



Born-digital archives

Thorsten Ries^{1,2} · Gábor Palkó³

Published online: 25 March 2019
© Springer Nature Switzerland AG 2019

The first special issue of *International Journal of Digital Humanities (IJDH)* is about born-digital archives, their preservation and research perspectives involving born-digital primary records in the humanities. This is not only a result of the collaboration between the journal's editor-in-chief, Gábor Palkó, Co-Director of the Centre for Digital Humanities at the Eötvös University, who is interested in the practice and theory of digital archives, and the editor of this volume, Thorsten Ries, who conducts research on born-digital *dossiers génétiques* with digital forensic methods at Ghent University. It is also meant to be a programmatic call to intensify cross-sectoral collaboration between galleries, libraries, archives, and museums (GLAM institutions), digital preservation projects, and humanities research working with digital primary sources.

The born-digital historical record of the present age poses great challenges for archival science, librarianship, museology, and information science on the one hand, and to humanities research on the other, next to offering exciting opportunities. Personal digital archives, legal, governmental, institutional, scientific, public, and non-governmental organisations' documentation records or datasets, public repositories of digital publications, web archives, and social media archives are incredibly rich, diverse and multi-faceted treasure troves for historians, political scientists, sociologists, philologists, literary scholars, art historians, digital humanists, and researchers from other humanities disciplines. The effort of long-term preservation, curator- and custodianship for these records and the development of setups, applications and application programming interfaces (API) to make them available for research has been subject of multiple large, successful international projects in archival science, librarianship, and information science. Landmark projects such as the archiving of the digital collections of Salman Rushdie at Emory University Library (Rockmore 2014; Waugh and Russey

-
- ✉ Thorsten Ries
Thorsten.Ries@ugent.be; T.Ries@sussex.ac.uk
- ✉ Gábor Palkó
palko.gabor@btk.elte.hu

¹ Ghent University, Ghent, Belgium

² University of Sussex, Brighton, UK

³ Eötvös Loránd University, Budapest, Hungary

Roke 2017); Hanif Kureishi at The British Library (Foss 2017), Friedrich Kittler at the German Literature Archive Marbach / Neckar (Enge and Kramski 2014), Franz Josef Czernin at the Austrian National Library (Catalogue ÖNB, accessed 2018) and the Thomas Kling Archive at Stiftung Insel Hombroich (Ries 2017, 2018), to name but a few, as well as national (e.g. UK, Germany, the Netherlands, Belgium, etc.) and international repositories and web archives (Internet Archive, etc.) with sophisticated frontends such as RESAW (REsearch Infrastructure for the Study of Archived Web materials), SHINE UK Web Archive and Wayback Machine, are just some of the most visible results of this broad development of born-digital archiving. Memory institutions, international archival, and information science projects are very active on addressing fundamental issues of born-digital archiving such as developing workflows for identification, selection, triage and bibliographic documentation. Management of the sheer data volumes and curatorship that caters for the fragility and obsolescence of legacy hardware, software and formats of complex, context-dependent digital records are ongoing challenges. Key research and development areas in this interdisciplinary sector are the development of preservation formats and workflows that ensure authenticity, fixity, physical as well as logical stability and accessibility by forensic imaging, virtualisation, emulation, migration and the development of environments, tools and API's for secure, controlled access to the archive for researchers. Currently, archival and information science, memory institutions, and archiving projects are working towards interoperable standards and making standardised workflows, protocols, expert resources, tools and infrastructure for born-digital curation available to archives, libraries, memory institutions, and projects of all sizes and all levels. The early beginnings of born-digital archiving practice and applications of digital forensic methodology in libraries and archives are mostly associated with the names of individual archivists, librarians, archival and information scientists, and humanists such as Susan Thomas, Kirschenbaum (2008, 2013, 2016a, b); Kirschenbaum et al. (2009, 2010); Jeremy Leighton John (2012); Duranti (2009); Duranti and Endicott-Popovsky (2010) and Doug Reside (2011a, b, 2017). Since then, we have seen an enormous growth of these efforts in archival research, development and professional practice, which today are orchestrated by large, national and international, often high-level projects such as InterPARES and InterPARES Trust (International Research on Permanent Authentic Records in Electronic Systems, Canada, Europe, international, since 1994, 4th phase), Digital Preservation Coalition, DPC (Europe, UK, international, 2002–today), PARADigM (Personal Archives Accessible in Digital Media, Europe, UK, 2005–2007), CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) and Digital Preservation Europe, DPE (Europe, 2006–2009), PLANetS (Preservation and Long-term Access through NETworked Services, Europe, 2006–2010), nestor (Kompetenznetzwerk Langzeitarchivierung, Germany, 2003–2006, 2006–2009, since 2009 self-sustained), PREMIS (Preservation Metadata: Implementation Strategies, USA, since 2003), the CLIR and OCLC research initiatives (Council on Library and Information Resources, Online Computer Library Center, Incorporated, USA, international, 2010: CLIR report, 2013 „Demystifying Born Digital“), VIMM (Virtual Multimodal Museum, Europe, Cyprus, 2016–2019) and the Computational Archival Science working group (international, since 2016). National repositories for born-digital publications, research infrastructures, and web archives are mostly hosted and run by the national library system of individual countries, and complemented by

supranational humanities research infrastructures such as DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure) in the European context. On a meso-level, however, there seems to be less institutions and projects that are enabling born-digital preservation, curation, and research at the level of smaller archives and individual researchers. We would like to highlight PACKED (Centre of Expertise in Digital Heritage, Belgium, since 2003), the DCC (Digital Curation Centre, UK, 2004–today), the BitCurator project (USA, 2011–2014, now BitCurator NLP) and BitCurator Access (USA, 2014–2016). It is encouraging to see that, at least every now and then, memory institutions reach out to humanities research in order to collaboratively identify in which digital formats, with which metadata and by which access tools born-digital records might be most useful for researchers and encourage them to find out about the possibilities. Excellent examples are the hands-on exhibition of Salman Rushdie's emulated computer at Emory Libraries (Rockmore 2014), the pilot of born-digital reading room at the British Library featuring materials from the Hanif Kureishi Archive (Foss 2017), the workshop on born-digital archives access at Wellcome Collection (Sloyan 2018), the inclusion of both humanities researchers and representatives of memory institutions responsible for web-archiving in the RESAW network (Winters 2018a), and the Personal Digital Archiving conference series (e.g. PDA conferences 2017 at Stanford University Libraries, 2018 at Houston, TX). This interdisciplinary and intersectoral collaboration between archival and humanities research, methodological development and practice is of crucial importance and the humanities certainly need to take Matthew Kirschenbaum's imperative to heart:

Digital archivists need digital humanities researchers and subject experts to use born-digital collections. Nothing is more important. If humanities researchers don't demand access to born-digital materials then it will be harder to get those materials processed in a timely fashion, and we know that with the born-digital every day counts. (Kirschenbaum 2013, 38)

Despite the fact that Kirschenbaum rather stated the obvious when he defined that „the concept of a primary record can no longer be assumed to be coterminous with that of a physical object“ and that „electronic texts, files, feeds and transmissions of all sorts are also indisputably primary records“ relevant to historical research (Kirschenbaum 2016b, 25:27), humanities researchers still seem to be rather reluctant when it comes to include born-digital primary sources into their research. There is probably no simple answer to the question why this is the case. If we look at personal digital archives, legal and ethical considerations concerning the protection of privacy and personal rights of the data subjects and of third parties as well as copyright are probably the most important reasons for the hesitation of humanities researchers (Carroll et al. 2011; Baker 2018). Jane Winters argues that “web archives, and other kinds of born-digital data, do bring the possibility of, and perhaps even necessitate, a radical reframing of humanities research – through their scale, their heterogeneity, their complexity, their fragility”, which might not be sufficiently accessible with “the tools and methods available to us at present” (Winters 2018b). Further concerns about born-digital archives, especially web-archives, might have to do with inherent biases and misrepresentations introduced through a focus on “significant and/or traumatic events, [...]

personal interest and enthusiasm or a serendipitous partnership” that comes with individual archiving efforts, triggered by events or specific research interests (Winters 2018b). Born-digital primary sources (and archives), according to Winters, are different from analogue ones in many ways, and she further makes the point that historians still need to embrace the fact “that a digit[al] manuscript is an object in its own right, with its own context of production” (Winters 2018a). For this delayed development among historians, she identifies disciplinary and sectoral boundaries as reasons, next to the methodological issues:

One explanation is that while digital history has embraced a range of historical sub-disciplines, and borrowed readily from subjects like archaeology and historical geography, it has largely failed to take account of developments in two crucial areas: library, archive and information studies; and digital preservation. Libraries and archives have necessarily been at the forefront of web archives research and practice. [...] (Winters 2018b)

This diagnose is indeed consequential. The gap between the progress in born-digital preservation development and archival science research, on the one hand, and (digital) humanities research on the other, needs to be closed, first and foremost in order to enable GLAM institutions, institutional networks and infrastructures to develop their born-digital collections in meaningful ways, improve preservation formats, curation workflows, repositories, services, and access for researchers. This can only be achieved by cross-sectoral and interdisciplinary collaboration to support active research on born-digital collections. This is precisely what this first special issue of *International Journal of Digital Humanities* seeks to encourage. The necessary collaboration will benefit from the new European General Data Protection Regulation (GDPR) guidelines, as these regulations provide an excellent basis for GLAM institutions and researchers to establish trust relationships with archive depositors and creators. This will, moreover, encourage them to enable research by having their materials preserved, archived, and made available to research in a secure, controlled, and authentic way with security procedures that empowers them as data subjects.

Kirschenbaum and Winters urge those in the field of humanities to embrace the born-digital historical record as an object and primary source in its own right – a claim which, of course, has precursors in media history and theory, historical bibliography, textual scholarship, and digital humanities (see Dahlström 2000; Manovich 2001; Gitelman 2006). This implies the critical appraisal of the born-digital primary record’s specific historical materiality, along the lines of philological and forensic disciplines such as diplomatics, palaeography, philology, and analytical bibliography. Since Kirschenbaum’s *Mechanisms. New Media and the Forensic Imagination* (2008) and Duranti’s introduction of the concept of digital diplomatics (Duranti 2009); Duranti and Endicott-Popovsky (2010), digital forensic methods and tools, especially bitstream-preserving imaging (also known as forensic imaging), became a standard practice in memory institutions for the preservation of digital storage media and born-digital records. Kirschenbaum’s seminal definition of formal and forensic digital materiality (2008, 10–11, see also Ries 2018, 389–401) – a selective focus on one dichotomic dimension in the spectrum of digital materiality (for an overview, see Drucker 2013) – conceptually enabled an analytical perspective on the materiality of the physical

characteristics of storage media and the bitstream of the historical born-digital record that reveals its digital history, embedded as latent forensic artefacts, recoverable data, and traces of processing and user interaction. While his work on digital materiality is certainly indebted to New Bibliography (Lebrave 2011) and oriented towards a relative physical stability of the forensic record, his more recent theoretical considerations of the born-digital archive seem to rather reflect issues of the instability, context-dependency, authenticity, and intangibility of the born-digital historical record as a logical digital object within formal materiality (Kirschenbaum 2013, 2016b).

But this also means that this data is fundamentally unstable in the sense that they rest upon the foundations of other data, what is quite literally in the trade known as metadata, in order to be legible under the appropriate computational regimens, which I have previously termed as formal materiality in my own work. (Kirschenbaum 2016b, 30:05)

In the terms I put forth in *Mechanisms*, each access engenders a new logical entity that is forensically individuated at the level of its physical representation on some storage medium. Access is thus duplication, duplication is preservation, and preservation is creation — and recreation. That is the catechism of the .txtual condition, [...]. (Kirschenbaum 2013, 16)

As questions of stability, authenticity, and technological context-dependency (on different concepts of authenticity relating to context, see Rogers 2015, 100) and materiality of the born-digital historical record become even more complex for preservation and research tasks, the role of archival custodian- and curatorship, digital signing (Blanchette 2012), digital forensic methodology, and context-preservation of complete operating systems by emulation or virtualisation and even computer hardware (Kirschenbaum et al. 2009) becomes even more prominent. In the light of inevitable ageing and obsolescence of hard- and software, bitrot, fading network contexts, and online services going offline, memory institutions already today have to decide according to which standards and criteria to select relevant materials. They have to decide which aspects of digital objects and their contexts are relevant to future research and have to be preserved in order to achieve an authentically preserved record, and what would be acceptable loss. Is it just the text or the content of a document that has to be preserved, metadata in the document or in the surrounding operating system, contextual material in file folders, the materiality of the complete operating system or file server – as „dead“ system in a forensic, fixed, bit-precise image or emulated at runtime –, or is the hardware or network context an important aspect to be preserved? Or is the experience of contemporary interaction a main factor that needs documentation? Some of the contributions to this special issue of *IJDH* revolve around these key questions of born-digital archives.

The archival and digital forensic perspective sheds light on the specific historicity of the born-digital record. Digital historicity does not only become apparent when one interacts with still functional legacy hard- and software in computing musea, experiencing the look and feel of historic operating systems and applications, the today unusual feel of thick cables, old port connectors and adapters, motherboards, controllers and storage media. The forensic materiality of the born-digital record, preserved in the form

of forensic images and other forensic formats, bears a highly specific signature of historical computing that can best be understood from the vantage point of Jean-Francois Blanchette's *A Material History of Bits* (2011). He remarkably takes the perspective of a historian who analyses historical hard- and software architectures, such as the processing and networking stack, principles such as layering and modularity of operating systems and applications, read as historical documents of design decisions taken by hard- and software engineers, programmers, and tech companies in their pursuit to overcome the physical constraints of computing by architecture abstraction and error-correction mechanisms to maintain an 'illusion of immateriality' (Kirschenbaum 2008, p. 135). Blanchette stresses that maintaining the illusion of immateriality of resources, and hiding their physical limitations and characteristics to programmers and users is in itself a resource-intensive, critical and error-prone task that is mostly implemented at the cost of technical 'efficiency trade-offs'.

This purported independence from matter would have two distinct and important consequences: (a) digital information can be reproduced and distributed at negligible cost and high speed, and thus, is immune to the economics and logistics of analogue media; (b) digital information can be accessed, used, or reproduced without the noise, corruption, and degradation that necessarily results from the handling of material carriers of information. [...] Yet, this abstraction from the material can never fully succeed. Rather, it stands in dialectical tension with the evolution of these material resources and with the efficiency trade-offs their abstraction requires. (Blanchette 2011, p. 1042)

Blanchette especially names the efficiency trade-offs implied by modularity, the efficiency cost of necessary garbage collection and error correction at runtime as 'design trade-offs inherent in abstracting from physical resources are rarely acknowledged in the computing literature' (Blanchette 2011, p. 1047). While some might want to nuance Blanchette's argument and note that modularity as a foundational principle of system architecture, code organisation, and programming language implementation is a necessity to ensure maintainability, manageability and extensibility of almost any larger system rather than be regarded as a performance penalty (which it can be), most will agree that overcoming the quirks of physical materiality is a resource-intensive task:

The digital abstraction can be maintained in spite of this "noise" because, as Kirschenbaum notes, through error-correction codes, buffering, and other techniques, computers can self-efface the static—scratches on a record, smudges on paper—that typically signals the materiality of media: [...] These mechanisms, formally described in information theory, are used throughout networked computing systems: the impact of media irregularities on hard drive platters can be mitigated through the use of error-correction codes; the unpredictability of network bandwidth can be mitigated through the use of buffering, ensuring smooth delivery of latency-sensitive content [...]. It is this ability to ceaselessly clean up after its own noise that so powerfully enables computers to seemingly sever their dependency on physical processes that underlie processing, storage, and connectivity. Yet the physical characteristics of a resource (be it computation, storage, or networking) cannot simply be transcended, and noise can only be

conquered at the expense of other resources. [...] error-correcting codes, widely used to protect against transmission interference, result in both data expansion (and thus, reduced capacity) and increased processing load. [...] Once again, then, independence from the material can only be obtained at the costs of certain trade-offs. (Blanchette 2011, p. 1047)

Blanchette's reasoning could serve as a foundation for a historical theory of digital forensics, an explanatory framework for many digital forensic phenomena, and the specific historicity of forensic digital materiality. Many phenomena that digital forensic tools and methods analyse are ultimately rooted in the mitigation of material constraints of hard- and software. Deleted data can be recovered because effective deletion through overwriting is a very resource-expensive task that would slow down a computer, which is why effective deletion does not take place by default. Often deleted data and documents "survive" on a system because of bugs, file system corruption, and system crashes: in CHKDSK error correction or hibernation files created by the operating system, in temporary and auto-recovery files not deleted because of system crashes. Temporary files are created on hard drives especially when a runtime environment runs out of physical RAM and has to swap memory with the storage medium. On some operating systems, automatic system snapshots are being created (e.g. VSS shadow copy partitions) in order to mitigate the risk of data loss through system instability. Files and file fragments are preserved in the so-called "drive slack" of data clusters because modern storage media are organised in blocks, which speeds up the process of data lookup and the navigation of large storage spaces on storage media with physical moving parts, such as conventional hard drives: it is the physical block size on the storage medium that determines where exactly a file is cut off. Fastsave artefacts in Microsoft Word documents and in temporary files are a result of a saving mechanism that was implemented to mitigate the relatively slow operation of early hard drives, at the cost of deleted text passages still present in documents and temporary files (Ries 2017, 2018). This incomplete list names just a few of the effects, mechanisms and design decisions that digital forensics is about and which are based on the computing-historical perspective that Blanchette describes. The digital forensic record, in turn, is deeply informed by designs that are specific for different types of hardware, versions of operating systems and application software, giving it a highly specific historicity that is accessible and readable through the forensic traces of digital processing. The latent digital forensic features of the born-digital historical record are not only of interest for philologists who search for hidden draft versions of a text. They are also relevant for historians and archivists who have to determine whether a historical record is authentic or might have been manipulated. Furthermore, they are relevant to the historian who investigates the history of the digitisation of society using original archived computing systems.

When we speak about the born-digital record, there is another aspect to be kept in mind, an aspect that is not in the foreground in this volume, but hopefully will be scrutinized in more detail during further issues of *IJDH*. As Blanchette rightly emphasizes, the historicity of born-digital phenomena is rooted in the material constraints of hard- and software, it is embedded in an infrastructure context without which it cannot be understood. The infrastructure of the digital archive, which serves as an interface between the researcher and their subject of research, requires attention in itself,

regardless of the fact that the research is based on born-digital or digitised materials. Michel de Certeau has pointed out in his seminal work *The Writing of History* (De Certeau 1988; Palkó 2019) that the computer, as an archive, forms a new apparatus for research and as such will fundamentally change the way historical documents will be formed. The materiality of the archive as medium of knowledge formation is one of the main research questions media archaeology focuses on (Ernst 2011; Parikka 2012). Parikka sheds light on the interdependence of problems current archiving practices face in a born-digital culture, and on the theoretical challenges of understanding how a digital archive as an apparatus forms our documents of the past and present.

the theoretical problems of recent media archaeologies of technical media and software along with a rethinking of the archive, go hand in hand with the practical challenges faced by cultural heritage institutions and professionals: how do you archive processes and culture which is based on both technical processes (software and networks) and social ones (participation and collaboration, as in massive online role-playing platforms as cultural forms). (Parikka 2012, 115)

The analyses of the institutional archiving practices have always been complicated for their medial and material mechanisms tend to stay in the shadow (Groys 2000; Palkó 2017). However, the analyses of the apparatus of the digital archive, which includes born-digital, processual, network- or environment-based material, is even more complicated. Although a lot has been done in the last decade to provide a stable digital object by forensic imaging on the level of forensic materiality, the actual documents extracted from a forensic image depend highly on the technical infrastructure (e. g. the chosen software and workflow), and requires technical skills that are normally not part of a humanities scholar's qualification. The same is true for the growing importance and complexity of searching the digital medium. As both digitized and born-digital records are available in a quantity impossible to fathom through the methodology of close reading, records relevant for a research question will mostly be gathered by using query services. Digital archives normally radically limit the possibility to use custom search tools and query languages, they only provide predefined and simplified options.

A lot has been done by national institutions and international projects both on technical, institutional, and discursive level to augment the traditionally analogue field of scientifically relevant material to the born-digital. Trusted formats, standards, methodologies, and services are available for GLAM institutions and researchers as well, but it remains an open question how the complexity of handling born-digital primary records and the thus established digital archives will be manageable for the humanists of the twenty-first century.

The current special issue of *International Journal of Digital Humanities* features articles by international researchers from the libraries and archives sector, as well as from the (digital) humanities that address born-digital archives on several levels, ranging from the digital forensic perspective on individual records (Archival Methodology: Digital Forensics), via personal digital archives and born-digital cultural heritage archives (Digital Culture and Literature Archives), web archives (Web Archives), to born-digital archiving in large digital infrastructures (Born-Digital Archives and Infrastructures).

Corinne Rogers (University of British Columbia, Vancouver, Canada) strikes the connection between digital forensics and born-digital archival science and

practice with a historical introduction to how digital forensics became a viable tool for digital curatorship.

Bénédict Vauthier (Bern University, Switzerland), after her studies on Robert Juan-Cantavella's born-digital dossier génétique of his novel *El Dorado* (Vauthier 2014, 2016), traces the inherent connection between Anglo-American textual scholarship and analytical bibliography on the one hand and the introduction of digital forensic methodology to archival science on the other, in an effort to find an explanation why European textual scholarship and philology seems to lag behind in this field. Vauthier also presents the results of her survey among Spanish-speaking writers about their digital self-archiving practice and their willingness to deposit their digital archives at memory institutions and make them available for research. Nicholas Schiller and Dene Grigar (Washington State University, Vancouver, Canada) provide an insight into their work at the *Electronic Literature Lab* (ELL) at Washington State University Vancouver on the process of archiving electronic literature, specifically about documenting the interactive experience with Sarah Smith's *King of Space* in the 'traversal' format. Schiller and Grigar's discussion and example show some of the important challenges of electronic literature archiving and the solutions practiced at ELL. Libi Striegl and Lori Emerson (University of Colorado Boulder, USA) describe their archival and 'anarchival' experience- and practice-based approach to research and research creation at the Media Archaeology Lab (MAL) at the University of Colorado at Boulder. As an example, they document the project on mesh-networked *One Laptop Per Child XO* laptops at MAL. The *One Laptop Per Child* initiative with its tailored technological ecosystem is an important educational inclusion project worth documenting, its use of mesh networks and hardware design introduced an innovative approach to local networking, network capacity sharing and solutions for operation under technologically difficult circumstances and infrastructure.

In the web archives section of the current issue, Trevor Owens, editor of Owens 2013a, b, and Grace H. Thomas (Library of Congress, USA) trace the history and functional changes of the Spacer GIF and the resulting challenges for web archiving. Eveline Vlassenroot (Ghent University, Belgium), Sally Chambers (Ghent University, Belgium), Emmanuel Di Pretoro (Haute École Bruxelles-Brabant, Brussels, Belgium), Friedel Geeraert (Royal Library and State Archives of Belgium, Brussels, Belgium), Gerald Haesendonck (Ghent University, Belgium), Alejandra Michel (Namur University, Belgium) and Peter Mechant (Ghent University, Belgium) discuss national and international web archives as a data resource for digital scholars in Europe.

In the *Born-Digital Archives and Infrastructures* section, Tibor Kálmán (GWGD Göttingen, Germany), Matej Ďurčo (Austrian Academy of the Sciences, Austria), Frank Fischer (Higher School of Economics, Moscow, Russia), Nicolas Larrousse (HumNum, Paris, France), Claudio Leone (State and University Library Göttingen, Germany), Karlheinz Mörth (Austrian Academy of the Sciences, Austria) and Carsten Thiel (State and University Library Göttingen, Germany) map the challenges, approaches and solutions of born-digital archiving and access, especially for born-digital research datasets, learning materials, services and software in the context of the European DARIAH research infrastructure, and beyond.

The special issue concludes with Peter Mechant's (Ghent University, Belgium) review of *Web 25. Histories from the first 25 years of the world wide web*, edited by Niels Brügger in 2017.

References

- Baker, J. (2018). *Outlook: Email archives, 1990–2007*, guest lecture at Ghent University, Belgium, 8 May 2018.
- Blanchette, J.-F. (2011). A material history of bits. *Journal of the American Society for Information Science and Technology*, 62(6), 1042–1057.
- Blanchette, J.-F. (2012). *Burdens of proof. Cryptographic culture and evidence law in the age of electronic documents*. Cambridge: MIT Press.
- Carroll, L., Farr, E., Hornsby, P., & Ranker, B. (2011). A comprehensive approach to born-digital archives. *Archivaria*, 72, 61–92.
- Catalogue ÖNB (2018). Catalogue entry at Austrian National Library, ÖNB. <https://www.onb.ac.at/bibliothek/sammlungen/literatur/bestaende/personen/czernin-franz-josef-geb-1952/>. Accessed 11 July 2018.
- Dahlström, M. (2000). Drowning by versions. *Human IT 4*, <http://etjanst.hb.se/bhs/ith/4-00/md.htm>. Accessed 11 July 2018
- De Certeau, M. (1988). *The writing of history*, (trans: Conley, T.). New York: Columbia UP.
- Drucker, J. (2013). Performative materiality and theoretical approaches to interface. *Digital Humanities Quarterly*, 7(1). <https://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html>. Accessed 11 July 2018.
- Duranti, L. (2009). From digital diplomatics to digital records forensics. *Archivaria*, 68, 39–66.
- Duranti, L., & Endicott-Popovsky, B. (2010). Digital records forensics. a new science and academic program for forensic readiness. In *ADFSL Conference on Digital Forensics, Security and Law*. <http://arqtuleufs.pbworks.com/w/file/fetch/94919918/Duranti.pdf>. Accessed 11 July 2018.
- Enge, J., & Kramski, H. W. (2014). “Arme Nachlassverwalter ...”. Herausforderungen, Erkenntnisse und Lösungsansätze bei der Aufbereitung komplexer digitaler Datensammlungen. Filthaut, J. (Ed.), *Von der Übernahme zur Benutzung. Aktuelle Entwicklungen in der digitalen Archivierung*. 18. Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen on 11–12 March 2014 in Weimar, (pp. 53–62) Weimar: Thüringisches Hauptstaatsarchiv.
- Ernst, W. (2011). Media archaeography method and machine versus history and narrative of media In: Huhtamo, E. and Parikka, J. (Eds.), *Media archaeology. Approaches, applications, and implications* (pp 239–255). Los Angeles: University of California Press.
- Foss, R. (2017). The British Library: *Learning from users of personal digital archives at The British Library. Research paper at conference Personal Digital Archiving 2017*, 29–31 Mar 2017, Stanford Libraries.
- Gitelman, L. (2006). *Always already new. Media, history, and the data of culture*. Cambridge: MIT Press.
- Groys, B. (2000). *Unter Verdacht. Eine Phänomenologie der Medien*. Hanser.
- John, J. L. (2012). *Digital forensics and preservation. DPC technology watch report 12–03 November 2012*. Digital Preservation Coalition. <https://doi.org/10.7207/twr12-03>. Accessed 11 July 2018.
- Kirschenbaum, M. (2008). *Mechanisms. New media and the forensic imagination*. Cambridge: MIT. University Press.
- Kirschenbaum, M. (2013). The .txtual condition: digital humanities, born-digital archives, and the future literary. *Digital Humanities Quarterly*, 7(1), <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html>. Accessed 11 July 2018.
- Kirschenbaum, M. (2016a). *Track changes. A literary history of word processing*. Cambridge: Harvard University Press.
- Kirschenbaum, M. (2016b): The transmissions of the archive. Literary remainders in the late age of print. Kirschenbaum, M. (Ed.), *Bitstreams. The Future of Digital Literary Heritage*. Lecture Series at KISLAK Center for Special Collections, Rare Books and Manuscripts, Penn Libraries. 14 March 2016. <https://www.youtube.com/watch?v=6TuA4dkRegQ>. Accessed 11 July 2018.
- Kirschenbaum, M., Farr, E.L., Kraus, K., et al. (2009). Digital materiality. Preserving Access to Computers as Complete Environments. In: *iPress 2009. 6th International Conference on Preservation of Digital Objects*. 5. October 2009. University of California. California Digital Library. http://www.escholarship.org/uc/cdl_ipres09. Accessed 11 July 2018.
- Kirschenbaum, M., Ovenden, R., Redwine, G., et al. (Eds.) (2010). *Digital forensics and born-digital content in cultural heritage collections*. Washington, D.C.: Council on library and information resources, Washington, D.C. Dec. 2010. <http://www.clir.org/pubs/reports/reports/pub149/pub149.pdf>. Accessed 11 July 2018.
- Lebrave, J.-L. (2011). Computer forensics: la critique génétique et l'écriture numérique. *Genesis*, 33, 137–147.

- Manovich, L. (2001). *The language of new media*. Cambridge: MIT Press.
- Owens, T. (Ed.) (2013a). *Preserving.exe: Toward a national strategy for software preservation*. National Digital Information Infrastructure and Preservation Program at the Library of Congress, Washington, D.C. http://www.digitalpreservation.gov/multimedia/documents/PreservingEXE_report_final101813.pdf. Accessed 11 July 2018.
- Owens, T. (2013b). *Historic iPhones: Personal digital media devices in the collection*. 15 November 2013. <http://www.trevorowens.org/2013/11/historic-iphones-personal-digital-media-devices-in-the-collection/>. Accessed 11 July 2018.
- Palkó, G. (2017). Media archaeology of institutional archives? *Studia UBB Digitalia*, 7(62), 75–82. <https://doi.org/10.24193/subbdigitalia.2017.1.05>. Accessed 11 July 2018.
- Palkó, G. (2019). Sites of digital humanities. About virtual research environments. In Kelemen, P. & Pethes, N. (Eds.), *Philology in the Making. Analog/Digital Cultures of Scholarly Writing and Reading* (pp. 221–230). Bielefeld: Transcript.
- Parikka, J. (2012). *What is media archaeology*. Cambridge: Polity.
- Reside, D. (2011a). “Last modified January 1996”. The digital history of RENT. *Theatre Survey*, 52(2), 335–340.
- Reside, D. (2011b, April 22). “No day but today”. A look at Jonathan Larson’s Word Files [Blog post]. <http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files>. Accessed 11 July 2018.
- Reside, D. (2017). West side story: The journey to Lincoln center theater. In L. MacDonald & W. A. Everett (Eds.), *The palgrave handbook of musical theatre producers* (pp. 359–367). New York: Palgrave Macmillan.
- Ries, T. (2017). Philology and the digital writing process. *Cahier voor Literatuurwetenschap*, 9, 129–158.
- Ries, T. (2018). The rationale of the born-digital dossier génétique: Digital forensics and the writing process: With examples from the Thomas Kling archive. *Digital Scholarship in the Humanities*, 33(2), 391–424.
- Rockmore, D. (2014). The digital life of Salman Rushdie. *The New Yorker* published 29 July 2014. <http://www.newyorker.com/tech/elements/digital-life-salman-rushdie>. Accessed 11 July 2018.
- Rogers, C. (2015). Authenticity of digital records. A Survey of Professional Practice. *Canadian Journal of Information and Library Science*, 39(2), 97–113.
- Sloyan, V. (2018). *Overview of a born-digital archives access workshop held at Wellcome collection*. Findings of a workshop held on 24th November 2017. <https://doi.org/10.6084/m9.figshare.6087194.v1>. Accessed 1 March 2019.
- Vauthier, B. (2014). Tanteos, calas y pesquisas en el dossier genético digital de El Dorado de Robert Juan Cantavella. In M. Kunz & S. Gómez-Rodríguez (Eds.), *Nueva narrativa española* (pp. 311–345). Barcelona: Linkgua.
- Vauthier, B. (2016). Genetic criticism put to the test by digital technology: Sounding out the (mainly) digital genetic file of El Dorado by Robert Juan-Cantavella. *Variants*, 12-13, 163–186.
- Waugh, D. & Russey Roke, E. (2017). *Second-generation digital archives: What we learned from the Salman Rushdie project*. Research paper at conference Personal Digital Archiving 2017, 29–31 Mar 2017, Stanford Libraries.
- Winters, J. (2018a). *Humanities and the born digital: Moving from a difficult past to a promising future?* Keynote at DHBenelux 2018, Amsterdam. 7 June.
- Winters, J. (2018b). Web archives and (digital) history: a troubled past and a promising future? In N. Brügger & I. Milligan (Eds.), *Sage Handbook of Web History* (pp. 593–606). Newcastle: Sage.