

Biscari Epistolography. From Archive to the Website

«DigItalia» 2-2023
DOI: 10.36181/digitalia-00090

Salvatore Spina

Dipartimento di Scienze Umanistiche – Università degli Studi di Catania

The digital turn has created an interrelated ecosystem between humans and machines capable of generating data and knowledge. This paradigm shift necessitates new archival approaches, as demonstrated during the digitization project of the “Corrispondenza” section of the Biscari Archive. This project showcased how, by leveraging artificial intelligence tools such as Transkribus and ChatGPT, the digitization process can effectively address the challenges associated with encoding archival materials. The digital edition and virtualization of document consultation on the Biscari epistolography website (<https://biscariepistolography.altervista.org>) have been specifically designed to facilitate the exploration of historical documents through the web. Moreover, recognizing that digital dissemination aims to preserve cultural heritage and enable access to scientifically relevant data for future generations of scholars.

1. Introduction

Digitality is the establishment of a digital ecological niche (DEN), that is an anthropological environment of “person2persons2machines” interplay, in which *Homo-Loggatus*¹ is enabled to generate data, to exchange and analyze them through tools that appear to be even more powerful than the abacus and the Pascaline, which allow the acquisition of new information, the inference of meaning, and the creation of new knowledge.

On the one hand, digitality responds to humans’ potential to navigate realms that appear to be challenging to analyze through an analogical lens —where «analogical» refers to the application of analogy. On the other hand, it is based on widespread dissemination on the internet network of what is generated and digitized, namely those real, paper-based, and material objects (archival documents, spaces and places, tools, and instruments) reproduced in a digital environment.

In this way, Computer Science becomes the keystone of interdisciplinary studies and a necessary linguistic system for unfolding.

The convergence becomes rigorous and almost necessary; the interdisciplinary approach in the field of research projects allows us to examine certain perspectives that, until now, remained in the shadows. Moreover, the Internet has ensured that most information is encoded in graphic interfaces and hyperlinks to be loaded on different servers and websites, guaranteeing the widest possible dissemination.

Therefore, the changes are monumental and digitality is no longer “tomorrow’s perspective”. It represents the *hic et nunc* of life.

AI and algorithms have constituted the humanists’ “laboratory” since the 1990s, when scholars referred

¹ Salvatore Spina, *Homo-Loggatus. The anthropological condition of historians in the digital world*: «Journal of Mathematical Techniques and Computational Mathematics», 2 (2023), n. 10, p. 431-437; Id., *Homo-loggatus? Come stare dentro la nicchia ecologica digitale*, in: *Informatica Umanistica e Cultura Digitale: il blog dell’AIUCD*, Billet: 2023, <<https://infouma.hypotheses.org/2224>>.

to them as «frontier and perspective». In the archival field, Computer Science has become the bridge between what this scientific area has always been regarded as in disciplinary and statutory terms, and what it is actually requiring in a digital and digitalized world where every person is measured, i.e. the need for innovative approaches in the management of new sedimentation.

The Web and data are dynamic, unlike what is “handed down” so far. The media that store them are designed to allow Web users to quickly capture and organize information based on their expectations in terms of function or search.

However, everything that has been archived up until now needs to be placed in the immense flow of the Web to meet multiple requests, from researching the exact determination of practices, to the integration of all the pieces of information we possess that allow us to understand the present world and the past —and this function is considered to be crucial.

Digital Archives and online repositories have become hybrid due to the fact that they collect heterogeneous documentation that requires re-evaluating the principles on which Archival Science stands. For instance, when we consider historical archives’ millennial role of preserving state authorities’ documentation, the practices of conservation and widespread dissemination must be combined. The latter is a cornerstone of the Web’s concept and design. It is the paradigm upon which² the concept of “narrativity of History” was built. Starting from 6 August 1991 —the date of the online launch of the first website— several tools have been developed, e.g. browsers like Mosaic Netscape and Netscape Navigator (already by 1994), which demonstrated the network’s potential of becoming a global informational structure. Remarkably, from that moment on people have rushed to put their stories on the Internet, thanks to the creation of websites dedicated to their favourite topics. The National Park Service and the Library of Congress published official websites on the most important historical topics. New computer technologies, such as JSTOR and PROQUEST, made full-text facsimiles of scientific journal articles accessible, enabling a broader dissemination of theories and ideas on various topics. Scientific libraries started to digitize their records to let Web users browse their collections, and, in the meantime, many digitization projects were initiated to allow researchers from around the world to access primary sources that were digitized using historical archives’ official websites or digital official repositories.

In those cases, too, further considerations are mandatory. Although praiseworthy, most historical document digitization projects have created websites where Web users can find images lacking computational importance. Photos and scans accompany digitization projects, but they remain facsimiles on which it seems challenging to perform computational analysis through linguistic tools, data mining, and machine learning. Undoubtedly, this way of digitization —photographing!— and even the diffusion on remote repositories has rapidly helped broaden research horizons. This perspective has also been crucial for the Archival field for an in-depth evaluation of its statute, which must modernize the system of conservation and diffusion to adapt it to the dynamism required by analogue and digital communities, that need digital data to be continually shared.

Internet circulation necessarily requires an organization that goes beyond archival classification and “registry entry”. On the Web, sources must be “accessible”, that is “computable”. They must be organized to respond to the need for consultation, but above all, to the founding rule of “digitalization”. Therefore, they need to be encoded in machine-readable texts to let calculators analyze them.

Archival sources must transform into meta-sources accompanied by metadata and encoded texts. For this reason: (1) they need to be transcribed; (2) XML files need to be developed to allow the application of tools and algorithms; (3) they need to be discoverable on the network; (4) they need to be prepared for the use of artificial intelligence and Semantic Web.

² Lawrence Stone, *The Revival of Narrative. Reflections on a New Old History*, «Past & Present», 85 (1979), p. 3–24.

2. The “Correspondence” section of the Biscari archive

In 2018, during conference proceedings in Viterbo, epistolography’s significant scientific value for historical and linguistic research emerged. Epistolography was mentioned as a testimony to the “encounter and dialogue among individuals and cultures, and among the societies that those individuals have in turn articulated and by which they have been articulated”³. Like all the great private archival heritages, the Paternò Castello family’s archive —preserved at the Catania’s State Archive— is the expression of that city’s political and administrative activities and of the role played by the kingdom of Naples in Sicily’s barony life, as described by Guzzetta⁴, Di Vita⁵, and Gazzè⁶, and in the culture of that place⁷.

On the one hand, the Paternò Castello family’s archive can provide answers to further questions about the persons who formed the princes’ network —many historians have tried to rebuild the network of the prince, but historiography has provided a weak definition, undoubtedly influenced by the limit of the close reading approach of the typographic age. It can also answer questions concerning places and events, e.g. the earthquakes that struck Sicily and that created enormous concern, as we can read in the epistles that Michele Maria Paternò sent to princess Anna Maria Morso Bonanno, wife of Ignazio, fifth prince of Biscari. The Biscari Archive contains documents belonging to the Paternò-Castello family (barons of Sigona), the Paternò Tornambene family (barons of Raddusa and Imbaccari), two National Guards of Catania’s protocol —commanded by Agatino Paternò Castello, son of Roberto (8th prince). The archive ends with some envelopes produced by the Moncada Paternò Castello and Leoni families and a small archival collection by the Tedeschi family.

In 1737, Vincenzo (the 4th Prince of Biscari) completed the archive’s first reorganization. The last archive grew and was enriched with new documentation on the branches grafted onto the Paternò Castello family tree. It was correctly described after being stored in the State Archive of Catania in 1975⁸.

If, on the one hand, family wealth (and then the archives) constitutes evidence of individual members’ strategies and how they constructed their genealogies, private correspondence and diaries represent an accurate snapshot of how dynasties related to communities to uphold their prerogatives. Therefore, analyzing such correspondence is crucial in constructing not only the family’s history, but also the history of an entire community and city, which leaves its mark on the families’ records.

³ Paolo Procaccioli, *L’epistolografia di antico regime. Convegno internazionale di studi. Viterbo, 15-16-17 febbraio 2018*, Sarnico: Edizioni di Archilet, 2019.

⁴ Giuseppe Guzzetta, *Per la gloria di Catania: Ignazio Paternò Castello Principe di Biscari*, «Agorà», fasc. VI, (2001).

⁵ Fabio Di Vita, *I Paternò Castello di Biscari. Una famiglia, un patrimonio nella Sicilia moderna, 1700-1734*, Torino: Giappichelli, 2007.

⁶ Lavinia Gazzè, *Fondare una dinastia. Di Ignazio Paternò principe di Biscari (1675-1700)*, «Archivio Storico per la Sicilia Orientale», 2 (2010), p. 11–46.

⁷ Ignazio Paterno principe di Biscari, *Ragionamento a madama N. N. sopra gli antichi ornamenti e trastulli de’ bambini di Ignazio Paternò Castello principe di Biscari ...*, 1781; Id., *De vasi murrini ragionamento d’Ignazio Paternò Castello principe di Biscari &c. &c. accademico della Crusca*, 1781; G. De Gaetani, *Le vicende del passaggio del «Museo Biscari» al Comune di Catania*, Catania: Est il popolo di Sicilia, 1931; Silvia Emanuele, *Il «Museo d’antiquaria» e il «Gabinetto d’istoria naturale» del principe di Biscari a Catania*, «Museologia», (1985), p. 5–26; Giuseppe Giarrizzo – Stefania Pafumi, *Oggetti, uomini, idee: percorsi multidisciplinari per la storia del collezionismo (Atti della tavola rotonda, Catania, 4 dicembre 2006)*, F. Serra, 2009; Anna Maria Iozzia — Cristina Grasso, *I viaggiatori del Settecento e la cultura antiquaria nelle lettere ad Ignazio Paternò Castello, V Principe di Biscari (1719-1786)*, in: Archivio di Stato di Catania. *Un millennio di storia tra le carte d’archivio: documenti dall’XI al XX secolo*, L’Almanacco Editore, 2003; Guido Libertini, *Il Museo Biscari*, 1930; Stefania Pafumi, *Museum Biscarianum. Materiali per lo studio delle collezioni di Ignazio Paternò Castello di Biscari (1719-1786)*, Catania: Alma Editore, 2006; Giuseppe Pagnano, *Le antichità del Regno di Sicilia. I piani di Biscari e Torremuzza per la Regia Custodia 1779*, Siracusa: Lombardi, 2001; Id., *Ignazio Paternò Castello custode della antichità*, «Archivio Storico per la Sicilia Orientale», 2 (2010), p. 47–51.

⁸ Gaetano Calabrese, *L’archivio della famiglia Paternò Castello principi di Biscari: inventario*, Catania: s. n., 2003.

The epic of Paternò-Castello, Princes of Biscari, began in the second half of the 16th century when the union of the Paternò and Castello families gave rise to a vast feudal estate located in various areas of Sicily⁹. Through family ties that relied on traditional (incestuous) agreements and economic alliances, the complex feudal system¹⁰ expanded from Catania, the location of the ancient family palace, to Biscari —now Acate—, a land of forests and farms.

Looking at this historical heritage from a digital methodological point of view, the outcome of the question changes significantly. That family heritage represents an invaluable complex of computable historical data whose heterogeneity can address various research perspectives, not only the research methodology in a digital environment. Moreover, its heterogeneity may support the hypothesis that proper digitization could transform that historical heritage into the “Big Data of History” complex¹¹.

In 2021, the “Archives and Big Data” research fellow project focused on some opportunities of digitizing the “Correspondence” section —creating a database and compiling the digital edition of the missives— in order to join, on the one hand, PNRR, TOCC and PND viewpoints, and on the other hand, to contribute to the dissemination of our national archival heritage, which is vast. Fortunately, the idea of introducing it into the network has fascinated scholars, archivists, and citizens over the years, leading to the blossoming of numerous public and private projects that have been carried out to provide access to complex archival materials to as many scholars (historians) as possible, intending to give them the most faithful proof —such as the family archives, as stated¹²—, of everyday life in order to let them reconstruct society’s institutional, economic, and religious aspects.

Inside this extensive archive, there are 2,000 folders of hundreds of thousands of sheets in which legal disputes, political decisions, and trade and personal letters are recorded. The “Correspondence” section consists of more than 42,493 sheets grouped into 84 archival units on a period ranging from the second half of the seventeenth century to the first half of the twentieth century.

In 1895, the archivist Santoro Rapisarda provided an inventory for identifying its parts. It contains both “private” and purely “scientific” and “administrative” correspondence. The reorganization work, carried out by Gaetano Calabrese, reconstituted the documentation into three subsets.

The first subset is composed of *copialettere* volumes containing administrative affairs. The second subset comprises “personal” and “scientific” correspondence. The third subset includes protocol registers, *pandette*, lists and notes, and various registers.

Within the “Correspondence” section, we can find different types of epistles, ordered by sender and chronologically, followed by a group of cards arranged according to the original alphabetical order and mainly relevant to administrative affairs. The section closes with other correspondence, organized according to the criteria established by the “Regolamento di Servizio pella Segreteria dell’Amministrazione”, introduced in 1845 by Roberto (eighth prince of Biscari) and his sister Marianna. In folder 1642 —which was chosen for creating the Biscari Epistolography digital edition and the website¹³—, we can find 366 epistles and a manuscript by Emile Rousseau (a total of 591 papers), covering a period from 1680 to 1844.

⁹ Lavinia Gazzè, *Fondare una dinastia*, cit.

¹⁰ Francesco San Martino De Spucches — Mario Gregorio, *La storia dei feudi e dei titoli nobiliari di Sicilia dalla loro origine ai nostri giorni (Ristampa)*, Lulu.com, 2013.

¹¹ Frédéric Kaplan — Isabella di Lenardo, *Big Data of the Past*, «Frontiers in Digital Humanities», fasc. 4, (2017).

¹² Antonio Saladino, *Introduzione, Archivi privati. Inventario sommario*, Roma: Ministero dell’Interno (Pubblicazioni degli Archivi di Stato), 1967, vol. 1-2.

¹³ <https://biscariepistolography.altervista.org>.

3. Digitization workflow

Suppose it is true that human reality is increasingly moving towards datafication. In that case, the idea that “digitization necessitates more complex attention and effort when it comes to encoding archival documentation” is even more accurate. This becomes particularly challenging when considering that archival material, especially from the modern era, primarily comprises manuscripts, which constitute a genuine obstacle to the application of computational analysis tools.

Fortunately, the development of Artificial Intelligence tools, e.g. Transkribus and ChatGPT, can, on the one hand, compensate for this intrinsic difficulty in human nature and, on the other hand, let us overcome the idea that digitization workflow is the mere photographic acquisition of a document. Most projects have actually been limited to this: photography. For example, the SAN (Italian National Archival System) has photographed over one and a half million civil registers, but these images remain unresponsive to prompts, queries, and googling. Research is conducted on analogically acquired data, making the portal inadequate for a proper digital perspective.

The archival system thus puts millions of documents online but, within the semantic web infrastructure and neural network, all these facsimiles are meaningless. However, the critical issue lies on another level: no computer can read and “grasp” all the information in a photographed text. Consonants, vowels, and all written signs are invisible to machines. No computer can read a text depicted in an image, just as no educated human can. To a computer, an image is a data set of colour codes, resolution, pixel count, location (for cameras equipped with GPS technology), and similar information.

In our case, to overcome the traditional digitization process perspective, it has been decided to use two innovative tools, Handwritten Text Recognition¹⁴ and GPT3¹⁵, which could approach the digitization workflow to the perspective of coding and processing by computers.

The document acquisition has been carried out using Nikon D610 equipped with the lens AF-S Nikkor 24-120mm f/4G ED VR.

At the State Archives of Catania, there are no publicly accessible spaces for the proper execution of digitization. Consequently, in order to address the challenges associated with the utilization of photographic equipment, as highlighted by the Federal Agencies Digital Guidelines Initiative (FADGI), the photographs have been captured according to the following parameters: (1) shutter speed of 1/6s, facilitating an extended exposure to natural light; (2) aperture set at F/22, accompanied by a dynamic focal point area of 39 points for optimal clarity. The ambient lighting was counterbalanced with warm light at 4200 K to achieve a neutral white balance and permit the camera’s CPU to regulate the white balance.

Subsequently, the various shots have been duplicated for further processing using the Adobe Lightroom software to enhance the contrast of black tones to prepare the files for flipbook creation.

All photos were collected in a Filemaker 19 database that allowed the entry metadata and index persons (sender and consignee) and places (containing GPS coordinates) into a relational structure.

¹⁴ Tomasz Adamek et al., *Word matching using single closed contours for indexing handwritten historical documents*, «International Journal of Document Analysis and Recognition», 9 (2007), n. 2, p. 153–165; The National Archives, *The National Archives - Machines reading the archive: handwritten text recognition software*, text: 2018 <<https://blog.nationalarchives.gov.uk/machines-reading-the-archive-handwritten-text-recognition-software/>>; Mohamed Cheriet et al., *Handwriting recognition research: Twenty years of achievement... and beyond*, «Pattern Recognition», 42 (2009), n. 12, p. 3131–3135.

¹⁵ Jianyang Deng — Yijia Lin, *The Benefits and Challenges of ChatGPT: An Overview*, «Frontiers in Computing and Intelligent Systems», 2 (2022), n. 2, p. 81–83; Aleksandra Fostikov, *First impressions on using AI powered chatbots, tools and search engines: ChatGPT, Perplexity and other – possibilities and usage problems*, (2023); Enkelejda Kasneci et al., *ChatGPT for good? On opportunities and challenges of large language models for education*, «Learning and Individual Differences», 103 (2023), p. 102274.

After meta-dating, each letter has been merged and exported in PDF format and then uploaded to the Transkribus READ servers for automatic transcription¹⁶. This step let us compile their digital edition (Figure 1) on the one hand and generate the PDF files necessary to build the “flipbook” format for website browsing, on the other hand. The project, in fact, tried to respond to different research needs. While the archivist’s role is preserving, maintaining, and ensuring access through the Web, historians, philologists, and linguists look at these documents from a different perspective. Thus, even though the digital edition, for the most part, is aimed at linguistic and philological analysis, the flipbook format will allow historians to read and live a virtual close reading experience.

Scholars can use various computer tools to compile a digital edition, such as the EVT (Edition Visualization Technology)¹⁷ open-source tool. This platform allows a certain freedom of linguistic annotation and the possibility to adapt the edition to different display devices. However, EVT requires ICT skills that humanists generally do not have for the configuration of servers. In addition to this, the need to keep a “machine-readable” text seems crucial; otherwise, the text must be transcribed manually. For this reason, the Transkribus tool was chosen¹⁸, a comprehensive platform that works both on the cloud (Transkribus Lite) and on desktop client solutions (Transkribus Expert) and allows both automatic transcription (with a service fee) and manual transcription (which is free). The transcripts can be annotated and browsed on the Web thanks to the “Read&Search” and the “Transkribus Sites” features¹⁹.

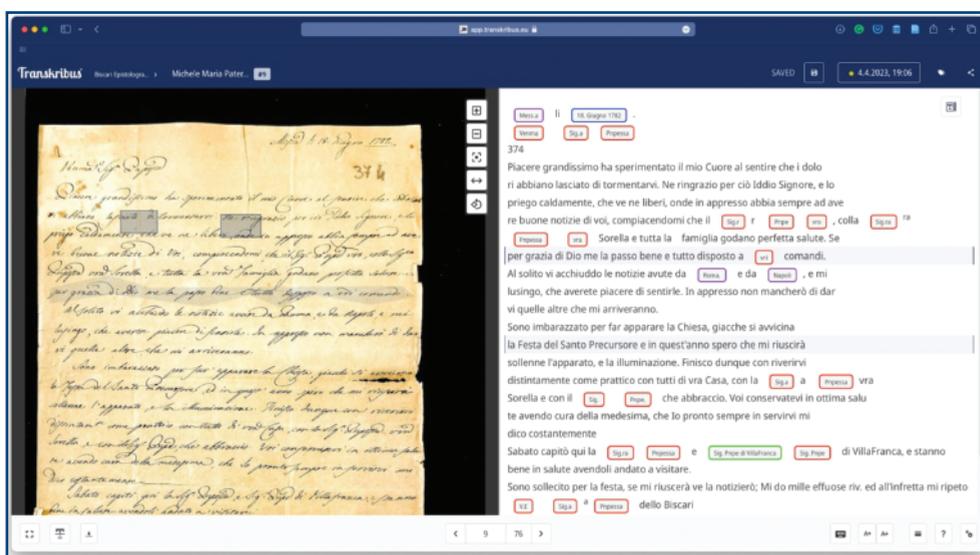


Figura 1. Michele Maria Paternò’s letters digital edition on the Transkribus website

¹⁶ Philip Kahle et al., *Transkribus. A Service Platform for Transcription, Recognition and Retrieval of Historical Documents*, 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017, p. 19–24; Nikolina Milioni, *Automatic Transcription of Historical Documents. Transkribus as a Tool for Libraries, Archives and Scholars*, 2020; Guenter Muehlberger et al., *Transforming scholarship in the archives through handwritten text recognition. Transkribus as a case study*, «Journal of Documentation», 75 (2019), n. 5, p. 954–976.

¹⁷ <http://evt.labcd.unipi.it>.

¹⁸ For details, Salvatore Spina, *Handwritten Text Recognition as a digital perspective of Archival Science*, «Aidainformazioni», 1–2 (2023).

¹⁹ <https://readcoop.eu/success-stories/>.

Additionally, Transkribus was allowed to launch a crowdsourcing initiative. In our case, this feature let some students who attended the “Digital History workshop” (at the Department of Humanities of the University of Catania) access the collection and correct and annotate Michele Maria Paternò’s epistles. One of the purposes of the HTR AI tool is that the transcribed document can be downloaded in different formats for personal research. It is possible to download a DOC file to use the transcribed text in an essay or typographical edition and, above all, an XML file that can be used for computational analysis²⁰.

4. Website and browsing

We archive data to ensure the proper functioning of governing bodies; archival documents also guarantee ownership of rights in rem. Nowadays, thanks to ICT development, we archive data to let researchers replicate findings and repurpose that data to answer new research questions.

To make it possible, researchers have to be able to discover and access the data. In the digital world, this new stance overcame the traditional use of archival catalogues and led archivists to develop online versions of archive registers. Archives usually provide catalogues... oftentimes, search engines such as Google provide metadata for harvesting, making data even more findable. Persistent identifiers such as DOIs or URNs are offered to shape uniquely identifiable and citable data.

Archives must set access conditions according to the new legal regulations and ethical standards. They must also have infrastructures in place that make access possible, like direct online access from the catalogue, or, for very sensitive data, access from a protected on-site room. Moreover, data archives seek to preserve data for future use, regardless of potential changes in technology or in the research practices. This condition leads archivists to monitor their environment very closely: will there be more changes to the software that renders data? Or will standard terminology evolve in ways that might im-



Figura 2. Biscari Epistology website

²⁰ R. Ahnert et al., *Networking Archives: Quantitative History and the Contingent Archive*, *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*, 2020, p. 397–419; Steven Bankes et al., *Making Computational Social Science Effective: Epistemology, Methodology, and Technology*, «Social Science Computer Review», 20 (2002), n. 4, p. 377–388; *Computational lexicography for natural language processing*, ed. by B. Boguraev, T. Briscoe, USA: Longman Publishing Group, 1989; Giovanni Camardi, *Computational models and information theory*, «Journal of Experimental & Theoretical Artificial Intelligence», 24 (2012), p. 401–417; Gregory Crane et al., *Participatory Philology: Computational Linguistics and the Future of Historical Language Education*, «Human Computation», 1 (2014), n. 2, p. 177–184; Rodolfo Delmonte, *Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution*, New York: Nova Science Publishers, 2009.

pact the ability to understand data in the future? If this is the case, archives have to plan new measures to ensure that data continue to be usable and understandable despite those changes, by migrating them into a new format or updating metadata, for instance.

Therefore, due to the above mentioned reasons, disseminating documents through the Web is not an easy task for archivists. Beyond AI or EVT tools, digitizing the national archival heritage still appears remote. However, among the boundless possibilities provided by technology, the increasingly rooted presence of archives in the network, also thanks to the work of individual scholars, must be taken into account.

Biscari Epistolography website is also this (Figure 2).

After the automatic transcription and annotation, each letter has been exported in PDF format to be converted into “flipbook” format to allow a close-reading virtual experience. The “flipbook” format is an interactive digital text based on PDF files that enables internal searching and enhances the reading experience with multimedia elements capable of taking the reading experience to hypertextual design dimensions (Figures 3 and 4).

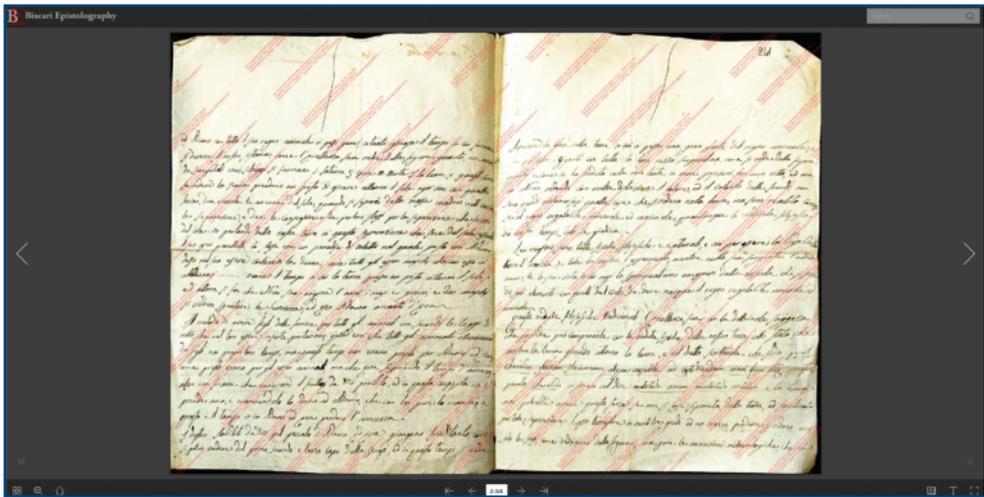


Figura 3. The flipbook formats

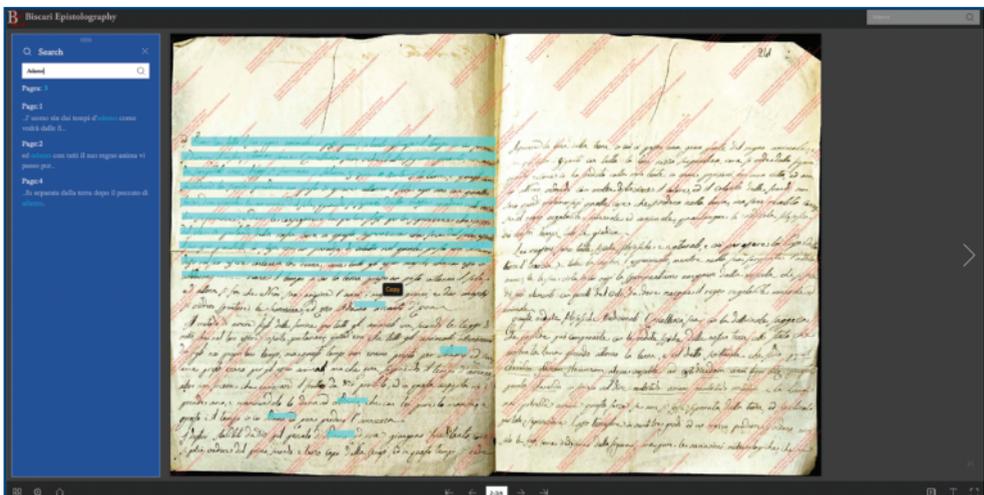


Figura 4. The flipbook and its interactive feature

Meanwhile, the XML-TEI file (Figure 5) is set for free download to let scholars conduct textual or linguistic analysis independently.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml">
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title type="main">Archivio Biscari, Corrispondenza, 1642, Michele Maria Paternò</title>
        <author>Salvatore Spina</author>
      </titleStmt>
      <seriesStmt>
        <title>salvatore.spina@unicit.it Collection</title>
      </seriesStmt>
      <sourceDesc>
        <bibl>
          <title type="main">Archivio Biscari, Corrispondenza, 1642, Michele Maria Paternò</title>
          <author>Salvatore Spina</author>
          <idno type="Transkribus">1020863</idno>
          <idno type="external">NA</idno>
          <note>Lettere manoscritte di Michele Maria Paternò alla principessa di Biscari, Anna Maria Morso Bonanno, moglie di Ignazio Paternò Castello, 5a principe di Biscari.</note>
        </bibl>
      </sourceDesc>
      <facsimile>
        <surface lra="2074" lry="3328" lxx="0" lyy="0" xmid="fac_1">
          <graphic height="3328px" url="https://files.transkribus.eu/GetFile-UQTXNUNSEEMAOQITUBJGALOC&fileType=view" width="2074px" />
          <graphic height="3328px" url="https://files.transkribus.eu/GetFile-UQTXNUNSEEMAOQITUBJGALOC&fileType=view" width="2074px" />
        </surface>
        <zone points="-120,220 120,2386 1719,220" rendition="TextRegion" xmid="fac_1_r2">
          <zone points="1035,303 1114,329 1170,314 1198,338 1313,309 1401,366 1492,326 1667,315 1666,255 1503,284 1436,237 1297,264 2185,266 1148,237 1102,271 1035,260" rendition="Line" xmid="fac_1_r2i2" />
          <zone points="1066,425 1110,447 1197,409 1237,432 1308,412 1419,431 1548,420 1635,388 1668,330 1538,351 1484,313 1444,329 1395,312 1210,339 1170,313 1065,340" rendition="Line" xmid="fac_1_r2i3" />
          <zone points="166,498 394,486 452,530 623,493 688,510 791,472 855,478 855,432 703,387 676,403 574,356 451,422 372,393 302,427 249,384 165,408" rendition="Line" xmid="fac_1_r2i4" />
          <zone points="143,603 762,553 1567,601 1696,548 1695,447 1258,455 1129,524 819,473 456,530 142,499" rendition="Line" xmid="fac_1_r2i5" />
          <zone points="113,664 1704,681 1703,593 118,605" rendition="Line" xmid="fac_1_r2i6" />
          <zone points="113,749 1683,763 1681,686 1470,647 1133,715 491,661 252,720 112,658" rendition="Line" xmid="fac_1_r2i7" />
          <zone points="119,863 606,831 683,876 1013,877 1543,842 1559,790 214,799 118,756" rendition="Line" xmid="fac_1_r2i8" />
          <zone points="216,942 1694,953 1693,864 1336,828 1199,879 579,839 430,889 216,873" rendition="Line" xmid="fac_1_r2i9" />
        </zone>
      </facsimile>
    </fileDesc>
  </teiHeader>
  <text>
    <p></p>
  </text>
</TEI>
```

Figura 5. XML-TEI example

The records of the digital epistolary, on the other hand, have been compiled thanks to the use of LLM ChatGPT — as described in Spina²¹ — whose internal structure and training enabled a necessary ability for entity recognition²². GPT-3.5 allowed the analysis of all the letters, extraction of various entities (names, dates, places, and events), and compilation of the searchable digital database on the website (Figure 6).

Epistolografia digitale

Le lettere sono parte della sezione "Corrispondenza" (busta 1642) dell'Archivio Biscari, depositato presso l'Archivio di Stato di Catania. I record del database sono organizzati per "Mittente" (in ordine alfabetico per "nome") e corredati dalle informazioni relative alla data di scrittura, destinatari, luoghi emissione e destinazione, e numero di foglio (la collocazione interna alla busta). Ogni lettera può essere consultata in formato Flipbook, e scaricata in formato XML-TEI. È, inoltre, possibile consultare l'edizione digitale sulla piattaforma Transkribus.

Dettagli
 SEGNAURA: Archivio di Stato di Catania, Biscari, *Corrispondenza*, b. 1642.
 SOGGETTO PRODUTTORE: Famiglia Paternò Castello, principi di Biscari
 SOGGETTO CONSERVATORE: Archivio di Stato di Catania
 PERSONE: Varie
 DATAZIONE: 1680–1844
 SUPPORTO: 599 fogli cartacei raccolti

Copy CSV Excel PDF

Visualizza elementi

Cerca:

Mittente	Descrizione	Destinatario	Descrizione	Data	Emissione	Destinazione	Foglio	Flipbook	XML-TEI
Michele Maria Paternò	Priore di Messina	Anna Maria Morso Bonanno	Principessa di Biscari	18 maggio 1784	Messina	Catania	390	Consulta	Scarica

- **Altri soggetti**
Padron Micale, Principessa di Cerami, Marchese Cardillo.
- **Altri luoghi**
Napoli, Spagna, Algeri, Civitavecchia.
- **Altre date**
11 maggio 1784, 14 maggio 1784.
- **Eventi narrati**
Donazione della pietra dallo schiavo liberato, Approdo di bastimento con uomini che muoiono dopo tre giorni dall'approdo.

Mittente Descrizione Destinatario Descrizione Data Emissione Destinazione Foglio Flipbook XML-TEI

Figura 6. Record example

²¹ Salvatore Spina, *Artificial Intelligence in archival and historical scholarship workflow: HTR and ChatGPT: «Umanistica Digitale»*, (2023).

²² Carlos-Emiliano González-Gallardo et al., *Yes but.. Can ChatGPT Identify Entities in Historical Documents?*, 2023; Salvatore Spina, *Artificial Intelligence in archival and historical scholarship workflow*, cit.

The internal search engine has been developed to enable a case-insensitive search and perform queries on terms distributed across all columns rather than individual lines, allowing to identify words even if they were not consecutively written in the records. For instance, the identification of an epistle from Ignazio Paternò to anyone named “Michele” is performed by simply typing the names “Ignazio” and “Michele” in the search engine bar to display all records where both terms are present. By adding terms to the search, such as a date and/or location, the engine is able to identify all records where those three terms are present (Figures 7 and 8). This setup makes the search much more straightforward and avoids using tags or restricting the search to metadata alone.

Visualizza 5 elementi

Cerca: ignazio michele

Foglio	Mittente	Descrizione	Destinatario	Descrizione	Data	Emissione	Destinazione	Flipbook
22	Michele Ardito	Non descritto	Ignazio Paternò Castello	5° Principe di Biscari	22 giugno 1782	Napoli	Non indicato	Consulta
519	Michele Vargas	Duca	Ignazio Paternò Castello	5° Principe di Biscari	09 agosto 1777	Napoli	Non indicato	Consulta

Vista da 1 a 2 di 2 elementi (filtrati da 367 elementi totali)

Inizio 1 Fine

Figura 7. “ignazio michele” search results

Visualizza 5 elementi

Cerca: ignazio michele

Foglio	Mittente	Descrizione	Destinatario	Descrizione	Data	Emissione	Destinazione	Flipbook
22	Michele Ardito	Non descritto	Ignazio Paternò Castello	5° Principe di Biscari	22 giugno 1782	Napoli	Non indicato	Consulta
519	Michele Vargas	Duca	Ignazio Paternò Castello	5° Principe di Biscari	09 agosto 1777	Napoli	Non indicato	Consulta

Vista da 1 a 2 di 2 elementi (filtrati da 367 elementi totali)

Inizio 1 Fine

Figura 8. “ignazio michele 1777” search results

The browsing is free, and it does not require subscriptions to the website to ensure the lack of traceability of the user and the dissemination of this historical heritage through the Web.

In addition to virtual consultation, the database structure allows different ways of browsing based on different kinds of information: sender, consignee, other named persons, dates (issues and any date in the body field), places (issue, destination and others mentioned), and narrated events.

5. Achievement

Digitality changes our perspective on Archival Science and History research fields. The term “digitization” has pervasively entered the vocabulary of our society. A new upper-boosted meta-source²³ emerged —aggregating different pieces of information. In the digital environment, archival heritage can become an open-source hypertext that guarantees multiple access. After a specific and laborious coding work (meta dating) of the data to be indexed, a digital edition will perform all those functions that traditionally rely on the long compilation of archival registers, letting archivists automatically index all information that makes epistolography complex²⁴.

The few design experiences have demonstrated the feasibility of creating hypertext editions of archival documentation; however, they have highlighted the difficulties of this path, which, without adequate measures and technologically advanced instrumentation, becomes a titanic task²⁵.

By welcoming Albonico’s ideas²⁶, the Biscari Epistolography’s digital edition was designed to fill the gap imposed on those Web users that are eager to access simple editorial products that virtualize analogue reality. This simple and functional solution, accompanied by its XML-TEI version, changed the edition into a dataset on which historians, linguists and philologists can conduct further computational analysis.

In addition to this, the final project represents a moment of evaluation of the prospects offered by Transkribus AI and its features.

This platform constitutes one of the best tools available to any scholar who operates in the world of archives.

Created as a tool for htr archival heritage, it offers all the options for creating XML-TEI formats without requiring Computer Science skills from users, the possibility of crowdsourcing, a non-scale configuration and the possibility of working in the cloud (Transkribus Lite).

From now on, the Biscari Epistolography’s digital edition will be available to the scientific community and constantly updated. Even though this is a “pilot projects”, it intends to be a plea for planning and building an IT infrastructure capable of guaranteeing access to the Italian archival heritage.

Beyond the pure analysis, the study seeks to rethink digitalization, digitization, and digital transformation. The digital turn leads archivists and scholars to redefine preserving and browsing²⁷. The Web and the Internet are the evidence of a society that aspires to build the digital Era of Culture and Knowledge²⁸. However, we are far from a digital world where all analogue data and information — documents and cultural heritage— are open-source, machine-readable, and correctly encoded. Most digitization projects, funded by governments and research centres, merely publish meaningless archival document facsimiles.

The digital revolution needs something different. Meta-dating alone is inadequate to boost the semantic Web and speak about *infosphere*²⁹. It is therefore necessary to overcome the analogue approach underlying most digitization projects and to go beyond the sterile storing approach.

²³ Domenico Fiorimonte, *Il documento immateriale. Ricerca storica e nuovi linguaggi*, «L’Indice», fasc. 5, Dossier 4 (2000).

²⁴ Enrico Garavelli, *Considerazioni per l’edizione digitale di un corpus epistolare, L’epistolografia di Antico Regime: Convegno internazionale di studi, Viterbo, 15-16-17 febbraio 2018*, Sarnico: Edizioni di Archilet, 2019, p. 322–330.

²⁵ Salvatore Spina, *Digital History. Metodologie informatiche per la ricerca storica*, Napoli: Edizioni Scientifiche Italiane, 2022.

²⁶ Simone Albonico, *Epistulae*. <<http://epistulae.unil.ch>>, *L’epistolografia di Antico Regime: Convegno internazionale di studi, Viterbo, 15-16-17 febbraio 2018*, Sarnico: Edizioni di Archilet, 2019, p. 315–321.

²⁷ Federico Valacchi, *Gli archivi tra storia uso e futuro*, Milano: Editrice Bibliografica, 2021.

²⁸ Salvatore Spina, *Digital History*, cit.

²⁹ Luciano Floridi, *The 4th revolution, how the infosphere is reshaping human reality*, Oxford; New York: Oxford University Press, 2016.

Il digital turn ha creato un ecosistema interrelazionale tra uomini e macchine, in grado di generare dati e conoscenza. Questa svolta paradigmatica richiede nuovi approcci archivistici, così come avvenuto durante il progetto di digitalizzazione della sezione "Corrispondenza" dell'Archivio Biscari, che ha dimostrato come, sfruttando strumenti di intelligenza artificiale quali Transkribus e ChatGPT, il processo di digitalizzazione può certamente affrontare le sfide dell'encoding dei materiali archivistici. L'edizione digitale e la virtualizzazione della consultazione dei documenti del sito web dell'epistolografia dei Biscari (<https://biscari-epistolography.altervista.org>), infatti, sono stati pensati per facilitare l'esplorazione dei documenti storici, attraverso il web; considerando, inoltre, che la diffusione digitale mira a preservare il patrimonio culturale e consentire l'accesso a dati scientificamente rilevanti per le generazioni di studiosi future.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2023

REFERENCES

Tomasz Adamek — Noel E. O'Connor — Alan F. Smeaton. *Word matching using single closed contours for indexing handwritten historical documents*. «International Journal of Document Analysis and Recognition», 9 (2007), n. 2, p. 153–165.

Ruth Ahnert — Sebastian E. Ahnert — Yann Ryan — Computational Humanities Research (CHR 2020). *Networking Archives: Quantitative History and the Contingent Archive, Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*, 2020, p. 397–419.

Simone Albonico. *Epistulae*. <<http://epistulae.unil.ch>>, *L'epistolografia di Antico Regime: Convegno internazionale di studi, Viterbo, 15-16-17 febbraio 2018*. Sarnico: Edizioni di Archilet, 2019, p. 315–321.

The National Archives. *The National Archives - Machines reading the archive: handwritten text recognition software* text. 2018, <<https://blog.nationalarchives.gov.uk/machines-reading-the-archive-handwritten-text-recognition-software/>>.

Steven Bankes — Robert Lempert — Steven Popper. *Making Computational Social Science Effective: Epistemology, Methodology, and Technology*. «Social Science Computer Review», 20 (2002), n. 4, p. 377–388.

Ignazio Paterno principe di Biscari. *De vasi murrini ragionamento d'Ignazio Paternò Castello principe di Biscari &c. &c. accademico della Crusca*. 1781.

Ignazio Paterno principe di Biscari. *Ragionamento a madama N. N. sopra gli antichi ornamenti e trastulli de' bambini di Ignazio Paternò Castello principe di Biscari ...*. 1781.

Computational lexicography for natural language processing, ed. by B. Boguraev, T. Briscoe. USA: Longman Publishing Group, 1989.

Gaetano Calabrese. *L'archivio della famiglia Paternò Castello principi di Biscari: inventario*. Catania: [s. n.], 2003.

Giovanni Camardi. *Computational models and information theory*. «Journal of Experimental & Theoretical Artificial Intelligence», 24 (2012), p. 401–417.

Mohamed Cheriet — Mounim El Yacoubi — Hiromichi Fujisawa — Daniel Lopresti — Lorette Guy. *Handwriting recognition research: Twenty years of achievement... and beyond*. «Pattern Recognition», 42 (2009), n. 12, p. 3131–3135.

Gregory Crane — Stella Dee — Anna Krohn. *Participatory Philology: Computational Linguistics and the Future of Historical Language Education*. «Human Computation», 1 (2014), n. 2, p. 177–184.

Giovanni De Gaetani. *Le vicende del passaggio del «Museo Biscari» al Comune di Catania*. Catania: Est il popolo di Sicilia, 1931.

Rodolfo Delmonte. *Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution*. New York: Nova Science Publishers, 2009.

Jiayang Deng — Yijia Lin. *The Benefits and Challenges of ChatGPT: An Overview*. «Frontiers in Computing and Intelligent Systems», 2 (2022), n. 2, p. 81–83.

Fabio Di Vita. *I Paternò Castello di Biscari. Una famiglia, un patrimonio nella Sicilia moderna, 1700-1734*. Torino: Giappichelli, 2007.

Silvia Emanuele. *Il «Museo d'antiquaria» e il «Gabinetto d'istoria naturale» del principe di Biscari a Catania*. «Museologia», (1985), p. 5–26.

Domenico Fiormonte. *Il documento immateriale. Ricerca storica e nuovi linguaggi*. «L'Indice», 5 (2000), n. 4.

Luciano Floridi. *The 4th revolution, how the infosphere is reshaping human reality*. Oxford; New York: Oxford University Press, 2016.

Aleksandra Fostikov. *First impressions on using AI powered chatbots, tools and search engines: ChatGPT, Perplexity and other – possibilities and usage problems*. (2023).

Enrico Garavelli. *Considerazioni per l'edizione digitale di un corpus epistolare, L'epistolografia di Antico Regime: Convegno internazionale di studi, Viterbo, 15-16-17 febbraio 2018*. Sarnico: Edizioni di Archilet, 2019, p. 322–330.

Lavinia Gazzè. *Fondare una dinastia. Di Ignazio Paternò principe di Biscari (1675-1700)*. «Archivio Storico per la Sicilia Orientale», 2 (2010), p. 11–46.

Giuseppe Giarrizzo — Stefania Pafumi. *Oggetti, uomini, idee: percorsi multidisciplinari per la storia del collezionismo (Atti della tavola rotonda, Catania, 4 dicembre 2006)*. F. Serra, 2009.

Carlos-Emiliano González-Gallardo — Emanuela Boros — Nancy Girdhar — Ahmed Hamdi — Jose G. Moreno — Antoine Doucet. *Yes but.. Can ChatGPT Identify Entities in Historical Documents?*. 2023.

Giuseppe Guzzetta. *Per la gloria di Catania: Ignazio Paternò Castello Principe di Biscari*. «Agorà», VI (2001).

Anna Maria Iozzia — Cristina Grasso. *I viaggiatori del Settecento e la cultura antiquaria nelle lettere ad Ignazio Paternò Castello, V Principe di Biscari (1719-1786)*, in: Archivio di Stato di Catania. *Un millennio di storia tra le carte d'archivio: documenti dall'XI al XX secolo*. Catania: L'Almanacco Editore, 2003.

Philip Kahle — Sebastian Colutto — Günter Hackl — Günter Mühlberger. *Transkribus. A Service Platform for Transcription, Recognition and Retrieval of Historical Documents, 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 2017, p. 19–24.

Frédéric Kaplan — Isabella di Lenardo. *Big Data of the Past*. «Frontiers in Digital Humanities», 4 (2017).

Enkelejda Kasneci — Kathrin Sessler — Stefan Küchemann — Maria Bannert — Daryna Dementieva — Frank Fischer — Gjergji Kasneci. *ChatGPT for good? On opportunities and challenges of large language models for education*. «Learning and Individual Differences», 103 (2023), p. 102274.

Guido Libertini. *Il Museo Biscari*. 1930.

Nikolina Milioni. *Automatic Transcription of Historical Documents. Transkribus as a Tool for Libraries, Archives and Scholars*. 2020.

Gunter Muehlberger — Louise Seaward — Melissa Terras — Sofia Ares Oliveira — Vicente Bosch — Maximilian Bryan — Konstantinos Zagoris. *Transforming scholarship in the archives through handwritten text recognition. Transkribus as a case study*. «Journal of Documentation», 75 (2019), n. 5, p. 954–976.

Stefania Pafumi. *Museum Biscarianum. Materiali per lo studio delle collezioni di Ignazio Paternò Castello di Biscari (1719-1786)*. Catania: Alma Editore, 2006.

Giuseppe Pagnano. *Le antichità del Regno di Sicilia. I piani di Biscari e Torremuzza per la Regia Custodia 1779*. Siracusa: Lombardi, 2001.

Giuseppe Pagnano. *Ignazio Paternò Castello custode della antichità*. «Archivio Storico per la Sicilia Orientale», 2 (2010), p. 47–51.

Paolo Procaccioli. *L'epistolografia di antico regime. Convegno internazionale di studi. Viterbo, 15-16-17 febbraio 2018*. Sarnico: Edizioni di Archilet, 2019.

Antonio Saladino. *Introduzione, Archivi privati. Inventario sommario*. Roma: Ministero dell'Interno (Pubblicazioni degli Archivi di Stato), 1967, vol. 1-2.

Salvatore Spina. *Digital History. Metodologie informatiche per la ricerca storica*. Napoli: Edizioni Scientifiche Italiane, 2022.

Salvatore Spina. *Handwritten Text Recognition as a digital perspective of Archival Science*. «Aidainformazioni», 1–2 (2023).

Salvatore Spina. *Artificial Intelligence in archival and historical scholarship workflow: HTR and ChatGPT*. «Umanistica Digitale», (2023).

Salvatore Spina. *Homo-Loggatus. The anthropological condition of historians in the digital world*. «Journal of Mathematical Techniques and Computational Mathematics», 2-10, (2023), p. 431-437.

Salvatore Spina. *Homo-loggatus? Come stare dentro la nicchia ecologica digitale*, in: *Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD*. Billet, 2023 <<https://informa.hypotheses.org/2224>>.

Francesco San Martino De Spucches — Mario Gregorio. *La storia dei feudi e dei titoli nobiliari di Sicilia dalla loro origine ai nostri giorni (Ristampa)*. Lulu.com, 2013.

Lawrence Stone. *The Revival of Narrative. Reflections on a New Old History*. «Past & Present», 85 (1979), p. 3–24.

Federico Valacchi. *Gli archivi tra storia uso e futuro*. Milano: Editrice Bibliografica, 2021.