Reconstructing a website's lost past Methodological issues concerning the history of Unibo.it

Federico Nanni <federico_at_informatik_dot_uni-mannheim_dot_de>, University of Mannheim

Abstract

This paper describes how to deal with the scarcity of born-digital primary sources while retrieving materials on the recent past of an academic institution. The case study is an analysis of the first 25 years online of the University of Bologna. The focus of this work is primarily methodological: several different issues are presented, starting with the fact that the University of Bologna website has been excluded for thirteen years from the Internet Archive's Wayback Machine, and possible solutions are proposed and applied. Moreover, this study aims at highlighting how web materials could give us new and distinct insights into the recent past of academic institutions, thereby becoming the starting point for several new studies.

Introduction

The University of Bologna is considered to be the world's oldest university in terms of continuous operation. Its nine-hundred year old roots can be traced back to the figure of the early recorded scholar, jurist and instructor Irnerius [Capitani 1987][1]. The study of specific aspects of the past of this institution has already offered researchers the unique possibility of digging deeper into the relationship between the university, its large student community and the city of Bologna itself [Brizzi 1991]; [Barbagli 2009]. Additionally, this kind of research has also allowed a better understanding of its key historical role in the Italian academic "ecosystem" [Brizzi 2007]. Several sources have been used to trace its past, starting from textual documents preserved in the university archive [Rea 1996]; [Romano 2007] to its collection of over six hundred portraits [Gandolfi 2011].

1

Since the introduction of the World Wide Web, a new and different kind of primary source has been available for historical research on the recent past of this academic institution: born digital documents, materials that exist only online and the role of which will increasingly be recognized as complementary to traditional-analogue and digitized sources [Brügger 2016]. However, due to their digital nature, these new sources are already more difficult to preserve, retrieve and analyze compared to traditional materials [Rosenzweig 2003]; [Brügger 2009]; [LaFrance 2015]. For this reason, during the last twenty years, public and private web archive initiatives all around the world have been preserving them for future studies.

2

Materials from the Internet Archive, the most important and comprehensive web archive [Gomes 2011], have already shown their potential as new primary sources in a few pioneering studies [Ben-David 2016]; [Milligan 2017]. Specifically talking about these new documents, Ian Milligan has recently raised a very provocative question: "Could one even study the 1990s and beyond without web archives?"[2]. Milligan, with this question, re-evokes an old teaching from Marc Bloch, namely the fact that "everything that man says or writes, everything that he makes, everything he touches can or ought to teach us about him" [Bloch 1949][3]. However, at the same time, Milligan's question addresses another, more methodological, aspect. Could one study the past, and more precisely our *recent digital past*, without web archives?

3

This study, which focuses on the recent past of the University of Bologna through its digital sources, explores both aspects of Milligan's question. It starts by considering the facts that (a) the University of Bologna's main website

4

(hereafter "Unibo.it")[4] was not accessible through the Internet Archive's Wayback Machine when this research was conducted (as described in [Nanni 2015]) and (b) Italy is one of the few countries in Western Europe that does not systematically preserve its National web sphere [Gomes 2011][5]. By doing so, the goal of this paper is to address the following research questions:

- Is it possible to reconstruct and study the past of a university website (namely the changes in its layout, structure and content, but more importantly the reasons that have caused them) without having at prompt disposal a collection of web archive snapshots?
- Could this study guide us to better understand the role the website has played in the interactions between the academic institution and its large and variegated community and, by that, could we obtain new insights in the recent past of the institution in itself?
- Will this research bring new materials *to the surface*, in ways that are useful for the research communities that, so far, have focused on the past and present of academic institutions?

Specific contribution

Before starting, it is important to note a few aspects of the paper presented here. First of all, this research does not set out to be a comprehensive overview on the methodological approaches used in the diachronic study of websites. On the contrary, the goal of this paper is to present a first-hand experience on the issues that emerge when employing born digital materials as primary sources and when considering websites as "objects of study" [Brügger 2009]. Presenting the problems encountered while retrieving born digital documents and the solutions adopted could be useful to other scholars who, in the near future, intend to use the same materials in their research. Additionally, while the focus of our first research question is primarily methodological (more explicitly: how a website's lost past was reconstructed), this study does not intend to be limited to what Blevins recently defined the "perpetual sunrise of methodology" [Blevins 2015] [Blevins 2016], which in his vision is the main characteristic of current digital history. In fact, the other two research questions emphasize how web materials could give us new and distinct insights into the recent past of academic institutions, thereby becoming the starting point for several new studies.

5

This paper is organized in five parts. First, a series of works on web archives and their use in humanities research, together with an overview on the research fields that study the recent past of academic institutions, are presented. Next, the types of sources employed in this work and their reliability are described and discussed. The reconstruction of the web past of the University of Bologna occupies the central part of the paper. Given the peculiarity of the exclusion issue that happened in the case of Unibo.it, the work which was conducted together with the Internet Archive to understand and solve the exclusion of Unibo.it from the Wayback Machine is then described. Finally, a new type of primary source that was collected during the research is presented, highlighting its usefulness for the communities that study academic institutions.

6

Related Work

The study presented in this paper places itself at the intersection of three different research fields. First, it discusses the consequences of the ways we preserve (or not preserve – as in the case of Italy) the web of the past. Second, it considers how web archival materials could be adopted in historical research. Finally, it highlights the impact that these sources will have on different research areas that focus on examining the recent past of academic institutions. In the next pages, a general overview of these research areas by presenting a series of related works will be offered.

7

Preserving the web of the past

In 1989 Tim Berners-Lee introduced his project at CERN, which later was identified as the "World Wide Web". In 1991 he created the first website, http://info.cern.ch/, and in the same year he publicly announced it in the Usenet newsgroup "alt.hypertext". The World Wide Web, after a slow start [Frana 2004], rapidly reached more than 16 million users[6], who were, already in 1995, the creators of a great amount of born digital traces. However, the first project that focused on the preservation of this new kind of information started only at the end of 1996 under the leadership of Brewster Kahle.

8

His utopian purpose was to archive the web in its entirety [Kahle 1997]. The project he presented under the name of "Internet Archive" has become, during the last two decades, a fundamental archival resource for the preservation of our digital past. Its crawlers started acquiring and preserving snapshots of web pages in November 1996, performing an ever-increasing uphill climb with the never-ending growth and continuous change of the web. In 2001, the Internet Archive introduced the Wayback Machine, the platform that permits the displaying and browsing (through a URL search tool) of the results of the crawl.

During the last twenty years, several other platforms, often inspired by the ideas behind the Internet Archive but with a more specific national focus, have been developed, such as Pandora in Australia (1996), the UK Web Archive (2004), Netarkivet in Denmark (2005) and the Portuguese Web Archive (2007). Moreover, in 2003, the Internet International Preservation Consortium (IIPC) was founded at the National Library of France[7] which, during the last decade, has coordinated national and international efforts to preserve Internet contents for the future. Today, with a General Assembly meeting every year since 2011 and organizations joining from 25 different countries, the IIPC has become the leading guide of these born-digital preservation projects.

The past of the Italian web sphere

Currently the national libraries of Florence and Rome are not a part of the IIPC and no project with the specific purpose of preserving the Italian web-sphere exists. In 2006, thanks to the effort of the project "Crawler" [Bergamin 2006]; [Tammaro 2006], which was supported by the "Biblioteca Digitale Italiana" (Italian Digital Library), Italy cooperated with the European Archive Foundation (now called "Internet Memory Foundation") and conducted its first wide-spread crawling of the ".it" domain[8]. However, no other project has been developed following this and the only part of its national web-sphere which has been constantly crawled and preserved are the PhD theses repositories of Italian universities, thanks to the activities of the "Magazzini Digitali" project.[9]

For these reasons, researchers interested in diachronically studying the Italian web sphere can currently rely only on the snapshots of websites preserved on the Internet Archive[10]. However, as Unibo.it was excluded from the Wayback Machine, this issue threatened to leave no trace of the web past of this academic institution.

Studying the web of the past

There are two different ways of considering web archive materials as primary sources in historical studies. The first, described for example by Brügger [Brügger 2012a], has its roots in the fields of media and Internet studies and aims at examining the web of the past by contextualizing and understanding changes in layout, structure, content and use. In 2010, Brügger edited the first book on the topic, and the title of the volume, *Web History*, clearly highlighted the research topic of the community [Brügger 2010]. A similar focus emerges upon reading the objectives of a new journal titled *Internet Histories*, recently launched by Niels Brügger and others.[11]

The second way of considering these materials as primary sources has, instead, a wider spectrum of applications. As already remarked by Roy Rosenzweig [Rosenzweig 2003], born digital materials (with their *scarcity* and *abundance*) will have an impact on the entire historical profession; for this reason, web archive materials could soon become new sources for political as well as social, cultural and economic historians. Recent works, such as the studies conducted by Anat Ben David [Ben-David 2016] and Sophie Gebeil [Gebeil 2016] already show how topics such as the Yugoslav conflict and North African immigration in France could be studied fruitfully from a web archive perspective.

In both areas, researchers have explored the potential of computational methods, such as text mining and network analyses, for extracting information from large web archive collections. Examples are presented in Milligan [Milligan 2012], [Milligan 2017], Hale et al. [Hale 2014] and Holzmannet al. [Holzmann 2016].

Reliability of Web Archives and Sources

A substantial number of articles focused on the reliability of web archives and web archival sources have been published in the last years. Howell [Howell 2006] analyzed how to use the Internet Archive in research and Murphy et al.

[Murphy 2007] established the Wayback Machine as a valid tool for identifying, among other information, web page contents, "website age" and updates. However, Brügger [Brügger 2008] highlighted a series of problems in web preservation and underlined the need of what he called a "web-philology" to deal with the reconstruction of partially archived websites. More recently he also defined the resources preserved in web archives as "reborn digital materials" [Brügger 2012b], which must be considered as different objects compared to the originals. Along the same lines, Dougherty et al. [Dougherty 2010] summarized the state of the art of web archiving in relation to researchers needs. Ankerson [Ankerson 2012] remarked that web historians need to "consider broadcast[ing] historiography scholarship that grapples with questions of power, preservation, and the unique challenges of ephemeral media". Finally Huurdeman and Ben David [Huurdeman 2014a] explored how to go beyond current limitations of search tools in web archives and, with Kamps, Samar and de Vries [Huurdeman 2014b], employed a new approach to analyze hyperlinks in web archives in order to deal with the reconstruction of the unarchived web.

Past and Present of Academic Institutions

While higher education has been a fundamental component of every advanced civilization, only in medieval Europe arose a new type of public institution of learning, which is now defined as "university" [Perkin 2007]. Considering this type of institution as a political, economic and social actor initially has attracted the interest of historians, who wanted to understand how its power, role and influence changed over time, especially in relation to other actors, such as the city, the church, the national government (e.g. [Bowosky 1973];[Clark 1976]). The massive four-volume book series *A History of the Universities in Europe*, commissioned by the European University Association, edited by Hilde de Ridder-Symoens and Walter Rüegg and published between 1992 and 2011, offers an unprecedented comprehensive overview of how universities have changed what they have taught and researched, how they have been institutionalized and how they have interacted with the society. In these studies, researchers adopt a large variety of primary and secondary sources, from university-archived materials such as matriculation and graduation statistics [Brockliss 1978];[Macleod 1978] to scientific publications [Richardson 1999], from public reports [de Wied 1991] to the results of large-scale statistical analyses [Finkenstaedt 2011]. Based on these data, scholars have described and drawn conclusions on the recent history of universities on a large variety of topics, such as the way universities have managed resources, the way the admission processes have changed before and after 1970, and how the major branches of knowledge have been taught and studied.

16

While universities have been largely examined as institutions that change and evolve in relation to each other and are conditioned by (while also conditioning) political and economic powers, they are also the physical place where scientists and humanities scholars conduct their work. For this reason, academic institutions have also been examined by historians of science and technology, interested in understanding how STEM (Science, Technology, Engineering and Mathematics) have been taught and studied in universities [Fox 1993], how scientific knowledge has moved back and forth between universities and the private sector [Mahoney 1988];[Guagnini 1988], how political, economic and social actors have influenced scientific research in academia [Pancaldi 2006] and how scientists work in their laboratories [Worboys 2011].

17

Another perspective on universities and their recent past is offered through the scientometrics discipline, whose goal is to measure and analyze the impact of publications, journals and institutes, and to produce indicators that would be adopted in policy and management contexts. The use of metrics such as citation and co-citation measures [Van Raan 1997] has attracted a lot of attention from university administrations, politicians, sociologists and quantitative historians (for further discussion, see [De Bellis 2009]); additionally, the quantification of the scientific output as a measure of evaluating and comparing academic institutions has had a huge impact on their recent past, influencing hiring strategies as well as the pursuit of certain research topics and practices. In addition to bibliometric measures, more recently, a series of publications has focused on the use of word-based and topic-based approaches in order to conduct scientometrics studies (see [Lu 2012]). These contributions have expanded the type of materials that could be analyzed (e.g. not just research papers, but also the content of grant proposal or awards as in [Nichols 2014]), the methods at disposal of the research community and the points of view that can be employed.

18

Studying the Past of the University of Bologna

The University of Bologna currently has two centers specifically dedicated to the history of academic institutions, namely the "Centre for the History of Universities and Science" (CIS) and the "Interuniversity Centre for the History of Italian Universities" (CISUI). These research groups continue the tradition of the oldest center in Italy for the history of universities, the "Centre for the History of the University of Bologna", founded in 1906.

CIS was initiated in 1991 at the Department of Philosophy, as a direct consequence of the IX Century anniversary of the Alma Mater [Pancaldi 1993]. Originally, CIS was a small research group with a focus on the history of science and universities. The first volume published, *Universities and the Sciences: Historical and Contemporary Perspectives* (Pancaldi, *ibidem*), clearly highlights the goal of the group to build a bridge between the history of universities and the history of science. Among the board of directors was Gian Paolo Brizzi, a professor of modern history with a strong research focus on the history of Italian universities.

After a few years, Brizzi started a second research center completely dedicated to the history of Italian universities, this time at the Department of History and in collaboration with other institutions such as the universities of Padova, Messina, Sassari and Torino [Pomante 2010]. In the two following decades CISUI and CIS differentiated their research topics, with the first becoming a coordination structure in Italy for the study of Italian academic institutions (see for example [Negrini 1998];[Dröscher 2002]) and the second moving increasingly towards the history of science, technology and the STS. However, a few doctoral students at CIS have continued to examine the recent history of universities and their interactions with political, economic and social actors (see for example [Serafini 2011];[Parolini 2013];[Piazza 2013].

This paper is also part of an ongoing PhD research project conducted at CIS, with a focus on adopting born digital materials as primary sources for studying the past of academic institutions. In particular, the specific goal of this work is to highlight new challenges that born digital sources present, to describe the way in which they have been dealt with and emphasize how the retrieved materials could be useful to the different communities that are studying the past and present of academic institutions.

Setting up the research

In this study, different types of primary sources have been adopted, offering an overview of the role of the website for the university and its community. As a first step, information related to the website was collected from the university yearbooks and through the analysis of the university's archived records. This provided an initial understanding of the administrative role of the website (through a top-down view) and indicated the people involved in its supervision. As a second step, interviews were conducted with those who have been managing Unibo.it during the last two decades. This helped, especially, in discovering the motivations behind specific changes and to trace down who, in the early 90s, created the website and for what reasons. The analysis was then consolidated by employing information retrieved from local and national newspaper archives, such as *La Repubblica* and *Il Resto del Carlino*, student forums and Usenet discussion groups. These materials facilitated a better understanding of the role that the website has played over the years as a "bridge" between the institution and its community.

The last step of the study aimed at restoring access to the previous versions of the website. In order to do so, information currently available on the live web was collected and its availability in foreign web archives was explored. The combination of these sources offers a comprehensive perspective on the changes of the website and the political (e.g. school reforms) and educational reasons behind specific choices and decisions.

Critically Assessing Sources

Library and archive materials.

In this study different materials from the university library and archives have been used as primary resources. One that has been very useful in different steps of our work is the university yearbook. The yearbook offers a general overview of the main activities of the university during the year, highlighting its management and indicating innovative decisions as well as presenting several statistics. Professor Fausto Desalvo has been in charge of the publication of the yearbook

since the early 90s. The yearbooks are accessible online (first edition available: 1994/95[12]) and at the library of the Department of History.

Even if, especially during the 90s, only a few pieces of information regarding the website would be mentioned in the yearbook, this source has nevertheless been an essential starting point for obtaining a diachronic overview of the official teams that were managing Unibo.it. When the different teams were contacted, the goal was to conduct interviews and to collect materials related to the website, such as archived documents as well as backups.

<div style="text-align: right">26</div>

The website has been managed by four different teams in the last twenty years. However, especially during the 90s, large parts of the website were directly modified by single departments and research groups. Very little analogue archival information has been preserved by the teams and researchers that have worked on the website and its sub-sections. Even more importantly, not even a single backup of the old versions of the website has been preserved. However, this initial research helped in identifying the key people to interview.

<div style="text-align: right">27</div>

Interviews.

Given the ephemerality of born digital materials and the general lack of their preservation by the teams that worked on the website, oral memories have played a key role in this research. These direct sources have been helpful for capturing the rationale behind the changing architecture of the website[13]. For this work, the different teams who managed the main website were interviewed, together with technicians and researchers who worked on the development of the pages of various departments in the past two decades.

<div style="text-align: right">28</div>

So much has been already written about the reliability and the criticism of oral memories (see for example [Hoffmann 1994]). In this research – especially given the fact that primary materials (e.g. backups of the website) were not at our prompt disposal – assessing the validity of the collected pieces of information has not always been an easy task. Therefore I proceeded by comparing the outputs of different interviews and, when possible, validated them by using other sources, such as newspaper articles. It is important to remark here that public and private backups of emails have often been used by the interviewees in order to recollect memories of their experience in working on Unibo.it and to confirm passages of the historical reconstruction. While email backups are "waiting to become" a new primary source for historians[14], the social and ethical implications of collecting, consulting and sharing their contents to sustain an argument still must be fully discussed.

<div style="text-align: right">29</div>

Newspapers and forums.

Another way of finding information related to previous versions of Unibo.it and its role for the University of Bologna is to search newspaper archives and retrieve articles that mention or describe it. The practice of using printed media to retrieve information about the web of the past has been already described, for instance in Brügger [Brügger 2011]. In this research the digital archive of the newspaper "La Stampa"[15] was used in order to retrieve specific articles published between 1996 and 1999 that described the general use of the web by Italian universities[16]. A great role in this study has been played by local and national newspapers (such as the digital archives of *La Repubblica* and *Il Resto del Carlino*), which especially during the 90s offered an overview on the new functionalities on the website (e.g. free email account for all students, online fee payments, etc.), together with university digital magazines (*Alma2000*, *AlmaNews*, *Unibo Magazine*). However, it is important to employ news articles critically and always consider how and why a specific piece of information regarding the website was selected and published in the daily edition of a general newspaper[17].

<div style="text-align: right">30</div>

Other sources that have been employed in this study include student forums (e.g. UniversiBo) and, to go further back in time, Usenet discussions preserved by Google. While academic forums offer new materials for historians of universities interested in better understanding student life, they also present the perspective of a very small and specific subset of the academic community. In particular, in the early 90s,these online forums were mainly kept running by students (together with researchers and professors) in STEM fields, whose departments were often the first to offer access to the web.

<div style="text-align: right">31</div>

Live web materials.

While the previous versions of Unibo.it were not available through the Wayback Machine when this research was conducted, at the same time the website has always been online, offering to the user a variety of primary sources. Live web materials reveal the current role of website in the university's organization and management (e.g. attracting national and international students and researchers, promoting collaborations with the private sector, etc). Additionally, by combining documents from the website and from social media pages of the institution (such as Facebook, Youtube and Twitter profiles), we can make reasonable assumptions about the digital interactions with the larger community. While materials from social networking websites will play a fundamental role in better understanding the multidirectional communication between academic institutions and their community, it is important to remember that their suitability for historical analysis is currently under scrutiny, as several issues have been raised [Webster 2015]; [Zimmer 2015].

32

Web archive materials.

Even if the University of Bologna homepage and all its subsections were not accessible through the Wayback Machine, its sub-domains were available on the Internet Archive and have been constantly preserved in the last twenty years (i.e. *Unibo Magazine*[18])[19].In addition, sources pertaining to Unibo.it were retrieved from other national web archives. The practice of retrieving primary sources related to an Italian university website in foreign web archives could sound strange as the goal of a national web archive is precisely to preserve the web of its country. However, as this preservation process is highly complex (as described by [Brügger 2009]), from time to time part of the non-national web will also end up being unintentionally preserved. For example, to archive national web spheres in an automatic way, archivists could set up crawlers with a maximum number of hyperlinks they can follow, with a specific set of starting points. A crawler which is set to go at most 10 links away from one of these URLs could also end up crawling non-national content, as it will systematically follow all the hyperlinks. For this reason, if the University of Bologna were to organize a Summer School and the University of Amsterdam had linked it from its website, the University of Bologna website (or at least part of it) would be unintentionally preserved in the Netherlands Web Archive.

33

The critical combination of the sources presented above provided the possibility of reconstructing the changing in Unibo.it structure, emphasizing the different roles that the institution assigned to its website during the years and the way the student community interact with the website so as to establish a dialogue with the university.

34

Unibo.it : 2015 – 2002, retracing steps

The narrative below follows the path in rediscovering the past of the University of Bologna website. This first part will take us back in time, starting from what is available now on the live web to a significant change in the content and structure of the website, which implied the removal of the majority of the materials published online during the 90s.

35

The website as it is structured today

The website of the University of Bologna is currently offered in two different linguistic versions: an Italian (which is available at the URL: http://www.unibo.it/it), and an English one. Moreover, as the university has five campuses, the website is consequently divided into five subsections (for example, http://www.unibo.it/it/campus-forli). As the English version offers the translation of only a part of the website, the focus of this research will be mainly on the Italian version.

36

Figure 1. Unibo.it homepage in 2016.

This website (see Figure 1) is currently managed by two different offices: "CeSIA - SettoreTecnologie web" that takes care of the structure (called "Sistema Portale di Ateneo"), and "AAGG — UfficioPortale Internet e Intranet di Ateneo"[20] that manages the content. If we consider its subsections, such as "Didattica" (educational information) and "Ricerca" (research), its sub-domains, such as the "Unibo Magazine", and retrieve current and old abandoned department web pages[21], we can obtain an initial overview of the current status and structure of the website. This allows us to notice that large pieces of information published online by the university between the early 2000s and 2015 are still available online (for example all the course programs, the descriptions of research projects and the contracts and grants published by each School). However, in order to retrieve them, a very basic "string-matching" search tool[22] is the only tool promptly available.

The moment of transformation

To understand why most of the materials from the early 2000s is still available online, while resources from the 90s seem way more difficult to retrieve, I contacted the people who have been involved with the management of the website during the last fifteen years.

Luca Garlaschelli was the Chief of the Information/Innovation Office (CIO) at the University of Bologna between 2002 and 2012. Under his supervision the "Sistema Portale di Ateneo" was created. This is a general interface to a hierarchical organization of all the digital resources of the university that are available online, with a specific focus on enhancing the accessibility of the information at disposal[23]. As it will be presented in the next pages, this has led to a revolutionary transformation of the digital presence of the institution, which made Unibo.it a reference point for all other Italian university websites. As a matter of fact, for three consecutive years Unibo.it received the "Osc@r del web" prize as the best Italian public administration website[24].

Among several improvements of the website, this transformation required that all departments and web pages which provided information on the various degree programs change their structure and adopt a common layout and organization of their content. As an example, the Department of Classic and Medieval Philology and the Department of Computer Science had to change their URL addresses to standardized ones ("abbreviation of the name of the department" + "unibo.it"). Thus, the first one changed from: http://www.classics.unibo.it/ to http://www.ficlit.unibo.it/ and the second one from http://www.cs.unibo.it/ to http://www.informatica.unibo.it/. This transition started in 2004 and often required the creation of completely new department pages. A few departments decided to keep the older version of their sub-domains online by adding a "2" after the "www"[25] (as an example, the previously mentioned http://www2.classics.unibo.it/), while the majority simply removed the old versions of their page from the live web, for example the Department of History, whose URL was: http://www.dds.unibo.it/.

As previously described, it is still possible to retrieve a sub-domain of the University of Bologna website (such as a department page) on the Internet Archive, but only if the original URL is known, which is not a trivial issue (as described in [Nanni 2015];[Ben-David 2016]). For example, the Department of Philosophy and Communication Studies were two different departments until 2012 and the Department of Philosophy used http://www.filosofia.unibo.it as a URL even before the transition to the "Sistema Portale d'Ateneo". However, in the 90s, this department used another URL for a couple of years, http://www.sofia.philo.unibo.it which, without the memories of the people who managed the sub-section at that time, would have been be very difficult to discover.

In summary, we can identify a specific turning point in the history of the University of Bologna website. With the "Sistema Portale d'Ateneo" project and in particular with the standardization of departments pages which started in 2004[26], the website has been completely re-organized and the majority of the previous content on these pages has been deleted from the live web. However, if they have maintained the same URL or if the previous URL is known[27], the subdomain materials and their structure can be retrieved from the Internet Archive.

Exploring the "Sistema Portale"

Mauro Amico, head of the web-technologies department at CeSIA, offered a collection of seven .png images (see Appendix) that capture the most important instances in the evolution of the organization of the homepage before the current layout (1996 – 2013)[28]. This was important as a starting point for understanding how the website structure and layout changed during the years. Looking at the snapshots after 2002 we can observe that, even if a few graphical adjustments were made (the Unibo-Magazine was introduced on the left in 2004; the search tool was repositioned in the center in 2006, etc.), the structure remained more or less the same until the July 2013, when the current interface was presented. The present organization of the "Sistema Portale d'Ateneo" is the first one to be completely created by CeSIA without the supervision of Luca Garlaschelli and, along with a new graphic interface, its main characteristic is the fact that it offers for the first time the possibility of surfing the website as a specific user (a prospective student, a student, a private company, etc.) and it proposes different contents accordingly, thereby allowing better optimization and personalization of the website.

Even if these .png images give us a first idea of the different interfaces, to be able to explore again the old versions of the website, other services have to be employed. As a start, the Internet Memory Foundation offers the results of the 2006 national ".it" crawl online, but only a single snapshot of Unibo.it homepage is available (archived on the 8th of May 2006[29]) and that is not even completely preserved. As previously noted, other national web archives could have captured, from time to time, parts of the Italian web sphere. Among them, it was discovered that both the Portuguese (Arquivo) and Danish (Netarkivet) web archives have preserved parts of Unibo.it several times from 2006. These snapshots (example in Figure 2) permit us, for the first time, to explore and examine the differences in structure and content of the previous versions of the "Sistema Portale".

Figure 2. Unibo.it (2009) in a snapshot from Netarkivet.

News from its recent past

The use of primary sources from the digital archive of the newspaper *La Repubblica* has been of significant help in this study. As a matter of fact, these articles present an overview of the interactions between the university and its community, as well as insights on the key role of the website as an intermediary. It has been found, for example, that in 2003 the university introduced, on its website, the digital edition of the student-guide of the city of Bologna, as also described in a news in the *Unibo Magazine*[30]. The guide was written by Umberto Eco, Carlo Lucarelli and other renowned professors and writers. This document provides a list of useful digital resources for new students, i.e. the platform "Flash Giovani", created with the support of the municipality of Bologna and focused on the cultural activities in the city and the website "Studenti.it" which has, in the last fifteen years, become one the most important Italian online communities for high school and university students[31].

45

As described in the *Unibo Magazine*[32], since 2004 each professor has had a personal page, in which they publish course programs (and all additional materials, such as slides), their research interests and publications list[33]. Digital sources relating to the recent years of the university also allow us to discover how in 2005 the future "Prorettore per la ricerca" Dario Braga underlined the importance of starting to teach courses in English (and also Chinese and Arabic) among his "proposals for the future"[34] or how five years later, during his administration, he actively discussed[35] in a Google Group newsletter the impact of the "Gelmini" school reform with a group of professors who were collectively termed as "Docenti preoccupati" ("worried professors")[36]. Moreover, these materials gave us insight into the activities of the "Centro Studi La Permanenza del Classico" whose director was the former Rector, Ivano Dionigi[37] and showed how the *Unibo Magazine* presented itself online in 2003 (with an interview[38] of the then Rector, Pier Ugo Calzolari, who spoke the scarcity of funding for higher education and research in Italy).

46

Among all these different resources, one source deserves special mention. In May 2007, a group of activists decided to create a copy of the Unibo.it interface. They were demonstrating against the European Credit Transfer and Accumulation System (ECTS) for the evaluation of the number of hours of study. They believed that the university website could be the perfect target for their protest, in order to attract the attention of the institution. At the URL http://www.unibologna.eu/ an identical version of the homepage was available, with the description of the reasons of the protest. In a couple of weeks, the website attracted a high number of visitors and most of all the attention of the university[39], which blocked the access to it from all its computers[40]. This source is not only important in our study as it

47

documents a different and innovative way of conducting a protest against an academic institution[41], but as the fake-website has been preserved by the Internet Archive it also paradoxically offers a preserved version of Unibo.it, so that we can browse and study (see Figure 3).
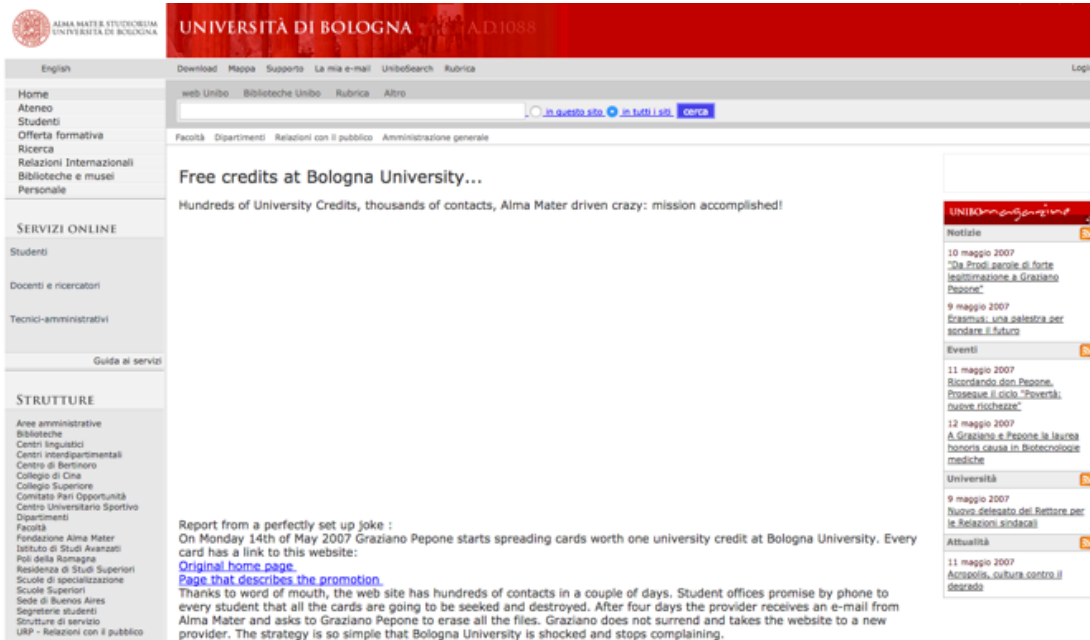

Figure 3. The cloned version of Unibo.it (2007).

The history of www.unibo.it: 2002- the early 90s

Neither material on the live web nor documents in other national web archives are available for the first ten years of history of this website. For this reason, the second part of this study will mainly employ information from local and national newspapers, which have often described new services offered by the university to its community and will combine it with archive resources (in particular from the university yearbooks). As before, a pivotal role in this study has been played by the collection and critical selection of oral memories.

Different ways of going back in time

In order to study the structure of the website before the "Sistema Portale d'Ateneo" several different sources have been employed, which will in turn help us in understanding what the website looked like, how it was used and how relevant it was in the academic digital "ecosystem".

In particular, as described earlier, the archive of the newspaper *La Repubblica* offers important information on how the website changed during the 90s. For example, it was discovered that the institution offered a free email account to all students from 2002[42] and it was the first Italian university which gave the possibility of paying fees online (2000)[43]; moreover since 1999 some departments also guaranteed the possibility of enrolling for courses and exams online[44]. Another interesting piece of news retrieved from the digital archive of *La Repubblica* is from October 2001, a few months before the project "Portale d'Ateneo" started. In those days, the University of Bologna website won the "WWW" prize from the Italian economic newspaper *Il Sole 24 Ore* for the best website in the category "School, university and research". At the ceremony Salvatore Mirabella, a technician who managed the website during the 90s, was also present[45]. However, as we can notice by looking at the images offered by CeSIA or by analyzing a few examples that are still available on the live web (i.e. http://www2.unibo.it/annuari)[46], before the "Sistema Portale d'Ateneo" project the homepage of Unibo.it was mainly an information page, presenting only a few links (see Figure 4).
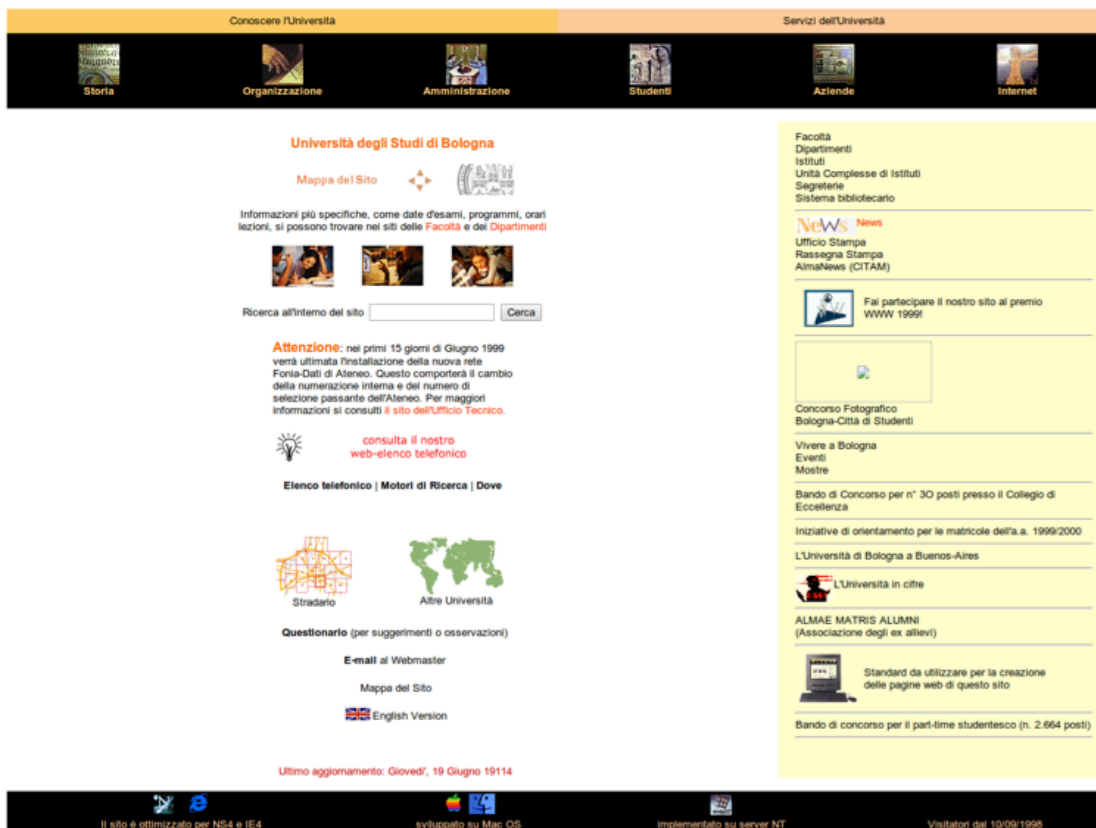
Figure 4. Unibo.it in 1999.

At the same time, consulting "The list and map of the Italian WWW servers"[47] created by Cilea and available from 1997 onwards on the Internet Archive[48], we can observe that several departments, faculties and research groups were already online and, as opposed to the relatively passive homepage, very active in the 90s. For example, we can retrieve all the information on courses in history since 1998[49], the organization of the university astronomical observatory[50] and of the Faculty of Engineering[51] since 1997, description on the inter-faculty library since October 1996[52] (the entire system was created in 1993[53]), the digitization of the students guide books carried out by the faculty of economics in 1994[54]. The Internet Archive has also preserved Unibo.it's old online magazine, *AlmaNews*, which offer several short videos of important events, such as the ceremony[55] for the first degrees in Business Administration and Political Sciences in 1997. These different pages, extremely useful for prospective and enrolled students[56], were continuously updated with new information by technicians, researchers, professors and, from time to time, also with contributions of students[57]. For these reasons, they all evolved differently during the 90s and they are now interesting instances on how the departments of this university approached the World Wide Web.

In addition to examining departmental pages, there are many other ways of looking at the second half of the 90s' history of Unibo.it. We could follow the information related to AlmaNET, the university internal Internet connection, which in 1988 was established to connect three departments and was highly improved in 1996[58], thanks to the collaboration of Telecom Italia and under the supervision of CeSIA (which was created in 1994)[59]. Another perspective on the recent past of this institution could be developed by the examination of the impact of the online service AlmaLaurea, presented in May 1998[60], which aimed at improving the relationship between the institution, its student community and the job-market. A third point of view might be focused on the relation and mutual influence between the university and the municipality, by considering the role played by the Internet. In fact, the city of Bologna and its citizens have a strong bond with innovation in computing technologies, with the municipality for instance having created one of the first civic-networks in the world in 1995, giving to all citizens free access to the Internet the very next year [Chiara 1998]. The early importance of the web for Bologna citizens appeared also in a 1996 article retrieved from *La Repubblica* digital

archive. As mentioned in the piece of news [Venturi 1996], in November of that year for the first time in Italy, a digital discussion was censored and an entire mailing list named "Lisa" was completely closed. This happened on the Unibo server[61]: CeSIA informed the professor of computer science Dario Maio of the presence of violent debates on the platform and the Department of Computer Science decided to drastically intervene. As the article reported, these digital conflicts were probably related to the internal discussion of an Italian association named "La città invisibile"[62] ("The invisible city"). This association, comprising early Internet activists, was interested in sharing the importance of digital cultures and rights. Among them there were also academics, for example Lucio Picci (currently professor of Political Economy[63]), who was at that time a young researcher at the University of Bologna[64].

It is evident that diachronically examining the digital alter ego of the institution and using these resources to extend our knowledge on its recent past is a complex challenge, which relies on both an interdisciplinary set of methodological approaches and specific research questions. The examples presented above only highlight some specific perspectives on the topic, which I have encountered while examining the collected sources. While each one of the aspects previously presented would offer a different insight on the early use of the web by the institution, the last part of this section will focus on a different task: tracing the origin of the university website. This will help us in understanding the process of creation of Unibo.it: who was responsible for it, for what reasons was it created and in which context. Old websites hold interesting stories on their origins, which are often placed at the intersection of academic research, curiosity in advanced digital technologies (together with aspiration of contributing to them) and mutual human desire of communicating with others. Unibo.it is one of them.

At the beginning of the digital era

Tracing the first online presence of an entity such as the University of Bologna has not been an easy task. In Italy a list of .it servers was initially maintained by the research center CNUCE (Centro Nazionale Universitario di Calcolo Elettronico) and is currently available on the website Registro.it[65]. However all early Italian websites (created before 1996) have a common creation date: 29-01-1996 (Figure 5).

Dominio		Dominio	
Dominio:	unibo.it	**Dominio:**	crs4.it
Stato:	ok	**Stato:**	ok
Data Creazione:	29-gen-1996 0.00.00 CET	**Data Creazione:**	29-gen-1996 0.00.00 CET
Data Scadenza:	2-mag-2015 CET	**Data Scadenza:**	29-gen-2016 CET
Data Aggiornamento:	18-mag-2014 0.53.11 CET	**Data Aggiornamento:**	14-feb-2015 0.52.48 CET

Figure 5. Unibo.it and Crs4.it (which is considered to be the first Italian website[66]) have the same starting date.

The research group "GARR-Network Information Retrieval" organized a series of annual meetings in the early 90s[67], dedicated to the spread of the World Wide Web in Italy at the university level. Consulting the proceedings of 1994, I learned[68] that Unibo.it was already active at least in the August of the year before. The fact that the university was fostering the use of the Internet and of web materials could be also deduced by consulting the 1993/94 yearbooks, where the importance of AlmaNet is mentioned, as well as the need of adopting emails as a form of communication and online databases as new resources.

During the first years of the World Wide Web Tim Berners-Lee curated a list of web-servers on the CERN website; the last update available is from late 1992[69]. Unibo.it is not mentioned in this list, but there is a link to another Italian research institution, the Physics Institute in Trieste. Later, on the NCSA website, a specific section called "What's New!" published a list of the new servers on the web each month (from June 1993 to January 1996)[70]. By consulting it, some interesting information about specific sub-sections of Unibo.it was found: for example the "Bologna Astrophysics

Preprints" has offered online, since November 1994, all the scientific publications of the Bologna Astronomical Observatory (OAB), the Astronomy Department of Bologna University (DDA), the Radio astronomy Institute of CNR (IRA) and the TESRE Institute of CNR (ITE). However, for what concerns specifically the creation date of the website, in December 1993 a link to a map of all Italian web-servers was published, but this link is not available anymore (it redirects to the 1997 version of the Cilea Map). Summarizing then, by consulting born digital materials as well as traditional archival sources, we know that Unibo.it was already available in the second half of 1993, and that the website was created after the end of 1992, according to the CERN web-server list.

University websites have usually been created by researchers who were already using the Internet in their work. For this reason, departments and research centers in computer science[71] and physics[72] are generally good starting points for discovering who created the website of a specific institution. However, in Bologna, the university website was created in a different place, namely at the Department of Mathematics, thanks to the collaboration between a Turkish professor who had at that time arrived from the United States, and a young Italian researcher.

<div style="text-align: right">57</div>

The story of the origins of Unibo.it emerged in an interview conducted with Renzo Davoli. Davoli is currently professor of Computer Science at the University of Bologna; in the early 90s he was working under the supervision of Ozalp Babaoglu, who arrived in Bologna in 1988 from Cornell University and wanted to use the Internet – among other things – to stay in touch with his colleagues and friends from abroad. Given the fact that Bologna did not have a Department of Computer Science at that time, Babaoglu and Davoli were working at the Department of Mathematics. In 1988 the two of them established the second Italian node to the Internet[73], from the Department of Mathematics to CNUCE, in Pisa[74] and Davoli became the person in charge of the University TCP/IP network. They then became part of AlmaNet, the internal network initially established between the Departments of Mathematics, Engineering and Physics. In the following years, AlmaNet played an essential role in connecting university departments, especially the ones located outside the city.

<div style="text-align: right">58</div>

Interactions between the departments were again improved thanks to the advent of the World Wide Web. Departments were the first to be online and, once again, this was accomplished thanks to Babaouglu and Davoli. As a matter of fact, in July 1993, the two researchers registered and created the web pages of the domains "cs.unibo.it" (Computer Science) and "dm.unibo.it" (Department of Mathematics). This helped colleagues in other departments understand the huge potential of the web. Initially, Davoli and Babaoglu managed the main website as well, which then passed under the supervision of the Public Relation Office, and in particular of Salvatore Mirabella[75].

<div style="text-align: right">59</div>

Working with the Internet Archive

In the previous sections, it has been highlighted how the digital past of an institution could be re-discovered without the prompt availability of Internet Archive snapshots. The following paragraphs will describe the work I conducted, for understanding why Unibo.it was excluded from the Wayback Machine.

<div style="text-align: right">60</div>

In order to understand the reasons of the removal of Unibo.it, the first step was to find, in the exclusion-policy of the Internet Archive, information related to the message "This URL has been excluded from the Wayback Machine", which appeared when "http://www.unibo.it" was searched. As described in the FAQ section of the Internet Archive, the most common reason for this exclusion is when a website explicitly requests to not be crawled by adding "User-agent: ia_archiver Disallow: /" to its robots.txt file[76]. However, it is also explained that "Sometimes a website owner will contact us directly and ask us to stop crawling or archiving a site, and we endeavor to comply with these requests. When you come across a 'blocked site error' message that means that a site owner has made such a request and it has been honored. Currently there is no way to exclude only a portion of a site, or to exclude archiving a site for a particular time period only. When a URL has been excluded at direct owner request from being archived, that exclusion is retroactive and permanent".

<div style="text-align: right">61</div>

When a website has not been archived due to robots.txt limitations, a specific message is displayed: "Page cannot be crawled or displayed due to robots.txt", which is different from the one that appeared when searching the University of Bologna website (as shown in Figure 6). Therefore, the only possible conclusion is that someone had explicitly

<div style="text-align: right">62</div>

requested to remove the University of Bologna website from the Archive.

Figure 6. Two different exclusion messages (left University of Trento, right University of Bologna).

Before moving on, it is essential to mention that, when a website is excluded due to robots.txt, its pages are not preserved by the Internet Archive. In the second situation, as it will be presented in the next paragraphs, it was instead discovered that the Internet Archive has continued to preserve the website (despite what is described in the FAQ section), which was simply not available for any kind of consultation through the Wayback Machine. Given the specificity of the exclusion message, I decided to consult CeSIA, the team that has supervised Unibo.it during the last decades, regarding this issue. However, they did not submit any removal-request to the Internet Archive and they were not aware nor had any trace in digital and archival documents of anyone submitting it. To clarify this issue, the Internet Archive team was then contacted. Thanks to the efforts of Mauro Amico (CeSIA), Raffaele Messuti (AlmaDL - Unibo), Christopher Butler (Internet Archive) and Giovanni Damiola (Internet Archive), a collaboration with the Internet Archive started at the end of March 2015. As we contacted Butler, he told us that the Unibo.it case was similar to another one that involved the New York government websites[77]. With their help, I discovered that a removal request regarding the main website and a list of specific subdomains had been submitted to the Internet Archive in April 2002. Thanks to this collaboration, the university website became available again on the Wayback Machine on the 13th of April 2015 (see Figure 7). This also gives us the opportunity of attesting that the Internet Archive has kept preserving Unibo.it in the last fifteen years; the website was simply not available for any consultation. Additionally, having the website at our disposal, once again, gave me the opportunity of re-evaluating the findings of this study.
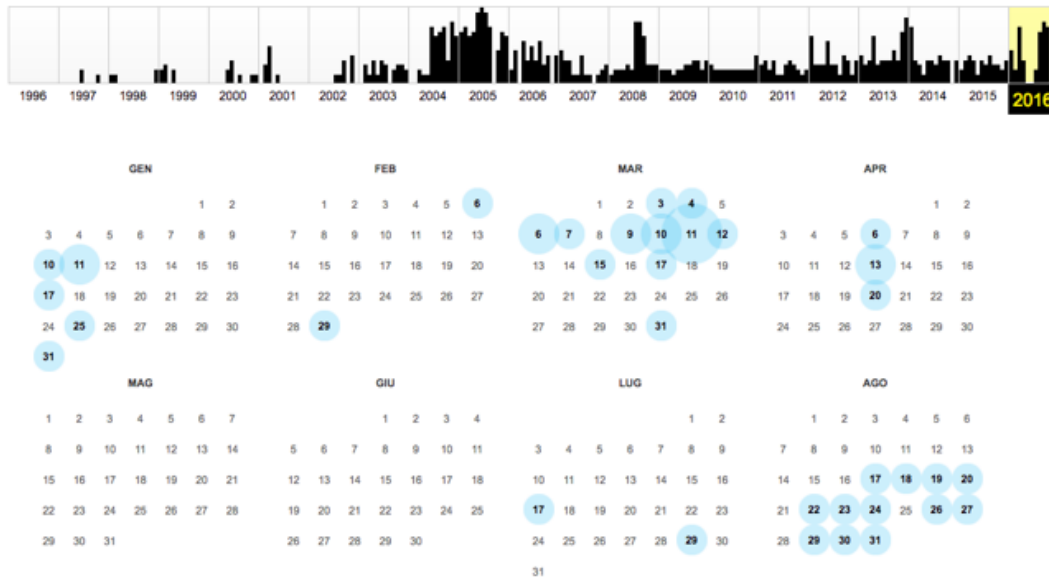
Figure 7. Unibo.it is once again available on the Wayback Machine.

While the exclusion issue was solved, it was necessary to investigate its causes further. As it has already been described, in 2002 the administration of Unibo.it completely changed, during a general re-organization of the digital presence of the university (the Portale d'Ateneo project). Therefore, while it is evident that this request was made by someone who was in the position to ask for the removal of the website[78] and who knew how the Internet Archive exclusion policy works[79], it still remains entirely unclear to us who, in that very same month, could have been in the position to submit this specific request, and the reasons behind it. Even though several years have passed by, it was assumed that someone involved in the administration of the website would have remembered (or had traces in a backup) this email exchange with a team of digital archivists in San Francisco. Between April and June 2015 a last series of interviews was conducted with several people involved in the Unibo.it website, pre- and post the 2002 reorganization. However, it was impossible to retrieve any information on this issue.

64

As the specificity of the request is the only hint that could help in identifying its author, I decided to analyze the different urls in more detail. The majority of them are server addresses (identified by "alma.unibo"), while the other pages are subdomains of the main website, for example estero.unibo.it (probably dedicated to international collaborations). A few questions therefore remain unsolved: why would someone want to exclude exactly these pages and not all the department pages, which were active online, at that time? Why exactly these four subdomains were selected and not the digital magazine *Alma2000* (alma2000.unibo.it) or the e-learning platform (www.elearning.unibo.it)?

65

A new primary source

In the previous sections, the paper has presented a) how the past of Unibo.it was reconstructed without having snapshots from the Wayback Machine at prompt disposal and b) how this process brought new light on the interaction between the university and its community. This paper also stressed that different research communities are currently studying the recent past of academic institutions and that web materials could open for them new perspectives. For this reason, in this final section I will highlight a specific type of primary source that was collected during this study that could become useful for obtaining new insights in these fields.

66

Syllabi

The history of universities has traditionally focused on topics such as the role of academic institutions in processes such as nation building as well as on the influence of governments (through economic and political decisions), on the experience and life of students as well as on the relationship of institutions and professional realities. Another topic that has attracted significant attention is the examination of universities as institutions of higher education, where the manner in which topics are taught and education is provided are also highly influenced by several political, economic and social factors.

Examining what has been taught within a given timeframe at a specific academic institution and understanding the global and local reasons that influenced specific changes is a topic that has attracted the attention of many historians of universities. For example, [de Ridder-Symoens 1992-2011], provides a complete overview of the most recurrent teaching topics in European academic institutions and how and why they have changed in the last centuries. Traditionally, historians collect this information by examining different sources, such as student transcript of lessons and professor's notes[80] as well as university yearbooks and the reasons behind widespread adoption of specific textbooks. In recent years, Dan Cohen [Cohen 2006]; [Cohen 2011] has considered the large and still unexplored potential that online syllabi could offer to the study of academic teaching. The first digital-historical work on the topic has been published in 2005 in the *Journal of American History* [Cohen 2005]. The goal of the paper was to show how the teaching of history in U.S. universities is still strongly based on textbooks. These findings, which are directly in contrast to a round table discussion published in the same journal four years prior [Kornblith 2001], were obtained thanks to a large study of around 800 syllabi available on the web.

While Cohen's work is a first step in this direction, the potential of online syllabi goes beyond his study and could rapidly affect the practices, topics and findings of historians of universities as well as the scientometrics community. During the reconstruction of the past of Unibo.it, it was possible to collect all syllabi published online by the institution, which in the case of the University of Bologna are in the number of thousands for each academic year[81]. This type of documents will further allow to conduct large-scale analyses on what have been, year after year, the topics that the University (and each single school and department) decided to focus on in its educational programs. Results of these studies will be useful for obtaining a wider perspective in a scientometrics setting (by considering both the input and output of an academic institution) and will also act as the starting point for addressing questions regarding the underlying cultural, economic and political factors that condition specific changes.

Conclusions

The aim of this paper has been to highlight both the issues and the potentialities of using born digital documents to study the recent past of the University of Bologna. The main focus was to describe the methodological approach employed in order to reconstruct its website (which has been excluded for the last thirteen years from the Wayback Machine). In doing so, the paper underlined how its history is divided into two parts (before and after the setting up of the "Sistema Portale d'Ateneo") and how different sources (the yearbooks, materials from foreign web archives, document preserved by CeSIA, articles on local, national and digital newspaper) have been useful to improve our knowledge on the metamorphosis of this website (specifically, on the role of department pages). This work also examined how born-digital sources can offer new insights on common research topics related to the history of this university and its relation with the students' community and the city itself.

The different issues presented in this paper highlight the need of an even more interdisciplinary approach for future historians. In the field of Internet studies and digital archiving, researchers are already discussing the importance of new ways of conceiving the retrieval, analysis, criticism and employment of born digital primary sources. As historians, we should openly join this discussion with both theoretical contributions as well as concrete examples. As a matter of fact, these materials will sustain traditional historical research questions and will lead to an infinite number of new ones.

Appendix

Figure 8. Unibo.it between 01/1996 and 01/1998.

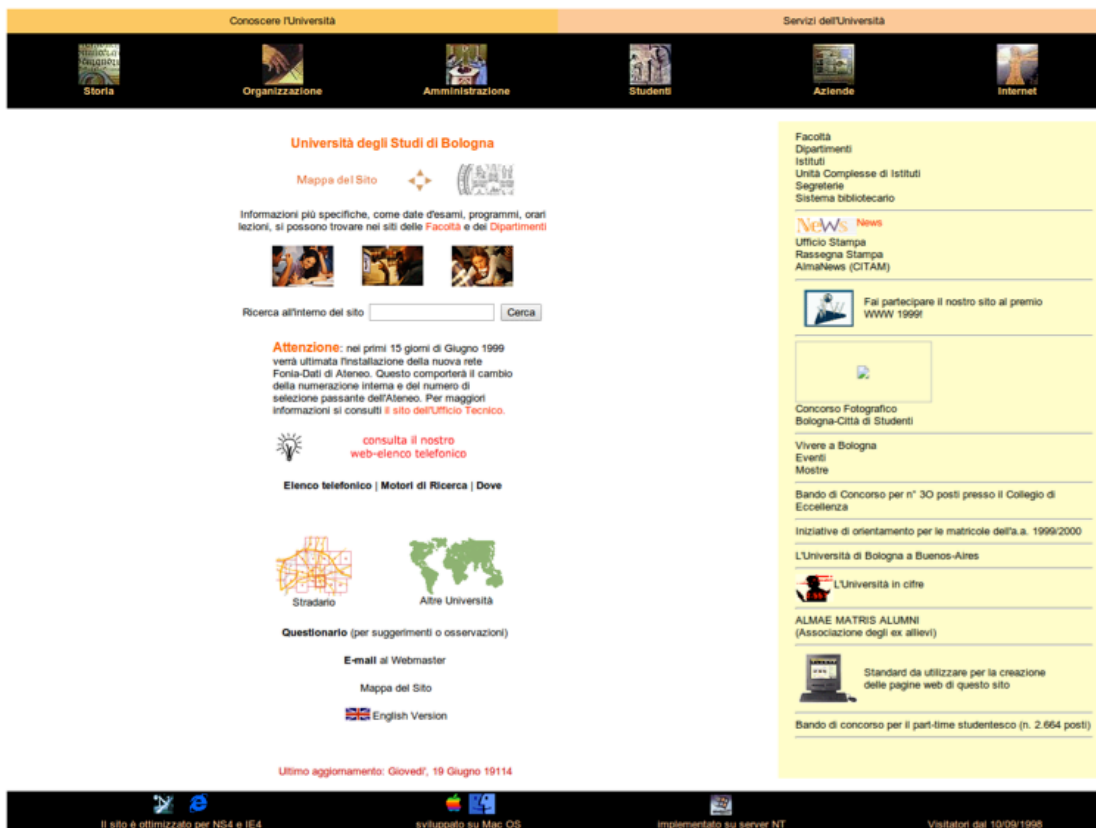**Figure 9.** Unibo.it between 01/1998 and 09/1998.

Figure 10. Unibo.it between 09/1998 and 07/1999.

Figure 11. Unibo.it between 2002 and 2003.

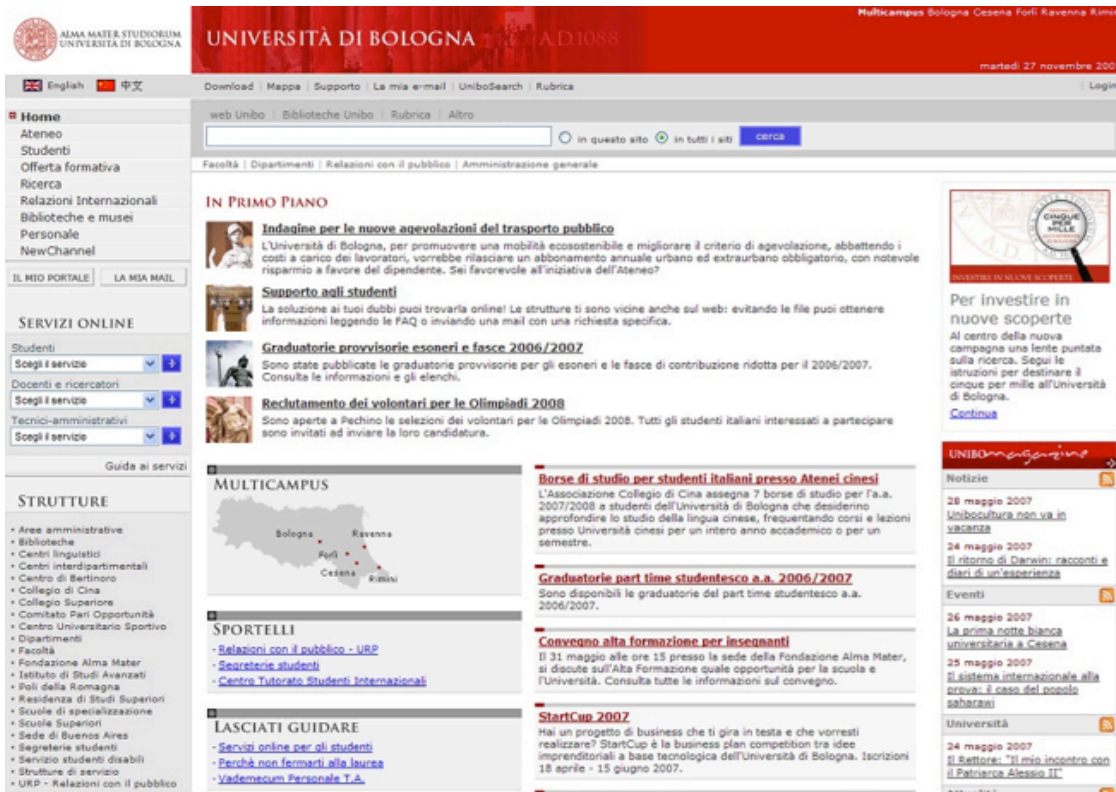**Figure 12.** Unibo.it between 2004 and 2006.

Figure 13. Unibo.it between 2006 and 2009.

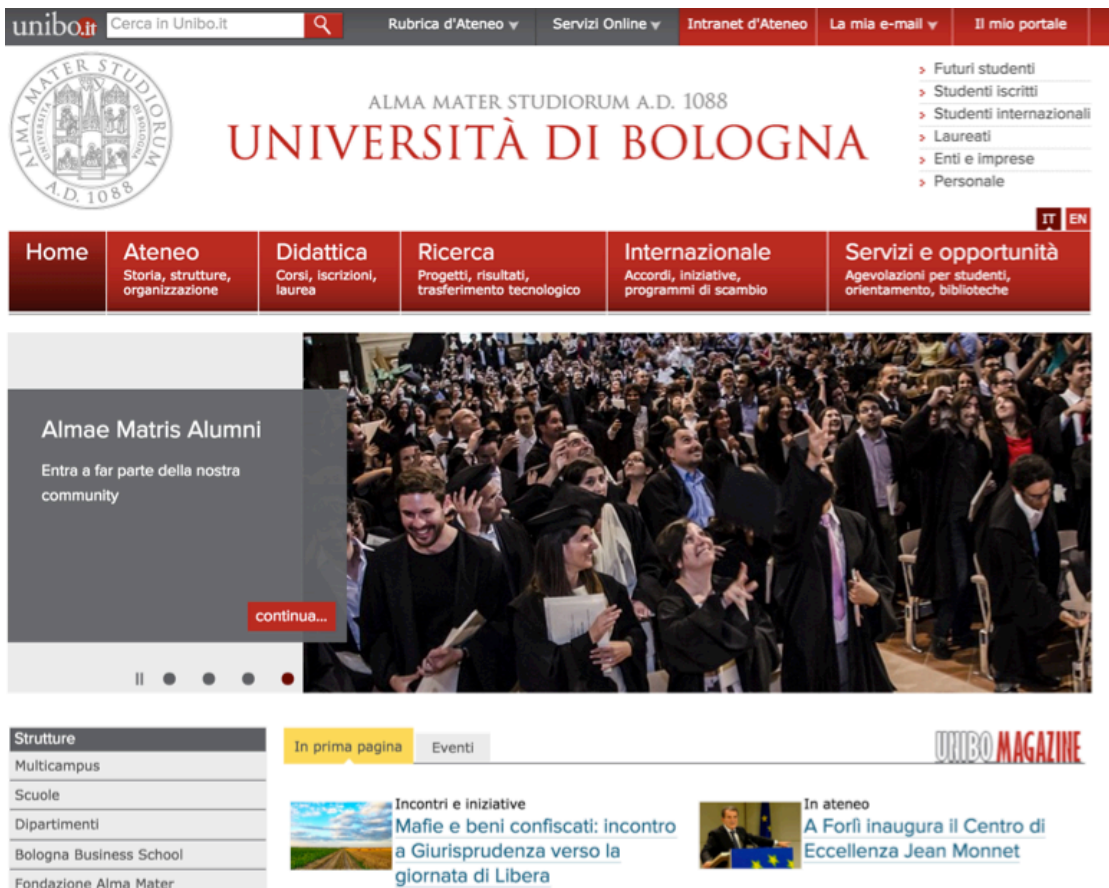**Figure 14.** Unibo.it between 2009 and 07/2013.

Figure 15. Unibo.it between 07/2013 and 03/2015 (Snapshot taken on the 17th of March 2015).

Notes

[1] Consider also how the university presents itself on its website:http://www.unibo.it/en/university/who-we-are/our-history/university-from-12th-to-20th-century

[2] Milligan raised the question during his talks, both at the 2015 meeting of the International Internet Preservation Consortium and at the 2016 meeting of the American Historical Association.

[3] Translation from: Bloch, M. The Historians' Craft, Manchester Univ. Press (1992).

[4] All the URLs mentioned in this research have been most recently checked on the 14th of October 2016.

[5] Gomes, Miranda and Costa have curated a Wikipedia page precisely on the topic: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives

[6] http://www.internetworldstats.com/emarketing.htm

[7] http://netpreserve.org/about-us

[8] This was conducted between May and June 2006, the snapshots are available here: http://collection.europarchive.org/bncf/

[9] http://www.depositolegale.it/

[10] Or, as it will be described later, in some specific cases on snapshots archived in other national web archives.

[11] http://www.tandfonline.com/loi/rint20

[12] http://www.unibo.it/annuari/

[13] At the 2015 International Internet Preservation Consortium General Meeting (IIPC2015), the importance of oral memories for web historical research has been emphasized both by Ahmed AlSum and by the author of this paper in two consequential presentations: https://www.youtube.com/watch?v=AHrxvRWf9OM.

[14] Dan Cohen [Cohen 2006] discussed it when considering the large abundance of sources that public administration will leave us in the next decades.

[15] http://www.archiviolastampa.it/

[16] E.g. the article "Anche l'università via Internet", written by Giovanna Favro and published the 14th of May 1998.

[17] For example this short article on the possibility of creating university email accounts in 2002: http://ricerca.repubblica.it/repubblica/archivio/repubblica/2002/10/09/mail-gratuita-per-gli-studenti.html?ref=search

[18] https://web.archive.org/web/*/http://magazine.unibo.it

[19] The reasons will be discussed in the final section of this paper.

[20] http://www.unibo.it/it/ateneo/organizzazione/amministrazione-generale/81380/817/index.html

[21] http://www2.stat.unibo.it/; http://www2.classics.unibo.it/

[22] http://search.unibo.it/UniboWeb/UniboSearch/Default.aspx

[23] As underlined by Garlaschelli: http://www.osservatoriosullacomunicazione.com/mezzi/internet/prontoweb/interviste/garlaschelli.php. And also presented in the "Annuario degli anni accademici 2003-2004 e 2004-2005", pp. 777-780

[24] http://www.magazine.unibo.it/archivio/2007/oscar_del_2007. In 2007 Luigi Nicolais, the Italian Minister of Public Administration, was also present to confer this honor.

[25] This could be due to a personal choice of the person who was managing each department page at that time and not to a decision of the CIO. On the Department of Classic and Medieval Philology homepage it is explicitly written that "the pages will continue to be available, but will be no longer updated".

[26] http://www.magazine.unibo.it/archivio/2004/dipartimenti

[27] A web page archived in 2002 could help us identify the URL of each department in that year: https://web.archive.org/web/20020224030346/http://alma2000.unibo.it/facolta/dipE.asp

[28] They cover the periods: 01/1996-01/1998; 01/1998 – 09/1998; 09/1998-07/1999; 2002-2003; 2004-2206; 2006-2009; 2009-2013.

[29] http://collection.europarchive.org/bncf/20060508021404/http://www.unibo.it/Portale/default.htm

[30] http://www.magazine.unibo.it/archivio/2003/09/22/guida-di-bologna

[31] Currently it is one of the 200 most visited websites in Italy: http://www.alexa.com/siteinfo/studenti.it#trafficstats

[32] http://www.magazine.unibo.it/archivio/2004/pagina-personale-docente/

[33] The main pages of professors have been also excluded from the Wayback Machine; other national web archives have preserved just a few of them.

[34] http://ricerca.repubblica.it/repubblica/archivio/repubblica/2005/04/13/proposte-per-ateneo-del-futuro.html?ref=search

[35] https://groups.google.com/forum/#!topic/docentipreoccupati/WqOJqQzkPmU

[36] Another interesting source to study the experience of Dario Braga as Prorettore and its run for the future Rettore of the university, will be his personal blog: http://www.dariobraga.com/blog

[37] http://ricerca.repubblica.it/repubblica/archivio/repubblica/2006/10/25/umanisti-scienziati-insieme-alla-stessa-lezione.html?ref=search

[38] http://www.magazine.unibo.it/archivio/2003/11/11/intervista-al-rettore

[39] http://www.magazine.unibo.it/archivio/2007/attacco_al_portale

[40] http://ricerca.repubblica.it/repubblica/archivio/repubblica/2007/06/13/clonato-il-sito-dell-ateneo-per-protesta.html?ref=search

[41] To know more on cyber protests see Van Laer and Van Aelst, 2009.

[42] http://ricerca.repubblica.it/repubblica/archivio/repubblica/2002/10/09/mail-gratuita-per-gli-studenti.html?ref=search

[43] http://ricerca.repubblica.it/repubblica/archivio/repubblica/2000/07/17/tasse-online-per-universita.html?ref=search

[44] http://ricerca.repubblica.it/repubblica/archivio/repubblica/1999/09/13/finalmente-offerta-portata-di-mouse.html?ref=search

[45] He was the head of "Urp – Servizio Web", as described here:
http://www2.unibo.it/Annuari/Annu9901/Indice/parte2/parte2sez1/parte2sez1.html

[46] It is important to note that the page http://www2.unibo.it is not available on the live web anymore and it was excluded from the Wayback Machine.

[47] This is a useful starting point for every researcher who is interested in the past of the Italian web sphere.

[48] http://web.archive.org/web/19971025045601/http://www.cilea.it/WWW-map/

[49] https://web.archive.org/web/19981206110539/http://www.dds.unibo.it/

[50] http://web.archive.org/web/19970114105744/http://www.bo.astro.it/

[51] https://web.archive.org/web/19970422153341/http://www.ing.unibo.it/

[52] https://web.archive.org/web/19961031164155/http://liber.cib.unibo.it/

[53] https://web.archive.org/web/20010424195903/http://www.cilea.it/collabora/GARR-NIR/nir-it-2/atti/cib.html

[54] https://web.archive.org/web/20010424195734/http://www.cilea.it/collabora/GARR-NIR/nir-it-2/atti/didbo.html

[55] http://web.archive.org/web/20050204125414/http://www.almanews.unibo.it/97_98/Html/Economia.htm

[56] https://groups.google.com/forum/#!search/%22esami%22$20%22unibo%22/it.reticiviche.bologna.universita/Ndj7QmZ2Iqc/1rTwBNK0gnMJ

[57] https://web.archive.org/web/19980117121355/http://caristudenti.cs.unibo.it/index.shtml

[58] https://web.archive.org/web/19990503060118/http://naomi.bo.astro.it/~federico/almanet/smds/smdspage.html

[59] Additional materials on this topic can be found in the bibliography dedicated to "Internet in Italy" edited by Riccardo Ridi on his website :
http://www.riccardoridi.it/esb/biblint/04.htm

[60] http://web.archive.org/web/20020225134330/http://www.almanews.unibo.it/Alma98.htm

[61] http://ricerca.repubblica.it/repubblica/archivio/repubblica/1996/11/24/troppe-parolacce-censurata-lisa.html?ref=search

[62] http://www.citinv.it/intro.html

[63] http://www2.dse.unibo.it/picci/

[64] http://www2.dse.unibo.it/picci/sfera/storia.html

[65] http://www.nic.it/

[66] The website of the "Centro di ricerca, sviluppo e studisuperiori in Sardegna" (CRS4 – www.crs4.it) was created between the second half of 1992 and November 1993. In different occasions the team behind it reported different dates, for example in [Colletti 2014]; [CRS 2015]; [Corriere della Sera 2015].

[67] http://web.archive.org/web/19971025045540/http://www.cilea.it/GARR-NIR/

[68] http://web.archive.org/web/20010424195903/http://www.cilea.it/collabora/GARR-NIR/nir-it-2/atti/cib.html

[69] http://www.w3.org/History/19921103-hypertext/hypertext/DataSources/WWW/Servers.html

[70] http://web.archive.org/web/19961220071416/http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/archive-whats-new.html

[71] https://web.archive.org/web/19970518021303/http://www.utexas.edu/teamweb/history/

[72] https://swap.stanford.edu/19911206000000/http://slacvm.slac.stanford.edu/FIND/default.html

[73] The first node is from the research institute CNUCE, in Pisa, as described here: http://www.30annidirete.it/.

[74] This information is also offered by the The Internet Engineering Task Force (IETF) : https://tools.ietf.org/html/rfc1117#ref-AA62

[75] http://www.unibo.it/annuari/Annu9597/final/c2/p1/sp1/index.html

[76] The Internet Archive follows the "robots.txt protocol", which is a convention of advising web crawlers and other web robots to access only parts of a website which is otherwise publicly viewable.

[77] http://www.villagevoice.com/news/why-were-new-york-government-websites-hidden-from-an-internet-archive-for-13-years-6721316

[78] The Internet Archive says "the website owner" and, even if they happened to be not absolutely rigid on this point, it has to be someone at least involved in the management of the website.

[79] As he/she explicitly declared a specific list of subdomains to remove (as described above, the Internet Archive excludes urls and their subsections – not subdomains).

[80] For example, a great resource to know what Professor Pasquini taught about Dante at the University of Bologna are the notes from its course from the academic year 1992-93.

[81] By analyzing the "Portale", it is possible to collect all syllabi since the academic year 2004-2005. Then, by using Internet Archive snapshots of retrieved department's pages, it is possible to identify all courses programs for several departments, for example all syllabi of courses in History since 1998-99.The dataset I collected is available at: https://federiconanni.com/syllabi-unibo/

Works Cited

Ankerson 2012 Ankerson, M. S. *Writing Web Histories with an Eye on the Analog Past*. New Media and Society (2012).

Barbagli 2009 Barbagli, M., Colombo, A., Orzi, R. *Gli studenti e la città: primo rapportosuglistudenti dell'Università di Bologna*. Bologna University Press (2009).

Ben-David 2016 Ben-David, A. *What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain*. New Media and Society (2016).

Bergamin 2006 Bergamin, G. *La raccolta dei siti web: un test per ildominio "punto it"*. DigItalia (2006).

Blevins 2015 The Perpetual Sunrise of Methodology (2015). http://www.cameronblevins.org/posts/perpetual-sunrise-methodology/.

Blevins 2016 Blevins, C. "Digital History's Perpetual Future Tense", *Debates in the Digital Humanities* (2016).

Bloch 1949 Bloch, M. *Apologie Pour l'Histoireou Métier d'Historien* (1949).

Bowosky 1973 Bowsky, W. M., Baldwin, J. W., Goldthwaite, R. A. *Universities in Politics: Case studies from the Late Middle Ages and Early Modern Period* (1973).

Brizzi 1991 Brizzi, G P. "La presenza studentesca nelle università italiane nella prima età moderna. Analisi delle fonti e problemi di metodo", In Brizzi and Varni Angelo (eds.), *L'università in Italia fra età moderna e contemporanea*. Aspetti e momenti, Clueb (1991).

Brizzi 2007 Brizzi, G. P., Del Negro, P., Romano, A. *Storia delle Università in Italia, I-III*. Sicania by Gem (2007).

Brockliss 1978 Brockliss, L. 'Patterns of Attendance at the University of Paris, 1400–1800", *The Historical Journal* (1978).

Brügger 2008 Brügger, N. "The Archived Website and Website Philology: A new type of Historical Document?", *Nordicom Review* (2008).

Brügger 2009 Brügger, Niels. *Website History and the Website as an Object of Study*. New Media and Society (2009).

Brügger 2010 Brügger, Niels (eds). *Web history*. Peter Lang (2010).

Brügger 2011 Brügger, Niels. "Web Archiving-Between Past, Present, and Future", In M. Consalvo and C. Ess (eds) *The Handbook of Internet Studies*. Wiley-Blackwell (2011).

Brügger 2012a Brügger, N. *When the Present Web is Later the Past: Web Historiography, Digital History, and Internet Studies*. Historical Social Research/HistorischeSozialforschung (2012a).

Brügger 2012b Brügger, N. *Web Historiography and Internet Studies: Challenges and Perspectives*. New Media and Society (2012b).

Brügger 2014 Brügger N. *Probing a Nation's Web Sphere: A New Approach to Web History and a New Kind of Historical Source*. WebSci (2014).

Brügger 2016 Brügger, N. "Digital Humanities in the 21st Century: Digital Material as a Driving Force", *Digital Humanities Quarterly* (2016).

CRS 2015 CRS4. *CRS4, Il Centro Di Ricerca Sardo Compie 25 Anni* – Comunicato Stampa, 30/11/2015.

Capitani 1987 Capitani, O. (eds). *L'Università a Bologna: Personaggi, Momenti e Luoghi dalle Origini al XVI Secolo*. Cassa di Risparmio (1987).

Chiara 1998 Chiara, Stefano. *La Telematica e la Città.* Il Progetto Iperbole a Bologna, MA Thesis in "Comunicazioni di massa", University of Bologna, A.A. 1997-98, (1998).

Clark 1976 Clark, B. R. and Youn, T. I. *Academic Power in the United States: Comparative Historic and Structural Perspectives*, Research Report no. 3 (1976).

Cohen 2005 Cohen, D. "By the Book: Assessing the Place of Textbooks in Us Survey Courses", *The Journal of American History*, 91(4), (2005).

Cohen 2006 Cohen, D. "When Machines Are the Audience", *Personal blog* (2006). http://www.dancohen.org/2006/03/02/when-machines-are-the-audience/.

Cohen 2011 Cohen, D. "A Million Syllabi", *Personal blog* (2011) http://www.dancohen.org/2011/03/30/a-million-syllabi/.

Colletti 2014 Colletti G. *Sardegna Apripista Di Internet. Pietro Zanarini, Pioniere Della Rete: "Il Web Italiano Nacque Qui"*, La Repubblica, 6/07/2014.

Corriere della Sera 2015 Corriere della Sera, Internet e Ricerca: I 25 Anni del Crs4, Centro d'Eccellenza in Sardegna, 23/12/2015.

De Bellis 2009 De Bellis, N. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Scarecrow Press (2009).

Dougherty 2010 Dougherty, M., Meyer, E. T., Madsen, C. M., Van den Heuvel, C., Thomas, A., Wyatt, S. *Researcher Engagement with Web Archives: State of the Art*. Final Report for the JISC-funded project "Researcher Engagement with Web Archives" (2010).

Dröscher 2002 Dröscher, A. *Le Facoltà Medico-Chirurgiche Italiane: 1860-1915: Repertorio Delle Cattedre E Degli Stabilimenti Annessi, Dei Docenti, Dei Liberi Docenti E Del Personale Scientifico.* Clueb (2002).

Finkenstaedt 2011 Finkenstaedt, T. "Teachers", In Walter Rüegg (eds) *A History of the University in Europe, Vol. IV: Universities since 1945*. Cambridge University Press, Cambridge (2011).

Fox 1993 Fox, R. and Guagnini, A. (eds) *Education, Technology and Industrial Performance in Europe, 1850-1939*. Cambridge University Press, Cambridge (1993).

Frana 2004 Frana, P. L. "Before the Web There Was Gopher", *IEEE Annals of the History of Computing*, 26(1) (2004).

Gandolfi 2011 Gandolfi, G. *Imagines Illustrium Virorum: La Collezione Dei Ritratti Dell'Università E Della Biblioteca*

Universitaria Di Bologna. Clueb (2011).

Gebeil 2016 Gebeil, Sophie. Les MéMoires De La Marche Pour L'éGalitéEtContre Le Racisme. Dans Les Archives Du Web. HommesEt Migrations. Revue FrançAise De RéFéRence Sur Les DynamiquesMigratoires(2016).

Gomes 2011 Gomes, D., Miranda, J., Costa, M. A *Survey on Web Archiving Initiatives*, International Conference on Theory and Practice of Digital Libraries. Springer (2011).

Guagnini 1988 Guagnini, A. *Higher Education and the Engineering Profession in Italy: The Scuole of Milan and Turin, 1859–1914.* Minerva (1988).

Hale 2014 Hale, S. A., Yasseri, T., Cowls, J., Meyer, E. T., Schroeder, R., Margetts, H. *Mapping the UK Webspace: Fifteen Years of British Universities on the Web.* WebSci (2014).

Hoffmann 1994 Hoffman, A. M., Hoffman, H. S. *Reliability and Validity in Oral History: The Case for Memory.* Memory and History: Essays on Recalling and Interpreting Experience (1994).

Holzmann 2016 Holzmann, H., Nejdl W., Anand, A. "The Dawn of Today's Popular Domains: A Study of the Archived German Web over 18 Years", *JCDL* (2016).

Howell 2006 Howell, B. A. "Proving web history: How to use the Internet Archive", *Journal of Internet Law* (2006).

Huurdeman 2014a Huurdeman, H. C., Ben-David, A. *Web Archive Search as Research: Methodological and Theoretical Implications.* Alexandria (2014).

Huurdeman 2014b Huurdeman, H. C., Ben-David, A., Kamps, J., Samar, T., de Vries, A. P. "Finding Pages on the Unarchived Web", *JCDL* (2014).

Kahle 1997 Kahle, B. "Preserving the Internet", *Scientific American* (1997).

Kornblith 2001 Kornblith, G. and Lasser, C. "Teaching the American History Survey at the Opening of the Twenty-First Century: A Round Table Discussion", *The Journal of American History*, 87(4) (2001).

LaFrance 2015 LaFrance, A. *Raiders of the Lost Web.* The Atlantic (2015).

Lu 2012 Lu, K. and Wolfram, D. "Measuring Author Research Relatedness: A Comparison of Word-Based, Topic-Based, and Author Cocitation Approaches", *Journal of the American Society for Information Science and Technology*, (2012).

Macleod 1978 Macleod, R. and Moseley, R. *Breadth, Depth and Excellence: Sources and Problems in the History of University Science Education in England, 1850-1914* (1978).

Mahoney 1988 Mahoney, M. S. "The History of Computing in the History of Technology", *Annals of the History of Computing* (1988).

Milligan 2012 Milligan, I. "Mining the 'Internet Graveyard': Rethinking the Historians' Toolkit", *Journal of the Canadian Historical Association/Revue De La Société Historique Du Canada* (2012).

Milligan 2017 Milligan, I. "Welcome to the Web: The Online Community of GeoCities and the Early Years of the World Wide Web", In *The Web as History*, London: UCL Press (2017).

Murphy 2007 Murphy, J., Hashim, N. H., O'Connor, P. "Take Me Back: Validating the Wayback Machine", *Journal of Computer-Mediated Communication* (2007).

Nanni 2015 Nanni, F. "Historical Method and Born-Digital Primary Sources: A Case Study of Italian University Websites", In: *Officina Della Storia – Special Issue "From the History of the Media to the Media as Sources of History"* (2015).

Negrini 1998 Negrini, D. *Repertorio Nazionale Degli Storici Dell'Università: 1993- 1997.* Clueb (1998).

Nichols 2014 Nichols, L. G. "A Topic Model Approach to Measuring Interdisciplinarity at the National Science Foundation", *Scientometrics* (2014).

Pancaldi 1993 Pancaldi, G. (eds) *Universities and the Sciences: Historical and Contemporary Perspectives.* CIS (1993).

Pancaldi 2006 Pancaldi, G. "Wartime Chemistry in Italy: Industry, the Military, and the Professors", In *Frontline and Factory: Comparative Perspectives on the Chemical Industry at War, 1914–1924.* Springer (2006).

Parolini 2013 Parolini, G. *"Making Sense of Figures": Statistics, Computing and Information Technologies in Agriculture and Biology in Britain, 1920s-1960s.* Doctoral Dissertation. University of Bologna (2013).

Perkin 2007 Perkin, H. "History of Universities", In *International Handbook of Higher Education*. Springer (2007).

Piazza 2013 Piazza, S. *La Valutazione Della Ricerca Scientifica: Uno Studio Empirico Nelle Scienze Umane*. Doctoral Dissertation. University of Bologna (2013).

Pomante 2010 Pomante, L. *The Researches on the History of University and Higher Education in Italy. A Critical Appraisal of the Last Twenty Years*. History of Education and Children's Literature (2010).

Rea 1996 Rea, L. S. (eds) *La Storia Delle Università Italiane: Archivi, Fonti, Indirizzi Di Ricerca*. Atti Del Convegno: Padova, 27-29 ottobre 1994, 30. Lint (1996).

Richardson 1999 Richardson, W. *Historians and Educationists: The History of Education as a Field of Study in Post-War England Part II: 1972–96*. History of Education (1999).

Romano 2007 Romano, A. (eds) *Gli Statuti Universitari: Tradizione Dei Testi E Valenze Politiche*, Atti Del Convegno Internazionale Di Studi Messina-Milazzo (2007).

Rosenzweig 2003 Rosenzweig, R. "Scarcity or Abundance? Preserving the Past in a Digital Era", *The American Historical Review* (2003).

Serafini 2011 Serafini, M. *Technological Innovation in Emilia-Romagna: Knowledge, Practice, Strategies*. Doctoral Dissertation, University of Bologna (2011).

Tammaro 2006 Tammaro, A. M. (eds) *Biblioteche Digitali in Italia: Scenari, Utenti, Staff E Sistemi Informativi*. Rapporto Di Sintesi Del Progetto Digital Libraries Applications, Fondazione Rinascimento Digitale (2006).

Van Laer 2009 Van Laer J, Van Aelst P. "Cyber-Protest and Civil Society: The Internet and Action Repertoires in Social Movements", In *Handbook on Internet Crime* (2009).

Van Raan 1997 Van Raan, A. "Scientometrics: State-of-the-Art", *Scientometrics* (1997).

Venturi 1996 Venturi, Ilaria. "Troppe parolacce, censurata 'Lisa'", *La Repubblica*, 24/11/1996.

Webster 2015 Webster, Peter. "Will Historians of the Future Be Able to Study Twitter?", *Personal website* (2015). http://peterwebster.me/2015/03/06/future-historians-and-twitter/

Worboys 2011 Worboys, M. "Practice and the Science of Medicine in the Nineteenth Century", *Isis* (2011).

Zimmer 2015 Zimmer, M. "The Twitter Archive at the Library of Congress: Challenges for Information Practice and Information Policy", *First Monday*, 20(7) (2015).

de Ridder-Symoens 1992-2011 de Ridder-Symoens, H., Rüegg, W. (eds). *A History of the Universities in Europe*. Cambridge University Press, Cambridge (1992-2011).

de Wied 1991 de Wied, D. *Postgraduate Research Training Today: Emerging Structures for a Changing Europe. Netherlands Ministry of Education and Science*. The Hague (1991).