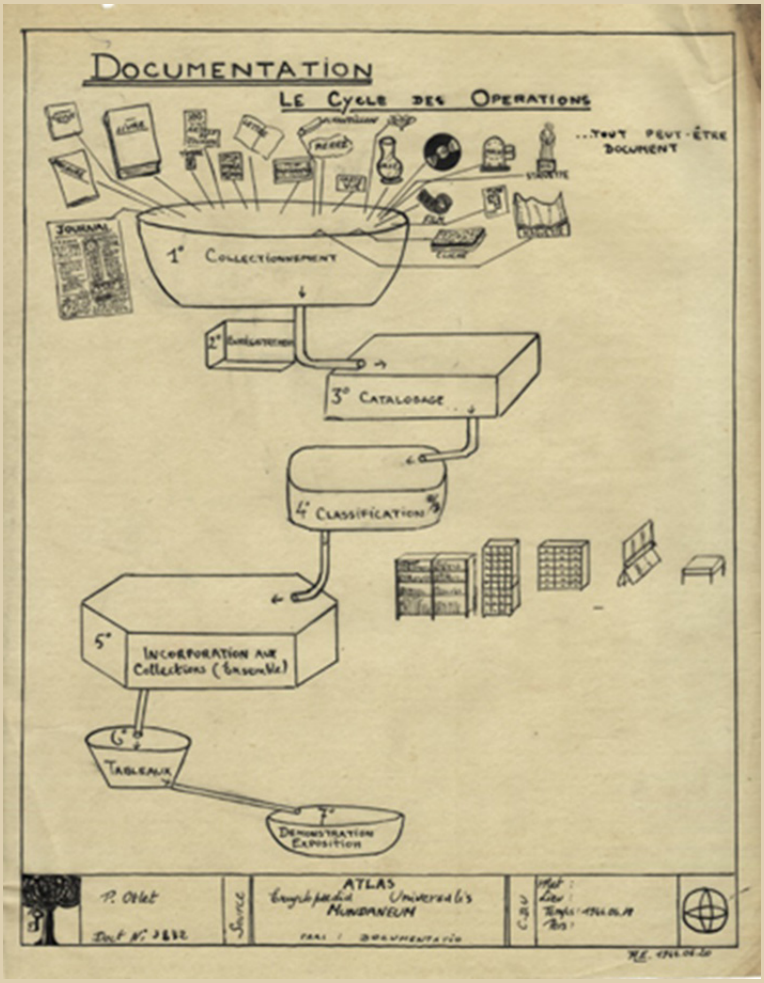
AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 3-4
ANNO 39
LUGLIO-DICEMBRE 2021



CACUCCI EDITORE
BARI

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:
Università della Calabria

Direttore Scientifico:
Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:
Fabrizia Flavia Sernia

Editrice: Cacucci Editore S.a.s.
Via D. Nicolai, 39 – 70122 Bari (BA)
www.cacucciedITore.it
e-mail: riviste@cacucciedITore.it
Telefono 080/5214220

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare topics quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita per le Aree 10 – Scienze dell'antichità, filologico–letterarie e storico–artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE –Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by SerialsSolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti dal comitato scientifico sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 39

N. 3-4 – luglio-dicembre 2021

CACUCCI EDITORE

BARI

© 2021 Cacucci Editore – Bari
Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
http://www.cacucciedit ore.it e-mail: info@cacucci.it

Sommario

Contributi

Contributi in memoria di Maria Pia Carosella

Note e rubriche

Contributi

AIDAinformazioni
ISSN 1121–0095
ISBN 979-12-5965-090-0
DOI 10.57574/5965090006
pag. 103-120 (luglio-dicembre 2021)

The digital age of historians

Salvatore Spina*

Abstract: History is about reconstructing experiences, records, and memories and thus investigating the meaning of human existence. History is recently setting up unpredictable and brand-new challenges for historians and their endeavour, the Digital History Era, in which exciting new tools are revolutionizing the way researching used to be conducted in the past. A unique quantitative approach along with brand new tools for reference sourcing is now pushing historians to redesign their working methods when it comes to seeking pieces of information within the archives.

Keywords: Computational History, Homologatus, dhSegment, HTR.

1. From History to the *Bit*

Man needs to improve himself, to overcome the limits imposed by Nature. This destiny led him to go through Time, marking his progress with several great revolutions, guided by intellectual spirit and genius.

Nowadays, a "revolution" is dealing with the 4.0 upgrade, that aims to create a man whose skills will allow him to control and face all future challenges, both in the scientific and cultural, industrial, and economic fields.

This is a new horizon that brings Mankind towards an epochal change, which will introduce new paradigms aiming to create a digital universe and implementing a Copernican techno-revolution that will invert the relation between technology and man, to put the latter and his *Ego* within the bits dimension, to forge the *HomoLogatus*[1], a "computable" person.

Inside the digital dimension, the HomoLogatus observes, and, thanks to algorithms and AI, he can overcome Physiology, Reason, and Life limits, which make him unable to grasp every single piece of information about History.

* Dipartimento di Scienze Umanistiche (DISUM), Università degli Studi di Catania. salvatore.spina@unict.it.

[1] *HomoLogatus* is the Latin for "logged in person". At the same time, the term indicates the "conformed men".

For this reason, it is necessary to train those who will replace us, to let historians collect all of the computer's algorithmic potentials, which are unique tools that, although lacking breath of life, can be "trained" to mine data from the Big Data of the Past, and turn them, rapidly, into historical information, in a quantity such that a single scholar would never be able to match.

Technology has already vastly overcome us as men, and We have been responsible for this situation since the beginning of human Time. The meaning of «revolution» lies in this: everything we have created, and continually create, is aimed not at improving our capabilities and performance – we have designed and built aircraft, but we still do not have wings –, but to make our lives increasingly smart, triggering a process of devolution that allows us to delegate our efforts to some tools. We did it for the purely physical dimension, and we are working for the intellectual, physiological, sociological, and aggregative ones.

The transformation/evolution of society towards the "augmented" dimension, invites scholars to revalue their skills, disciplinary and scientific paradigms, especially in that field which, up to now, looks to computer science and the technologies as tools for replacing paper and pen: Humanities. And this is even more true for historical research, which was put in a status of neutrality with the purpose of creating a framework of interdisciplinarity, where historians continue to ignore how to deal with the hyper-quantitative aspect of reality. «The historians ... didn't seem to have a good sense of how to wield quantitative data to answer questions, didn't have relevant computational skills, and didn't seem to have the time to dedicate to a big multi-author collaboration. It's not their fault: these things don't appear to be taught or encouraged in history departments right now» (Miligan 2012; Miligan 2019).

Yet, already François Furet looked with confidence at a quantitative "computational" approach to historical research, inviting scholars to put themselves into a control room to dominate the enormous amount of data, which would be characterized, thanks to the advent of personal computers, as the specific object of the historiographical future (Furet 1981).

This persistent neutrality status, nowadays – although notions such as "Big Data" and "Data mining" have pervaded Humanities Academy Departments –, freezes the use of technology in an embryonic state (Schiuma and Carlucci 2018). Still holding on to the "initial considerations" many historians look more at the implications than at the real opportunities of network-scholar integration, without considering the fact that, beyond the production of papers (which become "historical" after they come into being), man has always generated "data".

«Humanists – as Schiuma wrote – have always worked with data focusing mainly on their granularity and veracity, *i.e.*, trustfulness and truthfulness, rather than exploring and adopting approaches and tools for the creation and

use a great abundance of data. As a result, arts and humanities present a general lack of "big data culture"».

But, fortunately, the trend, today, is different. The opportunities offered by IT push humanists – even those who apply ancient methods – to synthesize new positions in research programs that, in recent times, have been labelled "Digital Humanities" (DH). DH can translate into multidisciplinary perspectives a dialogue within a scientific sector that, by its side, shows maximum flexibility, as Klein and Gold state, to the constant exchange with each sector of knowledge, without fearing that the continuous reference to other disciplinary systems could distort the ontological statutes, nor cancel science objectives (Klein and Gold 2019).

It goes without saying, therefore, that, in this perspective, Humanistic Informatics shows itself as the cardinal principle of the discourse, and the DH as the tangible proof of the union between Computational and Humanistic Sciences.

2. Digital historians

Infoglut society (Andrejevic 2013; Zeldes 2009) encourages historians to dare with complex and quantitatively vast narratives, where software, algorithms, and technological products will not be an additional tool, but a real *partner* allowing them, on the one hand, to digitize, in a short time, twenty-four centuries of human history (Perazzini 2012), recorded in millions and millions of manuscripts, and, on the other hand, to provide the solutions more suitable for the analysis of the data collection that is obtained.

It is the era of Big Data, where new technologies operate, and this must be taken into account by scholars, who, up to now, have developed theories, patterns, reconstructions, and experiments, solely based on data limited in space and time. It is, therefore, necessary to go beyond the assumption that «quantity» means «non-controllability», since the same technologies that allow us to extract the "much", allow us, at the same time, to control and process it.

However, a project about founding a standard interdisciplinary method, that looks at Humanities and IT, at the same time, is very distant from the common approach. Very few projects have seen some form of a convergence of knowledge, to take out essential IT skills from the "grey dimension", where they live as simple tools, and to train scholars to create a real bridge between information technology and humanities. For this reason, Automation and Algorithms must be the new way of considering and working on the problems of Humanities, accepting Julia Kristeva's assumption, according to which "multidisciplinarity" is not the «death of expertise» (Nichols 2019), but an «infinite discourse» capable of leading scholars into the *real* reality, both past, and present (Kristeva 2006).

Yet, still today, many scholars, including Perazzini, look at DH as a dimension with a euphemistically elusive and all-encompassing nature (Perazzini 2013).

When we deal with "multidisciplinarity", we have to look at information technology not as a Humanities' tool, but as of the latter, which can no longer look at technologies as a tool for reproducing/representing a cultural object on the screen, but as a source of information, both contemporary and historical.

This is a critical issue. Digital historians are called to demonstrate that the aim of Historical research always remains the same: to find the «true in the certain» (Giarrizzo 1999). But, nowadays, the upper-reproduction of a historical document makes the "certain" uncontrollable – because of the enormous quantity of digital information. This situation leaves historians in a condition of uncertainty; the uncontrollable information put them in a difficult position, because, on one hand, they do not want to consider the Internet – and tools like the "fact-checking algorithms" – as an *instrumentum historiae* (historical inquire tool) that can overcome their role. But, on the other hand, they acknowledge that IT and the Internet can create a new fundamental of "authenticity of the text", that can control and verify all historical documents within the Web. Digitization and its enormous capacity in terms of quantity and speed in reproducing historical sources threaten the slowness that has always characterized the work of the historian, who still hesitates in dealing with the heterogeneity of information travelling on the Internet with critical responsibility. Inside Big Data of the Past, not everything is true.

Anyway, digitization, Big Data, AI, can reduce the distance between Science and Humanities, creating opportunities, dialogue, and the positive acquisition of the idea that the scientific community should regroup within the network, as a new place that does not destroy the previous spaces.

What we are experiencing among historians – but, in general, among humanists – is an entrenchment into an integralist dimension that emphasizes old contrasts between philosophers and scientists, between *doxa* and *episteme*, in a pessimistic view that aims to end every discussion.

The question is not of little value, especially when historians look at the computer's possibilities as a dimension with little clarity. Although the Internet is becoming more and more uncontrollably complex, it offers new contents, different, endless culture, which is a different human product in relation with the past centuries. Therefore, a new epistemological approach is necessary, which has to emerge – but this is precisely the difficulty – from the network itself. Inside the Internet, we can find models and new points of view; inside the Internet, disciplinary boundaries can fall, thanks also to a new formal language that was universally acquired by the Internet community. Inside the Internet, Historical Science has to find a new theoretical foundation, aiming

at the control of the Big Information, to give way to a quantitative historical approach.

Even if historians have always tried to shut themselves away in a traditionalist opinion, avoiding the "Internet" phenomenon (Dollar and Jensen 1974; Krieger 1974; Glasco 1969; Green 1963; Janda 1965; Margolis 1970) as a new perspective for the Science of the Past, there has been no lack of comparative tests. But every time Historians spoke about the Internet, they had the only purpose of writing its history. They never legitimize its existence. They never look at the Web as an opportunity for Historical research or as a revolutionary moment capable of marking the beginning of a new age.

In this new world, Computer Engineering and digitization encourage historians to pay attention to new questions which, on the one hand, feed plausible fears, such as losing the primacy on historical heritage and, therefore, on the discarding and evaluation of what is relevant in History. But, on the other hand, digitization spreads the boundaries of reality, giving the historian the task of working with more information than in the past. This new condition pushes historians to definitively accept Ginzburg's assumption, according to which the idea that «historians must or can prove anything is outdated and ridiculous» (Ginzburg 2014), because the real meaning of Historical research is not only to create proof.

In the 1930s, *Annales* historians accepted the "quantity" challenge. Bloch, Febvre, and Braudel looked ahead to a new historiographic experience (Braudel 2003; Febvre and Cantimori 1966) in which historical time, considered as a "plurality of organized times that are placed on decomposable levels" (Cantù 1981), makes of the «longue durée» (Braudel 2003) and interdisciplinarity the cornerstones of this approach, in opposition to a historiography that has marked a scientific progression focused on objects circumscribed in time and space.

In this scenario, the concept of "quantitative method", based on calculations and correlation – the "serial History" (Braudel 2003) –, is reconstructed (Furet 1971), thanks to the calculator, which makes it possible to extend the interests towards those measurable and quantifiable dimensions, where the historical "datum" – of immeasurable quantity – finds its meaning only within a "series" that could explain its genealogy.

The historical sources crowd together and become meta-sources (Fiormonte 2000), which impose a new way to formulate questions about the Past. And the answers can create new narrations that can break historical continuity, but at the same time can recompose this continuity in the whole.

Within this Techno dimension, historians are enriched with new fundamentals and other historical brands which are added to those of the paper tradition. The changing media has imposed new tools on "making history" which push source analysis processes (Ortoleva 1999) into more difficult and

complex selection procedures, calling the scholars to pay attention – particular attention to the various software on which the Internet is based, which can condition both research and the validity of the document.

In addition, "quantity" generalists and specialists, concerning this matter, ask themselves how historians could overcome and handle a mass of data that they cannot simply define as "enormous".

If, on the one hand, the concept of "complete information" presses scholars to a forced interpretation to reveal those aspects of an event that had been obscured by the analogical (*ante* Internet) approach, on the other hand – nowadays –, historians are faced with tools that can bring back from the Internet everything that was set aside during a selection phase, but with the strong risk of losing control over the information, making it difficult to analyse the complex data without a computational tool that would confuse the historian's work. And it goes without saying that this condition is a reality rather than a risk.

For this reason, traditionalist historians seem to opt to maintain a *humanistic Middle Ages*, siding in favour of a position that does not undermine progress, but keeps technology within the niche of the pure instrument against the law of progress; and even if Stiegler (Stiegler 2014) is fully convinced that if progress were to definitively take the path that aims to alleviate the burden of cognitive functions, making algorithms carry out the processes of our knowledge, not everything would be left to the machine, since – as Enzensberger asserts – «the best search engine [would remain] the [human] brain» (Enzensberger 2004).

Therefore, to historians, the appropriation of new investigation tools is fundamental to control the contents of the Internet, to escape that destiny which, to date, would seem marked: completely lose control of the processes of cultural dissemination and the organization of historical memory, which would see the disappearance of the guilds of specialists.

The historian's work remains – as Giuseppe Giarrizzo states –, however, all in the «certainty», order and relevance of the facts. «Without them – continues the academician, quoting Ugo Foscolo (Foscolo 1958) – the genius of the historian would do nothing but poetry» (Giarrizzo 2018).

«Being historians [means] studying the transformations to which man undergoes»; and today, more than ever, scholars are pushed to think about new human acts, which are the intellect outcome that has enriched its vital spaces with technology, which is not an appendage or a stick, but an unfolding of its abilities.

A Historian cannot forget, even in the era of the dematerialization of culture, that his mission remains what Cantù defines «embracing the whole without neglecting the details; not to separate the telling of facts from the poetry of customs and thought; obtain regularity and yet leave wings to the imagination; group accidents without confusing them; to graft, the varied spectacle

of life with the profound metaphysical interest offered to us by the previous evolutions of the human spirit» (Cantù 1844). However, he is also responsible for the exact knowledge of all human phenomena, whether social or economic, which he must translate into a specific narrative that must not divert the knowledge of contemporaries or his "posterity" (Giarrizzo 2018).

Historians, however, do not realize that the 4.0-revolution lies within the progress of History, and he has to write about it, getting rid of the idea that computer engineering's products are simple machines for exhibitions. Historians have to consider computer machines as «cultural future», able to follow that path which, to Renan, already in 1863, allows us to «discover the secret of the universe» (Giarrizzo 2018).

Twentieth-century History will continue to be written, says Salarelli, after an elaborate selection of the documentation, but information pulled out from the digital dimension will increase the data used by historians (Salarelli 2001).

So, technology does not lead to pessimism. Rather, it is an opportunity (Riva 2013) to cross new levels of knowledge, higher and wider, which look at a conceptual complex that can only be created by a humanist who "becomes digital", and who becomes a pioneer of the new frontiers of "making history". Historical narrative, therefore, has to embrace new methods, without forgetting its fundamentals and the dust on the papers.

3. From typography to dhSegment

The virtual world is characterized by the decentralization of the "text" and the absence of introjection of the paper support – which always remains there –, the purpose of which, conceived and implemented since the end of the Twentieth century, is the erosion of the concept of "support" (Ortoleva 1999), which allows the former to move on a wider and more capillary dimension, and which has broken the walls of the academic niche, which, today, is no longer able to defend itself, advocating the position that the official nature of the story derives solely and exclusively from the stability over time of the "visible" support.

But this statement, which denotes, moreover, only the «typographic» phase (Ortoleva 1999), expresses a pessimistic and self-protective position rather than a mature cognition devised, perhaps, on confrontation with technology, which, indeed, can cross space and time more effectively than the "old support". Stability is not so much of the physical support, but the text; Johannes Gutenberg (1400-1468) looked at the text when he started to reflect on the typographic processes project (Braida 2000; Eisenstein 2011; McLuhan 2011; Mistretta 2008). Certainly, he did not look at the paper supports, and he did not show any remorse in annihilating the figure of the *amanuensis*.

The Gutenberg project, however, scared scientific and humanistic society. The "multitude of books" that would pervade the whole world, stirred up demonization and anxieties even in Conrad Gesner himself who felt his *Bibliotheca Universalis* threatened (1545), while Baillet stressed that the enormous quantity of books would have pushed societies into a state of a new barbarism (Baillet 1686). But to prevent the uncontrollability, Francis Bacon invited the scholarly community to «analysis and study», the only instruments able to select which books would be worthy of praise (Bacon 1961).

But the anxieties remained the same. The "information explosion" (Rosenberg 2003), from the 1650s up to 1700s, was a problem. The scholarly community was invited to find a tool to control and organize the Science system, capable of filtering information and giving them the necessary sense: the *Physica Sacra* (by Scheuchzer), the *Encyclopédie* (Diderot and d'Alambert), the *Cyclopædia* (Chambers) (Yeo 2003).

Modern Age acknowledges that "calligraphic time" arrived at the end of its experience, and it had to be replaced by an instrument that would guarantee the civilization that was – and still is! – the last purpose (Giarrizzo 2018) of the historian's work, which did not to undermine or confuse Gutenberg's intuition. And this basic principle is still valid today.

Nowadays, the process of "dematerialization" and transformation of the analogical into a digitized meta-product requires not only a new approach to scholarship, but a new and complex vision of the historical source, which is "data" itself, in its simply being a "layout".

The burgeoning digitization of historical sources, over the last ten years, has piled the Internet with an enormous amount of facsimiles of archive papers. These "meta-sources"[2] are full of information that does not exist in the paper version, so that historians – who must move in a thin line between philological investigation and deep learning –, to analyse them, are forced to use some tools that have never been part of their expertise.

DhSegment, for instance – an open-source implementation of a CNN-based (Hoo-Chang et al. 2016) pixel-wise predictor coupled with task-dependent post-processing blocks –, lets historians tackle document processing problems, such as page extraction, baseline extraction, layout analysis, or multiple typologies of illustrations and photograph extraction (Ares Oliveira, Seguin, and Kaplan 2018; Barbuti and Caldarola 2013; Barman et al. 2021).

Nowadays, Historical research needs to overcome the simple reading of a historical source text: historians need to work differently with digitized doc-

[2] This term does not mean the simple digital transcription of documents, but a different and more complex publication of the text, which is accompanied by a range of research tools – *regestas*, inventories, images, essays, bibliographies, databases, etc. – which are able to enrich the editorial product and determine new ways of reading and fruition (Fiormonte, 'Il documento').

uments. They need to know how to cut out the page of the manuscript, how to extract the illustration from the text, how to find the pages that contain a certain symbol, how to locate text in a digitized image, etc. So, they need to know how to analyse historical documents, and how to process them in a neural or semantic network; and how to automatically mine all the hidden pieces of information.

A meta-source – the digitized source –, is not the simple digital conversion of analogic support and of the text it contains, but it is a new element which, on the one hand, reproduces typical and fundamental information for the scholar's work, on the other, provides new elements to create a new and innovative historical narrative. Inside a meta-source, words become not only words, but interlinked instances in a knowledge base that can transform a bulk of data into connected information particles. A meta-source is a container of interlinked data and uncover hidden relationships.

Machine and Deep Learning, algorithms and AI tools allow the identification of all the elements of the meta-source layout (text, images, drop caps, etc.) and insert them within a system of interactions that can mine original data. Add to this, nowadays, the possibility of applying HTR systems (Handwritten text recognition) (Adamek et al. 2007; Bahlmann, Haasdonk, and Burkhardt 2002) to obtain automatic transcriptions, which is a new important possibility of being able to process numerous collection's texts, in a short time, and obtain, after a computational analysis, all the patterns that can provide new elements and information to historical research, which could come even closer to the *true* truth about the Past.

A document, therefore, beyond its physical dimension – paper or magnetic –, nowadays becomes an even more complex conveyed code, which creates, thanks to dhSegment, a bridge between information and scholars. A document, on its paper support, even if unhinged from its physical archive, remains *real* on the web server, in the same way as a book: once closed, the text does not vanish, but simply has no longer any relationship with the reader.

The historian will have to look at the digital and digitized sources, as dynamic elements, not as ephemeral cultural items, and mortal in the short term. The support's obsolescence, certainly, raises the question and animates debates, because of the speed with which they are continually updated, which is not comparable to the revolution that led scholars to leave behind the limited medieval handwritten system and embrace the typographic one. After this revolution, printed paper became the only support, and we still use it. In our digital age, the logic is still the same: copying on multiple supports to guarantee conservation and dissemination. Site mortality, for instance, stressed by Zanni Rosiello, cannot and must not be de-legitimizing, because the work of discarding/destroying/deleting the material accumulated in a server does not follow the same "definitive" logic adopted by archivists on reorganizing a

physical archive, where the action of selection prevents us from finding testimony and proof of the narration (Zanni Rosiello 2005). «Internet Archive» and «WaybackMachine» show, from this point of view, that conservation on the Internet follows the same principles as the material one, but with much higher possibilities. After the selection, cancellation, or closure (death) of a website, the documents are never destroyed, but they remain on a different level of the Internet, from where they can re-emerge – overcoming the documentary "amnesia" – much more quickly than on paper, which, in truth, once excluded and destroyed, can no longer emerge.

4. The invitation to considerations

Doing History, nowadays, means questioning everything about us; also, our projection into the digitized dimension where, however, the civilization, pedagogical, philosophical, and political role that Giarrizzo entrusts to History remain valid (Giarrizzo 2018).

Furthermore, the methodological exchange between History and IT could put a stop to the violent and sometimes arbitrary documentary mutilation, which creates a biased approach. And, above all, this interdisciplinary dialogue could give the entire non-scholarly community the possibility to be part of the historical research, to write the truth about our Past.

Even if we all agree that – not considering any "false history" on which there is plenty of historiography coming from ancient times – the "giudizio vero" ("true judgment") on events and ideas (Luciano 2010) is still made by historians, with internet History is also now being told by the people (HomoLogati) who come to play an active role in co-creating this "giudizio" (judgment) and its memory, though via their interpretative and methodological[3] drawbacks: it is the *public history* – that is what it can be called –, a new frontier of "telling events" no longer prerogative of academics, but of an active and broader audience aiming to collect new resources [4] and perspectives and by which historians are hoping to find brand new insights for a deeper understanding of the complexity of the past.

[3] A perfect example is the Societa Storica Spezzina experience, whose project of public narration construction has quiclky come to an end after its first enthusiasm, thus making it crystal clear that its endeavour is yet in need of a bigger effort on our national level to get to a real public history. Cfr. (Capezzuto 2018).

[4] For instance, this is the aim of the new Canadian project experiences of "history blogging" which have successfully supported new types of historical narration combining all excluded materials and created the conditions for a social participation with a dramatic involvement on the reconstruction of the history and identity of Canada. Cfr. ActiveHistory.ca (Beth Robertson), The Otter-La Loutre (Tina Adcock), Findings/Trouvailles (Stacy Nation-Knapper, Tina Adcock), the Acadiensis blog (Corey Slumkoski), and Borealia (Keith Grant). Also, consider (Adcock et al. 2016).

Historians, therefore, are responsible to tackle this revolution with rigour, which is an integral part of the flow of events that he is required to reconstruct and narrate, analysing the traces of the HomoLogati on the internet, where a lot of gap in the flow of events, destabilize the concept of "linearity" on which some historiographical theories are based, but which show how History is full of strong anachronisms, which oblige the historian to draw up a more articulated narration of past events, through a communication design (Capezzuto 2017) that can no longer be caged in paper.

And on all this, the uncontrollable Dark Web (Bradbury 2014; Chen et al. 2008; Chen 2020), the anarchist side of the Web that handles the power of dissolving all the certain inside the Internet. Dark Web and deep Web became a special digital archive, where sources are capable to break historical linearity. Historians, therefore, cannot overlook its content, but they need to become comfortable with all those tools that can lead them into this side of the internet dimension.

The historian, who will narrate the HomoLogatus, will have to deal with a multitude – and in any case a collective – that cannot be encaged into that classification that has marked Historiography up to now.

Edward Snowden (Barcelò 2018; Greenwald 2014; Snowden 2019) does not go down in History after his name has been bounced between one newscast and another; he goes down in History because he has shown how political power handles the Internet to control citizens (Greenwald 2019) – showing, also, the complexity and uncontrollability of the Web. Snowden passes into the annals of History because the latter, now, cannot be beyond the Net.

On the Internet, it is the most anonymous person who writes History. It is the HomoLogati like Julian Paul Assange who write an important document and live those digital events that will be the subject of historical inquire, in the future.

Historians, who seek the historical source on the web, must search all the deepest corners of the Internet, because, today as in the past, the most important events – such as those that led to the Protestant Reformation, thanks to a monk that "then" will go down in history as Marthin Luther – will be carried out by the anonymous agent, other "Luther", who will be more and more numerous and more and more uncontrollable within the Internet, where its dark side (Dark Web), for better and for worse, will be full of important traces, which historians will have to deal with, to the knowledge of the Past.

The difficulties of the historian's profession must be overcome. «History is not dead science – Giarrizzo quotes Balbo – like Astrology or Alchemy, which are not studied except for curiosity ... but Science and Language that are alive to be cultivated for the use of practice and progress» (Balbo 1858). And progress, nowadays, is the Internet.

Certainly, this is a real call out, but History cannot fail to reconsider itself, to return to being the holder of the depth of time; but it must do so within an irreversible process that arises within the Internet and finds its groundwork in it.

A new approach to History, therefore, aims to incorporate humanistic methodologies into informatics, through a reinterpretation of the methodological repertoire of the historian – new imagination, new investigation, digital analysis, digital comparison, critical interpretation (Balbo 1858).

This does not mean obscuring the role that historians have played over the centuries, but it wants to push historians to regain possession of the vision of the globality of human actions through the Internet, where everyone, nowadays, is the protagonist.

«The historian never goes out of time», as Braudel asserts, even if the Histories of crises and cycles have obscured the concept of "longue durée", the only paradigm that preserves the civilizing motion of many "present" that follow one another (Braudel 1974). The Internet can restore events to their place within the "series" – of Foucault's memory (Foucault and Bertani 2001) – of which they are part.

Society and humanists are suffering the consequence of this new "renaissance", which, as in the fifteenth century, enriches the heritage on which historians work, with a new immeasurable quantity of information.

Big Data threatens and fascinates sociologists and historians. Algorithms and AI redirect research towards perspectives that have been abandoned – such as "wide spaces" and "millenary periods" –, but with the possibility, today more than ever, of being able to reduce the great historical context to a specific visualization that could algorithmically demonstrate what – as Maguire asserts – has never been verified or falsified (Lane 2017). Thanks to the Internet, our Past could be investigated, nowadays, overcoming the ideological approach. The counterfactual analysis based on the quantitative comparisons could shed new light on past events and lead historians to a historical discourse nearer and nearer to the truth.

The Internet is open to digitized documentation of historical archives, aggregators, and accumulators (Armitage and Guldi 2016; Speth 1981; Edwards 2019; Hulme 2011; Miller and Edwards 2001; Tickell 1986; *Potential implications of trends in world population, food production, and climate* 1974; Williams 1978) of those "big data", on which, over the past seventy years, computational models were built, to analyse specific phenomena. All strictly off-line.

These enormous data assets, now, on the Internet, are the best proof of the effectiveness of the quantitative approach, and now they are a historical heritage that grows both in the contents and in the IT architectures that support them, which, nowadays, look carefully at the semantic web and at artificial neural networks (Mazzetti 1991; Milligan 2012; Milligan 2019), where algorithms will be increasingly able to identify – and then answer in a targeted and

exhaustive manner – and solve all the different problems that arise in specific research.

Historical and cultural heritages and technologies grow and attract scholars from many areas of interest, and, among all, contemporary historians, who look to those tools as an opportunity to regain possession of the "long duration" and the comprehensive storytelling; of the real possibility of measuring and showing, in a tangible way, the changes that have marked humanity and territories; of correlations between cultural, social, economic, political phenomena through time. The big data that can let historians move rapidly between "local" and "global", will probably release History from the dichotomies «micro/macro», «remembrance and oblivion», leaving it to future scholars to understand their validity.

References

Adamek, Tomasz, Noel O'Connor, Noel Murphy, and Alan F. Smeaton. 2007. "Word matching using single closed contours for indexing handwritten historical documents." *International Journal of Document Analysis and Recognition (IJDAR)* 9 (2-4): 153-165.

Adcock, Tina, Keith Grant, Stacy Nation-Knapper, Beth Robertson, and Corey Slumkoski. 2016. "Canadian History Blogging, Re*fl*ections at the Intersection of Digital Storytelling, Academic Research, and Public Outreach." *Journal of the Canadian Historical Association Revue de la Société historique du Canada* 27 (2): 1-39. https://doi.org/10.7202/1040560ar.

Andrejevic, Mark. 2013. *Infoglut. How Too Much Information Is Changing the Way We Think and Know*. Hoboken: Routledge.

Ares Oliveira Sofia, Benoit Seguin, and Frederic Kaplan. 2018. "dhSegment, A generic deep-learning approach for document segmentation." In *Frontiers in Handwriting Recognition (ICFHR), 2018 16th International Conference,* IEEE: 7-12.

Armitage, David, and Jo Guldi. 2016. *Manifesto per la storia, il ruolo del passato nel mondo d'oggi*. Roma: Donzelli.

Bacone, Francesco. 1961. *Saggi*, traduzione italiana a cura di Augusto Guzzo e Cordelia Guzzo, Torino: UTET.

Bahlmann, Claus, Bernard Haasdonk, and Hans Burkhardt. 2002. "Online handwriting recognition with support vector machines - a kernel approach." In *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 49-54.

Baillet, Adrien. 1686. *Jugemens des sçavans sur les principaux ouvrages des auteurs*, 4, 2, Paris : Dezallier.

Balbo, Cesare. 1858. *Pensieri sulla storia d'Italia*. Firenze: Le Monnier.

Barbuti, Nicola, and Tommaso Caldarola. 2013. "An Innovative Character Recognition for Ancient Book and Archival Materials, A Segmentation and Self-Learning Based Approach." *In Digital Libraries and Archives IRCDL 2012. Communications in Computer and Information Science* 354. Berlin: Springer.

Barcelò, Andrea Lattanzi. 2018. *Edward Snowden. La voce silenziosa del coraggio*, Area51 Publishing.

Barman, R. Raphaël, Maud Ehrmann, Simon Clematide, Sofia Ares Oliveira, and Frédéric Kaplan. 2021. "Combining Visual and Textual Features for Semantic Segmentation of Historical Newspapers." *Journal of Data Mining and Digital Humanities* n. HistoInformatics.

Bradbury, Danny. 2014. "Unveiling the dark web." *Network Security* no. 4: 14-17.

Braida, Lodovica. 2000. *Stampa e cultura in Europa tra XV e XVI secolo.* Roma: Laterza.

Braudel, Fernand. 2003. *Scritti sulla storia*. Milano: Bompiani.

Braudel, Fernand. 1974. *La storia e le altre scienze sociali.* Roma: Laterza.

Cantù, Cesare 1844. *Storia universale*. Torino: G. Pomba.

Cantù, Francesca. 1981. "Aspetti di metodologia della ricerca nella storiografia delle «Annales»." *Mélanges e l'Ecole française de Rome. Moyen-Age, Temps modernes* 93 (1): 433-455.

Capezzuto, Stefano. 2017. *Il design della conoscenza. Intervista a Jeffrey Schnapp*, *il lavoro culturale*, https://www.lavoroculturale.org/intervista-a-jeffrey-schnapp/stefano-capezzuto/2017/.

Capezzuto, Stefano. 2018. "Digital Public History e design della partecipazione. Riflessioni e proposte a partire dalla Società Storica Spezzina." *Umanistica digitale* no.3: 69-84.

Chen, Hsinchun. 2011. *Dark Web, Exploring and Data Mining the Dark Side of the Web.* Springer Science & Business Media.

Chen, Hsinchun, Wingyan Chung, Jialun Qin, Edna Reid, Marc Sageman, and Gabriel Weimann. 2008. "Uncovering the dark Web, A case study of Jihad on the Web." *Journal of the American Society for Information Science and Technology* 59 (8): 1347-1359.

Council on Environmental Quality (U.S.), *Global energy futures and the carbon dioxide problem*, Council on Environmental Quality : For sale by the Supt. of Docs., U.S. G.P.O., Washington, D.C. 1981.

Charles M. Dollar. 1974. *Historian's guide to statistics; quantitative analysis and historical research*. New York: R.E. Krieger Pub. Co.

Eamon Doyle. 2020. *The dark web*. Greenhaven Pr.

Edwards, Paul N. 2019. *A Vast Machine - Computer Models, Climate Data, and the Politics of Global Warming*. Goodwill BookWorks.

Eisenstein, Elizabeth L. 2011. *Le rivoluzioni del libro l'invenzione della stampa e la nascita dell'età moderna*. Bologna: Il Mulino.

Enzensberger, Hans Magnus. 2004. *Gli elisir della scienza, sguardi trasversali in poesia e in prosa*. Torino: Einaudi.

Febvre Lucien. 1966. *Studi su Riforma e Rinascimento e altri scritti su problemi di metodo e di geeografia storica*. Torino: Einaudi.

Fiormonte, Domenico. 2000. "Il documento immateriale." In *Il documento immateriale. Ricerca storica e nuovi linguaggi. Dossier 4 de «L'Indice»*, a cura di Guido Abbatista, Andrea Zorzi, v. 5. Torino: Indice Scarl.

Foscolo, Ugo. 1958. *Saggi di Letteratura Italiana*, Le Monnier 1958.

Foucault, Micheal. 2001. *Il discorso, la storia, la verità, interventi 1969-1984*, a cura di Mauro Bertani. Torino: Einaudi.

Furet, François. 1971. "L'histoire quantitative et la construction du fait historique." *Annales. Histoire, Sciences Sociales* 26 (1): 63-75.

Furet, François. 1981. "Il quantitativo in Storia." In *Fare storia. Temi e metodi della nuova storiografia* a cura di Jacques Le Goff e Pierre Nora. Torino: Einaudi.

Gardner, James B., and Peter S. LaPaglia. 2006. *Public history, essays from the field.* Malabar: Krieger Pub. Co.

Giarrizzo, Giuseppe. 1999. *La scienza della storia, interpreti e problemi*. Napoli: Liguori.

Giarrizzo, Giuseppe. 2018. *La storiografia della nuova Italia, Introduzione alla storia della storiografia italiana v.1*, a cura di Lina Scalisi. Roma: Edizioni di Storia e Letteratura.

Ginzburg, Carlo. 2014. *Rapporti di forza, Storia, retorica, prova.* Milano: Feltrinelli.

Glasco, Laurence A. 1969. "Computerizing the Manuscript Census." *Historical Methods Newsletter* 3 (1): 1-4.

Gold, Matthew K., and Lauren F. Klein. 2019. *Debates in the digital humanities*.

Green, Bert F. 1963. *Digital computers in research, an introduction for behavioral and social scientists*. New York: McGraw-Hill.

Greenwald, Glenn. 2014. *No Place to Hide - Sotto controllo, Edward Snowden e la sorveglianza di massa.* Rizzoli 2014.

Hulme, Mike. 2011. "Reducing the future to climate, a story of climate determinism and reductionism." *Chicago Journals 26* (1). doi:10.1086/661274.

Janda, Kenneth. 1965. *Data processings; applications to political research*. Evanston.

Kristeva, Julia. 2006. *Desire in language, A semiotic approach to literature and art*. Columbia University Press, S.l.

Lane, Richard J. 2017. *The big humanities, digital humanities/digital laboratories.* Routledge.

Margolis, Michael. 1970. "OSIRIS and SPSS: New Computer Packages for the Analysis of Social Science Data." *Computer Science*.

Mazzetti, Alessandro. 1991. *Reti neurali artificiali, Una nuova sfida tecnologica, Introduzione ai principali modelli e simulazione su personal computer*. Milano: Apogeo.

McLuhan, Marshall. 2011. *La Galassia Gutenberg, nascita dell'uomo tipografico*. Roma: Armando.

Miller, Clark A., and Paul N. Edwards. 2001. "*Changing the atmosphere, Expert knowledge and environmental governance.* Cambridge: MIT Press.

Milligan, Ian. 2012. "Mining the 'Internet Graveyard', Rethinking the Historians' Toolkit." *The Canadian Historical Association* 23 (2).

Milligan, Ian. 2019. *History in the age of abundance? How to web is transforming historical research*. London, Chicago: McGill-Queen's University Press.

Mistretta, Enrico. 2008. *L'editoria, un'industria dell'artigianato*. Bologna: Il Mulino.

Nichols, Tom. 2019. *The death of expertise, The campaign against established knowledge and why it matters*.

Ortoleva, Peppino. 1999. "La storia nel sistema dei media che cambia." *Contemporanea* 2 (3): 495-499.

Perazzini, Federica. 2012. "Oltre i confini del testo, Le Digital Humanities tra scienza e opportunità." *Quaderni Digilab* no. 2 (ottobre): 183-200.

Perazzini, Federica. 2013. "Words, Bytes and Numbers, Le Digital Humanities 'viste da vicino'." *StatusQuaestionis* 5: 171-196.

Riva, Elena. 2013. *Digital Humanities e Digital History, una nuova cittadinanza dei saperi*, Vita e Pensiero.

Rosenberg, Daniel. 2003. "Early Modern Information Overload." *Journal of the History of Ideas* 64: 1-9.

Salarelli, Alberto. 2001. "Con quali documenti si scriverà la storia del XX secolo? I sedimenti del «secolo breve»." *Bibliofilia* 103 no. 3 (settembre-dicembre).

Schiuma, Giovanni, and Daniela Carlucci. 2018. *Big data in the arts and humanities, Theory and practice*. Auerbach Publications.

Snowden, Edward J. 2019. *Errore di sistema.* Longanesi.

Stiegler, Berbard. 2014. *Prendersi cura, della gioventu e delle generazioni*. Salerno: Orthotes.

Tickell, Crispin. 1986. "Climatic change and world affairs." *Eos, Transactions American Geophysical Union* 67 (17): 425.

Williams, Jill. 1978. *Carbon dioxide, climate and society, proceedings of an IIASA workshop*. Institut international pour l'analyse des systèmes appliqués, Organisation météorologique mondiale, Programme des Nations Unies pour l'environnement, Comité scientifique chargé des problèmes de l'environnement Toronto: Pergamon Press, Oxford.

Yeo, Richard R. 2005. "A solution to the multitude of books, Ephraim Chambers's «Cyclopaedia» (1728) as «the best book in the universe»." *Journal of the History of Ideas* 64 no.1 (January): 61-72.

Zanni Rosiello, Isabella. 2005. "A proposito di web e del mestiere di storico." *Contemporanea* 8 (4): 743-755.

Zeldes, Nathan. 2009. "Infoglut." *IEEE Spectrum* 46 (10): 30-55.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 39

N. 3-4 – luglio-dicembre 2021

Contributi

Contributi in memoria di Maria Pia Carosella

Note e Rubriche

mundaneum

In copertina
Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-090-0
9 791259 650900

ISSN 1121-0095
9 770112 100950