INTERRUPTING DIGITIZATION AND THINKING ABOUT TEXT OR DIGITIZATION AND THE FORM OF DIGITAL TEXT [1]

di Geoffrey Rockwell

L'entusiasmo nei confronti della digitalizzazione nasconde la complessità dei processi interessati e il significato delle trasformazioni culturali dei nostri giorni. È importante approfondire i diversi modi in cui solitamente ci si interroga sul digitale, generalmente considerato come una interruzione del flusso continuo di informazioni in intervalli discreti. Ugualmente rilevante è chiarire come si possano usare i documenti digitali e quali differenze esistano rispetto agli stessi dati in formato analogico. La riflessione sul modo di concepire la differenza tra questi due modi di rappresentazione porta a comprendere che la maggior parte delle informazioni, nel mondo digitale, sta nel mezzo che si usa per codificare e rappresentare le informazioni. Quando si approfondisce il concetto astratto di digitalizzazione, in realtà si impara molto di più sulla nostra cultura che sui file dei quali si pensa di parlare. Da questo punto di vista può essere utile anche l'interruzione del funzionamento del nostro computer che può aiutare a comprendere che la macchina va considerata innanzitutto come una macchina.

Enthusiasm for digitization hides the complexity of processes involved and the significance of the cultural changes achieved in our time. It's worth examining again the ways in which we usually see the digital, often considered as an interruption of the continuous flow into discrete intervals. Equally important is to make clear how one can use digital documents and what is lost when digitizing the analog representation. Thinking about the way we conceive the difference between these two types of representation leads us to understand that in the

[1] This paper was first presented to the Digital Humanities seminar at the University of Virginia in 2002, an opportunity for which I thank John Unsworth and Johanna Drucker. The central idea about the discourse of the digital emerged in a conversation with Worthy Martin. A version was also presented at the University of Georgia, for which I thank Steve Ramsay.

digital world there is information encoded in the medium we use to codify and to represent content. When we consider the abstract concept of digitization we actually learn much more about our media and culture than about the files of which we believe we are speaking. Therefore the interruption of the computer can help us understand the machine as machine.

A machine may be defined as a system of interruptions or breaks (coupures) … Every machine, in the first place, is related to a continual material flow (hylè) that it cuts into.
Deleuze and Guattari, *Anti-Oedipus* (p. 36)

One of the *memes* of new media is that the form of communication determines the content. As McLuhan puts it *the medium is the message* [2], and therefore, as we digitize the evidence of human culture from the Roman forum to Hamlet we inaugurate not just a new edition of our knowledge, but a new knowing and with it a new way of thinking. This paper will not engage the question of technological determinism, instead it will assume that the enthusiasts are right and ask then *what is digitization*? or *what is the message of the digital form*? Asking such questions is an interruption in the rush to digitize everything; imagine the scanner has broken down for a moment letting us pause and ask if we really understand the digital, if we understand what is gained and lost, and if we understand the possibilities before us or how we are constrained.

In particular this paper will argue that in digitization there is first a sacrifice of information that can be attributed to the *coupures*: the splitting and measuring, sampling and coding. From there the paper will try to do something difficult, which is to talk about what is hidden by the enthusiasm for digitization. This is difficult, not just because I am struggling like most to master the intersection between the techni-

[2] M. McLuhan, *Understanding Media. The Extensions of Man*, Routledge, London 1964, p. 23: *In a culture like ours, long accustomed to splitting and dividing all things as a means of control, it is sometimes a bit of a shock to be reminded that, in operational and practical fact, the medium is the message*. The phrase is the title of the first chapter and this quote is the first line of that chapter. Note how the phrase is, in the original, given in the context of splitting and control.

cal and what can be said, but because our discourse itself has evolved to collapse distinctions around the digital. In other words it is hard to make a point about the language of digitization with the discourse of the digital.

1. WHAT IS THE DIGITAL?

There are many ways to ask about the digital. David Bell, in *An Introduction to Cybercultures* identifies three general approaches:

a) *Material stories*, stories about the hardware, in our case stories about signal processing and transducers;
b) *Symbolic stories*, stories about the software and what you can do with the flow of 0s and 1s;
c) *Experiential stories*, stories about what works and how to do things like digitizing a manuscript.

This paper concentrates mostly on the interface between the second and third sort of stories, though part of the point is how different stories are misleading when conclusions are drawn from abstractions in one story and applied at another level.

2. THE STANDARD STORY

Here is a summary from Lev Manovich, *The Language of New Media*, which is typical of how digitization is summarized so we can get onto more important things:

> *Digitization consists of two steps: sampling and quantization. First, data is* sampled, *most often at regular intervals, such as the grid of pixels used to represent a digital image. The frequency of sampling is referred to as* resolution. *Sampling turns continuous data into* discrete *data, that is, data occurring in distinct units … Second, each sample is* quantified, *that is, it is assigned a numerical value drawn from a defined range …* [3]

[3] L. Manovich, *The Language of New Media*, MIT Press, Cambridge - Massa-

Manovich identifies the standard two processes to digitization, and corresponding to them, two properties of the digital as symbolic. The first process is the sampling, or the interruption of the continuous flow into discrete intervals, and the second is the quantification of each interval so that it can be represented as a numerical value in binary code. I have two quibbles with this standard view.

My first quibble is that we have to expand, following Deleuze and Guattari [4], our view of what is digitization to include a process, actually more of an orientation, that precedes the sampling of the flow. To digress, the digital is often described by contrasting it to the analog, or the continuous. Digitization is described in a relationship with the analog or the continuous, a relationship of translating the analog or transducing it. It therefore follows that the first step of digitization is actually a preparation or turning-towards something in the world you chose to digitize. This is an orientation that comes before the interruption of the flow. The flow is not a given, nor is the machine. We treat something as a machine and something else as the analog to digitize, often without thinking. After all, when we turn to, pick up and use tools we don't think about the tool so much as the task. To use Heidegger's terms [5], something is a tool when it is treated as ready-at-hand in a totality of technology. The world is likewise treated as something that which could be interrupted. For example, a sound digitizing machine is oriented towards certain physical phenomena and by its very structure treats the confusion as a flow in time of a selected phenomena, sound. Therefore, my first qualification of the standard story is that digitization is first of all a technological orientation towards the world into a flow that could be interrupted at all. And that is also the first lost – the loss of the world as something other than that which can be digitized – a whole that when grasped by the technology of digitization is no longer a world, but something seen as analog.

---

chusetts 2001, p. 28.

[4] G. Deleuze and F. Guattari, *Anti-Oedipus. Capitalism and Schizophrenia*, trans. R. Hurley, M. Seem and H.R. Lane, Continuum, London 2004.

[5] M. Heidegger, *The Question Concerning Technology*, in Id., *Basic Writings*, ed. D.F. Krell, Harper & Row, New York 1977, pp. 283-318.

The second qualification I have with Manovich has to do with the second step in his definition of digitization: quantification. Strictly speaking quantization is not the same as quantification and, more importantly, digitization does not quantify, it encodes [6]. For each interval sampled in the flow a code is assigned from a set of available codes. Some of these codes represent numbers for which there is a larger number or a smaller one, in other cases what is assigned is simply a code from a code-table representing colours or letters. That 00000101 can represent the number 5 in decimal notation doesn't mean that a colour assigned to a pixel that is coded with that sequence of bits is a quantity; all that happens in the digitization is that the code for the colour closest to the sample is assigned. Another way of thinking about this is to think of a keyboard as a simple digitizer of gestures, specifically the gestures of our fingers we call typing. It is a motion capture device for specific finger movements. But what it does not do is assign a quantity to the different gestures it is oriented towards, it assigns a code that corresponds to the abstract letter you wanted.

In sum we can identify the following as the characteristics of the digital as symbolic:

Digital	Explanation
Discrete	The digital is made of discrete digits. Each digit can have a limited number of settings.
Binary	The digital has two possible settings for each digit.
Code	The digital is not essentially numeric. From the binary one can build collection of digits like a byte that represent a code word that is meaningless by itself.
Code-Table	The codes only make sense in a context where they can be deciphered by the computer.
Computing	Codes are built on codes. ASCII is a code for representing alphanumeric characters. On it can be built other codes.

[6] To be fair, Manovich was probably simplifying as the difference between *quantization* and *quantification* is not relevant to his argument.

3. What are the advantages of digitization?

A second way to ask about digitization is to look at the experiential stories, the stories about what you can do with the digital and how great digitization is. Most discussions in new media texts glosses over what happens at the material level to jump to all the benefits and convince you that all the work of digitizing is worth the interruption. The focus is switched to the potential experiences of the digital, what you could be doing with it. Here are some of the outcomes touted, for example, by the techno-literati like Nicholas Negroponte in *Being Digital*.

1) Digital data is easy to discipline with *error correction*. This might strike you as a strange benefit to start with, but it underlies many of the other benefits for Negroponte.
2) For example, error correction allows us to *reliably duplicate* the digital without loss of information. With analog information there is typically a loss of information with every copy generation so that no photocopy matches the original. With the digital it is questionable as to which is the original and which is the copy thanks to error correction that preserves the exact symbolic sequence. Is it the original that is moved to the new spot and the copy left behind, or the copy moved? Does it make sense to ask this question about the digital? No, the story says.
   a) Therefore, there is *no authentic original* to digital media, there are only conditions for representation over time.
   b) Also, and related to 2. a, there is *no material presence* to digital media only the representation through an output device.
3) Digital media, thanks to error correction, can be *transmitted without loss* of information. And, because what matters in digital media transmission is the sequence of bits not the atoms, digital media can be transmitted at the speed of light, at least when using optical transmission technology.
   a) Thus digital works have *no necessary physical location*.
4) Digital data can *represent anything* that can be sampled and coded. Thus text, images, audio, video, and 3D spaces can be represented in digital code on a computer. For that matter, anything that can be

formally described can be represented digitally, including gestures and processes, continuous or not.

5) Digital media can be *manipulated using computer processes*. Error correction is one of many processes, though a guarantee against corruption. Because the digital is code (where each discrete digit can be combined into discrete code words corresponding to symbols or quantized samples), we can use cyphering processes to manipulate the codes substituting them for other codes.
   a) It could be argued that the processes or programs, which are themselves stored as data, are *yet another medium* that can be represented in digital form, a medium called programs that has a special relationship to the machine and the other digital media. In the symbolic story the computer is a machine of layers upon layers of symbolic systems, some of which are used for manipulating the others.
6) Digital media can be combined into new digital media or *multimedia* works. Since all digital media are sequences of bits they can be combined into meta-media works. The digital not only subsumes other media, it can weave them together. In particular we can combine media that in analog form are incompatible, especially time-dependent media like audio and video with time-independent media like text and images.
7) Finally, digitized media can be *converted to other media*. A text can be transcoded into audio and vice versa. Such conversion is but one way you can manipulate the digital, but one that leads Negroponte to the conclusion that the digital message has no particular medium in the traditional sense. The digital can be prepared so that it can be represented in different forms. The same bits can be synthesized into voice or output as letters on a screen. We might ask what the message is of a new media with no predetermined medium?
   a) Thus we can say that *digital media has no form* other than their being digital and that it is not, except as a matter of convention, comparable to any particular traditional media. It is all media and none. For example, a digital text is not a digital text until rendered that way. In its essential form it is code.

These observations about the digital are connected. One could derive them all from the discrete and coded nature of digital information in

the symbolic story. Because digital information is coded one can represent other media through processes of encoding; that is what codes are after all, enciphered messages. Because the digital is code, you can process it using coding processes; enciphering and deciphering are, after all, just reliable processes for changing messages. Because the digital is made up of discrete codes, it is easy to store, access, and copy without loss of information thanks to redundancy and error checking processes. Because all media are ultimately represented in the same coded form, 0s and 1s, you can easily combine or translate (*transcode*) what to us appear as different media, a miracle difficult to perform with analog media technologies.

To digress, we might ask why *binary*? Binary digits seem the most primitive and inefficient form of code; one where each digit can have only two settings and anything meaningful has to be generated from longer code words like bytes for ASCII. One answer is to switch to the material story and point to the engineering convenience of the binary which has proven the most reliable implementation of inscription for use in the everyday noisy environments we live in that are full of dust and children with sticky fingers. To assign only two possible settings to each digit minimizes the chances of errors in the reading and writing of code in material environments.

4. The Loss of the Authentic

From this story of benefits we can immediately identify a second loss that is obvious in the arts and to those concerned with intellectual property and copyright; it is the loss of the material presence or the loss of the authentic original. The question of the authentic original is not new to digital media. I take this to be the point Walter Benjamin makes in *Art in the Age of Mechanical Reproduction* [7] and a point that applies even more to digital media than mechanically reproducible media like

[7] W. Benjamin, *The Work of Art in the Age of Mechanical Reproduction*, trans. H. Zohn, *Illuminations*, ed. H. Arendt, Schocken Books, New York 1968.

photography. With digital media we do not have unmediated access to the digital file, it is always mediated by the computer, which translates it from the form it is stored to an output device like a screen that we can see [8]. Thus every digital representation that we perceive is a copy or rendering that has been transmitted and translated from one physical form to another. With nothing but copies, there is no authentic work.

An example of this is browsing a digital text like a webpage; every webpage we see has been copied and transmitted to be viewed, even if off your hard drive. It has then been rendered or interpreted by the browser for the screen. There is no original site for viewing, no place for viewing that has a special distance from the work. What you see is what you get, not where you get it. The digital is potentially everywhere at the same time. When a work of art is accessed through reproductions (photographic prints) the work, according to Benjamin, loses its *aura* which is its unique place in time and space. With a digital work there is no unique location or place in a tradition of ownership that can make a work authentic and its copy not.

The loss of authenticity when described as a loss of physical presence in space and time leads to a change in how we treat digital media. Without a privileged physical presence there is nothing to own the way you can own a painting which, of course, complicates copyright, but that is another story. Instead there is only access, which can be licensed. Further, there is no market for unique works the way there is for art because there is no unique work. For that matter, there is really no distinction between copies and originals because a copy is identical logically to the original. Which instance would be the original? Thus we are tempted to say that there are no originals and copies, there are really only conditions for reliable representation [9].

---

[8] About these themes see also I.V. Kerlow, and J. Rosebush, *Computer Graphics for Designers & Artists*, 2nd ed., Van Nostrand Reinhold, New York 1994; W.J. Mitchell, *The Reconfigured Eye. Visual Truth in the Post-Photographic Era*, MIT Press, Cambridge - Massachusetts 1992.

[9] One of the conclusions of the *InterPARES 1* project that looked into the authenticity (in the archival sense) of electronic records was that *It is not possible to preserve an electronic record, it is only possible to preserve the ability to reproduce it*. (From a

5. Digitization as Acquisition

Digitization is not only discussed theoretically, but is also treated in *how-to* learning materials that typically concentrate on the process of acquisition; how you can acquire a digital text by scanning and optical character recognition. When we teach digitizing we spend a lot of time on the practicalities of image scanning, text recognition, audio digitizing or video digitizing. Moreover, in such contexts we distinguish the digital works resulting from digitization by scanning from the analog from those that are *born digital*, namely those texts that typed in directly. Needless to say, the practice of digitization is very different if you are digitizing an analog work as opposed to creating one from scratch on the computer. To better understand this we can review how digital information is created or acquired:

1) *Born Digital.* Digital media can be created directly on the computer using an entry or input system like a keyboard or mouse. Thus the email you type would be considered a born digital text.
2) *Digitized from Media.* We can use media specific input systems like image scanners, audio digitizers, and video digitizers to create digital representations of existing works already in some media. Thus a page image scanned from a manuscript can be called a digital text.
3) *Communicated and Remixed.* We can also select and acquire digital media already in digital form off the Internet or from removable media like a cd-rom. Thus an email message sent to me is a digital text acquired. This last form of acquisition wouldn't normally be called digitization, but it is the most common way we acquire digital information.

Digitization usually refers to the second practice since we don't usually think of keyboards as a way of digitizing. The born digital is not really digitized where we are trying to represent a work in an analog media. The point is we create a native digital work using specialized input mechanisms optimized for creation on the computer. Nonetheless, I

---

presentation given by Luciana Duranti in February, 2002). In other words you can't determine if you have the original file, but you could determine whether the representation you have was the result of a reliable process of transmission and maintenance. See www.interpares.org.

would tend to collapse the first two categories because we can think of keyboards and mice as haptic digitizers meant to acquire specialized movements (stroking keys) from analog sources like humans. This allows me to contrast two forms of digital acquisition not usually contrasted:

- digitizing non-digital media, events and gestures, and
- downloading already digitized media.

Including downloading as a means of acquisition may not seem important, but Lev Manovich rightly, to my mind, points out that *selecting* already prepared objects is a defining process in the practice of computing. In the third chapter of *The Language of New Media* he identifies three operations as characteristic of new media: selection, compositing, and teleaction. At first this choice seems bizarre, but he makes a good case regarding how so much of what we do on computers is to select prefabricated digital objects to process and combine into new works (at a distance). We should, however, beware that this hides important differences in practice and skill. The practices of digitizing existing media involve all sorts of decisions about resolution and format. The practices of selection and remixing involve issues around copyright and practices of browsing.

It is also worth pointing out that digital media can be categorized into two types of digital representations connected again to practice. To borrow the language of digital images, they can classified as *raster* representations that are scanned or *vector* representations that are drawn (or born digital). The first type we consider to be *digitized* media proper, where an analog source was sampled and quantized, though it is possible to create such bit-map images from scratch on the computer. The second are representations where the media is represented by structured objects of different levels of abstraction. The first type tend to produce more faithful representations of analog sources, the second more efficient files that save space. The second are also often the more useful representations in that abstractions are closer to what we want to manipulate, but they are representations that demand more processing to output. A text represented on the computer as a page of black and white dots (pixels) is harder to manipulate than one represented by the type of objects we think of a constitutive of text, namely letters

in a sequence. It can be argued that these categories of digital media can be applied to all media, not just images: text (digitized page / marked-up text), audio (digitized audio / MIDI), and moving images (digital video / animation).

This broad generalization about the ways we can talk about the digital should remind us of an important aspect of the digital: that binary digital codes can be used to represent higher, or more abstract, data structures. As mentioned before, we can build codes built on codes starting with the binary digit. The more complex codes can represent more complex abstractions of what we think is important. If ASCII is a code that assigns bytes to alphanumeric characters, HTML is a code built on ASCII that can represent a formatted page. These higher order formats can be open or proprietary, they can involve lossy or lossless compression, and they can be used to represent different features in the phenomena being represented. In principle, however, any one format can be translated into any other format, though the results may vary [10]. Each code system involves choices about what is important and what isn't. In the representation of a page is it important to get the sequence of letters or the arrangement of ink on a page? These choices show and hide features of the world.

Nelson Goodman in *Languages of Art* [11] would point out that the absence of an authentic original and of a unique physical presence is due to coding in the first place; the use of formal languages in which we encode digital media. The point of formal languages with discrete symbols from a predetermined symbol set is that they are unambiguous and can represent information logically so that it can be reliably processed. In the formal languages it is clear what is essential (what is coded) and what is not (what is discarded or doesn't fit) in the digitization. Formal languages like binary digital code carry their own messages as they are technologies to make possible the unambiguous identification of what is considered part of the work for purposes of trans-

[10] Obviously someone working with digital media needs to learn about the different formats and there are a wealth of online and print resources available. A good starting point is the *Theoretical Foundations of Multimedia.*

[11] N. Goodman, *Languages of Art. An Approach to a Theory of Symbols*, Bobbs-Merrill, New York 1968.

mission and representation and what is not. The choice precedes any particular instantiation because it lies in the definition of the formal language and the choice between formal languages; this is an important site of choice and loss. A technology like HTML hides in plain sight a view of what a text is, that is a formatted sequence of characters (and not, for example, a designed page such as would be represented in PDF).

What is not important at this symbolic level is the material. The fact that a digital file is inscribed on a cd-rom with bumps and pits is not essential to the digital file which is conceived of as a sequence of binary codes that can be inscribed any way the system wants as long as it can be read back. The irrelevance of the physical form is a property of representation through formal languages as a means of communication, a property which enables the infinite and perfect reproducibility of information across time and space, at least until it breaks down. The danger is that we forget the limitations of formal languages.

To return to digitization, what is interesting about this process of digitizing from the perspective of practice is that:

1) When we digitize we are forced to choose what to digitize and what to not input. We may not understand what we are choosing, but the process of digitization always involves interpretation as to what it is that is to be represented digitally and thus the digital representation always reflects, in part, the knowledge (or ignorance) of the digitizer. The choice is not just what object to digitize but what aspect of the phenomenon to digitize: do you want an image of the page or the text as a sequence of letters and punctuation? The choices made during digitization are the first place where knowledge informs the digital and these choices constrain any further enrichment of a digital representation. As the old saying goes *garbage in, garbage out.* What you digitize is all you get.
2) The tools of digitization also constrain our choices. The absence of accented characters on my keyboard makes it difficult for me to input French, though there are work-arounds. The technologies make assumptions about what we want and thus guide us in certain usually unexamined ways.

3) When we digitize material artifacts we always lose information. Analog works have an indeterminate amount of information; when we sample and code we are always discarding the information in between samples and we are reducing the possible values of a sample to a finite set from which to chose a particular value for the sample and therefore loosing the intermediate values 12.
4) When we digitize we are creating a coded representation of a work. The coded nature of the digital is what makes it amenable to computer processing. The codes are chosen from a predefined codebook so we lose those values not describable in that formal language. If you chose to assign only one byte to each pixel you can only have one of the 256 colours available in the chosen lookup table. If you assign an ascii code to each letter you lose the shape of the letter on the page.
5) To expand on this point, the most interesting aspect of the process of digitization is that much of the knowledge in a digital representation is not actually found in the digitization itself, but in the technologies (*transducers*) that convert analog signals to digital data and back. Both sampling and coding are dependent on predefined lookup tables, codebooks, and hardware implementations. This knowledge is needed in the system to represent the digital data back through an output device in a form recognizable to a user as a representation of the original. You don't get a reliable representation from the digital file itself, but from the file as performed by a system. Thus much of the knowledge in the digital is in the cluster of technologies that make it possible to digitize and represent media through output devices. The file contains only codes, the deciphering of which takes a system that has wired into it knowledge about

[12] Worthy Martin has argued that in principle it should be possible to represent the atomic (or subatomic) structure of an analog work and therefore have no loss of information. While it might be possible scientifically, such science fiction is not what we do in real digitizing. Further, there is a problem of observation: we can only digitize what can be observed and measured. No matter how fine the digitization there is no way of knowing for sure if you have not missed something unless you are able to make an even finer measurement to compare the digitized version to. Thus you can only know what you have lost; you cannot know for sure that you have lost nothing.

outputting. For that matter it also takes trained operators who can get the system to do the representation. With training we can make informed choices among systems, but few of us can develop new systems where existing ones are inadequate; what if you wanted to capture the feel of paper with coffee stains?

These points are worth dwelling on. Let us take the digitization of a page of poetry. First, there is interpretation on the part of the person doing the scanning when they choose which page and when they choose digitization settings like the resolution at which to scan (the DPI or dots-per-inch), colour depth to scan at, and the format to save the file in. These choices are one way in which human knowledge enhances (and constrains) digital representations. Part of that choice is a choice of what not to scan or what measurable information to discard.

Second, the information in the codes comes not from the values of the samples themselves but from the relationship of the values and the possibilities for representation when these values are processed by a transducer that has access to the relevant codebook. It doesn't really matter if a green point is represented by one code or another, what matters is that green is represented by a different code than blue and that there is a codebook that is used consistently for both digitizing and output. The same is true for text. It doesn't matter what code point is assigned in Unicode to the *latin letter a*, what matters is that it is a unique and standardized code point. The digital file is a cipher that, with the appropriate decoding system, can be displayed reliably as something recognizable as a digital representation. The digital image file does not stand by itself – it is not the digital representation as we see it – it is a cipher that can be translated into a representation that we can see [13]. That it is in a code unreadable by humans is its versatility for logical machines that can decypher it in many ways. Its encryption is its potential for change. Another way to put this is that what is stored

[13] It is interesting that E.H. Gombrich, *Art and Illusion. A Study in the Psychology of Pictorial Representation*, Princeton University Press, Princeton N.J. 1969, makes a similar point about painting. In the first chapter – *From Light into Paint* – he quotes Winston Churchill (of all people), *The canvas receives a message dispatched usually a few second before from the natural object. But it has come through a post office en route. It has been transmitted in code* (p. 39).

as a file is not a representation of the analog, but a set of differences that the computer can decipher so that we can make sense when seeing.

An important point to be made about this, which has to do with loss, is that digitization is dependent on a culture of computing and computer support. This is what I take Heidegger to be talking about when he talks about the totality of technology. Tools like computers are tools because they are in a totality of supporting technologies which in turn are understood this way in our technological age. The ciphers we exchange and treat as digital representations can only stand in for texts if we have available systems at hand that can re-present them to us as readable texts. For these systems to work we need an economy that creates them, fixes them, and so on. Much of the knowledge that we think of as being a text on the screen is not encoded in the file but is in the hardware, operating system, software, and culture of computing.

6. Analog and Digital

Let us return to the analog. I have so far been using the word *analog* in contrast to the *digital* which is how we generally use it. This distortion of the term is a use that has emerged in the discourse of computing for want of a better term; the discourse of digitization needs another into which to bundle everything else which is *not digital*. However, this distinction of the digital and analog does not hold when examined closely. The term analog comes from media studies to describe media where there is a physical analogy between the signal recorded and the recording. Thus the shape of the groove of a vinyl record is physically analogous to the shape of the electronic signal which in turn is analogous to the sound wave, at least as sound wave are represented to us on oscilloscopes. The analogy, needless to say is rather tenuous. Most of us don't see in the groove of a record anything like what we hear which raises questions about the status of the visual wave as some sort of real representation of sound. Digital media, in this story of contrast,

are different because they sample and code the signal representing the signal in a fashion that has no physical resemblance to the original. Without the codebook and hints about the format you could not reconstruct the original signal, which you might be able to do with an analog representation. This too is a tenuous hypothesis, but one that points to a difference in the sophistication of the engineering. It would be more accurate to say that it is a lot easier to build a technology to render analog works than digital ones.

A further point to be made about the analog is that the term is usually applied to rhetorical artifacts that need technological mediation to be accessed. A vinyl record needs a record player, an audio cassette needs a cassette player, but a book needs «only» a reader. Thus the analog properly is only a subset of rhetorical technologies which are not digital.

The problem with using the analog / digital distinction is that it hides the distinction that we want, that between digital representation and other types of representation, and it hides it behind the uninteresting observation of analogy in certain media. The distinction once accepted hides the possibility that many traditional media may in fact have more in common with the digital than with each other. Another way of saying this is that the distinction encodes as an either / or proposition what is really a more complex and continuous phenomenon. As mentioned above, all sorts of non-digital works are not analog media. A text transcription of a speech act, for example, is not analog since there is no proportional relationship between the marks on paper and the original sound wave. The reconstruction of the speech act depends on a semiotic system for deciphering the marks, a process that is closer to the coding of the digital than analog media. The text is a code, just not one that is as formalized or as dependent on mediating technology. Text needs only the human deciphering machines while digital text needs the human machines extended with decoding tools. This problem of usage illustrates what will be the closing point in this paper. When we compare the digital to other media we tend to compare the digital as symbolic to other media as material. This move, to say there is the digital on this side, and everything old over there, com-

presses the reality of the variations in media and rhetoric [14]. Look at it this way:

	Analog (Non-Digital)	Digital
Material	Discourse of Arts	
Symbolic		Discourse of Digital Media

In my earlier list of the advantages of the digital, a list that summarizes much of the hype in the discourse around digital media, many of these advantages come from comparing the digital as symbolic works to other media as material objects. A digital work can only be treated as something with no physical presence in time and space if we are thinking of the abstract or logical level, not its inscription on physical media. We can complain about the loss of information in the copying of traditional media compared to digital media only if we are comparing the reproduction of the physical painting to the reproduction of the sequence of symbols (codes) that make up a digital image.

The loss of information in any single digitization is a reminder of the larger choice of the community to bracket the physical off for the engineers that leaves us with a way of speaking and thinking that can treat the ideal artifacts as real. We have chosen this path and like any choice there is a loss of possibilities, but to truly appreciate that which we gain we need to remind ourselves of that which was lost.

7. Digital Logic and the Material

I want to illustrate my point by asking about the logic of those observations and return to the earlier story told about loss of the material in digitization. Remember the apparently obvious assertion that there is no original material presence for the digital, an assertion which is both the source of benefits and the source of artistic loss. The argument is

[14] A very readable discussion of the different types of codes around from Braille to bar-codes can be found in C. Petzold, *Code. The Hidden Language of Computer Hardware and Software*, Microsoft Press, Redmond, Washington 2000.

that digital files have no material presence and can be copied without loss of information or quality [15]. Negroponte's *being digital* revolves around this difference between *bits* and *atoms*. You can do all sorts of things with these modern bits while the material atoms are difficult and intransigent, the way the old is. This, at a material level is clearly not true. All digital files have a material instantiation on a storage device or in memory. Without a physical presence the file would not exist except as a possibility. When one copies a file from, let's say, one flash drive to another, you do not have an identical physical copy. Not only are the flash drives different, the file is probably arranged differently on the drive. Even the signals have differences that, thanks to the engineering we don't have to deal with.

Two files in reality are no more identical than two books from the same print run. Likewise the same file displayed on two different monitors, or on the same monitor at two different times, will not appear the same. The complexity of the output process for a digital image or other media is such that it is impossible to guarantee exactly the same output twice in a row. Subtle variations in light, viewing position, temperature, window location and so on make it impossible, as Heraclitus pointed out, to step in the same representation twice. And I haven't even asked about the conditions of the viewer, the state of the perceiver at the moment when they think they see the representation.

Why then do we consider the digital to have as one of its primary virtues (and limitations) the absence of material presence and consequently the capacity for infinite and perfect reproduction, transmission and representation? Why do we talk this way of something so physical as the light in this room?

The answer lies not in the technology, but in our human culture of computing and a long tradition going back to Plato of abstracting form. We are taught to talk of digital representations as symbolic objects

[15] N. Negroponte, *Being Digital*, Alfred A. Knopf, New York 1995, is one of the most concise expositions of the difference between bits and atoms, the logical and the physical. I don't want to suggest that Negroponte is unaware of the physical dimensions of computing, but he chooses to argue that the digital should be thought of as logical and that when thought of this way, by comparison to information in atoms (read books), one can see a glorious future for digital media.

made up of code not material objects on storage devices; even our language for digitization makes it difficult to distinguish in which sense we are talking about the digital. This does not mean that we do not talk of the infrastructure that makes possible digital representation, rather we tend to isolate discussions of hardware, drivers, backups, and so on from our discussions of the digital in a way that allows us to treat the digital as purely logical and distinct from the physical. We worry about things like the type of hardware we have and whether information is backed up, but we worry about it in conventional ways and in different contexts than when we discuss the digital and the glorious future of electronic texts. It need not be this way; we don't do this when it comes to sculpture and its material inscription. We could imagine an alternate culture of computing where digital media were discussed as physical objects, where the floppy on which a file resides had a special value as the *original* and copies on other devices were treated as derivative castings because the inscription was not the same (even if we couldn't see the difference). But, of course, such a culture of computing would be awkward and not a culture of computing at all. We would probably reinvent the discourse of the logical just because it is so convenient.

This evolution in the discourse of computing technology is something Jacques Ellul discusses in *The Technological Society* [16] and it is perhaps the most important loss in the digital in contemporary culture. The culture of computing has evolved to make it possible to consider the logical independently from the material engineering issues, a possibility that that speaks to the transparency of the engineering. One of the things that makes this possible is the binary character of the digital because the binary character of the digital makes it possible to inscribe a file on just about any material that can unambiguously record sequences in two different states. The binary enables the engineering of the physical so that we can ignore the physical. Computing culture has evolved so as to isolate the engineering from the logical. Were it not so,

[16] J. Ellul, *The Technological Society*, Vintage Books, New York 1964. See also Id., *Perspectives on Our Age. Jacques Ellul Speaks on His Life and Work*, ed. W.H. Vanderburg, Anansi, Concord, Ontario 1997.

these tools would not be general-purpose media tools and they would not be accessible as media appliances that can be used without being an engineer, a difference that makes a difference.

This communal abstraction called *digitization* says more about our culture and our needs than it does about digital files. It is the message of digitization, a message hidden very close in plain view in each digital representation. Computing evolved not just as engineering solutions to particular problems but as a way of implementing useful ideas about the nature of human knowledge and ways of logically representing knowledge so that it can be manipulated, compared and aggregated. The culture of computing evolved to make possible a concrete context in which to treat knowledge and the rhetorical artifacts / media in which we communicate knowledge as logical representations. Landow [17] has pointed out how computer-based hypertexts implemented features of literature that literary theorists had drawn our attention to; likewise digitization implements the dreams of philosophers and logicians for a universal language amenable to formal procedures. It has proven useful to have a domain in which we can treat all messages as logical objects or codes which have no original and which can be copied, aggregated, compared and translated. That the universal logic of the digital has not led to wisdom is beside the point, I'm not sure we really ever wanted to be wise, just distracted by the appearance of wisdom [18].

In this message I have emphasized loss, perhaps overemphasized it as a caution, but generally we only think about what is lost in digitization when the machine breaks down interrupting what we were doing. Which is another way we can read the opening quote from Deleuze and Guattari: the understanding of the machine lost in everyday use is defined by its breaking down. It is defined, in the sense of its outline being made clear, by the fact that it can break down and by the

[17] G.P. Landow, *Hypertext. The Convergence of Contemporary Critical Theory and Technology*, John Hopkins, Baltimore 1992.

[18] This is, of course, a play on Benjamin's point about mechanical media providing a distraction from the system of property, but it is also a point made by Plato in the *Phaedrus* about the danger of confusing innovations in information technology with the pursuit and love of human wisdom.

amazing variety of ways it does break down. When, late at night the machine breaks down and I lose a file, I curse it, and at that moment I am aware of it as a machine instead of seeing through the technology at hand to some purpose. The experience of bugs, errors, and hardware failures defines the machine as a machine and opens an opportunity for reflection on the machine. The interruption in our everyday life is an opportunity, a turning of the machine towards us for thinking not work. It is the way the machine reveals itself to us to be thought about, even if all you do is try to understand the technology to fix it.

So … next time you experience the breaking down, think of it as a welcome interruption that can provoke thought rather than an inconvenience. Think about how the machine has interrupted the flow of your life through breaking down and welcome its refusal to let you ease your way through tasks less important than thinking.