ABSTRACT

The dataset contains metadata describing the Princes of Biscari (Paternò Castello) correspondence, one of the most important noble families in Sicily during the modern era. The dataset was created with the digitization of letters preserved in the Biscari Archive, held at the State Archive of Catania, and this digitization initiated a Historical Network Analysis research project aimed at reconstructing the family’s relationships and how its members interacted with the political class, the religious class, and the monarchy. After digitization, the letters were curated in a database and analyzed.
(1) OVERVIEW

Repository Location: https://doi.org/10.5281/zenodo.8340192

CONTEXT

In 2021, the “Archives and Big Data” research fellow project focused on digitizing the “Correspondence” section of the Biscari Archive (Paternò Castello family) located at the State Archive of Catania (Italy). I created a database and compiled the digital edition of the missives in order to support the PNRR (National Recovery and Resilience Plan) philosophy, and contribute to the dissemination of the Italian archival heritage, through a website.

Inside the extensive archive are 2,000 folders composed of hundreds of thousands of documents (legal disputes, political decisions, trade records, and personal letters). The “Correspondence” section consists of more than 42,493 sheets grouped into 84 archival units, representing a wide range of dates from the second half of the seventeenth century through to the first half of the twentieth century. Within this section, there are different types of epistles, ordered by the sender and chronologically, followed by cards arranged according to the original alphabetical order and mainly relevant to administrative affairs. The section closes with other correspondence, organized by criteria established by the “Regolamento di Servizio pella Segreteria dell’Amministrazione”, introduced in 1845 by Roberto (eighth prince of Biscari) and his sister Marianna (Calabrese, 2003).

In folder 1642, which was chosen for creating the Biscari Epistolography digital edition and the website, there are 366 epistles and a manuscript by Emile Rousseau (a total of 591 papers), covering a period from 1680 to 1844. The documents have been documented in several papers published in the scientific journals «Umanistica Digitale» (Spina, 2023a) and «Aidainformazioni» (Spina, 2023b). The essays highlight contemporary methodologies available to historians, such as artificial intelligence tools like Transkribus and ChatGPT, as well as some tools in Computational Linguistics, such as Keyphrase Digger.

(2) METHOD

STEPS – THE AIM

Digitization necessitates more complex attention and effort when it comes to encoding archival documentation. This is particularly challenging when the archival material primarily comprises manuscripts, which pose a genuine obstacle to the application of computational analysis tools. When discussing “digitization”, it is important to consider encoding processes that create a machine-readable text, not simply a photographic acquisition (Spina 2022a). It is also important to consider the difficulties historians face when manually transcribe thousands and perhaps more archival documents. Fortunately, thanks to the development of two artificial intelligence tools, Transkribus and ChatGPT, scholars can move beyond mere photographic acquisition of a document, a serious limitation in the field (Adamek, O’Connor, and Smeaton, 2007; Cheriet et al., 2009; Archives, 2018; Deng and Lin, 2022; Kasneci et al., 2023; Fostikov, 2023).

STEPS – DIGITIZATION WORKFLOW

I captured the document photographs with a Nikon D610, equipped with a AF-S Nikkor 24–120 mm f/4G ED VR lens. To address the challenges associated with this photographic equipment, as highlighted by the Federal Agencies Digital Guidelines Initiative (FADGI), the photographs were captured utilizing 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 were duplicated for further processing using Adobe Lightroom software to enhance the contrast of black tones and to prepare the files for flipbook processing.
**STEPS – THE DATABASE AND THE AUTOMATIC TRANSCRIPTION**

All photos were collected in a Filemaker 19 database, which afforded entry of metadata and indexing of all entities (sender, consignee, places, date) into a relational structure.

After meta-dating, each letter was merged and exported in PDF format and then uploaded to the Transkribus READ servers for automatic transcription (Kahle et al., 2017; Muehlberger et al., 2019; Milioni, 2020). This application compiled the digital edition and created PDF files that let me create the “flipbook” format for website browsing, and the XML-TEI files for computational and linguistic analysis.

**STEPS – CHATGPT**

To correct the transcriptions, and to compile the records of the digital epistolography, I elected to test the LLM (GPT 3.5), whose internal structure and the training support entity recognition (González-Gallardo et al., 2023; Spina, 2023a). GPT-3.5 facilitated analysis of all the letters, extraction of various entities (names, dates, places, and events), and compilation of the searchable digital database on the website.

**STEPS – THE WEBSITE**

The https://biscariepistolography.altervista.org website was designed to meet the expectations of the historical community, who demand increased internet accessibility to archival material for their research, allowing them to navigate through digital documents. For this reason, each epistle has been exported in PDF format and converted into a “flipbook” format to enable a close-reading virtual experience. A flipbook is an interactive digital text based on PDF files that facilitates internal searching and enhances the reading experience with multimedia elements, thus expanding the reading experience into hypertextual design dimensions.

On the other hand, XML-TEI files have been made available (open access) to all scholars who would like to go beyond simple reading and wish to analyze the documents using algorithms and tools in Computational Linguistics. For instance, Keyphrase Digger (Moretti, Sprugnoli, and Tonelli, 2015) has been utilized to extract key concepts from all the epistles, aiming to identify valuable information for reconstructing events experienced by members of the Paternò Castello family.

**(3) DATASET DESCRIPTION**

- **Object name** – Prince of Biscari Network
- **Format names and versions** – Biscari_Network.CSV, Version 1
- **Creation dates** – Start: 2021-06-04, End: 2023-02-26
- **Dataset creators** – Salvatore Spina, Department of Humanities – University of Catania
- **Language** – Italian
- **License** – Creative Commons Attribution 4.0 International
- **Repository name** – Zenodo
- **Publication date** – 2023-09-13

**(4) REUSE POTENTIAL**

By embracing Albonico’s ideas (Albonico, 2019) on data interoperability and the need for the development of computer tools enabling rapid data exchange, the digital edition of Biscari Epistolography, which is essential for elucidating the political, economic, and cultural facets of the Baroque and Enlightenment periods (Procaccioli, 2019; Spina, 2022b), was meticulously designed to bridge the gap between demand for accessible editorial products that simulate their analog counterparts. For this reason, the straightforward and practical solution of creating the epistles’ digital edition, coupled with its XML-TEI version, transforms the paper edition into a dataset conducive to additional computational analyses for historians, linguists, and philologists.
The Computerized Age is founded on several pillars: (1) prompt access; (2) global dissemination; (3) interoperability; (4) entities interplay; (5) open access; and (6) sharing and participation. Specifically, in line with the latter principle, Archives and Libraries need to initiate digitization projects for their paper heritage to create digital complexes that can be shared.

However, the Internet is not merely a platform for visualization. The digital world is a space not only for sharing but also for processing and computation. Furthermore, “sharing” should not only concern images but also ensure adequate encoding of documents in a machine-readable format to guarantee data processability. This need has led to platforms like Zenodo, GLAM labs, and AI4LAM, which enable participation in a “person2persons2machines” community or a “Digital Ecological Niche (DEN)”, where everyone is Homo-Loggatus (Spina, 2023c) and can create new knowledge. In the case of the Prince of Biscari Network, the open-access Zenodo dataset is the most effective means to showcase how scholars are constructing the digital Era of Culture and Knowledge (Spina, 2022a), empowering everyone to enhance their research projects and move beyond a sterile storing approach.

The “Biscari_Network.CSV” dataset can guide historians in analyzing Sicilian society through the “Big Data of Sicilian History.” On one hand, it allows for the re-evaluation of historical narratives about Sicily, and on the other hand, it facilitates the description of the Sicilian nobles’ network (Brügger, 2013; Erickson, 1997) and their role at the court of the king of Naples (Guzzetta, 2001; Gazzè, 2010; Iozzia and Grasso, 2003; Muscolino, 2015; Pagnano, 2001; Alberghina, 2010; Di Vita, 2007; Giarrizzo and Pafumi, 2009; Giarrizzo, 1978; Giarrizzo and Aymard, 2006).

COMPETING INTERESTS
The author has no competing interests to declare.

AUTHOR CONTRIBUTIONS
Salvatore Spina, Research Fellow, University of Catania/Data Curation, Formal Analysis, Methodology, Software, Writing/“Archivi e Big Data” Research fellow project.

AUTHOR AFFILIATIONS
Salvatore Spina orcid.org/0000-0001-6367-8183
Department of Humanities, University of Catania, Catania, Italy

REFERENCES
Cheriet, M., Yacoubi, M. E., Fujisawa, H., Lopresti, D., & Guy, L. (2009). Handwriting Recognition Research: Twenty Years of Achievement... and Beyond. Pattern Recognition, 42(12), 3111–35. DOI: https://dx.doi.org/10.1016/j.patcog.2009.03.014