002, Conclusions).

Similarly, McCarty (2002, p. 103) argues that a main result of the attempts to model non-digital originals is that the failures – "an inevitable feature of modelling" – reveal important insights into the very nature of these originals. Computational modelling remained a core theme and was the central dogma throughout the history of humanities computing, defining what should be counted as part of the "interdisciplinary academic area" (Hockey, 2004). So, the idea that modelling was always imperfect was inevitably bad news. For Unsworth, it meant "that all humanities computing projects today are involved in some degree of charlatanism". The good news was the argument that the degree of charlatanism could be measured

> by the interactivity offered to users who wish to frame their own research questions. If there is none offered, and no interactivity, then the project is probably pure charlatanism. If it offers some (say, keyword searching), then it can be taken a bit more seriously. (Unsworth, 2002, Abstract)

Unsworth's idea is to establish an "evaluative scale of relative charlatanism" because "no perfectly exemplary project exists". The scale suggested is an ordered list of features that should be included, such as keyword searching, variable parameters and new calculation algorithms. Today, many of these functions are delivered as standard facilities in a growing range of software. Still, the idea of defining levels of human interactivity in relation to the functions of the computer is interesting. It also reveals that the computer – unlike previous mechanical devices – has a variable functional architecture and, therefore, remains open for the implementation of new epistemologies and ideas in the programmes that define the functional architecture (Finnemann, 1999a).

Any list of criteria for inclusion of new features will be relative to computational practices at a given time. Humanities computing was sensitive to such developments. New features were gradually incorporated and new areas of study were included. Originally centred on the digitisation of texts for linguistic and literary studies, the scope has been extended to include other disciplines and a wider range of semiotic formats and even film and electronic media. There is openness toward including new materials but only within the original constraint of the concept of computation and of finite entities, works, for which there is a non-digital original.

Humanities computing is pioneering the development of software-supported methods for the study of digitised materials but maintains a narrow perspective, rooted within the mainframe-based interpretation of computers as the rule-governed, automatic and deterministic machines of the 1940s, 1950s and 1960s. This perspective fits well with the classical concept of the work as a closed – if complex – fixed and finite unit of analysis – if not as a literary work, an oeuvre, then as a linguistic corpus. In retrospect, it is noteworthy that that the bridge between the formalist epistemology and the finite work was built in the 1960s, when the idea of both computation and closed works was questioned.

The development of new concepts of computers can be traced back to the LISP programme from the 1950s in which instructions were sequenced as lists rather than logical chains (Berkeley and Bodrow, 1964). More steps followed in the 1960s with Ted Nelson's concept of hypertext (1965) and J.C.R. Licklider's and Douglas Englebart's contributions to the development of interactive digital media in the 1960s and 1970s (Barnet, 2013), anticipated only by Turing's marginal remark that the universal computing machine would turn out to be a "choice machine". The potential of this machine was unfolded as the operators (users of all sorts) increasingly used random access to retrieve and combine any set of sequences independently of any overarching operational logic.

On the literary side, the notion of the work was questioned in the 1960s by Eco's *Opera aperta* (1962), Derrida's *De la grammatologie* (1967), and Genette's works on intertextuality, transtextuality and hypotextuality (1967ff). None of these influenced the core concepts of humanities computing. There are two main reasons for this: first, that the reinterpretation of digital media centred on the concept of hypertext, interactivity and link–node relations, which later developed into the human-computer interaction paradigm, while the digital humanities were rooted in the classical idea of the computer as a rule-governed, deterministic machine; second, that the core concepts in the emerging poststructuralist interpretation of texts centred on the ideas of the open work, intertextuality, hypotextuality and paratexts; contrary to this humanities computing was rooted in classical notions of fixed texts and closed works, which ought to be digitised if they should be studied by computational methods.

The reinterpretations of the computer in the 1960s did not attract much attention outside a small proportion of professional IT communities at universities. The 1960s were the heyday of administrative computing centred on mainframes. However, during the 1980s, new applications of hypertext, as well as new bottom-up concepts of artificial intelligence, human-computer interaction (HCI), object-oriented programming (OOP), computer-supported cooperative work (CSCW) and participatory design (PD) entered the scene. The mainframes gave way to microcomputers, also known as personal computers, modifying the computer into a toolbox in the hands of a growing number of professions and disciplines. The aims were dominated by efforts to bridge the rifts of alienation between man and machine by building user-friendly, graphical interfaces (GUI). The computer was regarded as a designable artefact (Norman & Draper, 1986; Ehn, 1989) or as a malleable machine in the hands of man (Bolter, 1991). The hardware dominance of IBM was followed by the software dominance of Microsoft. These developments allowed humanities computing to expand as the machinery became cheaper, faster, easier, and available on the researcher's desktop. At the same time, emails started to allow a huge increase in international communication and also initiated the use of computers to deal with non-computational textual content, which later developed on bulletin boards, in Unix user groups and, finally, with word processors in the 1980s. Furthermore, hypertextual, interactive and multimodal features were added to the conceptual array of humanities computing. Unsworth

suggested his hierarchy of interactive formats to measure the degree of scholarly perfection, and hypertext was also included in the conceptual repertoire (Landow, 1992). In both cases, though, this was only done as part of the methodological repertoire of humanities computing – not as part of the materials studied, which were almost exclusively digitised versions of non-digital originals.

The HCI paradigm indicated significant innovations. First, the control of the processes was moved from internal operations to the growing array of usages. Thus, a growing range of human needs and desires were added as significant drivers of IT development, also including the influence of a wider range of disciplines in the ongoing development of what was now called IT or ICT rather than computers. Second, the interfaces allowed the operators to change the functional architecture of the machine.

For some scholars – referring to Alan Kay's notion of the computer as a meta-medium (Kay & Goldberg, 1977), this meant that the computer was transformed into a medium (Andersen, 1986). According to Alt (2011, p. 279), this was primarily the result of object-oriented programming (OOP), which affected the fundamental operational logic of the machine: "computation became a medium when the concepts of medium and interface were implicitly embedded in computation at the material level of the programming language itself." One might as well argue that the computer was a medium from the very beginning or, as in mainstream media studies, that it was only turned into a medium when it entered the field of modern mass media and part of the overall global media system (Finnemann, 2011). In any case, the PC and the HCI paradigms helped to break down the idea of the rule-based machine as the overarching interpretation. The formalisms were reduced to syntactical devices and made subject to hypertextual, interactive, and multimodal modifications operated by means of graphical user interfaces, allowing a variety of semantic regimes in developing new programmes and practices.

The public breakthrough of the Internet in the 1990s inherited a more radical transformation of the technology as it paved the way for the transformation of the computer into a growing range of different but networked digital media. The scope and reach of hypertext, interactivity and multimodal communication were widened, and three new scales of seamlessly variable, communicative reach were added: local-national-global, private-public, and who-to-whom (Finnemann, 2005).

Until the 1980s, digital media were in the hands of a narrow range of specialists, primarily serving military, scientific and administrative needs – though the array of professional competences involved in the implementation of new ideas in the functional architecture has been steadily growing. During the 1990s, the development of digital media was put in the hands of all sorts of agencies, including professional IT experts, professionals in other fields, and people who were motivated by all sorts of commercial, institutional, civil or even personal interests. The range of needs, motives and longings contributing to the use and further development of the technology was, thus, dramatically widened. This and the lowering of the access threshold to public communication are major components contribut-

ing to the production of increasingly heterogeneous digital materials. The implications are amplified by numerous new variables ranging from production agencies and authorship formats to purposes, software genres, interface genres, rhetorical genres, communicative styles, areas of usage, usage practices, circulation, and remix practices. "Big data", which was previously encapsulated within corporations (Zuboff, 1989) and public institutions, was now flooding the whole of society. No wonder that there were also a variety of new approaches in the sciences, social sciences and humanities. Humanities computing merged with computer-mediated communication (Journal of CMC, 1995), new media studies (Journal of NMS, 2000), media archaeology (Huhtamo, 2011), cultural studies and media studies (Liu, 1994), just to mention a few. To adapt to these transformations, humanities computing became part of a less consistent but more inclusive digital humanities.

## Digital humanities – stretching the tent

"Digital humanities" is today a widely-recognised term signalling a possibly ground-breaking paradigm within and around the humanities. It has also become an umbrella term for a variety of different epistemologies and associated methodologies. As a consequence, there are many different interpretations of how digital humanities might or might not contribute to a "digital turn", a renewal of the humanities, or a new epoch of "e-science" or "i-science" – or whether digital humanities are part of a fundamental breakdown of the humanities as we knew them in the 20th century.

Patrick Svensson (2012) identifies five different conceptualisations. One is articulated by the US National Endowment of the Humanities Office of Digital Humanities, which emerged from the humanities computing tradition with a classical interpretation of digital media as computing machinery. The technology is seen as a deterministic driver, impacting the humanities in many ways, but also as an instrument to ask new questions – for instance, focusing on buzzwords such as "big data" or "big humanities". A second approach refers to the soft-core perspective (Presner and Johanson, 2009), with a focus on creative professional and civic usages. "Digital humanities" is conceived as an umbrella term, and the computer is referred to in the plural as a growing array of digital media.

Svensson identifies a third position represented by a junior scholar, Whitney Trettien, who did not find a place within digital humanities for her PhD thesis pursuing "an interest in the digital as an object of inquiry". We might call this 'a surfer perspective', which she pursued (in her own words) "conscious of its methodology", like "a twitter blogger who follows her passions across interdisciplinary boundaries, the facebooker who makes the personal political and doing so humanizes the humanities." (Trettien, 2010; Svensson, 2012, p. 46). Even though the methodologies are not yet fully developed, it is clear that we have a new kind of digital process that is different from the ideas of rule-governed computation and the closed work. Hypertext and interactivity are taken to new levels.

A fourth position is taken from a debate at the Modern Language Association convention in 2011 discussing a variety of different epistemologies and practices. The voice chosen by Svensson reflects both the breadth of the landscape and takes a clear stand against that very same breadth. First, it is stated that "digital humanities may mean anything from media studies to electronic art, from data mining to edutech, from scholarly editing to anarchic blogging, while inviting code junkies, standards wonks, transhumanists, game theorists, free culture advocates, archivists, librarians and edupunks under its capacious canvas" (Ramsay, 2011; Svensson, 2012, p. 47). Since it is difficult to identify a shared basis, one might look for a single defining feature, as did Steve Ramsay, claiming you have "to make code" to be a part of digital humanities. For some, this means simply that one should build some sort of application while, for others, it is part of a more far-reaching transition from an interpretative to a productive mode of operation in the humanities. "Digital humanities" is defined here as a "making code" perspective or as a "community of practice".

The "making code" criterion appears to be clear and strong, but the coding of computers today is highly diversified. It is often a complex process involving different types of competence. Those who perform the analysis of relevant needs or desires, those who define the purposes, provide the categories, those who come up with suggestions of possible data structures and flows, and those who actually write line-by-line codes draw on very different competences. Collaboration is needed to such a degree that Hayles (2012) regards collaboration as one of the six characteristics of her digital humanities. Much coding today is also generated by means of graphical user interfaces, by drawing on a screen, and possibly based on visual, literary or auditory aesthetic principles and the modification of previous coding. Coding can be based on a variety of semiotic regimes. Nobody covers anything like the full range of coding traditions. On the same MLA panel, Ramsey's "individual essentialism" was also refuted by Alan Liu (2011) as it undermined the idea of collaboration and networked knowledge production. Liu represents a fifth perspective, claiming that the digital humanities should be able to "move seamlessly between text analysis and cultural analysis", not only to join "the mainstream humanities, […] but to take a leadership role". (Svensson, 2012, p. 48; Liu, 2011). In order to claim leadership, digital humanities need to be relevant to the humanities at large by connecting a narrow, somewhat introverted concern with the technicalities of corpus production with wider analytical and critical perspectives.

If digital media studies call for interdisciplinary collaboration, it is difficult to come up with clear criteria for digital computation or coding. So, it seems that there is a need for the "big-tent" metaphor to include different as well as new approaches. This metaphor has been supplemented or supplanted by the idea that the various approaches constitute a privileged "trading zone" for interchanges between the various approaches. (McCarty, 2002; Svensson, 2012). The two metaphors refer to two different types of relationship, with the latter referring to relationships that can be verified. However, it might be as relevant to trade with partners outside the tent – for instance, trading with the wider community of scholars studying the same phenomena with other methods. Similarly, media studies,

including media archaeology, would have to bring digital media into the general history of media and, possibly, elaborate a theory of media capable of including old as well as new media.

Still, there might be some shared criteria for approaches dealing with software-supported methods in the study of digital materials. Hayles (2012) suggests that digital humanities are defined by certain characteristics collected by interviewing outstanding exponents of the main US-UK traditions. The candidates proposed are: collaboration, scale, productive/critical theory, databases, multimodal scholarship, and code. Each of these themes is treated in accordance with the ambition of the book "as it attempts to intervene in locally specific ways in the media upheavals currently in progress" and building on the claim that "people – not the technologies in themselves – will decide" how we think (Hayles, 2012, p. 18). Still, the book provides a very broad and inclusive concept of digital humanities by embracing two major tendencies: on one hand, an "assimilation strategy", which extends existing scholarship into the digital realm, and, on the other hand, a "distinction strategy" emphasising new methodologies, new kinds of research questions, and the emergence of entirely new fields (Hayles, 2012, p. 46). The two strategies are exemplified in several ways, often based on rival or non-related conceptualisations. For instance, Hayles (2012, p. 33) juxtaposes Timothy Lenoir's big-data radicalism ("follow the data streams") with Ramsey's "data must lead to meaning", a controversy that illustrates a major epistemological difference. It is more difficult to see the trading zone in between. At the very least, it is doubtful whether it would make sense to include both in the same community of practice.

"Big data" is also subject to different interpretations. Some argue, like Lenoir, that such studies are data-driven without hypotheses and uninformed by theoretical assumptions because they look for patterns in huge amounts of data, which we cannot look into in other ways. For Victor Mayer-Schönberger and Kenneth Cukier (2013), it is not a matter of amounts of data but a matter of epistemology since, in their view, "big data" is about replacement in research of causality with statistical correlations. This again relates to data that cannot be analysed in traditional causal ways or interpreted as representative samples but only as messy datasets, which might allow valid interpretations of noisy but recurrent patterns. It sounds similar to issues originally addressed by Claude Shannon (1949) in his concerns with different forms of redundancy: as a function of the message, of the physical medium and of the coding (Finnemann, 1998). At least, we are in some sort of noisy statistics. Others argue that such studies depend on axioms and assumptions implemented in the definition and construction of the corpus, in the algorithms of the search routines for analysing the data, and that the findings can be no more than statistical – and the significance of these statistics is more often than not rather insecure (Snijders, Matzat & Reips, 2012; Marth and Scharkow, 2013). These criticisms are in accordance with the literature on "data doubles" (Lyon, 2007).

Other dichotomies are constructed differently, such as the dichotomy between the idea that history is a record of what happened in time versus, say, Ethington (2007), who

argues that history is a record of what happened in places and spaces. Whatever position one might take on this dispute, it should have no particular relation to the concept of digital humanities, unless it could be argued that networked digital media are used to change the relationship between time, space and place. This might well be the case; and, indeed, it is the case. But it is a matter of fact, an empirical issue, and not a matter of epistemology. The issue can be explored from the perspectives of both Ethington and his opponents. It might also turn into a methodological issue, as digital media allow us to develop new ways of analysis, representation and storytelling. For instance, historians now have the tools to present space and place relations visually (still or live) in ways that were not previously possible. They may want to privilege place and space to time in telling their story; this would be an issue of positioning the narrator of the story. The narrator is a construction of the author's perspective and, possibly, his representation in the story – his avatar, one might say – while the author who constructs the narrator's perspective is still embodied in a world of inseparable time, space and place, as is any reader. Digital media should not be regarded as being in support of any particular epistemology. They are neither modern nor postmodern. Instead, they permit new configurations of the relationship between researcher and presentation and, in a wider perspective, between author and text (discussed further below).

Diversity is king. The range of positions includes different ideas of the computer, different ideas of the humanities, different epistemologies, different ideas of the role of technology in culture, different goals for activities, different disciplinary backgrounds, different practices in the employment of traditions inherited from the humanities as it were, and different ideas of the defining characteristics of the digital humanities. But what about digital materials? Do they have any shared characteristics?

## Digital materials – outlining a new field

Current debates within digital humanities focus primarily on methodologies, whether quantitative or qualitative. Both issues need to be further informed by a more elaborate conceptualisation of digital materials. There are two main reasons for this. First, digital materials are today increasingly important and often constitute unique source material. Second, they are also increasingly heterogeneous. This is the case in society at large and in the sciences, social sciences and humanities. Both aspects are important for historical documentation and methodological potential with intricate implications for the notions of texts and documents.

Today, it is safe to predict that, in the 21st century, an increasingly significant part of political, cultural and social life will be articulated in digital genres performed on networked digital media platforms. It is a hypothesis, but it is not unlikely that it will turn out to be one of the grand narratives of the 21st century; and it may well serve as a valuable guide for the humanities and social sciences for many years to come. The production of digital materials takes place today in virtually all spheres of society, ranging from data collection from outer

space to scanning the interior parts of our bodies, including the brain, and everything in-between. We will still have other sources of documentation. Old media are seldom completely replaced by new media. They may still be useful for documenting important parts of what we want to document and to preserve. However, the adoption of digitisation in an increasing number of areas will make digital materials a unique source for anyone wishing to study public, political, cultural and social affairs in the 21st century.

Software-supported methods were an option in the age of humanities computing, but they are today a necessary precondition for a growing range of old and new disciplines because the source materials of contemporary society are increasingly born-digital-materials – to be analysed and presented using software-supported methods. A digital turn is inevitable due to the increasing importance of born-digital materials in all spheres of society.

The amounts of data are growing, but so is their heterogeneous character. This is partly a result of the complexity of the world, which is reproduced in the processes of digitisation to include a growing range of different needs and desires articulated by all sorts of commercial, civic and institutional agencies, including the development of more complex methodologies employed in political, cultural and social communication, including research and development practices.  The complexities of modern society alone would explain the heterogeneous character of digital materials, but there is also a second dimension that stems from the variable functional architecture, which can be modified by means of messages sent in the same medium.

The heterogeneity of digital materials is intrinsic to digital media because there are no invariant distinctions between programmes and data. The programmes of today are also data, and the data of yesterday may be turned into a programming feature altering the functional architecture of the machine. As a consequence, new ideas can always be projected into the functional architecture, which remains open for the implementation of ideas from an increasing range of human concerns, needs, interests, desires and longings. This is the fundamental property behind the development from mainframe machines via PCs to networked digital media.

In humanities computing, hypertext and the web were acknowledged as new tools but regarded as being external to the works concerned. For born digital materials, hypertext and interactivity are inherent parts of the materials. They are not simply objects but are part of the grammatical repertoire of digital media. So, in terms of their grammatical nature, born-digital materials differ from digitised materials, the latter being defined by their non-digital originals (Brügger and Finnemann, 2013). Conceptually, the most intriguing aspect is that web materials may include not only hypertextual, interactive and multimodal and multi-semiotic features but also the ever-growing array of scripts implemented in, say, Facebook and elsewhere.

The heterogeneity of digital materials transcends the notion of databases and finite works. The notion of the database refers to homogeneous datasets. Databases are

extremely important, as they largely serve to reduce the heterogeneity of data materials by ordering them in terms of a limited set of parameters. This also explains why 'the database' is not suitable as a concept of data materials in general. The heterogeneity also goes beyond the notion of the work as an overarching concept for organised texts. Digital materials and software-supported methods will acquire different forms due to their particular production histories, the provenance of the data in question, and their history of circulation and preservation. This heterogeneous character needs to be further analysed, and the outline provided below is only preliminary with a view to identifying how four significant constituents of digital media materials are manifested: the notions of authorship, text, hypertext and, fourth, the way different relations to data materials are expressed in different research traditions. Even these dimensions are too complex to be analysed in detail. So, the conclusion will be that the growing heterogeneity of digital materials requires their study to become a field in its own right.

*Authorship forms.* Authorship forms have always included a heterogeneous set of modes. The types of relation between authors and texts – be they religious, legal, literary, public or private texts – have changed throughout history. The modern paradigm was formed around the nexus of the individual author and a finite text or as a series of individual works collected into an oeuvre. In 20th-century theory, a narrator was inserted in the story between the real-life author and the story. The narrator represents the author's perspective, the point of view to be found in the text. The author might also be hiding in a pseudonym outside the text.

Even though the fixed relation was widespread and dominant, it was never the only paradigm. The individual author instantiates some sort of agency in the world, writing on behalf of a person, institution or organisation aiming at certain goals. The complexity of the world, thus, articulates itself in digital media as a complex set of authoring agencies. Digital media allow a wider array of authoring agencies, on one hand, and a wider array of relations to the texts produced on the other. Thus, there is an array of new types of authorship primarily made possible by random access to stored digital materials. The array of such authorship forms or avatars is open and can only be presented as a list of manifest occurrences. It has to be a historical and not systematically-defined list.

To mention a few examples:

First, any user of digital media generates a series of user profiles ranging from simple access data to a variety of anonymous, semi-anonymous and non-anonymous profiles with more or less elaborated self-presentations that develop over the years into a sort of non-intended 'life logging'. To this, one might add fictional profiles, an avatar in a virtual (game) world, and the data doubles based on more or less noisy traces, which are left when using digital media. Thus, over the years, changes in age, personal preferences, tastes and interests, identities and social affiliations are reflected in an ever-growing chain of self-profilings.

In this way, people today are more or less spontaneously involved in documenting or even narrating their biographies as their lives unfold.

A second kind of authorship is created as "computer-generated" texts, which are generated by individual requests to a service provider (for instance, a Google search or searches in other database types and constellations). Such texts are created as products of individual authors, who define the search criteria and select the possible set of sources without knowing the content. These authors may be the only ones who read the text, which may exist for only a very short time span. Each of these texts may be important; and, in terms of the numbers of daily searches and the general role of searching that is integral in digital media, they constitute a significant new type of authorship. There are a variety of related forms of coproduction and crowd sourcing involving collaboration between authors who may have a shared goal and other forms such as remix and viral communication in which a text may be transmitted and transformed, independent of the original author. In some cases, authorship evolves over many years in the form of chains of authors who may succeed each other. It is often not possible to reconstruct the authorship of such texts, but it may be relevant to keep some sort of trace; since digital texts are dynamic and historical in their inner nature, their provenance can only be established in the form of dynamically aggregated histories. As the notion of authorship has to adapt to the new hypertextual and interactive potential of digital media, so does the notion of text.

*Text.* The heterogeneity of digital materials does not fit well with the notions of text and document if they refer to a set of generally-shared characteristics for all texts or documents (or whatever term is used for finite chunks of content).

Hypertextual linking, remixing, ongoing and multiple authorship forms, build-in scripts and editing practices are among the features that undermine our previous conceptualisations of texts, works and documents, which are normally conceived as delimited either in time or in fixed graphical space. As argued by Aarseth (1997), the use of (some) hypertextual and interactive features leads far beyond Eco's notion of an open work (Eco, 1962) and also far beyond Iser's notion of empty spaces in the text, which remains open for the reader's interpretations (Iser, 1980). However, they also go beyond Aarseth's notion of 'ergodic cybertexts', which still define the text as a finite work and consider the computer as a calculational, rule-based device, which leaves to the reader only a risky choice of probability of probability.

On one hand, the notion of cybertext is described as more limited than hypertext, since it only includes "non-trivial" hypertexts while, on the other, it is presented as a more generalised concept including both texts and (some) hypertexts. Thus, the concept fails to include all possible kinds of digital hypertexts, including those mentioned above. The criteria for being 'non-trivial' also remain obscure.

Still, there is room for the notion of text-as-work because it is still possible to declare any constellation of sequences to be a finalised work. It is a format that can be chosen from

among other formats. For instance, this is the case when one studies digitised versions of non-digital originals. These types of digitised material exist in a digital format, which is defined *a posteriori* in relation to the original format, such as cultural heritage materials from previous epochs. For born digital materials, a work can be defined by declaring a deadline for the final edition even if they include hypertextual and interactive sequences.

Digitised and born digital materials differ. Digitised versions of non-digital originals will always be distorted in some respects, and born digital materials may also include programme scripts, hypertextual features, and interactive features. Such features can only be non-original additions to digitised materials. They belong to the grammar of digital media, and they can be exploited in many different ways. Their use may also be very limited if so desired. This is often the case with research-defined materials (Pohorec et al., 2013). Born digital materials can be created in their own digital format and recreated/converted into other formats; the latter is the case, for instance, with archived web materials, which constitute one of the most complex sets of data material. Such recreations and modifications may also take place in very simple situations, such as when you convert a word document file into a PDF file. Digitised materials still allow us to refer to previous delimitations and categories because they are replicas of them, adding only digital facilities around the source. This is not always the case for digitally born materials.

The notion of document or text becomes even more complicated because a text may be incorporated on a website, which is both a text itself and a chunk of texts and links used most often as one of a number of related units. Units of expression of any size and level may be blended, separated, recombined and mixed deliberately, even at the level of bits, which – similar to the letters in written language – are units defined independently of semantic content. In some cases, limitation issues can be solved by specifying a time span that has to follow a given sequence on its further course and history. We may freeze such time spans, turn them into chunks or textual units in storage, preserved from the ongoing flow of bits on the Internet. But the flow itself, in its many different forms, is a significant part of digital media.

Perhaps, we should abandon the idea of correspondence between the document that is produced and the documents that are read, scanned or browsed. The sequences defined by writers may never be read in full but are broken up and mixed with other sequences, implying that the reader needs to compose her own sequences retrieved from a variety of different works. Still, at the end of the day, the reader has traversed or even composed a sequence of materials.

Digital media bring the writer and reader to positions very close to each other, but they also facilitate a greater distance as the reader composes her linear sequences from an open-ended array of possible sources. Sometimes, she alternates constantly between the reader and writer positions. As a consequence, the notion of a text or document has to be modified. Hypertext and interactivity has to be inscribed into a growing array of semiotic functions and semantic definitions within and in between delimited texts. The question

is how to decide whether a sequence constitutes a text. And if we introduce a time span as part of the delimitation, how do we decide about the rights to this sort of information?

*Hypertext and the modalities of link-node relations.* The notion of hypertext has a strange history, coined in the 1960s, gaining widespread use only in the 1980s, and almost disappearing soon after the breakthrough of www protocols including the hypertext transmission protocol http (Barnet, 2013). However, the history of hypertext is a much more fundamental history of the functioning of computers, because the link-node relation is the basic mode of digital processes, more basic than the binding of a series of steps together with the help of algorithms. Any computational process will take off and proceed by ongoing requests (links) for chunks of bits (nodes) in storage to be processed. The point of departure is always the actual node, the anchor point, from which the operator may choose the destination of the next step. Insofar as these processes are combined into repeatable sequences, they can be used as a programme or chunk of content of some sort, or as blends of programme and content, by inscribing dynamic features in the material. The link-node relations are, in principle, arbitrary relations with the link being prompted by a single click whether it is retrieving a letter in the alphabet, a high-order logical sequence as a programme to be executed according to a specific set of criteria, or a huge array of data that may be composed based on mathematical, linguistic, auditory or visual criteria or which may be the scanned results of physical processes and phenomena.

So, hypertext may be part of a text, a feature built into a work, or part of a relation between texts linking them together. It can also be utilised in semantically very different forms. The hypertextual link-node constellations are syntactical in nature, and the array of hypertextual structures has grown into a heterogeneous set. Take, for instance, a webpage. Links may appear *as menus for the navigation* of a website, as a meta-structural device, *but links may also connect any single unit on a webpage* and any other single unit on the same page *or on other pages on the same site* or on *other sites*. Links can be applied *on all levels* and *between different levels*. They may also *vary in respect to the motive*, the idea of the link. Some are purely navigational; some are part of the semantics of the message. *Some are created as part of the* message; *some are added later*, years later, perhaps. And to make things even more complex: you cannot be sure that the anchor leads to the same content at the destination that was present when the link was established. The basic feature of hypertext is that you are forced into a modal shift between ordinary reading modes and link modes (Finnemann, 1999b).

On any given computer screen or webpage, you will always have to follow a link to continue. This can be done by choosing a predefined destination or by specifying a destination using a search function. Search is yet another word for exploiting the fundamental anchor-link destination that characterises digital media. It is central and constitutive because it is the basis of the fundamental – and closely interconnected – hypertextual and interactive properties of digital media.

*Data materials defined and redefined by research.* Even though digital materials always come in the form of binary sequences, they are also marked by their source, what they are about, in what format they were originally stored, and how they were treated. In short, the provenance of data is of relevance for its interpretation, whether it is treated mechanically by software tools or hermeneutically by human interpreters. To establish a preliminary typology, four different types of data material are proposed:

First, cultural heritage materials, which are digitised in formats that are defined by scholars or archivists *a posteriori* to the original non-digital source. Depending on the original source, such materials are distorted in one way or another, as well as being converted into a homogenising format of some kind. Scripts, hypertext and interactivity are used to deal with the source but are not part of the materials.

A second type of digitised material comprises research data that are produced by (and often only obtainable by means of) digital media – whether the data source is outer space, social space or inner bodily states (scan) or produced by questioning a selected group of respondents such as research-defined databases organising a selected dataset that are digitised in a format specified by researchers or archivists (e.g., evidence, survey data, visualisations, etc.). In these cases, formats are most often defined *a priori* to the collection of data – even if the formats can be made sensitive to the responses. Data may be both nonresponsive and responsive; but, in most cases, it is defined as part of a particular type of research questions. Both these types are derived from non-digital processes and phenomena. Research-defined formats might also be defined *a posteriori* to the production of data, when defined, for instance, by a selection of a corpus due to criteria that differ from the intentions of those who created the data. This is true of big data studies.

A third type can be characterised as digitally-born materials, which are created by using a variety of digital media for some sort of political, cultural or social purpose and for more or less private and more or less public purposes. This data is also marked by its provenance but differs from the other types because it may include scripts, hypertextual, interactive and multimodal features in its grammatical repertoire. In the 21st century, this kind of material seems to become the most important source material by far within the humanities insofar as they are concerned with contemporary phenomena. Among born digital materials, web materials constitute one of the most heterogeneous sets of data material ever created. The non-proprietary network structure of the Web allows everybody to publish software applications as well as any kind of content – possibly, personalised by the user or the content provider, which means that web materials over the years become increasingly heterogeneous in character. The difference between digitised copies and born-digital materials will impact all areas within the humanities as they have different characteristics and have to be treated and analysed with the help of different digital tools even within the same disciplines, whether within historical studies, ethnographic studies, literature studies or media studies.

Archived web materials, finally, constitute a fourth type, as they will always be distorted and incomplete. Some of the formats and features that make the Web superior to other media cannot be archived (Brügger and Finnemann, 2013).

These four main types do not exhaust the list of possible forms of digitised and digital material. In other words, the typology suggested here is preliminary, but it is of relevance for the methods possible and for the building of research infrastructures for the studying of digital materials. Further elaboration will be a precondition for study within and of 21st-century history.

## Conclusions

The history of the computer cannot be written independently of the history of the ever-growing number of different conceptualisations, purposes, needs and longings that have been projected onto the device, ranging from the centralised mainframes of the 1950s and 1960s to the spread of personal computers in the 1980s to the contemporary array of networked digital media. The development of humanities computing into the digital humanities is part of the same story.

This article has focused on the history of digital materials. It has been argued that this is primarily a history of increasing importance and heterogeneity. Both aspects are seen as the result of two main drivers: first, the adoption of digitisation processes in a growing range of areas by a growing range of different expertise and by commercial, institutional, civic and personal agencies in society; second, the properties of the 'choice machine', random access, allow human operators to change the functional architecture and to widen the semiotic regimes used in operating digital media.

It has been shown that digital materials cannot always be adequately conceptualised and captured within the notions of text, document and documentation, notions derived from dealing with messages conveyed in previous types of media, each with its own set of characteristics. Digital materials have a more complex history than media materials in other media because the ideas expressed may be combined with changing digital formats and remixed, modified or even distorted if they are not removed from continuous recirculation. The provenance of media materials mattered for old media materials; it matters even more fundamentally for digital materials.

As the processes of digitisation penetrate an increasingly broad array of societal affairs, both the digital materials and the software-supported methods used will acquire different forms depending on their production histories, the provenance of their data and their history of circulation and preservation. Due to the heterogeneous character of digital materials, it is safe to conclude that the provenance of digital materials is highly relevant. History counts.

Due to their characteristics, a systematic account of digital materials is not feasible. This is where history merges into information theory. History is back on the micro level of data

because the provenance of data is an intrinsic part of its meaning, although it can always be processed without taking notice of those meanings. On the macro level, history returns as a history of changing conceptualisations that are projected onto the machine, thus transforming the history of ideas of the computer into a history of the functional architectures of digital media.

The mapping of digital materials based on authorship, textual and hypertextual patterns and various forms of research-defined data is preliminary both with respect to these parameters and even more with respect to the internet parameters of local-global, public-private and who-to-whom, all of which provide options on a scale of seamless variation. However, this mapping should provide sufficient documentation to prove that the study of digital materials deserves to be established as a field in its own right. If, as argued above, the digital humanities cannot be conceptually unified above the level of studying digital materials with the help of software-supported methods, this might provide a platform for shared reference. Different approaches may attach different values to this, but they might also serve as an important contribution to the future organisation of research infrastructures required by the heterogeneous character of digital materials and methods.

The notion of digital media needs to be elaborated by specifying further the intrinsic relations between materials, search methods and communicative reach. Due to the role of digital media in society at large, this will also include the incorporation of digital media into the wider history of media and the very of notion of media.

## References

Aarseth, Espen. 1997. *Cybertext. Perspectives on Ergodic Literature*. Baltimore. John Hopkins University Press.

Alt, C. (2011). Objects of Our Affection. How Object Orientation Made Computers a Medium. In E. Huhtamo, & J. Parikka (Eds.), *Media Archeology. Approaches, Applications, and Implications* (pp. 278-301). Berkeley: University of California Press.

Andersen, P.B. (1986). Semiotics and Informatics. In P. Ingwersen. *Information Technology and Information Use*. London: Taylor Graham.

Barnet, B. (2013). *Memory Machines. The Evolution of Hypertext*. London: Anthem Press.

Berkeley, E.B. & Bobrow, D.G. (Eds.), 1964. *The Programming Language LISP: Its Operation and Applications*. Cambridge, Massachusetts, and London, England: Information International, Inc. The M.I.T. Press, Massachusetts Institute of Technology.

Bolter, Jay D. (1991). *Writing Space. The Computer, Hypertext, and the History of Writing*. Mahwah, New Jersey: Lawrence Earlbaum.

Brügger, N. & Finnemann, N.O. (2013). The Web and Digital Humanities: Theoretical and Methodological Concerns. *Journal of Broadcasting and Electronic Media. Vol. 57, (1)*, 66-80.

Busa, R.S.J. et al. (2006). *Corpus Thomisticum – Index Thomisticus*. Web edition by Eduardo Bernot and Enrique Alarcón. English version. Consulted on: 13 December 2013, http://www.corpusthomisticum.org/it/index.age.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Derrida, J. (1967). *De la grammatologie*. Paris. Éditions de Minuit. (Eng.: *On Grammatology*).

Eco, U. (1962) *Opera aperta*. Milano: Bompiani. (Eng.: *The Open Work*)

Ehn, P. (1988). *Work-Oriented Design of Computer Artifacts*. Stockholm: Arbetslivscentrum.

Ethington, P.J. (2007). Placing the Past: 'Groundwork' for a Spatial Theory of History. *Rethinking History. Vol. 11, No. 4*, 465-493. Including comments. Consulted on: 31 October 2013, http://www-bcf.usc.edu/~philipje/PDFs/Ethington/Ethington_et%20al_Placing_the_Past.pdf.

Finnemann, N.O. (1998). On the Notions of Rule-Generating & Anticipatory Systems. In Proceedings from the 1st International Conference on Computing Anticipatory Systems (CASY'97). August 11-15 Liege, Belgium. (http://www2.ulg.ac.be/mathgen/CHAOS/IJCAS/IJCAS_CONTENT.htm#IJCAS1).

Finnemann, N.O. (1999a) Modernity modernised. In Paul A. Mayer (ed.) *Computer Media and Communication – A Reader* (pp. 141-160). Oxford: Oxford University Press.

Finnemann, N.O. (1999b). *Hypertext and the Representational Capacities of the Binary Alphabet*. Working Paper 77. Center for Cultural Research, Aarhus University. Consulted on: 9 December 2013, http://www.hum.au.dk/ckulturf/pages/publications/nof/hypertext.htm.

Finnemann, N.O. (2005). The Cultural Grammar of the Internet. In Jensen, K.B (Ed.) *Interface://Culture – The World Wide Web as Political Resource and Aesthetic Form* (pp. 52-71) København: Samfundslitteratur/Nordicom

Finnemann, N.O. (2011). Mediatization Theory and Digital Media. *Communications 36*, 67-89.

Genette, G. (1967-1970). *Figures I-III*. Paris: Editions du Seuil.

*Guidelines for Electronic Text Encoding and Interchange (TEI)*.  Consulted on: 9 December 2013, http://www.tei-c.org/Guidelines/ and http://www.tei-c.org/About/history.xml.

Hayles, N.K. (2012). *How We Think. Digital Media and Contemporary Technogenesis*. Chicago: The University of Chicago Press.

Hockey, Susan (2004). The History of Humanities Computing. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *Companion to Digital Humanities. Blackwell Companions to Literature and Culture*. Oxford: Blackwell. Consulted on: 15 September 2013, http://www.digitalhumanities.org/companion/.

Huhtamo, E. & Parikka, J. (Eds.), (2011). *Media Archeology. Approaches, Applications, and Implications*. Berkeley: University of California Press.

Iser, W. 1980 (1974). The Reading Process: A Phenomenological Approach. In Jane P. Tompkins (ed.), *Reader-Response Criticism: From Formalism to Post-Structuralism*, 50-69. Baltimore: Johns Hopkins University Press.

Jakobson R. (1960). Closing Statement: Linguistics and Poetics. In T. Seboek (Ed.), *Style in Language*. Cambridge, MA: The MIT Press.

Kay, A. & Goldberg, A. (1977). Personal Dynamic Media. *Computer 10(3)*, 31-41. Reprint consulted on: 31 October 2013, http://www.newmediareader.com/book_samples/nmr-26-kay.pdf.

Landow, G.P. (1992). *Hypertext : The Convergence of Contemporary Critical Theory and Technology*. Baltimore: Johns Hopkins University Press.

Liu A. (1994). *The Voice of the Shuttle*. The remains Consulted on: 24 October 2013, http://vos.ucsb.edu/.

Liu, A. (2011). *Where is Cultural Criticism in the Digital Humanities?* Blog entry, paper presented at The History and Future of the Digital Humanities panel at the MLA Convention. Consulted on: 24 October 2013, http://liu.english.ucsb.edu/where-is-cultural-criticism-in-the-digital-humanities/.

Liu, A. (2012). The State of the Digital Humanities: A Report and a Critique. *Arts and Humanities in Higher Education 11(1-2)*, 8-41. Consulted on: 23 September 2013, http://ahh.sagepub.com/content/11/1-2/8.

Lyon, D. (2007). *Surveillance Studies: An Overview*. London: Polity.

Mahrt, M. & Scharkow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media 57 (1)*, 20-33.

Mayer Schönberg, V. & Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

McCarty, W. (2002). Humanities Computing: Essential Problems, Experimental Practice. *Literary Linguistic Computing, 17(1)*, 103-125.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the Structure of Behavior*. New York: Henry Holt, 1960.

Nelson, T. (1965). The Hypertext. in: *Proc. World Documentation Federation Conf. 1965*.

Nelson, T. (1965). A File Structure for the Complex, the Changing and the Indeterminate, *ACM 20th National Conference*. Pp. 84-100, 1965.

Norman, D. & Draper, S. (Eds.) (1986). *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pohorec, S.,  Zorman M., & Kokol, P. (2013). Analysis of Approaches to Structured Data on the Web. *Computer Standards & Interfaces 36 (2013)*, 256-262. Consulted on: September 23, 2013, http://dx.doi.org/10.1016/j.csi.2013.06.003.

Presner, T. and Johanson, C. et al. (2009). *The Promise of Digital Humanities – A White Paper*. White Paper from the UCLA Center for Digital Humanities.

Ramsay, S. (2011). *Who's In and Who's Out*. Blog entry, paper presented at The History and Future of the Digital Humanities panel at the MLA Convention. Consulted on: 31 October 2011, http://lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html.

*Rethinking History. Vol. 11, No. 4*, 465-493. Consulted on: 31.10 2013, http://www-bcf.usc.edu/~philipje/PDFs/Ethington/Ethington_et%20al_Placing_the_Past.pdf.

Schreibman, S.; Siemens, Ray; & Unsworth, J. (Eds.). (2004). *Companion to Digital Humanities*. *Blackwell Companions to Literature and Culture*. Oxford: Blackwell.  Consulted on: 15 September 2013, http://www.digitalhumanities.org/companion/.

Shannon, Claude, (1949) 1969. *The Mathematical Theory of Communication*. Univ. of Illinois Press, Urbana.

Snijders, C., Matzat, U., Reips, U.D. (2012). 'Big Data': Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science 2012 7 (1)*, 1-5.

Svensson, P. (2012). The Digital Humanities as a Humanities Project. *Arts and Humanities in Higher Education 11(1-2)*, 42-60. Consulted on: 23 September 2013, http://ahh.sagepub.com/content/11/1-2/42.

Trettien, W. (2010). *Digital Humanities vs. The Digital Humanist*. Blog entry. Consulted on: 24 October 2013, http://blog.whitneyannetrettien.com/2010/04/digital-humanities-vs-digital-humanist.html.

Turing, Alan, M., (1936) 1965. On Computable Numbers – with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, ser. 2, vol. 42*, 230-265 with corrections in vol. 43, 1937, 544-546. Reprint in Davis, (ed.) 1965, pp. 115-154.

Unsworth, John (2002). What is Humanities Computing and What is Not? *Jahrbuch für Computerphilologie 4*. Consulted on: 16 October 2013, http://computerphilologie.uni-muenchen.de/jg02/unsworth.html.

Wardrip-Fruin, N. (2011). Digital Media Archaeology – Interpreting Computational Processes. In E. Huhtamo & J. Parikka (Eds.) *Media Archaeology: Approaches, Applications, and Implications*. (pp. 302-322). California: University of California Press.

Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*, New York: Basic Books.

## Note

1    The project was sponsored by IBM from 1949 until the 1970s. A web version (Busa et al. 2006) was published in 2005.

*Niels Ole Finnemman, DPhil*
*Professor*
*Royal School of Library and Information Science*
*University of Copenhagen, Denmark*
*sxj957@iva.ku.dk*