Doing Media History in 2050

Dr. Gabriele Balbi

(gabriele.balbi@lu.unisi.ch)

Draft version, please don't quote without the permission of the author

«Now we can say we've been there»
Letter of Walter Benjamin to Theodor Adorno (1940)

Abstract

The evolution in both the content and storage of information will change the way in which media historians approach their task: to study the history of the mass media, telecommunications and new media of the late twentieth century, scholars will have to interact with digital sources, giving rise to a whole new series of concerns and questions. It is not clear if, how and in which format digital data will be available. What is the relationship between *old* and *new* sources? Who should be responsible for preserving the digital heritage? What should be preserved? How should digital data be preserved? All these questions are crucial for the study of media history now, and above all in 2050.

Introduction

Digitalization is one of the most consequential phenomena to have influenced Western culture in the last decades (Jenskins 2006; Grant and Wilkinson 2009). "Digitalization" is impacting and modifying culture and the preservation of our past in at least three different ways. Firstly, there is the gradual and material digitalization of *old* analogue data: hundreds of people are involved, and will probably continue to be involved in the conversion of paper and magnetic material into digital format, multiplying, facilitating access to, converting and saving *old* data in new forms (Saksida 1997). Secondly, our culture is producing a huge and continuously growing amount of "born digital" data: contemporary working practices, contemporary communication, contemporary culture is based on the production, use, re-use and storage of binary data. Thirdly, born analogue (and digitalized) and born digital data are stored in special archives, called

1

digital archives or digital libraries: they introduce a complex set of problems and challenges regarding their use, preservation and maintenance (Borgman 1999).

The term "digitalization" therefore also signifies a turning-point in the way our culture is stored, recorded and preserved and this alters and will increasingly alter culture itself making it "digital dependent" (Ross 2000, 3). This cultural change affects a whole range of entities: national governments, private companies and broadcasters are investing money in digital archives, trying to preserve their past and make it available in the future. Future media historians will play an important role in this cultural change (and in this business). They will have to read, listen to, see, select, cut off and re-build our digital past: to tell the story of the mass media, telecommunications and new media of the late twentieth century in 2050 scholars will have to interact with digital sources. This paper attempts to summarize why and how the digitalization of historical sources will affect the work of future media historians and what challenges they will face.

New sources, Old sources

Old sources are analogue data that historians have used and will continue to use in the future. By *new* (historical) sources we mean digital data that will become digital sources in the near future.

Old sources have three distinct characteristics: scarcity, stability and availability. Before beginning field research and collecting analogue data media historians (like all historians) know that they have to fully understand, search and "burrow" into different collections and, if they are lucky, they find a few interesting documents for their research: first of all *old* sources are generally **scarce** and one of the most important skills that a historian must have is a "nose" for them.

Media historians know something else: *old* sources are **stable**. They are generally physical objects such as papers (documents, pictures) or magnetic material (films, audiocassettes, recorded radio and TV broadcastings). They are single, static pieces of data that do not change over time; when an *old* source is recorded it is frozen forever. They are stored according to established and defined archiving techniques: the paradigm of "written archives" was created hundred of years ago and has remained more or less the same ever since.

Another characteristic of *old* sources is **availability**: indeed, they are less available than digital ones because media historians have to travel, gain access (generally by "negotiating" with the archivists) and, finally, check if the document is actually in the collection. On the contrary they are also more available for two reasons. First of all, archivists help researchers in their work, they listen to their needs and thus create a kind of archive-researcher interaction, a human-machine that can help us understand and get to grips with the structure of the archive. Secondly, during their search, media historians may find "unexpected" documents, sources they were not searching for: this kind of serendipity is more difficult in *new* digital archives.

The digital revolution has completely changed these aspects typical of *old* sources. First of all, with the digital revolution, there is no longer a scarcity of sources. Indeed this has now been replaced by an **abundance** of material: contemporary digital is characterized by the continuous production of data that could represent future historical sources. If truth be told, this is a pre-digital phenomenon which began with the advent of radio and television that produce a never-ending flow of sounds and images along with a series of problems for their archiving. Digitalization is radicalizing this tendency and propelling Western societies from a culture of scarcity to a culture of abundance (Rosenweig 2003). Perhaps the greatest difference between *old* and *new* sources is their **volatility** (Schloman 2003). We have said that *old* data and *old* historical sources are static while new digital materials are instable, volatile and fragile. First of all, old sources have a physical unit: every paper, cassette, etc. is a single piece of a document linked to others in the same folder, but with its own materiality and unity. New digital data have **no singularity**: they are heterogeneous and fragmented; they are interrelated and fragmented because «the user experience is not confined to a single object at a time, but is constructed out of multiple fragments from different sources and of different types» (Mackenzie Owen 2007, 48). If you preserve a single Facebook profile without considering its connections with, at least, all the friends' profiles, the posted links, and many other applications, you cannot understand the networked nature of Facebook, a medium of personality in which interrelation is the central character.

This volatile element was also a problem for radio and TV broadcasting. Broadcasting is typically volatile as a radio or TV show dies immediately after its transmission: this

means that it could not be conserved without specific machines and, indeed, only a few broadcasted shows have been conserved. Digital preservation is changing the nature of broadcasting, transforming the flow into texts: with podcasts, for example, radio and TV shows can be frozen and transformed into single texts. This is quite different to traditional broadcasting (one-to-many, fixed appointments, little possibility of interaction with audiences); it is a new media approach to listening to the radio or watching television: one-to-one, on-demand, comments and cuts by the users. However, thanks to digitalization, broadcasting paradoxically seems to have become more stable. The volatile nature of digital data does not regard their textuality but their instability, especially when published. For example, a website is a publication but it is different from a paper one: it is undefined, a temporary intellectual product that could disappear and can always be changed. Indeed, the quality of a website is measured also through its continuous refreshing (Ortoleva 1999). This is another reason why new media have been defined "underdetermined" (Poster 1999): always modifiable by its designers and, increasingly, also by its users. This instable nature of digital data affects the way in which these materials can be stored and preserved. First of all, many documents can be and often are already **lost forever** (Kuny 1997), without the possibility of ever finding them again: if a website changes its structure, content, pictures, *old digital* data disappear without any possibility of retrieving them. This means that they will not be able to reach future media historians because they disappear forever and scholars will be unable to understand the web-style, language, cultural significance, historical development of websites and the Internet. Secondly, the volatility and changeability of digital products forces historians and archivists to rethink their preservation philosophy. There is an intrinsic paradox involved here: we want to **permanently preserve** (and aim to study) objects and data that are **ephemeral by nature**, that were born to die quickly. In fact, the specificity of digital data is interaction, the continuous intervention of users; what William Uricchio called the participatory culture (Uricchio 2007). They are fluid/dynamic objects that change over time also because they are interactive and collaborative. For memory institutions it is extremely difficult to save this "infinite heritage" because digital data «are dynamic and always in progress rather than having a final and complete state» (Uricchio 2007, 22). We cannot store and preserve them in a fixed manner and then study them in this new form: this fixed and frozen source is

something completely different from the "underdetermined" datum. Indeed, what we should try to preserve are the uses and processes, not the materials and artefacts (MacKenzie Owen 2007). For example, let's consider a page of Wikipedia: the page can potentially always be changed by users with consequences on the present and future history (Rosenzweig 2006) and, indeed, the power and popularity of this project is the fact that definitions can be refined, improved day by day by a "collective (and connective) intelligence" (Levy 1997). So what Wikipedia page should we preserve? Which ones and at which point of the timeline will show future historians how a concept was considered in the past?

In *The Bias of Communication*, Harold Innis (1951) distinguished between durable and portable media and, paraphrasing him, we could suggest a history of stability: it seems that, in the evolution of communication history, data and documents have become increasingly volatile, passing from stone, to paper, to magnetic, to digital. These supports are evermore portable, inexpensive and easy to access, but can be more easily lost and erased and, above all, they are more difficult to store. This is the process of dematerialization and "lightening" of culture. But it is also a process of "**structural weakening**" of culture because the technical support has become more and more fragile and breakable. Stone, paper and magnetic are in order less difficult to damage; for example you had to use brute force with the stone or fire with the paper while a digital collection can be completely destroyed by damaging just one bit: just one letter or one unit of a repository could make the repository unusable.

Digitalization has also changed the availability of data: compared to the past, sources appear **more available but with an inherent risk**. On-line digital documents are accessible to a huge number of people and, in this sense, digitalization has democratized access to historical sources. With the Internet, non only the source, but also historical essays are always available, anywhere in the world (Cohen 2004, 294); you can download them – maybe with password access – when and where you need them and historians no longer have to spend money to physically reach the archives. Moreover, access to the source is direct and non-mediated, an old dream of historiography which has finally come true: accessing the past directly, without mediation.

On the other hand, the absence of mediation, generally represented by the archivist, risks leaving historians out on their own and, above all, without basic information. For media historians, the relationship with the archivist is fundamental because he/she provides valuable information on the research object, on new and innovative keys of research, on the structure of the archive and, consequently, on the past "archival mentality".

The digital revolution is modifying the supply, stability and availability of the historical sources but there are many similarities and continuities with the past that should be mentioned. First of all, **the *old* doesn't disappear**. Old archives and old materials remain: paper publishing continues and libraries and paper archives are, and will continue to be relevant for humanity and media historians. Secondly, a real challenge of the present and of the future is the **integration** between *old* archives and *new* digital "environments". In practice this integration has already begun: for example many archivists have started to provide e-mail reference services to historians and scholars: this is the best way of letting media historians know the contents of archives and maybe avoid futile journeys. In general, for historians, «the greatest promise of the WWW is not that it replaces the need to use the original material, but rather that it supports researchers in becoming aware of and locating archival materials to be used in situ» (Duff, Craig & Cherry 2004, 22). So, for media historians the greatest potential of digital is to expand and "democratize" access to physical repositories, paper archives and old structures in which data are stored.

At the end of this and every paragraph of this paper we will summarize and delineate how the change in the materiality of sources has affected and will continue to affect media historians. First of all, we said that digital sources will be more abundant than analogue ones: this means that researchers will have more difficulty in selecting *relevant* materials and in not being "fetishists of the source". They will have to endeavour to maintain the critical distance between the object of research and the source. Secondly, due to the volatility and instability of digital data, historians may lose significant sources in their quest to understand the digital past of contemporary society. Thirdly, considering that the most relevant characteristic of digital data is their continuous change and refresh, future media historians could find frozen digital

documents that don't correspond to the original changeable and participatory ones. Finally, digital documents are more readily available and historians can easily reach and make searches (e.g. with Google Scholar) in historical essays. The automatic search function, on the other hand, changes the relationship with the archivist, giving rise to new methods, new forms and a new loneliness for media historians.

Who preserves?

Analogue data have usually been preserved by specific subjects having two principal characteristics. First of all they are structured at a **national level** and so they generally preserve documents useful for remembering the nation's past and for recreating national unity. Secondly, especially in the twentieth century, these subjects have either been **public institutions** or **private companies**. Public institutions preserve the past for 3 main reasons: first of all because their aim is to have a monopoly over national representation; secondly because national libraries, archives and museums have the specific "public service function" of preserving data; finally, a public service corporation like the BBC «is legally required to record its TV and radio output off-air to enable the corporation to answer complaints from the listening and viewing public» for 90 days and 42 days following the initial broadcast respectively (Smith 2005, 17). Private companies have conserved data and created archives for the re-use of data: for example, professional firms, private broadcasting companies and universities have archived radio and TV shows in order to sell them, re-use them and re-broadcast them for educational purposes.

With digital data, many of these assumptions no longer apply. First of all, digital **data nationality** is an issue that **doesn't make sense**. A website has no nationality, it sometimes has a national language but English language websites are frequently not even produced in Anglophone countries because English is the lingua franca of the web. Digital data and, in perspective, digital historical sources erode national boundaries and establish new international relationships. The second aspect is more pertinent and concerns **who has the duty to conserve** the digital heritage. Unlike the case of analogue data, public companies are not interested in conserving data that are not national and that cannot be re-used for national purposes. Private companies have

launched many projects concerning the archiving of the digital past; we have chosen two examples. The Internet Archive (http://www.archive.org) is a project that began archiving the web in 1996 with the aim of preserving the Internet and especially text, audio and audiovisual images appearing on the web. The second example is case of the *New York Times* (http://www.nytimes.com): the online version of this newspaper provides free access to its current contents but it charges for articles that are older than one week. There are important questions that inevitably affect the future media historian: is it right to leave the past in private hands even if it is still seen as a public asset? Will the politics of these two private companies change in the future? Will the Internet Archives or the NYT online decide to delete archives older than x years? Who can stop them? «Can they safeguard in the long term materials that have lost their commercial value?» (de Jong and Wintermans 2007, 6) How much can a private company charge a user or historian to access old sources? Who should set the price?

There are possible solutions to prevent these digital catastrophes. First of all, **digital heritage preservation** could be set as a **policy goal** (Smith 2003, 7), not only at a national level, but also with international recognition. What's more, national legislation is ineffectual in preserving copyright and this is another issue in understanding who should have the duty of preserving data. Another solution is a **public-private partnership**: public companies have the duty of preservation for humanity and private companies are specialized in providing services and could contribute their know how to creating user-friendly solutions.

With digital data there is also a new and unexpected subject who could be involved in the preservation process: **the user**. Users now produce content and store it in specific repositories: for example YouTube is an audiovisual archive of private content, videos taken from the web and, frequently, snippets of broadcasted and recorded shows. This is another meaning of the so-called web 2.0: the web becomes a medium of preservation of old sources that form part of popular mentality. The user, from this point of view, can be considered a sort of archivist because he/she store videos from his/her personal collection and shares them with other unknown people. Users are a crucial link in the chain of digital preservation, because the «preservation of digital content must be a collaborative effort that involves the professional archivist, the technology expert, the

user, and creating and producing entity» (Ide, MacCarn, Shepard and Weisse 2002, 12). Users are pivotal also in Bearman's view, the author of what we consider to be the most interesting solution to identifying the subjects tasked with digital preservation. Bearman maintains that museums, libraries and archives have different traditions, have saved different things and have been interested in different aspects of the same sources. This model could be applied and improved also in the digital environment. In other words, the same digital datum could be preserved by different subjects (including the users) for different purposes: each of these subjects preserves a particular aspect of the same data that could be of interest from different historical perspectives. To do this, "we will need to create and maintain multiple representation of digital objects to satisfy different aspects of reliability and authenticity" (Bearman 2007, 30). This is a collective model that assigns the **entire society the duty of preserving** pieces of different digital pasts: the collection will not be complete but at least it will reflect different interests and instances of the society.

There is a final issue related to the subject tasked with preserving the data: **power** (Schwartz and Cook 2002). With analogue data and sources specific institutions were set up to preserve the past heritage. If we schematically consider the history of preservation, until the invention of the printing press in the sixteenth century there was a link between preservation and religious power: from the pyramids in ancient Egypt to the Roman churches and monasteries it was always religious institutions that had a monopoly over the preservation of the past. This power became more secular in the eighteenth and nineteenth century: public museums and archives were created to re-use the national past. In the twentieth century the power of memory was shared by public institutions and private companies, for example broadcasting companies. With digital data, the "preservation power" seems to be at stake again and the very concept of power afforded by preservation will continue to change. The term "archives" comes from the Greek *arkhein* (to command) and, until now, conserving and archiving have always been synonymous with power (Sorlin 2007, 16). However, when information is scarce, as in the past, its value increases; instead, when it is abundant as in our digital era, when it is easily available and the only problem is that of selecting what is relevant, its value (and notion) inevitably decreases (Esposito 2001). The transition from analogue to

digital data and sources could mask a similar change in power: in an era when producing information is an inexpensive and widespread habit, the task of preserving old digital information could be assigned to a less powerful institution.

Who preserves the data is important for media historians because it could affect their work in at least three different ways. Firstly, it could affect **source accessibility**. Public and private companies have different responsibilities but private companies do not have the duty to preserve data even if they do it. What claims could a historian make of private companies? If the New York Times decides to erase its digital past how will media historians be able to study the way the newspaper saw and reported, for example, 9/11? If the Internet Archive decides to stop its project tomorrow, who will take it over and, above all, who will guarantee that all the gathered sources will be available for historians? These questions deal with the possibility of understanding contemporary society. A second topic is what we call the **risk of a one-dimensional past**. We wonder: what kind of digital data will reach future historians? The risk is that media historians will work on digital sources conserved by just a few institutions – generally the "dominant ones" and the richest private companies – that preserved *this past* for specific aims and purposes. Another risk is that, sold on the market, these sources and access to particular repositories will be available only to the "richest", most famous and influential historians. Finally, will media **historians** become increasingly **involved in** the process of **the re-use of data** held by private companies? For example, in the future broadcasting companies (if this term will be still used) could involve media historians in the interpretation of their old digital sources and in giving them an historical structure in order to re-broadcast or re-sell them. Will this lead to a change in historians' work? For sure, they will have to learn new practices, but will they maintain their freedom of research and expression or will professional history be involved in a more general process of mediatization of the society? Media historians will have to understand which institution is their main interlocutor in order to be free.

Preserving what?

A central question in understanding digital data and in transforming them into historical sources for future historians is: what should be preserved? There are two main approaches, two main schools of thought in what to preserve in the digital era (Phillips

2005, 58). The first one is the *old* one: in the past, national libraries and archives were typically comprehensive and their aim was to collect at least one copy of everything published. Supporters of this approach want to continue this tradition of conserving the **entire (digital) heritage**. The second school of thought maintains that online publishing is completely different from the traditional one and so requires a **selective methodology**. Using poker jargon, we could call the first the "all in approach" and the second the "small play approach".

If a poker player goes "all in", he/she bets all the chips left on the table: translating this into our language and interests, this approach aims to preserve everything produced in the digital era. There are three main advantages in the all in approach. First of all, with this approach the entire digital domain will be available for future researchers and humanity. Secondly, this full back-up is performed automatically and so **human intervention** (human errors and choices) is minimized. Finally, due to the decreasing costs of memory hardware, for the first time it is actually possible and quite **cheap to store** huge amounts of data, maybe everything.

The all in approach is also criticized for four main reasons. First of all, the main disadvantage is that it appears to be a **utopia**. Storage back-up is performed at intervals (for example every two months) and during these intervals new digital data originate and die because the production of digital data is incessant. Furthermore, due to its volatile and connective nature, the digital heritage is potentially infinite and interconnected: it is extremely difficult, almost impossible, to store each and every connection. A second disadvantage is that **quality control** is impossible: for example, inaccessible websites (with password) are automatically rejected and are not saved. Thirdly, one crucial aspect is that of **metadata**: the quality, accuracy and richness of metadata are essential for retrieving "born digital" data (Wactlar and Christel 2002) and the all in approach does not allow complete and efficient control of their excellence. Finally, this approach implies a critical change in the filing clerk's work methodology: if we let the machine archive everything the "**duty of selection**" vanishes.

The opposite school of thought is what we call the "slow play approach". A poker player plays slow, with small bets, in order to get to know his opponents and elaborate a

strategy. Similarly, due to its increasing volume, the digital heritage could be conserved in small, representative "pieces", selecting them from the entire digital heritage. This approach has two advantages. First of all **accuracy**: focusing on a small sample you can check the completeness (with all the links), the evolution of the website pages considered and the quality of the metadata. In other words, the stored data are fully catalogued. Secondly, **access for users** is generally easier because if you store a small amount of data, you can negotiate access permission with the publishers. This is true also for acceding inaccessible data: they could be identified and negotiated with publishers.

There are three main disadvantages to this approach. The most important one is the risk of **rejecting what could be useful** for future media historians. What are the selection criteria? Is it right to define these criteria *now* for future historians? «Do we really know what will be important for future researchers»? (Phillips 2005, 61). For example, gossip blogs and websites could appear to be a superfluous and rejectable part of our digital heritage but these sources could be crucial for a cultural and social historian of the late twenty-first century. Another example is spam: we see them as rubbish or, better, digital rubbish, but maybe future historians will see them as important sources, a significant product of contemporary society and therefore something to study. Cohen and Rosenzweig think that it is possible to assess the long-term value of a document by asking questions such as: is the information held unique? And how significant is the source for research? (Cohen and Rosenzweig 2006, 226). But, on one hand the concept of uniqueness contrasts with that of digital reproducibility and, on the other, who knows the significance of a datum for future researchers? Furthermore, and above all, who has the duty to choose what is relevant and what is not? The second disadvantage of the small play approach is also one of its positive aspects: the limited amount of material preserved allows better and complete storage but it also represents a **limited part** of the huge volume of digital data produced. The final disadvantage of this approach is the risk of including data **out of context** «and often not including other resources to which it is linked» (Phillips 2005, 61): for example, if you preserve a set of websites, you could perhaps exclude a semantic link that you don't know.

As we have seen, both the "all in" and the "small play" approach have potential advantages and disadvantages in preserving the digital heritage and for this reason

hybrid approaches have been experimented (Phillips 2005, 63-70). Our aim above was to present the challenges and main problems of a very simple question: what should we preserve of our digital present? Another question we aim to answer in this final paragraph is: how does what we preserve affect the work of media historians?

First of all we have to understand the needs of a media historian now and in the future. For example, do **historians need *everything*** of our digital present in order to study it? An historical source is something that should help historians complete, verify or refute a theory. But the work of an historian is much more difficult than just reading sources and blending them into a "story"; they come up against different theories, they interpret facts according to a reference theory, they want to comb through sources and read them only as a sign of the past, and the meaning of this past cannot be explained only basing on these contingencies. They will probably not need all the series of *The Big Brother* to understand the social and cultural significance of *The Big Brother* in late twentieth century. Maybe one or two episodes will be enough. But the real question is: which ones?

Secondly, in the past one of the historian's aims was **to look at "everything"** that happened in their period. In this new era of abundance this will quite simply be **impossible**. Media historians will have to select the relevant things after having seen and checked only a part of the total amount. Preserving the entire digital heritage means preserving volumes of sources that future media historians will not be able to consult. But each media historian will look at a small, and probably different, piece of digital data: preserving everything means preserving many possible lines of research.

The problem of selecting or rejecting material has always been a crucial topic for historians but the digital era increases the scale of the problem. This is a **methodological and technical change** too: even if not everything but only a small part of the digital heritage is preserved, historians in any case will have to learn how to make searches in digital repositories, for example using automatic tools, and in general they will have to select and reject much more than in the analogue past.

How to preserve?

So far we have discussed the major changes in the materiality of digital sources compared to analogue ones, who should have the (new) duty of preserving digital data,

and what kind of digital data should be preserved and transformed into digital sources for future historians. However, last but not least, comes the problem of how to preserve this new data and make them available for future researchers.

The most important problem in this respect is what we could call the longevity of the digital heritage. Media longevity problems exist both for analogue and digital but the scale is different: stone remains available and readable longer than printed books, and printed books last longer than magnetic equipment (Conway 1996). Digital data and records age (much) faster than all other technological supports because of both the hardware and the software. Moore's law which claims that information technology hardware becomes obsolete every 18 months is significant. We don't intend to analyze the truth of this law here; we merely want to draw attention to the rapid obsolescence of digital hardware or, better, the support on which digital data and sources are stored. Digital data may age rapidly also due to their organization: they are structured in a format designed for an application program that may no longer exist: designed and organized for a specific software that could become obsolete in a few years or even just a few months.

Howard Besser (2000) has identified at least five key factors that pose digital longevity problems. The first is the so-called **viewing problem**. «When we discover older analogue works, at least we can view them and their structure even if we had lost the ability to decode their language» (Besser 2000): for example we can read cuneiform tables even if we don't know what they mean. It is more difficult to read digital information because it requires specific software and hardware: for example, in general we cannot read 8" floppy disks because our personal computer is not equipped to read them and, even if we could find an appropriate hardware, we would still have the problem of deciphering the file format. The second difficulty is the so-called **scrambling problem** which involves three different concepts. There are problems created by compression: the long-term effects of compression are still not fully understood and what we compress today may preclude the use of these digital data in the future. There are also problems related to the encoding of a file: it may be difficult for future archaeologists and media historians to decipher these digital sources. Finally, there are problems related to the ways in which digital files are protected and encrypted: the digital heritage could be difficulty decrypted by future historians. Another difficulty

is created by **inter-relational problems**: as we have said, a central characteristic of digital data is their interrelation, their potentially infinite linkability. In order to better preserve a digital product it is important to try to establish the boundaries of the work: this is difficult for the digital heritage. Besser has identified a fourth problem: the **custodial** one. This is what we discussed in the second paragraph: while we wonder how our digital past should be preserved, it is also crucial to understand which organization(s) should take responsibility for their long-term preservation. The fifth and last problem is **translation**. Translation was necessary also in the past: analogue data had to be translated into other forms (also digital ones) but this is even more crucial in the digital work. We could believe that, because we can make identical copies of digital files, then the digital-digital translation will not create problems. This is a mistake because «though the bits in the file's contents may be identical, the application environment used to view the file most certainly will be different. In fact the very reason for converting the file is because we are unable to successfully sustain that application's environment over time» (Besser 2000). Media historians have to be aware that, if translated into other support and technical environments, digital sources change but at the same time translation is indispensable in preserving documents and giving future historians the possibility of reading, seeing, listening to them.

While these are the major problems that, according to Howard Besser, have to be taken into consideration in order to understand *how* to preserve our digital heritage, Caroline Arms (2000) has developed a conceptual framework that shows at least five different preservation methods, five avenues for saving our digital present. Media historians have to know these methods in order to fully understand how digital sources are stored now and how they will probably be available in the future.

The first long-term preservation method is the so-called **better media**. Longevity and technology independence of the digital heritage is ensured also by the medium on which it is stored: this doesn't mean that a single medium is *the* best one for all digital sources, but that there are media that are more suited than others to preserve specific data. It entails an ability to transfer the information to new media (Abdelaziz 2007). A second and related avenue in long-term digital preservation is the **migration of content** and this implies converting digital data into different and more evolved digital formats. The

first two methods for long-term digital preservation are time-consuming and expensive and are also paradoxical. As noted by Su-Shing, digital preservation is subject to a fundamental paradox: «On the one hand, we want to maintain digital information intact as it was created; on the other, we want to access this information dynamically and with the most advanced tools» (Su-Shing Chen 2001, 3). In other words – and this aspect is of fundamental importance for media historians – we want to maintain the originality of the digital source that implies either the originality of the content or the originality of the environment, but we also have to deal with hardware and software obsolescence and, if we maintain digital info without modifications, accessing them will eventually become impossible. What's more, we want to use the most advanced machinery for optimal search in these sources.

A series of **emulation** techniques has been created to respect the original environment of utilization: this third long-term preservation method implies that new technologies simulate to function like and reconstruct the ambient of previous generation technologies. In this way digital sources can be read, seen, listened to in their original environment, even if this originality is an artefact. The fourth method of long-term preservation is called **refreshing bits** and «includes any operation that simply involves copying a stream of bits from one location to another, whether the physical medium is the same or not» (Arms 2000): this means that if you frequently and automatically copy *old digital* sources in new repositories and with new formats you will probably contribute to their preservation. The final method considered by Caroline Arms is digital archaeology and is used only if «all else fails» (Arms 2000): if a future historian is unable to access a digital past in any way, he/she will have to pick up pieces, traces, fossils of this digital past like an archaeologist does with dinosaurs.

I would add a sixth avenue, a sixth way of preserving data that could ensure the long-term availability of digital data: the **multiple (potentially infinite) copy**. As pointed out in the second paragraph, users will increasingly become a "fragmented institute of preservation", because they have personal repositories and sometimes share them. A decentralized memory with redundancy and continuous multiplication in copies could ensure long-term digital preservation. Millions of decentralized personal computers own small or huge pieces of cultural heritage and, if they were linked, a network-repository would be created and this network would enhance the long-term availability of digital

sources. This is a fairly old preservation strategy. Above all in the analogue world the existence of multiple copies of a work stored at different geographic locations helped preserve this work: copies of the books stored in the Alexandria Library helped saved the cultural heritage also after the Library was destroyed. The multiplication of copies could be an analogue strategy, maybe *the* analogue strategy, best suited also to the digital case.

The way in which digital sources are preserved obviously affects the future media historians' work. Historians of 2050 studying our present media history will have to use different methodologies and approaches depending on how data are stored. For example, to study digital forms they will have to abandon the idea of reading on paper or, at the very least, they will have to follow the inextricable connection between paper and bits. In fact, you can print a website and read it on paper but you automatically lose its interactive and hypertextual character. Another major problem in studying digital sources on paper is the impossibility of the automatic search typical of digital data: when you have to search within a huge number of e-mails you need to use automatic functions; if you print these e-mails you miss out on a large potential and a peculiarity of the digital heritage. This is the greatest danger for historians: sacrificing «the original form, which may be of unique historical, contextual, or evidential interest» (Rothenberg 1999).

Abandoning paper is a small example of a necessary change in mentality for historians approaching digital media history. This change implies that historians will stop using the web like a printed repository guide or the telephone book (Tibbo 2002, 9) or printing resources rather than consulting them on line (thus maintaining the better medium and the better original context in which the source was created). This change of mentality implies that historians will have to understand new search methods and learn new skills. Huge amounts of data could be stored in a non-readable format in the near future: this is a danger but it also implies that future media historians will also be digital archaeologists. This change in mentality implies that historians will consider – in McLuhan's words – either the message or the medium: in other words, they will have to be aware of and understand how the message is and was conveyed and how the

hardware affects the way in which the message is and was perceived (evaluating the migration and emulation techniques).

The way in which digital sources are preserved and will used by historians also entails new dangers; for example, the danger of **decontextualization**. In *old* analogue archives documents were collected in folders and so many papers were *physically* close to each other. In other words, each document was part of a more complex sphere and this singularity was influenced by multiplicity (Vitali 2000). In digital repositories – due to the huge amount of data stored, to the immateriality of digital data and to the structure of digital archives – media historians risk decontextualizing documents: it is the automatic search that produces singular data, singular digital documents not linked with other (maybe linked) digital objects.

Conclusion

The aim of this paper was to try to summarize why and how the digitalization of historical sources affects the work of future media historians and what challenges they (will) have to accept and overcome.

We have tried to point out that, on digital preservation, there are more open questions that solve ones. Compared to the long-standing experience of preserving the *old* analogue heritage, it is still unclear who has responsibility for preservation, what should be preserved and how it should be preserved. It is much clearer that the digital heritage has different characteristics to the analogue one and this difference poses many questions and challenges for future media historians.

First of all, as pointed out several times in this paper, media historians will have to be willing to accept major changes in their work. They will have to approach the digital heritage differently from the analogue one because digital data are volatile, interconnected and abundant; they will have to search with new tools in order to manage this great amount of data, «Archives of the future will be different and researchers will adopt new, and more technology dependent, ways of working» (Ross 2000, 11); they will have to interact with new institutions and they will have to learn new methods of access to the sources; they will have to learn new methods of reconstructing the digital past of these sources; finally they will have to tackle new difficulties such as accessibility, ownership, fragility, originality and decontextualization of digital sources.

New digital sources will not only lead to changes in the media history profession, but in certain ways new data will probably not change old habits. A first example: historians will still have to select from among material, maybe they will have to learn new search strategies but they will still have to interpret data according to personal and specific research questions. In other words, the digital revolution doesn't seem to have anything to do with the duty of selection and interpretation typical of media historians' work. A second example: history has always been an interpretation of reality (indeed there have always been different readings of the same historical event) and it will probably continue to be an interpretation of the (huge amount) of available data.

Maybe we are entering a new preservation paradigm and we need to reconsider all our certainties about "preservation" (Abdelaziz 2007). What is certain is that we have to act now in order to **preserve the memory of our times** and to let historians have sources to interpret our digital past or, better, our present. Deciding now which institutions have responsibility for preservation, what should be preserved and how is the only way of giving future media historians traces of our digital media past. To let them study and to let us know late twentieth century memory and, finally, say «Now we can say we've been there» (Adorno 1999, 326).

References

Abdelaziz, Abid. 2007. "Safeguarding our digital heritage: a new preservation paradigm". In *Preserving the digital heritage. Principles and policies, eds. Yola de Lusenet and Vincent Wintermans*, 7-14. Selected papers of the international conference organized by Netherlands National Commission for UNESCO, November 4, 2005, The Hague, The Netherlands, http://www.knaw.nl/ecpa/publ/pdf/2735.pdf (accessed April 17, 2009).

Adorno, Theodor W. 1999. *The complete correspondence, 1928-1940 / Theodor W. Adorno and Walter Benjamin*. Cambridge, Mass.: Harvard University Press.

Arms, Caroline R. 2000. "Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress". *RLG DigiNews*, 4/3, June, http://webdoc.gwdg.de/edoc/aw/rlgdn/preserv/diginews/diginews4-3.html (accessed April 17, 2009).

Bearman, David. 2007. "Addressing selection and digital preservation as systemic problems" In *Preserving the digital heritage. Principles and policies*, eds. Yola de Lusenet and Vincent Wintermans, 26-44. Selected papers of the international conference organized by Netherlands National Commission for UNESCO, November 4, 2005, The Hague, The Netherlands, http://www.knaw.nl/ecpa/publ/pdf/2735.pdf (accessed April 17, 2009).

Besser, Howard. 2000. "Digital longevity". In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, ed. Maxine Sitts, 155-166. Andover MA: Northeast Document Conservation Center.

Borgman, Christine L. 1999. "What are digital libraries? Competing visions", *Information Processing and Management: An International Journal*, 35/3, May: 227-243.

Cohen, Daniel J. 2004. "History and the Second Decade of the Web". *Rethinking History*, 8/2, June: 293-301.

Cohen, Daniel J., and Roy Rosenzweig. 2006. *Digital History. A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press.

Conway, Paul. 1996. *Preservation in the digital world*. Washington, DC: Commission on Preservation and Access.

Dalbello, Marija. 2006. "Scholarly Editions, Historians' Archives and Digital Libraries: The Pragmatics and the Rethoric of Digital Humanities Scholarship", September, http://dlist.sir.arizona.edu/1473 (accessed April 17, 2009).

de Jong, Annemieke, and Vincent Wintermans. 2007. "Introduction". In *Preserving the digital heritage. Principles and policies*, eds. Yola de Lusenet and Vincent Wintermans, 1-6. Selected papers of the international conference organized by Netherlands National Commission for UNESCO, November 4, 2005, The Hague, The Netherlands, http://www.knaw.nl/ecpa/publ/pdf/2735.pdf (accessed April 17, 2009).

Duff, Wendy, Barbara Craig, and Joan Cherry. 2004. "Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age". *The Public Historian*, 26/2, May: 7–22.

Esposito, Elena. 2001. *La memoria sociale: mezzi per comunicare e modi di dimenticare*. Roma: Laterza.

Grant, August E. and Jeffrey S. Wilkinson (eds.). 2009. *Understanding media convergence: the state of the field*. New York: Oxford University Press.

Hedstrom, Margaret. 1998. "Digital Preservation: A Time Bomb for Digital Libraries". *Computers and the Humanities*, 31: 189-202.

Ide, Mary, MacCarn, Dave, Shepard, Thom, and Leah Weisse. 2002. "Understanding the preservation challenge of digital television". *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*, http://www.clir.org/pubs/reports/pub106/television.html (accessed April 17, 2009).

Innis, Harold A. 1951. *The bias of communication*. Toronto: University of Toronto Press.

Jenkins, Henry. 2006. *Convergence culture: where old and new media collide*. New York: New York University Press.

Kuny, Terry. 1997. "A Digital Dark Ages? Challenges in the Preservation of Electronic Information". Paper presented at the 63RD IFLA Council and General Conference, September 4.

Lévy, Pierre. 1997. *Collective intelligence : mankind's emerging world in cyberspace.* New York: Plenum Trade.

Mackenzie Owen, John. 2007. "Preserving the digital heritage: roles and responsibilities for heritage repositories" In *Preserving the digital heritage. Principles and policies*, eds. Yola de Lusenet and Vincent Wintermans, 45-49. Selected papers of the international conference organized by Netherlands National Commission for UNESCO, November 4, 2005, The Hague, The Netherlands, http://www.knaw.nl/ecpa/publ/pdf/2735.pdf (accessed April 17, 2009).

Manoff, Marlene. 2006. "The Materiality of Digital Collections: Theoretical and Historical Perspectives". *Libraries and the Academy*, 6/3, July: 311-325.

Ortoleva, Peppino. 1999. "La rete e la catena. Mestiere di storico al tempo di Internet". *Memoria e ricerca*, 31/3, http://www.fondazionecasadioriani.it/modules.php?name=MR&op=body&id=76 (accessed April 17, 2009).

Phillips, Margaret E. 2005. "What Should We Preserve? The Question for Heritage Libraries in a Digital World". *Library Trends*, 54/1, Summer: 57-71.

Poster, Mark. 1999. "Underdetermination". *New media & society*, 1/1: 12-17.

Rosenzweig, Roy. 2003. "Scarcity or Abundance? Preserving the Past in a Digital Era". *The American Historical Review*, 108/3, June: 735-762.

Rosenzweig, Roy. 2006. "Can History be Open Source? Wikipedia and the Future of the Past". *The Journal of American History*, 93/1, June: 117-46:

Ross, Seamus. 2000. *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship*. National Preservation Office, http://www.bl.uk/npo/pdf/wigan.pdf (accessed April 17, 2009).

Rothenberg, Jeff. 1999. *Avoiding technological quicksand: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources*. Washington, DC: Council on Library and Information Resources.

Saksida, Marino. 1997. "The Information Society in the 21st Century: Converting from Analogue to Digital", *The International Information & Library Review*, 29/3-4, September: 261-267.

Schloman, Barbara F. 2003. "Information Resources: Now you see it, now you don't: The Ephemeral Nature of Digital Information". *The Online Journal of Issues in Nursing*, 8/2, May, http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeri odicals/OJIN/TableofContents/Volume82003/No2May2003/EphemeralNatureDi gitalInformation.aspx (accessed April 17, 2009).

Schwartz, Joan M., and Terry Cook. 2002. "Archives, Records, and Power: The Making of Modern Memory". *Archival Science*, 2/1-2, March: 1-19.

Smith, Bernard. 2005. "Preserving Tomorrow's Memory: Preserving Digital Content for Future Generations". *International Preservation News*, 29, May: 4-10.

Smith, Cathy. 2005. "Building an Internet Archive System for the British Broadcasting Corporation". *Library Trends*, 54/1, Summer: 16-32.

Sorlin, Pierre. 2007. "Historians at the crossroads: What can we do with visual archives?". In *Heard-Seen. The Uses of Digitised Archives for the Sciences*, ed. Herausgegeben von Kurt Deggeller, Ursula Ganz-Blättler un Ruth Hungerbühler, 16-23. Baden und Lugano: Hier und Jetzt.

Su-Shing, Chen. 2001. "The Paradox of Digital Preservation". *Computer*, 34/3, March: 24-28.

Tibbo, Helen R. 2002. "Primarily History: Historians and the Search for Primary Source Materials". In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*. Portland and Oregon: 1-10, http://portal.acm.org/citation.cfm?doid=544220.544222 (accessed April 17, 2009).

Uricchio, William. 2007. "Moving beyond the artifact: lessons from participatory culture". In *Preserving the digital heritage. Principles and policies*, eds. Yola de Lusenet and Vincent Wintermans, 15-25. Selected papers of the international conference organized by Netherlands National Commission for UNESCO, November 4, 2005, The Hague, The Netherlands, http://www.knaw.nl/ecpa/publ/pdf/2735.pdf (accessed April 17, 2009).

Vitali, Stefano. 2000. "'Archivi on line': qualche riflessione metodologica". Paper presented at the Workshop Archivi storici e archivi digitali tra ricerca e comunicazione, Florence, 20-21 October,

http://www.dssg.unifi.it/_storinforma/Ws/archivi/vitali.rtf (accessed April 17, 2009).

Wactlar, Howard D., and Michael G. Christel. 2002. "Digital Video Archives: Managing Through Metadata". *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*: 84 – 99, http://www.ri.cmu.edu/publication_view.html?pub_id=3964 (accessed April 17, 2009).