HUMANITIES OF THE DIGITAL: MEDIA PHILOLOGY

[Extended pre-version of keynote for *Digital Humanities* conference Budapest, Eötvös University, DH Centre, May 2018]

Media philology as critical inquiry of Digital Humanities practice
New methods: The message of DH
Good-bye, "archive - towards dynamic data retrieval
The insistence of "humanities"
Rooting DH: techno-mathematics
Big textual data and the distant symbol-processing gaze (Moretti's *distant Reading*)
New options of big data retrieval
Signal instead of text criticism: sound recording *avant la lettre*
"Forensic" media archaeology of a digital "archive" (the storage architectural element ROM)
Counting by numbers instead of story-telling: Markov chains
The archive and the question of "open access"
"Active archives"
The self-organizing archive
"Hermeneutics" after Shannon
Algorithmic hermeneutics
Media *archivology*: Kittler's case
"Digital humanities" *avant la lettre*? 'New Archaeology' and Peirce's archaeological semiotics

Media philology as critical inquiry of Digital Humanities practice

Digital Humanities (DH) as research practice is not only about digitising documents, neither is its focus only on "open access" to "big data", but (methodologically more challenging) by the application advanced software tools (called "cultural analytics" by Lev Manovich). This orientation requires a new kind of (second order) media philology: the critical analysis of the underlying algorithms, such as the codecs applied for temporarily compressing "big data".

In computational culture, we are dealing with technologies in its literal sense: not simply the symbolic order of textual *lógos* (be it the alphabet, be it alphanumeric algorithms), but with their implementation in matter (*techné*). Therefore, next to critical code studies, the other side of media philology is actual computer hardware "forensics" (like Kirschenbaum's analysis of the hard disk drive), as the contemporary version of what archaeologist Eduard Gerhard in mid-nineteenth century had coined "material philology"[1].
Media archaeological analysis, complementary to media philology, embraces the material component as technical criterium to otherwise software-focused data analysis as well.

---

1 Eduard Gerhard, Zur monumentalen Philologie. Vortrag vor der Philologenversammlung in Berlin im J. 1850, in: Archäologischer Anzeiger 1850, 203 ff.

With the concept of emulation, the re-enactment of an obsolete computer architecture within a present computer turns former hardware engineering itself from a material object into an alphanumerically coded text, therefore accessible for critical media philology.

New methods: The message of DH

Digital Humanities is employing analytic tools which have not been available "before" to the traditional canon of humanities (that is *avant la lettre*, in fact the letters of the binary alphabet). An example is the tracking of changes in digital texts, such as the successive versions of entries in the online encyclopedia Wikipedia, by application of the "diff" operation, which is way of text retrieval well known from "genetical" philology.

The Internet search engine Google includes a tool called the Ngram Viewer for graphically plotting the frequency of terms over time. "An *n-gram model* is a type of probabilistic language model for predicting the next item in such a sequence in the form of a(*n*-1)-order Markov model."[2]

The complex techno-logical operation of sampling *alias* digitization of signals (text, sound, image) into binary data, combined with mathematical and logical intelligence called "software" (effectively brought together in the von-Neumann architecture of stored-program computing) is the very media archaeological condition of possibility for DH performances - kind of Kantean *a priori*, understood here in a techno-mathematical sense.

Digital Humanities have been criticized for emphatically browsing "bid data" by statistical analysis and producing just trivial results. But evidence such as the predominance of the colour red in late eighteenth century paintings is only the superficial aspect of Digital Humanities practice. In its incubation period, it is less the (adminttedly thin) content that counts for DH, but its medium message (in McLuhan's sense) which is experimenting with algorithms (for which the spread of "labs" is indicative[3]), the application of stochastic (rather than simply statistical) analysis. For film studies, this goes beyond Juri Trivan's *cinemetrics* which still refers to the *Einstellung*, the camera shot sequence[4] within a single film, it rather calculates with the single frame, and more radically: with every pixel in the digitized frame to identify its entropy on the micro-scale[5], or big numbers of films on the macro-scale.

Only the algorithmic media-active archive - the central agency of so-called Digital Humanities - can ensure processing and computation of such "big data".[6]

---

2 Wikipedia entry "*n*-gram", version 17 Juli, 2014
3 See Jussi Parikka / Lori Emerson / xxx Weshler (eds.), The Lab Book, xxx
4 See xxx Heftberger, xxx
5 See Lev Manovich, How to xxx one million images, in: David Berry (ed.), xxx
6 See Morgens Jacobsen / Morten Søndergaard (eds), Re-Action. The Digital Archive Experience, xxx (Aalborg University Press) 2008

„How to compare one million images?", Lev Manovich asks.[7] Attention is not given exclusively to the single text or artifact any more like in traditional (film) philology, but to patterns instead - either combinatorial or statistical. Statistics (symbolical archival order) transforms into pseudo-random stochastics in the world of computed signals (DSP).

As an epistemological heritage of nineteenth century administrative statistics / physical thermodynamics, in so-called *cultural analytics* (Manovich), "entropy describes the degree of uncertainty in the date"[8], resulting in a rather un-archival order in fluctuation. Training in the application of algorithmic tools, which is mathematical analysis, is the real effect where Digital Humanities modify traditionally hermeneutics-based Humanities.

Dynamic algorithmic access (such as the Google search engine) nowadays replaces the static classification of the traditional catalogue in libraries. From this results the need for a flexible tool which allows for the coexistence of different orders without destroying the material structure - *relational* databases and *random search* (familiar in "hashing" in the the administration of computer storage, a kind of *order in fluctuation*) which is the radical temporalization of order itself.

The answer lies in discovering and reflecting upon (and techno-mathematically realizing) new technologies of memory by flexible access with the signal-oriented search functions not being limited to the alphabetic address exclusively.

Good-bye, "archive - towards dynamic data retrieval

From the digitization of vast amounts of records - mostly by necessity of preserving the data against progressive material obsolescence - arises a creative chance which is progressively performed by so-called "digital humanities": applying creative algorithms to experiment with new forms of navigating enormous amounts of archival signals and data from within (be it textual or audio-visual), resulting in new insights by mathematical intelligence like stochastic analysis and similarity-based retrieval and information as measure of Shannon entropy.

In "nineteenth century" (which itself is a rather cloudy term, symbolically given shape by historiographical narrative), thermodynamics co-originated with "social statistics": Quetelet's *homme moyen*. And in Tarde's social statistics, information theory and sociology, for once, converge in the non-metaphorical "thermal" concept of stochastic probabilities: the statitistian, like the archeologist, „jette sur les faits humains un regard tout abstrait et

---

7 In: David Berry (ed.), Understanding Digital Humanities, xxx
8 Lev Manovich, How to Compare One Million Images?, in: Understanding Digital Humanities, edited by David M. Berry, Basingstoke (Palgrave Macmillan) 2012, 249-278 (266)

impersonnel“[9] - which in present Digital Humanities returns as "social analytics" (Lev Manovich).

As expressed by the think tank of the Norwegian National Library in Oslo, the archive is "in motion". A current case is the NSA discussion: "predictive analytics" does not accumulate data from a long past but collect present data to predict immediate future profiles - thus historicizing the future already (the "future in the past"), resulting in new forms of time manipulation - true chronopoetics.

The insistence of "humanities"

In Medieval scholasticism, the kernel of what was to become the academic university, four "scientific arts" – music, arithmetic, geometry and astronomy respectively astrology - were known as the Quadrivium; the remaining three arts (the Trivium) constituted the "humanities" – grammar, logic, and rhetoric. In that two-fold form the seven *artes liberales* were taught. In DH, "science" and "humanities" fuse (at least fold) into one. In the Renaissance, the old Trivium has been re-christened into *Studia humanitatis*, with new emphasis on poetry. Since then, humanist disciplines like history and related academic domains study subject matters that the experimental method does not apply to - "and instead mainly use the comparative method"[10]. In DH, comparative research itself becomes *algorithmically experimental*.

"As humans and data machines become equal partners in cultural practice, social experience, and humanistic research, the humanities may no longer look like 'the humanities.'"[11] "Digital post-Humanism" is no more *Geisteswissenschaft* (Dilthey); in that sense, DH actually (and partly in the sub-conscious epitemological memory) re-invents the informational aesthetics of cybernetics (Moles) - rather an up-dating than an exorcism of *Geisteswissenschaft*.

Against the suggestive term "Digital Humanities", in media theory there is a rather humanistic awareness (*Geistesgegenwart*) of algorithmic knowledge. It is not simply the quantitative increase of big data processing due to available storage capacities (Moore's Law) which escalate in so-called digital culture, but its combination with a different quality of data processing: algorithmic programming as techno-logical (mathematical) "intellectualising" in its double technical and philosophical (enlightenment) sense of computational *intelligence*.

Once the harvesting of "big data" turns into epistemogenic operations, quantities of digitized cultural sources become qualitative *humanities*; "DH" methods require critical reflection as has been traditionally cultivated within

---

9 Gabriel Tarde, Les lois de l´imitation, Paris 1890, chap. IV (Qu´est-ce que l´histoire?), Absatz „L´Archéologie et la Statistique“, 99 u. 114
10 Entries "Liberal arts (education)" and "Humanities" in https://en.wikipedia.org/wiki, accessed August 7, 2017
11 Anne Burdick / Johanna Drucker / Peter Lunenfeld / Todd Presner / Jeffrey Schnapp, Digital_Humanities, Cambridge, MA (MIT Press) 2012, 105

"Humanities of the Digital" (Jan Claas van Treeck) - even infused with "Geisteswissenschaft" (which is not identical)?

Media science of computational practice is appropriately located in the Department of Humanities and not in Computer Science only, since it asks different questions. This is an inheritance of philosophical reasoning, as expressed in the way Martin Heideggers questions technology: "The essence of technology is nothing technical."[12]

Rooting DH: techno-mathematics

The meaning of "humanities" in the term DH stretches back to early modern humanism as been primarily text-based knowledge and communication, resulting in academic "humanities" (German "Geisteswissenschaften", as defined by Dilthey, and "Belles-lettres" in France).

[For the study of technology, this is accompanied by material antiquarianism for material analysis of relics from the past, which cooriginated with textual humanism in the Renaissance.]

Textuality allows for time-invariant knowledge transmission - the alliance between code (alphabet) and technology (the printing press). Humanist communication took place in metahistorical con-temporality. With computing, this option has extended from exclusively human performance into the operative machine: "Once a software-based system is working, it *should* work forever (or at least until the underlying hardware breaks down)."[13] Software "in theory never breaks down" <ibid.>; only then it becomes "ready-at-hand" (an issue of theoretical contemplation) in Heidegger's sense.

While the turingmachine (*alias* computer) is the first theory-born technology, DH sprang from computing. "The roots of computational work in the humanities stretch back to 1949 when the Jesuit scholar Roberto Busa, working in collaboration with IBM, undertook the creation of an automated approach to his vast Index Thomisticus, a computer-generated concordance to the writings of Thomas Aquinas. <...> Other early projects included the debut, in 1966, of Computers and the Humanities, the first specialized journal in the field. Seven years later, the Association for Literary and Linguistic Computing (ALLC) was founded, with the Association for Computers and the Humanities (ACH) following in 1978."[14] But "algorithmicized humanities" extends beyond the literary text and its "distant reading" (Moretti), sorting images and sound as well *from within* (instead of subjecting them to verbal, "literary" descriptions and metadata). This results in the "Active Archive" (as called by the Bruxelles-based media research group Constant).

Media philology - and here it differs from textual philology - "literally" operates *avant la lettre*. Computer code is a kind of writing indeed, and software development is "a form of literary production", but its actual implementation in

---

12 Heidegger, in: The Question Concerning Technology, xxx
13 Ensmenger 2009: 88
14 Burdick et al. 2013: 123

computing is electronics.[15] The ease with which computer code can be written, modified, and deleted" questions the durability of the underlying document. The "palimpsest" <ibid.> of record traces on a hard drive requires media philological forensics indeed. There is a palimpsestuous con-temporaneity in computing, a coexistence of different "languages" within the von-Neumann architecture of computing. Today, "there are still more than 240 million lines of computer code written in Cobol, first introduced in 1959.[16]

- What happens inside a computational signal processor requires both physical (engineering) and philological (code) skills: measuring and deciphering the alpha-numeric code (hexadecimal / binary).

- Media philology is an auxiliary science for Media Studies (just like media archaeology). The technical term "archaeography" originally referred to applications of non-numeric computing within the discipline of archaeology.[17]

Big textual data and the distant symbol-processing gaze (Moretti's *distant Reading*)

"[T]o see through computer 'eyes'"[18] is "a powerful mechanism of defamiliarisation <...> - a device for seeing what we could have not noticed previously"[19]; culture-free (Pias) technical scanning - recalling philosopher Friedrich Nietzsche's "passion for distance" - is a non-hermeneutic optics / option to extract a different kind of knowledge.

"The computational allows us to perform what literary scholar Franco Moretti has termed 'distant reading' – a practice that moves away from the close, hermeneutical reading of texts in favor of an algorithmic approach that presents overarching structures and patterns. For Moretti, distance is 'a condition of knowledge' because it allows a scholar to 'focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems'."[20]

According to Moretti, one can not study a large archive in the same way one studies a signle text. Individual texts have been written to "speak" to the reader, and so, provided s/he knows how to read them hermeneutically, they always end up telling something; "*but archives are not messages that were meant to address us*, and so they say absolutely nothing until one asks the

---

15 Nathan Ensmenger, Software as History Embodied, in: IEEE Annals of the History of Computing 31 (1), 2009, 88-90 (90)
16 M. Swaine, Is Your Next Language COBOL?, in: Dr. Dobbs J., 18 Sept. 2008 (as referred to in Ensmenger 2009: 90)
17 Gundlach 1968 "Maschinelle Philologie als historische Hilfswissenschaft", 232, referring to journal *Computer and the Humanities* New York 1966 ff.
18 Lev Manovich, How to Compute One Million Images?, in: ebd., 249-278 (276)
19 Manovich 2012: 276
20 TS Presner, http://www.joodsmonument.nl/?lang=en, 23, referring to: Franco Moretti, "Conjectures on World Literature," in: New Left Review (January / February 2000), 54-68 (57)

right question"[21]. There is a fundamental difference, though, between Shannon's mathematical theory of communication as ntentional channel-coding, and the administrative archive as depository of legal claims; still, such archives are (mis-)read by historians as if provided with an intended message from the past to the present, to be transformed into a narrative by historiography. "Archives are not messages that were meant to address us, and yet meaningful information can be extracted from them. This is an exciting idea when the archive is the entirety of literary history, a chilling one when it's our private internet activity or phone records."[22]

"Such archives and methods and technologies permit scholars today to ask different kinds of questions about the materials they investigate. Moretti describes the approach to literary analysis that he develops across his essays as a 'quantitative formalism'; like all formalisms, the point is not just the objective information derived — in this case, through counting — but rather what that counting reveals."[23]

New options of big data retrieval

For librians as metadata experts, classification is still essential in data retrieval (Thomas Hill[24]). But as media archivist have started digitizing and processing media such as film and sound recordings, the classification component of such projects has transformed from external to internal, from *within* the digitized records. Signal processing is replacing discourse in media culture, creating a new kind of algorithmic archive. Familiar cultural *analysis* is increasingly replaced by big data cultural *analytics*. There are new options of information retrieval, based on the physical signal qualities, not limited any more to its logocentristic transcriptions (subject to metadata).

Signal instead of text criticism: sound recording *avant la lettre*

Patrick Feaster succeeded in re-sonifying Léon-Scott's 1859 phonautogram of the children song *Au Claire de Lune*. Such a recovery of the acoustic past is possible by means of highly sophisticated algorithmic filters only which becomes itself the active archeologists of signal intelligence.[25]

---

21 Kathleen Fitzpatrick, The Ends of Big Data, in: Los Angeles Review of Books, June 27th, 2013, on Franco Moretti's collection of previous essays *Distant Reading*; http://lareviewofbooks.org/article.php?type=&id=1801&fulltext=1&media=#article-text-cutpoint
22 Fitzpatrick 2013
23 Fitzpatrick 2013
24 See interview W. E. by Thomas Hill, Vassar College, Dept. of Art, Poughkeepsie, N. Y., on occasion of W. E., *Digital Memory and the Archive*, in the academic radio program *Library Cafe*, http://library-cafe.org
25 For the recovery of early 30line television images from so-called "Phonovision" recording by means of algorithmic image processing, see Donald McLean, Restoring Baird's Image, xxx

"Digital humanties", in the media-archaeological sense, means "algorithmic hermeneutics": The application of computational software as active archaeologist of cultural knowledge hidden within techno-phyiscal signals. This is not only relevant for harvesting "bid data", but for close analysis of material technology.

Algorithms allow for critical signal studies. By optical reading of signals and application of digital filters, it is possible to digitally trace past acoustic signals from records.

The first officially archivized record of sound in Norway is a tinfoil, flattened to a "document" and annotated by a remark by its former collector, as exhbited within a frame at the Technical Museum of Oslo. The Sound Archive Project at the School of Engineering Sciences in the University of Southhampton attempted a digital restauration: "The whole artefact's surface topology is mapped to high precision using optical sensors, and the audio recovered by applying signal and image processing methods to the measured data. The measurement process for this artefact took three weeks of continuous scanning. Initial attempts at audio recovery from the surface data using existing processing techniques were largely disappointing, leading to the development of a more sophisticated methodology based on feature tracking through the groove. Out of six short tracks found on the foil, four contained significant audio portions featuring both music and speech, the remaining two tracks were both short and contained negligible content."[26]

Finally, the extracted (and reproduced) signal results in true media philology, falsifying the accompanying alphabetic annotation: "The extracted audio <...> was not the expected psalm singing as documented in the contemporary sources, but a mixture of shorter extracts. Features of the grooves and the extracted audio may confirm that the foil is a small portion of the recorded foil, and that portions of the remaining foil could have been distributed to other guests of the event, consistent with contemporary practice" (ibid.).

- From digitally enhanced restauration of the ealiest (archived) sound recordings we expect sound, but really what we primarily hear is noise. In the case of the earliest remaining sound recording from Norway, a tinfoil flattened to a „document“, an annotatation by the former collector claims this has been the first Norwegian recording of music on Edison cylinder. But the digital reading of this record (at a laboratory in Southampton) resulted in nothing but noise - which in terms of communication theory might be a message itself, that is: the communication of the recording medium.

Sound: https://www.nrk.no/kultur/xl/kan-verdens-eldste-opptak-av-edison-ha-ligget-i-en-norsk-kjeller-siden-krigen_-1.13727285; accessed November 13, 2017

---

26 P. J. Boltryk, J.W. McBride, L. Gaustad, Frode Weium, Audio recovery and identification of first Norwegian sound recording, lecture at JTS 2010 conference in Oslo (Digital Challenges and Digital Opportunities in Audiovisual Archiving); *online* xxx

Different from conventional historical research in archival records, real media philological criticism is not related to the contextual metadata but, paradoxically, derives insight from critical, "forensic" (Kirschenbaum) signal analysis itself.

"Forensic" media archaeology of the digital "archive" (the storage architectural element ROM)

Media philology even escalates when not only computing is applied to other sources, but to computers themselves. In order to extract code from an obsolete Read Only Memory within a micro processor, it requires both physical analysis and software to extract the bits.

"The chip itself is using a known architecture and a published assembly language, so the only reverse engineering required is to recover the actual instructions stored in the ROM." By electro-physical signal analysis, "data" become clearly discernible" (ibid.).

Fig.: Aperturelabs

If the unknown bits is put through a disassembler, they may make code sense again, restauring them for hermeneutics. Media philology and its twin media archaeology is both hardware and software hacking. While software hacking can be destructive on the symbolical level, tinkering with circuits that are directly connected to mains electricity can be dangerous in a bodily sense. Material de-construction of computer chips is driven by material criticism: "[...] trying to do something like reset a fuse to allow reading/writing of protected areas or probe a data track to observe data being processed by the chip, or even trying to figure out the actual logic of a proprietary chip by viewing and reverse engineering it's construction."[27]

One specific media-archaeological (or -archival) target is to restore the program code that is stored in a masked Read Only Memory (ROM) chip:

Counting by numbers instead of story-telling: Markov chains

In the cybernetic premise, literary texts and categories like the author "style" can be identified by computation words and letters.[28] What in principle had been DH *avant la* lettre became, with massively increased processing power and available data in computing today. Stylometry (based on the Deep Learning data recursions) is a radical challenge to traditional hermeneutics. According to a basic law of technological media, such a form of analysis can flip into active synthesis itself.

---

27 Fun with Masked ROMs - Atmel MARC4, http://adamsblog.aperturelabs.com/2013/01/fun-with-masked-roms.html; accessed 10th July, 2014

28 See Wilhelm Fucks, On mathematical analysis of style, in: Biometrica 39 (1952)

A radio play has been composed by the protagonist of informational aesthetics Max Bense (with Ludwig Harig) *and* an early electronic computer as co-author. The plot is aut(h)o(r)-executing: A girl is found knocked unconscious on the beach after a shipwreck. In hospital she starts to produce a monologue which from meaningless fragments of artikulation slowly emerges into meaningful bits of a narrative. In *Der Monolog der Terry Jo* (1968), informational is introduced into an apparently meaningless, "noisy" sequence of letters by application of a Markov chain algorithm. "[T]he basic cultural technique is not anymore to decipher the meaning of a chain of signifiers but to identify a message" - literally in this case - "in an ocean of noise (filtering). Human intelligence becomes subordinate to signal intelligence."[29]

The measuring unit of mean probabilities (entropy) in discrete communication engineering called *bit* [30] has enabled "information" to raise beyond the constraints of matter and energy, actually treating sequences of letters not as writing any more but literally like numeric *calculi* (pebbles). Markov chains deal with conditional probabilities, where "the likelihood of a given future state, at any given moment, depends only on its present state, and not on any past states"[31].

"Information" in the sense of mathematical theory of communication can only appear in relation to some other signals (like de Saussure's linguistic concept of phonetic articulation). The single letter "A" is no information at all. As part of the conventional alphabet consisting of 26 letters (in Germany), it has a higher "informational" value (that is, of surprise) than as part of a technical "alphabet" like the binary one, consisting only of the states "A", and "B".

By tabulating the sequence of vowels and consonants in Puschkin's poem *Eugen Onegin*, A. A. Markov in 1911 emancipated the alphabet from oral poetry (beyond Homer) which once induced the ancient Greek modification of Phenicean syllabic writing into the musicality of the phonetic alphabet (Barry Powell 1992). All of the sudden, a literary text is not a symbolic reflection of world states any more, but becomes operational part of it.

If the cultural idea of being "human" is anthropologically linked to story-telling, it is about to be reconfigured in the digital matrix.[32] The notion of the "digital human" is an oxymoron: What looks, sounds or behaves like human is itself unseparable from the machinic, as indicated by the subtitle of Norbert Wiener's

---

29 Bernhard Siegert, Turn Meaning On/Off: The Flip and Flop of „Understanding“ Media, Ziegler Lecture 2016, St John's College, University of British Columbia, Vancouver; typescript

30Claude Shannon / Warren Weaver, The Mathematical Theory in Communication, xxx [*1948] xxx

31 Editorische Anmerkung zu: Abraham A. Moles, Cybernetics and the Work of Art [*1965], in: Margit Rosen (Hg.), A Little-Known Story about a Movement, a Magazine, and the Computer's Arrival in Arts. New Tendencies and Bit International, 1961-1973, Karlsruhe (ZKM) / Cambridge, Mass. (MIT) 2011, 217-225 (225)

32 A central hypothesis of the MAC International Symposium *Story, Hypermedia, Digital Human*, Yonsei University Seoul, Institute of Media Art, May 2004

Cybernetics, or communication and control in the animal and the machine from 1948. According to Gregory Bateson, in creative art man experiences himself as a cybernetic model. This becomes even more acute when man is coupled to uniquitous computing. In this context, media science identifies the digital human in its very being language-coded already.

The archive and the question of "open access"

Documentary record archives *strictu sensu* are administrative and legal, (often state- or copyright-related) institutional agencies which radically differ from the experimental algorithms for processing "big data" in the Digital Humanities laboratories.[33] When algorithmic access replaces the static classification of the traditional library catalogue, statistical probabilities replace particular knowledge according to information theory, and pattern recognition replaces alphabetical identification), as dynamic articulation of *implicit* record-knowledge.

Archives in times of "digital humanities" can not be reduced to the question of "open access" to "big data". Jean-François Lyotard once formulated the political challenge of *La condition postmodene* (1979): *who* gets access to data banks. In times of DH, this transforms into the question of *how* to access the archives.

The question of "open access" does not only refer to big data banks but has a material aspect: the shrinking intervals of obsolescence in both hard- and software generations which become incompatible. "Access" of cultural heritage remaining from the digital age becomes a radical media-archival challenge.

- With(in) the computational *l'rchive* (in Foucault's sense), the familiar historicist order of cultural time, the chronological sequence, "as the emptiest of all kinds or order in which stored things are to be put, could be replaced by an order of co-presence once their combinatory connections were located."[34] Such operations in computational space are epistemologically productive since they do not destroy the material integrity of the existing record. "Digital archiving could break up the alliance that the institutional archives have maintainted with historiography and historicism since 1800"[35],

An even more radical media archivology results from experimenting with "digital born" archives, as performed by the research art collective Constant in Bruxelles with their project "Active Archive" (Nicolas Malevé / Michael Murtaugh).[36] Within the computational context of Digital Humanities, criteria

---

33 See Joanna Drucker, SpecLab, xxx
34 Friedrich Kittler, Museums on the Digital Frontier, in: Thomas Keenan (ed.), The End(s) of the Museum, Barcelona (Fondació Antoni Tápies) 1996, 67-80 (75)
35 Kittler 1996: 75
36 The Constant Association for Art and Media runs its Active Archives project since 2006. For a case study, see Geoff Cox / Nicolas Malevé / Michael Murtaugh, Archiving the Databody: Human and Nonhuman Agency in the Documents of Erkki Kurenniemi, in: Joasia Krysa / Jussi Parikka (eds.), Writing and Unwriting (Media) Art History. Erkki Kurenniemi in 2048, Cambridge, Mass. (MIT Press) 2015, 125-141

borrowed from communication engineering like informational entropy make more sense than the out-dated architectures of memory classification.

"Active archives"

Algorithmicized epistemology is experimenting with the computational *a priori* of data organization which is software.

In the case of Roberto Simanowski's edited book *Digital Humanities and Digital Media*, "[t]he cover image is a visualization of the book's text. Each interview was algorithmically assessed, paragraph by paragraph, for relevance to "politics", "culture", "aesthetics" and "literacy" and the result plotted as a streamgraph. All of the streamgraphs were overlayed to create a composite image of the book", created with Gensim and Matplotlib.[37]

Kaplan's "cliometric" *Venice Time Machine* is based on a graphic visualization of data structures from the millenium-old Venice city archive, exciting historical simulation environments which do not actually represent the past "as it really was" in terms of historical research; "instead, they foreground [...] experimentation, allowing new research questions to be asked and hypotheses to be tested using a wide range of variables. For instance, one may employ time-sliders to visualize when and where certain buildings came into existence [...]."[38]

The promise of animating the archive "stands for a series of strategies for launching that afterlife from the very moment of archival processing. This implies a user-centered approach to the construction of archives"[39] - which is actually a misunderstanding of the archive as legal institution. All of the sudden, in a rather political than simply conservative sense, a counter-strategic defense of the *archivium secretum* (the inaccessible archive) becomes attractive again, against the prevailent "social web" ideology of "open access" to big archival data, as expressed: "Techniques such as automated metadata generation, user-tagging, and crowd-sourcing must be employed to expedite availability for user communities. The user-centered - not document- or object-centered - archive must become the rule. Gone is the era of the archive as a Fort Knox."[40]

By digitizing the archivally protected, "read-only records", they can be algorithmically 're-discovered' anew without invading the material source, as applied by Actives Archives to Finnish artist-enginneer Erkki Kurenniemi's multi-

---

37 Legend to the cover image (David Ottina 2016 cc-by-sa), in: Roberto Simanowski (ed.), Digital Humanities and Digital Media: Conversations on Politics, Culture, Aesthetics and Literacy, London (OPEN HUMANITIES Pr.) 2016
38 Burdick et al. 2012: 44; see as well the diagrams in the chapter "Temporal Modeling", in: Johanna Drucker, SpecLab. Digital Aesthetics and Projects in Speculative Computing, Chicago / London (University of Chicago Press) 2009, 37-64
39 Anne Burdick / Johanna Drucker / Peter Lunenfeld / Todd Presner / Jeffrey Schnapp, Digital_Humanities, Cambridge, MA (MIT Press) 2012, 48
40 Burdick et al. 2012: 115

media memory, based on hundreds of hours of audio recordings, video diaries, photographs, preserving the objects of his everyday life, shooting 8mm films and digital videos. Separate the institutional archives from what the Brussels-based media art collective Constant, more in Foucault's sense, calls the "active archive" (Nicolas Malevé / Michael Murtaugh). "[T]o Constant, archives are understood as a collection" - that, rather *library* - "of material that is not merely readable but also writable and exectuable"[41] - which is exactly the contrary of the "dusty repository of fixed meanings" (126) which is in fact the *virtue* of the archival *katechon*.

Algorithmic analytics in the Digital Humanities sense has a transitive relation to the archive, when "the tools that are used for archiving can be registered as a part of the archive, [...] through the extensive use of / software repositories/archives such as Gitorious, with all scripts carefully documented for future modification and reuse"[42].

The experimental application of creative algorithms to a set of "big data" is a kind of software "carpenty" (Ian Bogost). Such a process-oriented ontology results in digital enunciations for which traditional hermeneutics did not pose the right question. By what Geoff Cox calls *The Speaking Code*[43],"the material is provided with the ability to 'speak' for itself" (134). Addressing the archive with creative algorithms uncovers the archive "for what is is: essentially, a collection of data"[44]. But a metonymic shift takes place when the attention is focused from the archival rules (implicit algorithm) to the archive-as-records.

Instead of mistaking the previous medium (archival taxonomy) the "content" of the archival digital medium, the real "message" of the new archive is algorithms.

[Contrary to the current claim for keeping the archive "open", for "instant" and "open access", though, there are arguments for preserving temporally sheltered records ("Sperrfrist") in *online archives*
- even in terms of computational science where there is the "protected mode" for embedded code.]

Software tools allow for engaging differently with structured and / or stochastically distributed sets of digital records. Digital taxonomy allows for the co-existence of different orders without violating the archivally protected files.

Active Archive algorithms produce random knowledge; inanimate objects are made to articulate.[45] In the case of the Kurenniemi estate (Central Art Archive

---

41 Geoff Cox / Nicolas Malevé / Michael Murtaugh, Archiving the Databody: Human and Nonhuman Agency in the Documents of Erkki Kurenniemi, in: Joasia Krysa / Jussi Parikka (eds.), Writing and Unwriting (Media) Art History. Erkki Kurenniemi in 2048, Cambridge, Mass. (MIT Press) 2015, 125-141 (125)
42 Cox et al. 2015: 126 f.; see http://gitorius.org/kurenniemi
43 Geoff Cox / Alex McLean, Speaking Code. Coding as Aesthetic and Political Expression, Cambridge, MA (MIT Press) 2012
44 Cox et al. 2015: 135
45 See e. g. http://kurenniemi.activearchives.org/spectrum

at the Finnish National Galery), the nonhuman approach[46] has been suggested by the artist himself: to be decoded in 2048.[47]

The self-organizing archive

The combination of both the human and electronic approach to textuality results in the "fuzzy archive" ("disorderly" order, similarities rather than exact differences, stochastic probabilities instead of rigid indexes and inventories) which works best when being combined with the "exact" archive (still needed for academic and other reasons) in a complementary way.

Algorithms are getting the archive in motion, which is the optimism of so-called Digital Humanities. Creative algorithms, such as for face recognition, have been developped for data surveillance first, such as by the NSA.

Either sound and images are still humanly or automatically tagged by textual metadata (which is the traditional archival inventory) as an organization of the archive as database. This belongs to the symbolic order of what is properly called the archive. An alternative approach does not produce metadata for the ordering of such records but sorts them from within, which is the signal-based approach. One can either tag an image by, e. g., the painter's name, or one can treat the same image as complex signal which allows for sorting it according to, e. g., colour values or shape detection within. The metadata approach belongs to the familiar archival symbolic regime, whereas the signal approach is truly oriented at the materiality of sound and images; more complicated when images are not recorded, e. g., as analog photography or electronic video signals but digitally sampled. This results in a symbolic regime in an even more fundamental sense and re-introduces the archival order; this fundamental archive of digital sound and images is strictly techno-mathematical and numerical, not metadata in the traditional sense which subjected sound and images to logocentristic key-terms expressed alphabetically

The self-organizing map (SOM) is the core concept of the Kohonen algorithm which represents one of the strongest models for the similatity-based "signal" approach. Media artist George Legrady, in his installation *Pockets full of Memories* (which was extended to the "social web" by his up-dated version Cell Tango) combined both approaches: the algorithmic sorting of objects in the media-archaeoloical way (as self-organizing map) on the one hand, and the subjective, personal tagging of objects by the individual participants (the human approach, focused on emotional semantics). This combination tries to get out the epistemologically "best" from both cultures.

"Hermeneutics" after Shannon

---

46 http://activearchives.org/wiki/Archiving_the_Data-body:_human_and_nonhuman_agency_in_the_documents_of_Kurenniemi
47 See Lars Bang Larsen, Erkki Kurenniemi. Einführung. dOCUMENTA (13): 100 Notizen – 100 Gedanken, No. 007, Ostfildern (Hatje Cantz) 2011

Media archaelogy unfolds implicit knowledge from within technological configurations, as a material equivalent to hermeneutics un-covering layers of hidden meanings in biblical texts.

Techno-logical hermeneutics (like "material semantics" applied to media art preservation[48]) is tracing the implicit knowledge within embodied signal processing and its circuitry diagrams. Inductive experimentation with knowledge-inviting (epistemoxenic) things is the media-scientific alternative to textual hermeneutics, such as Nam June Paik's magnetical distortions of the electronic TV image cathode ray tube.

For the academically trained *literary* scholar McLuhan, *Understanding Media* (1964) is not about interpretation, but uncovering the technologically induced message.

1985, one year after the *The Talking Heads* released their concert movie *Stop Making Sense* Schreiber's article called „Word-Engineering“ has been a manifesto about the state of literature in the age of signal processing and information theory: "Frequencies, amplitudes, rhythms, cell tissue, matrix connections, electrodes, multi-channel-recorders. There is no more singing, no more staking of heads, but only recording, because it is not about understanding anymore [weil es nicht ums Verstehen geht]. Below understanding there is transmission and recording."[49]

The basic techno-hermeneutic operation is not to transform a chain of signifiers into meaning anymore but to identify a message "in the presence of noise" (Shannon), such as by signal filtering *alias* communication "intelligence". While in hard-wired technology, the syntax of the signifier replaces the "semantic", in coded symbol programming stochastics replaces the individualizing reading. According to Claude Shannon in the introduction to his *Mathematical Theory of Communication*, "[t]he semantic aspects of communication are irrelevant to the engineering aspects.”[50] In its deterritorialized form information knows only engineering aspects such as discrete sets of signals, frequencies, entropy, redundancy, channel coding.

Epistemological curiosity rather than hermeneutic "understanding" drives media archaeology; electro-physical action can not be "understood". Still, all technology is nature (physics) encoded by cultural knowledge; therefore humanities asks different questions to technology than engineering and informatics do.

Still, the language of technology is diagrammatic circuitry and mathematical logics. The re-turn of the text within computing is alpha-numeric, not semantic.

Conventional "hermeneutics" belongs to the (holy) scriptural regime (the symbolic order of the alphabet), as opposed to signal analysis (such as spectrograms). Media "philology" refers to signals for "analog" technologies, but returns to symbols in the case of digital code.

---

48 See Monika Wagner, xxx
49 Jens Schreiber, Das Symptom des Schreibens, xxx
50 Shannon/Weaver, xxx

Media archaeology is a heuristic, methodological *temporary* suspense from the hermeneutic imperative. After that "epoché", is it therefore time for the reboot of a hermeneutics in the sense of understanding non-human agencies ("authors") such as algorithms, *algorithmic hermeneutics* as exercized by Active Archive (Constant)? Rebooting hermeneutics leads to historization; media archaeology rather copes with the hon-historical temporality of media operativity.

Algorithmic hermeneutics

The core of the techno-mathematical media definition in terms of communication engineering (Shannon 1948) is what happens in the channel of signal transmission, the actual "medium" (Shannon); within that *delta-t*, data are suspended from cultural semantics. Already in the technologically coded telegraphy differed from the written alphabet.

In algorithmic analysis instead of hermeneutic understanding, audio recordings from the the past are not just cultural objects any more, but become items in an experimental laboratory of "archived presence". Semantic emphasis can be identified as a function of tonal pitch in the recorded voice, just as Max Planck - in a recording from 1939 in the Lautarchiv collection "Stimmen berühmter Persönlichkeiten"[51] raises (in German: "erhebt") his voice with the very German word "erhebt" itself, and lowers it with rhetorical skill at the end of his phrase in the last word "Gelehrten" (scholars). The techno-mathematical analysis of intonation, performed by Nikita Braguinski with the software Sonic Visualizer, reveals Planck's application of quasi-musical phrasing and thereby bridges the gap between semantics and affect:

Fig. POWELL-HEXAMETER-SPECTRUM.jpg

Humans almost irresistably interface to images in an iconologic way and to texts in a hermeneutic way. But there is a kind of knowledge instead which can be uncovered from within the visual, acoustic or textual endo-data: entering the digitized record itself (data-immersion); the media-archaeological analysis is performed by algorithmic machines of information processing better than by human perception. Such *informatized* organization of knowledge generates diagrams (the Deleuzean intepretation of the Foucaultdean *archive*) - infomapping. Occidental culture is still dominated by semiotically iconic, musically semantic, of literally hermeneutic ways of seeing, hearing, reading; the twenty first century, though, allows for genuinely computer-generated information aesthetics which is closer to processual diagrams than to figurative phenomena within the audio-visual (or textual) regime.

[In an exemplary way, the book cover of Friedrich Knilli's analysis of the radio play *Das Hörspiel. Mittel und Möglichkeiten eines totalen Schallspiels* (Stuttgart 1961) does not display an allegory of communication but by technical drawings: psychoacoustic parameters and diagrams.]

---

51 See Web site of the Lautarchiv = B8-29 Max Planck

Diagrams eventually enable unprecedented types of the generative archive (rather than representations) which demands description as "archaeography" which is the indication of non-discursive media tempor(e)alities: its governing principles, archaic essentials.

Text criticism has been a core operation of reading sources from antiquity in the age of Humanism. In times of algorithmicized Humanities, critical code studies is necessary - though beyond mere reading like in print culture, since executing software (different from its written algorithm) is a dynamic object. As expressed by George Dyson: 'You can't predict how software will behave by inspecting it'; the only way to plausibly analyse software is 'to actually run it."[52]

Media *archivology*: Kittler's case

The driving mind who once radicalized Foucault's archaeology and McLuhan's media theory into media science, late Friedrich Kittler, has become a memory address himself, with his written papers, self-designed electronical toys and experimental software code now being located at the German Literature Archive in Marbach. To answer the question in which way computing once shaped Kittler's research in the 1990, media archivology is required.

Even if Friedrich Kittler has become an archival subject himself, let us not historicize his legacy. The German Literature Archive at Marbach instead has creates a software for navigating his estate, the "Indexer", which non-hermeneutically, rather algorithmically searches for patterns in Kittler's conventional writings, as well as in his electronic diagrams and source codes.[53]

Being a specially designed search engine, the Indexer, after having copied the hard drive and storage discs of Kittler's computer in sector images, allows for the sub-hermeneutical, chronologically simplest and statistically most reliable search option of looking for modification times of his digital files - a dynamic parameter rather than the historicist focus on straightforward origins (creation time). A media-philological, "forensic" warning: Even if the Indexer offers a search option for creation times, these are not historically reliable as they rather depend on the inner time (*eigenzeit*) of the storage devices themselves.[54]

[Media *archivology*, in an analytical sense, refers to the archive of computing itself - with *l'archive*, once more in Foucault's sense, naming less the institution

---

52 http://www.wired.com/magazine/2012/02/ff_dysonqa/all/1 (accessed September 27, 2012); same argument by Tobias Matzner, Grasping the ethics and politics of algorithms, https://medium.com/@t_matzner/grasping-the-ethics-and-politics-of-algorithms-c2932804fa9d#.i1oymdxrg, accessed February 2017

53 See Susanne Holl, Friedrich Kittler and the Digital Humanities: Forerunner, Godfather, Object of Research. An Indexer Model Research [= Friedrich Kittler's Digital Legacy. Part II], demnächst in: Digital Humanities Quaterly

54 See Susanne Holl, Friedrich Kittler and the Digital Humanities: Forerunner, Godfather, Object of Research. An Indexer Model Research, in: Digital Humanities Quarterly (2016), note 2

for record memory which in French would always be expressed in the plural: *les archives*. *L'archive des médias* rather refers to the material and logical conditions of possibility for any kind of technical articulation. Methodologically, the approach *from within* technology expresses the media-archaeological, that is: non-human point of view, distant from the cognitive or bodily perception of "media" which humans experience from interfaces like the computer screen. For such an investigation, media archaeology necessarily departs from the familiar historical research. Radical media archaeology is not simply another variance of historiography but an alternative way of dealing with temporal evidence resulting from times past; it is rather radical historicism. A term like "historical media archaeology" (as coined by Kittler) therefore is an undecided *oxymoron*.]

"Digital humanities" *avant la lettre*? "New Archaeology" and Peirce's archaeological semiotics

Challenges to traditional scholarship and new opportunities derive from so-called "digital humanities".[55] Computer-augmented algorithmic analysis, as applied in classical archaeology among the first disciplines within the "humanities" department, leads to database aesthetics instead of narrative (as expressed by Lev Manovich in his *Language of New Media*), to pattern-matching *versus* hermeneutics, to new forms of search with stochastic data analysis instead of traditional statistics (which still refers to the classical archive).

"The mere use of digital tools for the purpose of humanistic research and communication does not qualify as Digital Humanities. Nor [...] is Digital Humanities to be understood as the study of digital artifacts, new media, or contemporary culture in place of physical artifacts, old media, or historical culture. [...] Digital Humanities understands its object of study as the entire human record, from prehistory to the present. This is why fields such as classics and archaeology have played just as important a role in the development of Digital Humanities as has, for example, media studies."[56] In media archaeology respectively media philology, both fields converge.

It is not by coincidence but by epistemological necessity that archaeology has been among the first disciplines within the humanities to employ computing and statistical techniques ("Digital Humanities" *avant la lettre*), but:

"Even the beneficial contribution of such 'hard' science such as radio carbon determinations of date or ground penetrating radar to archaeological interpretation, rely on operators having a close empathy with archaeological material, the context of discovery and the role of post-depositional processes"[57]

---

55 See David M. Berry (ed.), Understanding digital humanities, xxx

56 Burdick et al. 2013: 122; see further Susan Hockey, A Guide to Computer Applications in the Humanities (London, 1980)

57 E-mail Peter Rauxloh (Information Strategy Manager, Museum of London), July 2002

The techno-mathematical application of stochastic analysis in "computational" humanities is not simply a special method in classical archaeology (as expressed in journals like the Italian *Archeologia e calcolatori*), but can be identified upside down as the *archaeological* element in mathematics itself

Foucault's *Archaeology of Knowledge*, should not be taken metaqphorically in the sense of an archaeological dig, but at its implicit mathematical face value, which is the propositional logic of enunciations.[58]

"Post-processual" archaeologist Hodder stays close to Charles S. Peirce, in not reducing semiotics to (de)coding, but semiosis as agency; the past can be "read" exactly because material culture is not text: text is just a metaphor, not an analogy for material culture.[59]

Manuscripts, for Peirce, are not immediately "documents", esp. as long as they are not yet deciphered. For ancient history, even manuscripts are first of all: monuments.[60]

Peirce started from the assumption of the materiality of any sign.

In the chapter "Über Methodenprobleme der Klassifikation"[61] of his *Minutiöse Logik* (1902), Peirce explicitely refers to Flinders Petrie, founder of pre-historical archaeology of Egypt: his system of sequential chronology as quantitative archaeology (genealogy of ceramics); as entries in lists they constitue series. On paper stripes entries of pre-dynastic ceramics, numbers: relative seriation.[62]

[This coincides with Egyptian mathematics itself which did not apply a calculus but lists: Results from were listed, especially complex ratios.[63]]

---

58 See Martin Kusch, xxx
59 Hodder / Hutson 2003: 169
60 Charles S. Peirce, in his unpublished *A History of Science*, chap. "The Logic of Drawing History from Ancient Documents, especially from Testimonies" (1901), 146
61 In: Charles S. Peirce, Semiotische Schriften, Bd. 1
62 See Franziska Lang, Klassische Archäologie, Tübingen / Basel 2002, 139
63 James Ritter, Jedem seine Wahrheit. Die Mathematiken in Ägypten und Mesopotamien, in: Michel Serres (ed.), Elemente einer Geschichte der wisenschaften, Frankfurt/M. 1995, 89