Models of Data Extraction and Architecture in Relational Databases of Early Modern Private Political Archives
edited by Dorit Raines

Historical Research and Archival Sciences in a Digital Perspective
Relational Database, Data Architecture and Data Extraction in Graziani Archives Portal

Dorit Raines
Università Ca' Foscari Venezia, Italia

Abstract Historians often struggle with organizing unstructured data, as existing metadata standards and relational databases do not fully capture the complexity of historical narratives. While archival sciences have adapted to digital environments, historians face challenges in structuring data, particularly for fragmented private collections that do not conform to standard models. This article examines key decisions in a Digital Humanities project on historical sources, including platform selection, data structuring, and metadata integration. Using the *Nuncio's Secret Archives* project as a case study, it highlights strategies for standardizing, contextualizing, and interlinking data to enhance searchability and facilitate customized analysis.

Keywords Relational database. Web portal. Data architecture. Data ingestion. Front-end customization.

1 Introduction

Unlike archivists, historians today often struggle with organizing unstructured data. Archival sciences, like library and information sciences, have long established metadata standards for their respective cultural artefacts and have further advanced these standards

Edizioni Ca'Foscari

Studi di archivistica, bibliografia, paleografia 8
e-ISSN 2610-9093 | ISSN 2610-9875
ISBN [ebook] 978-88-6969-919-1 | ISBN [print] 978-88-6969-920-7

Peer review | Open access
Submitted 2025-03-11 | Accepted 2025-04-28 | Published 2025-05-22

DOI 10.30687/978-88-6969-919-1/010

with the advent of digital technology, adapting them to the digital environment. The development of type-specific databases and, subsequently, relational databases that connect digital objects through their properties has significantly advanced archival, library, and cultural studies over the past two decades in unprecedented ways (Kamal, Golub 2025; Aliaga, Bertino, Valtolina 2011; Dallas 2004, 283). Historians, however, have remained somewhat at an impasse. While they have made progress with network analysis, they have faced considerable challenges in representing the complex contents of objects (whether records or other sort of narration) within a binary system.

Historians are primarily concerned with the content of cultural objects and their significance, in addition to the provenance metadata associated with these objects. Existing digital portals that offer tools for constructing customized relational databases are based on conceptual models that describe resources, yet they lack a corresponding conceptual model for content. As a result, even if historians do not attempt to encode the full complexity of historical narratives within a relational database and instead focus on extracting recurring entities to facilitate key access to the text, the task remains complex. Difficult decisions must often be made to ensure conceptual clarity, including standardizing, disambiguating, or even interpreting entities within their historical contexts.

A substantial body of literature currently addresses various aspects of the digitization and indexing of archival sources (for example, Gilliland, McKermmish, Lau 2017; Opgenhaffen 2022; Hawkins 2022). However, while debates within the Digital Humanities community have largely focused on the digitization of entire collections held by cultural institutions or on the development of digital libraries (Agosti, Ferro, Silvello 2011, 17), or on the archive as a research laboratory and its role in the changing digital scenery (Trace 2022) relatively little attention has been given to private archival collections. These collections, often fragmented across multiple institutions and private owners due to historical contingencies, do not always conform to existing conceptual models. Consequently, adjustments are necessary to enable both a diachronic and synchronic virtual reconstruction of such collections.

This article seeks to highlight the decision-making processes that teams of historians navigate when undertaking a Digital Humanities project centered on historical sources, particularly private archival collections.[1] Key considerations include the selection of an appropriate hosting platform, the curation and processing of materials, the chosen data architecture, challenges encountered during data structuring, and the solutions implemented. Most importantly, the project

1 In doing that I follow the suggestion made by Blaney et al. 2021, 27.

aims to incorporate and interlink the maximum possible amount of extracted data – standardized, disambiguated, and contextualized – to facilitate user searches and enable the generation of customized datasets based on specific queries. As a case study, this article examines the three-year, nationally funded project *Nuncio's Secret Archives: Papal Diplomacy and European Multidenominational Societies Before the Thirty Years' War*,[2] illustrating the workings of a historian's digital laboratory.

2 The Source – The Graziani Archive

This article presents a project focused on the Graziani archive, which has been partially explored by scholars in recent years[3] but has never been comprehensively studied or inventoried in its entirety. Three main factors have contributed to this gap in research:

1. the archive is dispersed across multiple institutions in different countries and continents, making comprehensive study challenging;
2. the largest portion of the archive remains in private ownership, held by the Graziani family in Vada, near Livorno. Although the family actively supports scholarly research and facilitates access to the materials, the archive's location is not conducive to systematic and continuous study;
3. the complex history of the archival materials has posed significant obstacles to undertaking the time-intensive and costly process of organization and inventorying.

Before outlining the scope of the project, it is useful to provide a brief historical overview of the Graziani archive and its founder Antonio Maria Graziani.[4] The Graziani family belonged to the nobility and was known as Graziani di Sansepolcro, originating from a small town near Perugia, where they had been established since the

2 Principal Investigator: prof. Elena Bonora, University of Parma. The other three units were: the University of Padua led by prof. Antonella Barzazi, the University of Modena and Reggio Emilia led by prof. Matteo Al-Kalak, and the University Ca' Foscari of Venice led by prof. Dorit Raines, in charge of the Digital Humanities part. For team participants, their roles and credits: `https://grazianiarchives.eu/s/graziani-archives/page/chi-siamo`.

3 Among others: Corsini 2000; Jaitner 2004; Moretti 2012; 2015; 2018; Bonora 2019; Moretti 2021; Jaitner 2021; Moretti 2023; Bonora 2023.

4 The following reconstruction of the history of the archive and his whereabouts is based on research made both in the Graziani archives, by the Soprintendenza archivistica e bibliografica della Toscana (`https://siusa-archivi.cultura.gov.it/cgi-bin/siusa/pagina.pl?TipoPag=comparc&Chiave=209524`), by Marsili 2002; Mariani 2022; Raines 2024.

thirteenth century. This branch of the family descended from the Graziani lineage in Arezzo, Tuscany.
Antonio Maria's father, Giulio Graziani, was a military officer who initially fought with the Venetians in the Holy League in 1509 and later served under the Florentine *condottiero* Francesco Ferrucci in the Battle of Gavinana near Pistoia in 1530. Giulio died in 1543, when Antonio Maria was six years old. Following his father's death, the family ensured that he completed his education before placing him, in 1560, under the protection of Giovanni Francesco Commendone, the future cardinal. At that time, Commendone was tasked with presenting the convocation of the Council of Trent to Emperor Ferdinand I. Graziani remained in his service until the cardinal's death at the end of 1584. As Commendone's secretary, Graziani was responsible for managing his archive. Over time, during his extensive travels in Eastern Europe, he continued to utilize these documents and gradually integrated item with his own materials. His position as one of the secretaries to Pope Sixtus V enabled him to expand his archive further, incorporating additional documents, particularly those related to Poland, when he was appointed in 1587 as the pope's envoy to the region.
In 1592, Pope Clement VIII appointed Graziani as Bishop of Amelia in Umbria. He spent three years in the city before being tasked by the pope with organizing an anti-Turkish league. The success of this mission led to his appointment as nuncio to the Republic of Venice in February 1596. By mid-1598, exhausted and suffering from gout, Graziani was granted an exemption from his duties as nuncio. He subsequently retired to Amelia, where he dedicated himself exclusively to his episcopal responsibilities. He passed away in that city on 16 March 1611.
Following his death, his political archive was supposed to be transferred to the Holy See by his heirs. However, in practice, this extensive collection – containing valuable documents spanning over fifty years (from the 1560s until his death) on sensitive diplomatic relations between the Papacy and various European sovereigns – remained in the possession of the Graziani family, despite repeated efforts by the Holy See to reclaim it.

The archive remained housed in the Graziani residence in Città di Castello, near Perugia. Over the course of the eighteenth century, some manuscripts found their way into the possession of the Spanish-born Jesuit Girolamo Lagomarsini. These manuscripts later became part of the historical archives of the Jesuit *Collegio Romano*, now the *Pontificia Università Gregoriana*. Lagomarsini, who taught Greek at the college until his death in 1773, was deeply attached to its library. His strong opinions about its structure and management are

reflected in manuscript 1487, written in 1758, in which he expressed sharp criticism of the library's organization (Al Kalak 2024, 675-91).[5]

In 1864, a sales catalogue of parchments and manuscripts from the Graziani family archive in Città di Castello was published in Florence. The vendors were identified as "the nobles Niccolò and Teresa Libri", and the catalogue was compiled by Pietro Berti (1864). The Libri family were indirect heirs of the Graziani lineage. In fact, Antonio Maria Graziani's branch became extinct at the beginning of the eighteenth century, after which the archive passed to another branch of the Graziani family. A generation later, the last member of that branch, Porzia Graziani, married Guido Guerra. Four generations later, the last descendant of the Guerra family, Teresa, married Niccolò Libri.

Although the sale did not ultimately take place, some materials were likely dispersed and found their way into various Italian public libraries. Meanwhile, the archive came into the possession of Giovanni Magherini (1852-1924), a local historian with a keen interest in its contents, who had married Melania, the daughter of Niccolò Libri.[6] Magherini expanded the collection by adding documents and manuscripts related to the history of Città di Castello.

By 1904, the archive remained largely intact, as indicated in the description provided by the librarian Giuseppe Mazzatinti in his monumental work *Archivi della storia d'Italia*, published that same year (Mazzatinti 1904). At that time, the collection had been divided into two distinct sections:[7]

1. The *Graziani Archive*, which comprised 143 parchments (dating from 1232 to 1624) and 388 additional items, including manuscripts, files, and bundles of records;
2. The *Magherini Graziani Archive*, which contained 194 archival units.

Subsequently, both collections were transferred to Florence. On 3 January 1941, Niccolò Magherini Graziani formally reported the presence of these two archives in his residence in Palazzo Roffia situated in Borgo Pinti to the Prefect of Florence.

Due to a dispute between Giovanni Andrea Magherini Graziani (1907-1975) and his brother-in-law, Niccolò Mels-Colloredo (1897-1966), the family decided to formally divide their property. By

5 See also https://www.unigre.it/archivioimg/APUG_Topografico/1487_1.jpg and MANUS online, record identification number: CNMD\0000224330.

6 http://www.san.beniculturali.it/web/san/dettaglio-soggetto-produttore?id=10909.

7 The family archive is a separate archive that keeps all the family records from the thirteenth to the twentieth century.

1943, the Graziani collection had already been returned to Città di Castello (albeit without a portion of the collection – mostly Commendone's and Graziani's correspondence – that remained in the hands of Giovanni Andrea Magherini). In 1996, the Ferri Graziani family, heirs of Maria Teresa Magherini Graziani, relocated the Graziani archive to their residence in Vada (Livorno).

The Magherini Graziani collection (along with the Commendone-Graziani correspondence) was transferred from Florence to the Villa di Poggitazzi in San Giovanni Valdarno (near Arezzo). On 26 May 1944, the latter was officially declared to be of significant historical interest. However, that same year, the collection suffered damage when the villa was occupied by German forces. In 1953 and 1954, inspectors from the Superintendency conducted multiple visits to the villa, which was soon to be sold. By that time, little remained of the archive, apart from a few scattered pieces. Further investigations in the 1980s, prompted by reports of documents being offered for sale by the antiquarian Perlini of Arezzo, revealed that between the 1950s and 1960s, the Magherini Graziani collection had been dismembered. Some of its materials resurfaced at the University of Kansas, at least one document was identified in the New York Public Library, and other parts were dispersed among various institutions and private collectors.

The Graziani collection at the University of Kansas is a substantial assemblage of letters and letter-books, consisting largely of correspondence to or from Antonio Maria Graziani and Giovanni Francesco Commendone. It also includes documents, reports, historical texts, and notes, many of which appear to have belonged to Antonio Maria Graziani. The majority of these materials, which primarily concern Commendone's nunciature in Poland, were acquired through the Polish-American bookseller Alexander Janta between 1967 and 1971. Additional items were obtained from the Institute for Canon Law at Boalt Hall, University of California, Berkeley, and from the *Libreria Antiquaria Mediolanum*.[8]

The Graziani initial archive was a family archive that, at a certain point, became integrated with the political archive of one of its members, bishop and nuncio Antonio Maria Graziani. The latter, in turn, incorporated parts of the political archive of Commendone and other high-level prelates transforming this collection into a valuable collection jealously kept by the family. Over the centuries, the archive has been modified and reorganized, partially dismembered

8 I thank Elspeth E. Healey, Special Collections Curator at the Kenneth Spencer Library, University of Kansas, for this information presented in the international conference, "La Chiesa di Roma e l'Europa multiconfessionale nella prima età moderna: attori, politiche, esperienze" (Parma, 17-19 aprile 2024).

and scattered due to family disputes, and ultimately preserved – albeit with unquantified losses – across multiple institutions.

3 Scope(s) of the Project

As previously noted, the three-year, nationally funded project *Nuncio's Secret Archives: Papal Diplomacy and European Multidenominational Societies Before the Thirty Years' War* serves here as a case study for the historian's digital laboratory.

When undertaking a Digital Humanities project that involves data extraction from historical narratives or correspondence, researchers must recognize that multiple approaches exist for processing texts and presenting results. Moreover, existing textual analysis techniques do not fully align with the specific needs of historical research, which prioritizes establishing facts, describing events, analysing complex short- and long-term developments, and integrating findings within the broader framework of scholarly inquiry. Therefore, the first consideration must stem from the project's historically driven research objectives.

In the case of the *Nuncio's Secret Archives*, the stated objectives were twofold:

1. enhance this exceptional and stratified private political archive through the creation of an 'open access' research portal Graziani Archives, which virtually unites the different sections of the original archive (today in Italy, in the US and in Poland), allowing scholars to use them in an integrated and full-scale manner;
2. make an innovative contribution to the understanding of the early stages of papal diplomacy in the aftermath of the Peace of Augsburg (1555) through a complete divulgation of Graziani's 'secret' archives, bringing to light an extraordinary and unexpected range of perceptions, knowledge, and orientations developed by Roman mediators in the face of heresy and the European multi-denominational space.[9]

These two objectives were complementary; however, the project's abstract made it clear that the primary focus would be the creation of the Graziani Archives portal as a research instrument:

9 PRIN: PROGETTI DI RICERCA DI RILEVANTE INTERESSE NAZIONALE – Bando 2017, Prot. 2017JMPYTA, part A, p. 4. The project commenced in March 2020 with an initial duration of three years. However, due to the COVID-19 pandemic, the deadline was extended by an additional twelve months, culminating in March 2024 with the launch of the Graziani Archives portal. The project's concluding conference and the official presentation of the portal were held from 17 to 19 April 2024 at the University of Parma.

> Through a pioneering approach integrating Digital Humanities and traditional research, the project aims to reorganize for the first time an extraordinary private archive, now divided between Italy and Kansas, decisive for reconstructing the history of papal diplomacy in the crucial period between the peace of Augsburg (1555) and the Thirty Years War. The archive, created by two of the most high-ranking diplomats of the late sixteenth century, G.F. Commendone and A.M. Graziani, will be valorized through the portal Graziani Archives, which, in addition to providing a reference model for the interrogation and study of private political archives of modern age, will give access to unpublished documentation, mostly informal and very different from the official and already known one, describing with unprecedented richness and depth the vast networks activated by papal diplomats, their relationship with the regular orders and the clash between papacy and multi-confessional space. The construction of the portal, combined with a rigorous historical-archival investigation, will activate a research laboratory that will give an impulse to a new interpretive perspective, capable of finally including the history of papal diplomacy in the European historiographical debate on multi-confessionality.[10]

The project thus aimed to render the extensive collection of records preserved in the correspondence of Commendone and Graziani accessible through the creation of an online, open-access research portal, Graziani Archives. The research team emphasized in the proposal that this tool would not merely replicate the functions of traditional resources such as paper catalogues, inventories, or thematic bibliographies. Instead, it was designed as an advanced, innovative instrument that would serve as a reference point for the enhancement, interrogation, and study of private political archives from the early modern period.

More specifically, the portal aimed to guide scholars, as well as any interested users, through six distinct yet interoperable search pathways (or sections) accessible from the homepage:

1. *Lettere* – The correspondence of Commendone and Graziani;
2. *Inventari* – A virtually reunified inventory of the Graziani Archives, currently dispersed across multiple institutions in Europe and the United States. This section provides the history and present location of each manuscript, as well as the full inventory of the Graziani family archive (held separately from the nuncio's archive) that shed light on the history of the latter archive;

10 Citation from the project's proposal: PRIN: PROGETTI DI RICERCA DI RILEVANTE INTERESSE NAZIONALE – Bando 2017, Prot. 2017JMPYTA, Part A, p. 3.

3. *Persone* – A database of individuals mentioned in the correspondence and their relationships and titles or occupation;
4. *Luoghi* – A geolocated visualization of the places referenced in the correspondence, displayed on an interactive Google map;
5. *Viaggio di G.F. Commendone* – An interactive map tracing Cardinal Commendone's journey (1560-62) to the bishops and princes of northwestern Europe, undertaken to announce the convening of the Council of Trent. This section includes biographical profiles of individuals he encountered, images of key figures and locations, and textual excerpts drawn from both the diary of his entourage member, the Bolognese Fulvio Ruggieri, and Commendone's own dispatches to Rome;
6. *Biblioteca di A.M. Graziani* – A digital reconstruction of Graziani's personal library.

Figure 1 Homepage of the Graziani Archives portal

When a team initiates a project – meaning when it undertakes the realization of a proposal – it must clearly define the 'who', 'when', and, most importantly, the 'how'. It is useful to consider that "a process is a sequence of tasks [...], through which an output is produced by operating on an asset" (Dunn, Hedges 2012, 21). For instance, if a process involves cataloguing, the 'who' refers to individuals with expertise in cataloguing. The 'when' depends either on the project's established deadline or on the internal timeline set by the team, particularly as other collaborative tasks may be conducted in parallel. However, the most critical aspect is the 'how' – the organization of various processes into complementary tasks within a coherent and feasible timeline.

Given the diversity of institutional holdings, the lack of information regarding the whereabouts of certain materials that were once part of the archive, the time constraints of the funded project, and the financial and human resources available to the team, several critical decisions had to be made by the *Nuncio's Secret Archives* project team:

1. Project Scope – *prioritizing research output or adopting a broader approach to address the research needs of related user communities*.
    - While the production of research results is a fundamental requirement of any funded project, the predetermined timeline often limits the possibility of an online publication of the structured data collected, processed, and utilized during the research, thereby restricting access for other scholars. It seemed to the team that it should take into consideration that the creation of an online portal takes longer than the publication of a book. It was therefore advisable to adopt a realistic approach toward the amount of material and work needed for the realization of the web portal;
2. Digital Tool – *developing a custom-built tool tailored to the team's specific needs or utilizing an existing, more generic solution*.
    - Digital tools for storing structured data range from simple Excel files to complex online platforms designed to establish relationships between different properties, apply markup tools to annotate uploaded digital images, or integrate properties into interactive maps. The selection of an appropriate digital tool depends primarily on financial and time constraints, as well as the desired level of online accessibility and interaction;
3. Online Open-Access Communication – *choosing a website or platform capable of providing a user-friendly interface*.
    - The more a project aims to publish collected data through an intuitive, user-centric interface, the greater the need for a platform that can efficiently generate a seamless frontend from the backend module used for data ingestion and relational structuring;
4. Long-Term Preservation – *ensuring the maintenance and updating of both the digital tool and the data in a suitable repository*.
    - Currently, most projects require a three- to five-year Data Management Plan and accept the online publication of data in formats such as Excel or Comma-separated values (CSV) in an open-source repository. However, this approach often leaves the software or platform used for

storing linked data without further financial support for maintenance and updates beyond the project's duration. As a result, while the data remains searchable at the item level, the relationships between properties become inactive, thereby limiting the potential for advanced research applications.

4 Choice of the Right Platform

The option of building an online database from scratch to meet the project's specific requirements was immediately ruled out by the research team, as it was deemed too costly and time-consuming for a three-year project primarily focused on analytical deliverables.

A publicly funded project must adhere to a set of principles established within a working framework developed by decision-makers at the national and, in this case, European levels. The *European Science Agenda* identifies three distinct levels of data science: data, services, and governance (Ganguly, Budroni, Sánchez Solís 2017, 203-10). These levels are supported by digital infrastructures, which form the foundation for data preservation by ensuring that data are managed and curated effectively. The development of services for data uploading and reuse is based on the principle that infrastructure serves as a cornerstone, aligning with the European Commission's approach to structuring these three levels. Services must be user-friendly and adaptable to various use cases, as defined by data producers. Lastly, governance provides the overarching framework through which functional and publicly articulated policies establish an institutional format for data preservation.

These three levels guide the technical choices of any project, requiring a research team to identify a solution that ensures open access, a user-friendly platform, and a high degree of data preservation. Given the range of available open-source tools capable of combining a content management system with a flexible organization of item-level metadata (Hardesty 2014, 80-4; Tritch Roman 2018), the *Nuncio's Secret Archives* project's team initially explored the possibility of using the PHAIDRA platform.

PHAIDRA (Permanent Hosting, Archiving, and Indexing of Digital Resources and Assets), as stated in its mission, helps organizations manage and preserve the full value of their digital assets: from diverse media files, digitized artifacts, and collections; to research, data, and analysis".[11] Built on Fedora open-source software,

11 https://phaidra.org/. Other similar free and open-source digital preservation systems are DSpace (https://dspace.org/) developed in 2002 by the Massachusetts

which offers a modular architecture,[12] the platform was first developed in 2008 by the ZID – Zentraler Informatikdienst and the Library of the University of Vienna[13] to promote the principles of Open Science, Open Data, and Open Access (Montanaro 2016-17, 12).[14] To enhance interoperability with other platforms, PHAIDRA soon collaborated with Europeana, OpenAIRE, and OpenAIREplus, adopting the FAIR principles[15] and the Data Management Plan (Miksa et al. 2016) as its guidelines. Additionally, it established a network with academic institutions in Austria, Italy, Serbia, Montenegro, and Bosnia-Herzegovina.[16]

The decision to use PHAIDRA as a repository for the project's data was further supported by the fact that both the University of Padua and the University of Venice, which hosted research units involved in the *Nuncio's Secret Archives* project, were also partners of the PHAIDRA platform.[17] Moreover, after reviewing the detailed documentation on PHAIDRA provided online by the Library System of the University of Padua, the research team concluded that the data ingestion process was relatively straightforward and could be mastered with the help of a brief training seminar.

However, upon closer examination of PHAIDRA and after discussions with the University of Padua librarians responsible for managing the platform, several concerns emerged regarding its suitability for the project. PHAIDRA was primarily designed for the reconstruction of digital libraries and archives, ensuring their accessibility and long-term preservation. Consequently, its data architecture was centred primarily around digital objects and their metadata (Cappellatto 2015-16, 55). By contrast, the *Nuncio's Secret Archives* project pursued different objectives. It sought to extract data from a selected body of correspondence, enrich that data, establish relationships with other sources – such as the Graziani inventories – and ultimately use part of the dataset to design an interactive map narrating Cardinal Commendone's journey (1560-62) to the bishops and princes

Institute of Technology (MIT) and Hewlett-Packard Labs (Lewis et al. 2010) or Archivematica (https://www.archivematica.org/en/docs/archivematica-1.17/getting-started/overview/intro/#intro), developed by Artefactual Systems and University of British Columbia Library in 2012 (Sprout, Romkey 2013, 257-68).

12 https://fedorarepository.org/. Fedora is the acronym of Flexible Extensible Digital Object Repository Architecture.

13 https://phaidra.univie.ac.at/.

14 The PHAIDRA project followed the UNESCO Memory of the World report: Bradley, Lei, Blackall (2007, 3), that argued for simplification in the digital and archival preservation system.

15 https://www.go-fair.org/fair-principles/.

16 https://phaidra.org/community/overview/.

17 https://phaidra.cab.unipd.it/.

of northwestern Europe to announce the convening of the Council of Trent.

The exclusion of PHAIDRA as an optimal tool for achieving the project's objectives led to a reassessment of its priorities. It became evident that the chosen solution should rely on an existing, free, open-source platform capable of describing linked properties while adhering to established data standards. Following a brief but targeted search, the research team determined that the most suitable solution was Omeka S, a platform released in 2009 by the Roy Rosenzweig Center for History and New Media, funded by multiple organizations.[18]

This choice was further reinforced by the fact that Omeka S meets the Data Management requirements set by funding institutions. It offers modules for importing resources from repositories such as Zotero, Zenodo, Fedora, DSpace, CKAN, Dataverse, and Invenio and allows data export in multiple formats, including CSV, JSON-LD, JSON-table, ODS, TSV, and TXT (Maron, Feinberg 2018) – ensuring long-term data preservation (Morton 952-3).

Although Omeka S has been criticized by some Library and Information Science (LIS) researchers for not adequately clarifying and supporting the Dublin Core metadata standard on which it relies (Maron, Feinberg 2018), it is important to recognize that Omeka S was not designed solely for the description of properties and values from a LIS perspective. Rather, it was conceived as a "free, open-source platform that, like blogging software, offered an easy-to-use administrative interface, provided syndication for sharing content, and extended the core function of publishing content with a flexible plugin architecture and rich design theme API".[19]

From a historian's perspective, this flexible approach is particularly advantageous for several reasons. First, it enables non-LIS experts to utilize the platform effectively, leveraging their expertise in other fields to develop highly complex projects based on collections that may not conform entirely to standardized classification systems. Second, it allows users to incorporate ambiguous expressions that are typically restricted by the Dublin Core's rigorous approach, thereby avoiding forced disambiguation in cases where insufficient evidence is available. For example, the Dublin Core's 'levels of granularity', which seek a clear, consistent, and standardized format, struggle to provide a satisfactory solution for date values such as *circa 1940*. Should this date be disambiguated on a decadal basis (e.g., 1930-40 vs. 1940-50) or on a yearly basis (e.g., 1939-40 vs.

18 https://omeka.org/s/.

19 https://omeka.org/about/project/.

1940-41) (Maron, Feinberg 2018, 682)? Historians are usually very attentive to these kinds of problems.

Finally, Omeka S also facilitates the extraction and contextualization of complex, incomplete, or even contradictory data from multiple sources while ensuring compliance with Dublin Core standard metadata.

5 Data Architecture

The research team firmly believed that an archive's contents could be effectively represented through dataset models and the creation of essential descriptive categories, facilitating a certain degree of cross-searchability both within and across institutions. This conviction led to the selection of the Omeka S platform. A key challenge was transforming content into a new form of information infrastructure that is user-centred and capable of supporting both content management tasks and those related to communication and collaboration.

As previously explained, the Omeka S platform prioritizes display and utilizes an unqualified Dublin Core metadata standard (Maron, Feinberg 2018). It accommodates multiple sites that draw from a shared pool of resources and offers considerable flexibility, allowing users to design their own projects.

The following discussion provides an overview of our approach to modeling information on the backend. Before determining the appropriate data architecture model, it is essential to conduct a thorough examination of the content to be structured. The process can be better understood through the framework of the Graziani Archives portal.

The first section of the portal concerns a collection of more than 3,000 letters, primarily related to the diplomatic activities of Commendone and Graziani within the Holy Roman Empire and Poland. Another significant portion consists of letters addressed to them by various correspondents. This body of correspondence was perhaps the most crucial focus of the project, as it promised to

> give access to unpublished documentation, mostly informal and very different from the official and already known one, describing with unprecedented richness and depth the vast networks activated by papal diplomats, their relationship with the regular orders and the clash between papacy and multi-confessional space.[20]

20 See note 9, *supra*.

A preliminary review of the letters and dispatches yielded several observations further underlining the complexity of the correspondence content:

1. the majority of the letters follow a recurring pattern, in which the senders report information obtained during their encounters with envoys, prelates, high-ranking officials, or sovereigns, specifying the source of each piece of information. Alternatively, the information is derived from other sources, such as letters or hearsay;
2. the content of the letters pertains to a wide range of individuals, primarily within the political or ecclesiastical spheres, either in Rome or in the territories of the Holy Roman Empire and Poland-Lithuania;
3. some letters contain references to past or unfolding events that, within the context of a single letter, are not always easily identifiable or fully understood;
4. the correspondence mentions various locations, often linking officeholders, prelates, or other individuals to specific places;
5. neither Omeka S nor any other database based on the Dublin Core metadata vocabulary, the RDF data model, or the more recent Records in Contexts-Conceptual Model (RiC-CM)[21] can adequately capture the richness and complexity of the Commendone-Graziani correspondence as a form of historical narration.

These observations, along with the awareness of time and resource constraints, led to the following rationale: given that the research team's primary objective was to provide users with access to this source, and considering that the primary users would be historians, three distinct yet complementary types of access points were proposed.

1. Full Digitization and Metadata: the first access point involved the complete digitization of more than 3,000 selected letters and their online publication, accompanied by appropriate metadata;
2. Data Extraction and Record Cards: the second access point entailed the extraction of data from each letter, focusing on individuals, places, and organizations. A dedicated 'record card' was created for each entity, along with references to the letter(s) in which it was mentioned. This feature was designed to assist users in navigating the full digitized collection and contextualizing the extracted data. However, tagging extracted properties directly onto the digitized material was deemed

21 https://www.ica.org/resource/records-in-contexts-conceptual-model/.

impractical and overly time-consuming. This decision was influenced by the expertise of the researchers recruited for data extraction and ingestion, who were highly trained historians with foundational knowledge of Digital Humanities but limited familiarity with descriptive markup tools;

3. Summaries for Contextualization: given the density and richness of the historical data, the research team determined that providing users with an abstract for each letter would be valuable. Consequently, most letters were accompanied by a 'short summary' (*Regesto veloce*) highlighting key events, individuals, and meetings. However, for a subset of letters containing particularly detailed information, considerable length, or complex interrelations among people and events, a more 'detailed summary' (*Regesto approfondito*) was also provided alongside the brief one. This additional layer of contextualization was considered essential for enabling users to efficiently identify relevant materials.

At this stage, these decisions needed to be translated into a data architecture model based on Omeka S functionalities and on the conviction that the relationships to create are not only of a vertical type: each and every item is to be linked to others following its attributes.[22] The first critical question concerned the nature of the items and their relationships within the architecture: should 'people' function as a central node, serving as the pivot of the structure, or should 'letters' be treated as the primary item, representing the physical objects whose images would be uploaded and linked to each individual letter entry?

The latter approach was ultimately chosen, as it not only represented a tangible object but also needed to be associated with its corresponding collection and hosting institution. The initial proposal for this structure was then visually represented, as shown in figure 2 (where the 'letters' are represented by 'Document').

22 Although, naturally, the letters had to be linked to their current corresponding collection and hosting institution in a 'tree' like architecture. On the complex problem of archival description in the digital hierarchical environment (Michetti 2013, 1002-10).

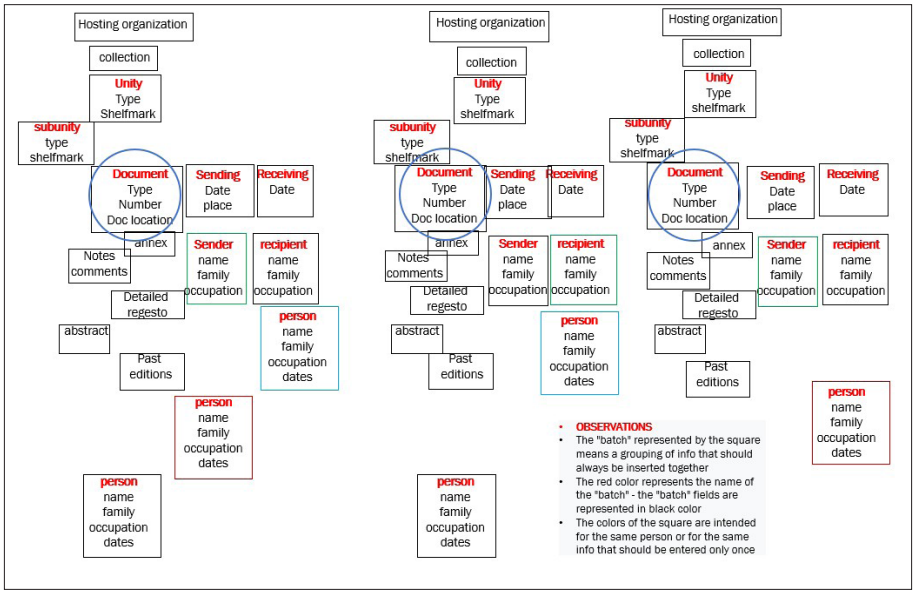


Figure 2 First proposal of data architecture for the *Lettere* section

As observed, the letters were to be linked to their current hosting institutions and collections. Following the prevailing standards of archival description, the data architecture also incorporated additional metadata beyond the hosting institution, collection, and shelf mark, which were essential for each 'letter' item. These included the date and place of sending and receipt, the names of the sender and recipient, and, where applicable, any annexes. All other items were connected either to the letter's contents (such as persons, abstract, or 'detailed summary') or to supplementary information (e.g., past editions of letters, notes, and comments).

An additional step involved the development of resource templates, which were renamed by the research team but corresponded to specific classes based on the Bibliographic Ontology in Omeka S **[fig. 3]**.

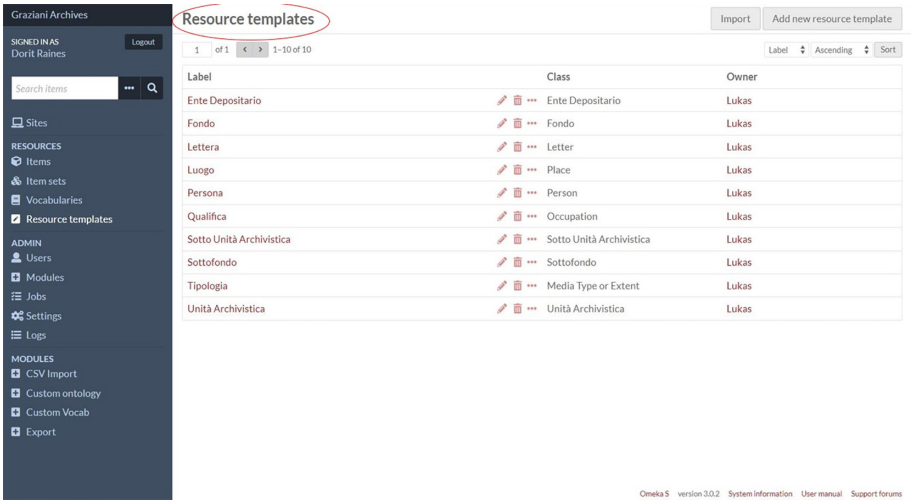


Figure 3 The Resource templates based on the Bibliographic Ontology in *Omeka S*

The creation of resource templates facilitated the final design of fields within each template, drawing on the available modules in Omeka S, including Dublin Core, Bibliographic Ontology, Friend of a Friend (FOAF), and Schema. While some templates were relatively straightforward to create, requiring only a limited number of fields, two proved to be particularly challenging: *Persona* ('Person') and *Lettera* ('Letter'), both of which, as previously noted, were central to the project.[23]

As illustrated in the figure below, the *Persona* template is linked not only to other templates – such as *Lettera*, which establishes familial or marital relationships with other *Persona*, and *Qualifica* ('Occupation') – but also to other sections of the Graziani Archives platform. These include the inventories of the two Graziani archives (Antonio Maria Graziani's archive and the family archive) and the *Itinerario* ('Itinerary') template **[fig. 4]**.

23 One should differentiate between the names given to a portal section and those given to a template: while a section is called *Lettere* or *Persone* (in the plural), the templates were named in the singular form: *Lettera* or *Persona*.

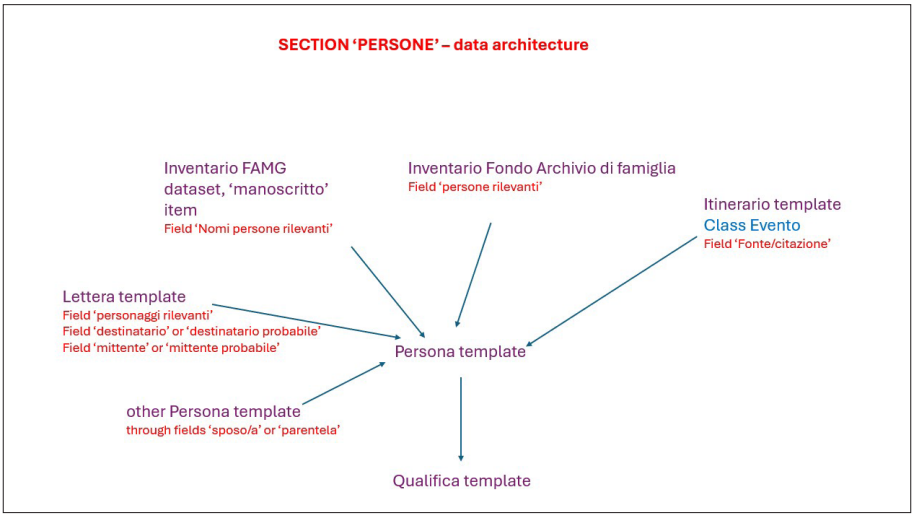


Figure 4 Data architecture for the *Persone* section

The *Lettera* template resulted in an even greater number of connections to other templates. First, it is associated with its current location and shelf mark, including the sub-collection or collection and the archival unit or sub-unit to which it belongs, as well as its present hosting institution. Second, it incorporates metadata specific to each individual letter, such as the names of the sender and recipient, as well as the locations and dates of sending and delivery.

Additionally, the *Lettera* template records all relevant person names and locations mentioned in the letter, along with a specialized vocabulary designed to describe the characteristics of each archival unit. These descriptors include details such as the presence or absence of annexes, whether the document is a copy of the original, whether it is written in cipher, a draft (*minute*), and other relevant classifications **[fig. 5]**.

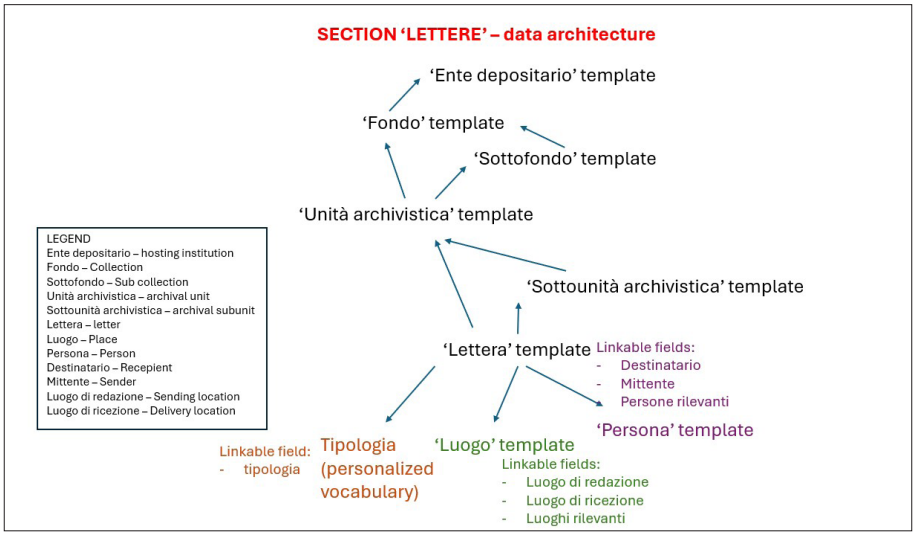


Figure 5 Data architecture for the *Lettere* section

The *Lettere* section on the platform was also linked, through the *Persona* template, to two other sections: *Viaggio di G.F. Commendone* and *Inventari*.

As previously explained, the research team considered the narration of Cardinal Commendone's journey (1560-62) to be an integral part of the project. This journey, undertaken to inform the bishops and princes of northwestern Europe about the convening of the Council of Trent, is reconstructed in an interactive map. The map traces Commendone's travels across Eastern Europe based on a diary attributed to a member of his entourage, as well as on the dispatches he sent to Rome. Each stage of the itinerary is supplemented with biographical profiles of the individuals he encountered, images of people and places visited, and textual excerpts from both the diary and the dispatches **[fig. 6]**.

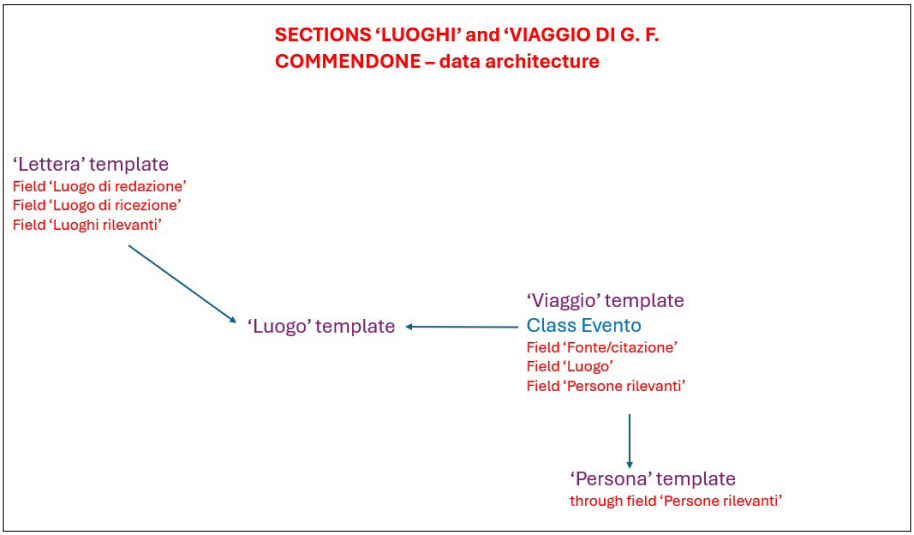


Figure 6 Data architecture for the *Luoghi* and for the *Viaggio di G.F. Commendone* sections

The *Inventari* section presented distinct challenges to the research team. As previously discussed, the current state of the Graziani material is complex, as the collection is housed in three major locations, each associated with a different hosting institution. Additionally, the Vada portion is divided into two separate sections: the first is the Antonio Maria Graziani archive, which is connected to the Graziani manuscripts in Kansas, New York and Poland, while the second comprises the family archive, entirely preserved in Vada.

To provide users with the most comprehensive information possible, the team decided that, in preparing an inventory for the Antonio Maria Graziani archive – one that would virtually encompass all manuscripts historically associated with it – it would be valuable to include links to two key inventory catalogues: the one compiled by Pietro Berti in 1864 and the one prepared by Giuseppe Mazzatinti in 1904. The dual challenge of this archival fragmentation, combined with the objective of virtually reconstructing the Antonio Maria Graziani archive as it existed in the early eighteenth century (prior to the dispersal of its holdings), led to the intricate solution **[fig. 7]**.

The two Graziani sections in Vada were linked through the *Ente depositario* ('Hosting Institution') template, while the Antonio Maria Graziani archive was virtually connected to the Kansas, New York and Polish collections, as well as to the two historical inventory catalogues (Berti and Mazzatinti), via a *Data set* class template named *Inventario integrato* ('Integrated Inventory'). This approach enables

users to determine the past and present locations of each manuscript at any given time.

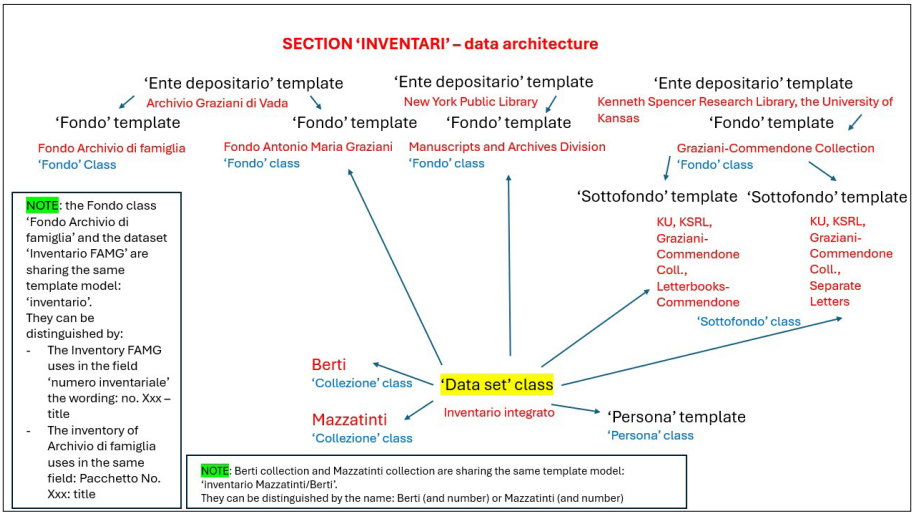
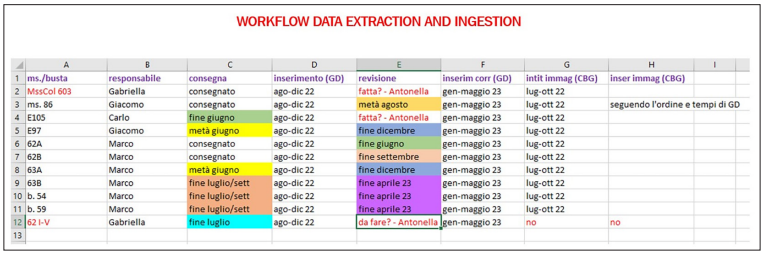


Figure 7 Data architecture for the *Inventari* section

The library section (*Biblioteca di A.M. Graziani*) was addressed separately. The data architecture applied to this section is explained in detail by Luca Iori (*infra*).

6 Data Extraction, Data Ingestion

Parallel to the elaboration of the data architecture for the Graziani Archives portal, the research team set to decide the types of recurrent data to be extracted from the selected Commendone-Graziani letters. As explained beforehand, the decision to provide each letter with a short abstract of its content and in certain cases, to publish also a detailed summary, led the team to focus on the extraction of individuals' names and locations. Four postdoctoral researchers were chosen to read the letters, prepare for each an abstract, and if needed, a detailed summary, locate past (whole or partial) editions of the letter and extract all relevant names and locations. The four were assigned each a number of manuscripts that contained the correspondence and a workflow plan was designed to respect deadlines **[fig. 8]**.



WORKFLOW DATA EXTRACTION AND INGESTION

	A	B	C	D	E	F	G	H	I
1	ms./busta	responsabile	consegna	inserimento (GD)	revisione	inserim corr (GD)	intit immag (CBG)	inser immag (CBG)	
2	MssCol 603	Gabriella	consegnato	ago-dic 22	fatta? - Antonella	gen-maggio 23	lug-ott 22		
3	ms. 86	Giacomo	consegnato	ago-dic 22	metà agosto	gen-maggio 23	lug-ott 22	seguendo l'ordine e tempi di GD	
4	E105	Carlo	fine giugno	ago-dic 22	fatta? - Antonella	gen-maggio 23	lug-ott 22		
5	E97	Giacomo	metà giugno	ago-dic 22	fine dicembre	gen-maggio 23	lug-ott 22		
6	62A	Marco	consegnato	ago-dic 22	fine giugno	gen-maggio 23	lug-ott 22		
7	62B	Marco	consegnato	ago-dic 22	fine settembre	gen-maggio 23	lug-ott 22		
8	63A	Marco	metà giugno	ago-dic 22	fine dicembre	gen-maggio 23	lug-ott 22		
9	63B	Marco	fine luglio/sett	ago-dic 22	fine aprile 23	gen-maggio 23	lug-ott 22		
10	b. 54	Marco	fine luglio/sett	ago-dic 22	fine aprile 23	gen-maggio 23	lug-ott 22		
11	b. 59	Marco	fine luglio/sett	ago-dic 22	fine aprile 23	gen-maggio 23	lug-ott 22		
12	62 I-V	Gabriella	fine luglio	ago-dic 22	da fare? - Antonella	gen-maggio 23	no	no	
13									

Figure 8 Workflow chart for the data extraction and data ingestion phases

In addition to the extraction and ingestion workplan the team had to face another question, currently surfacing when more than one person is working on data extraction: allow each of the four postdoctoral researchers directly ingest the data extracted and the short and detailed summaries into the Omeka S *Lettera* template and eventually correct the verified data in a second moment; or terminate the extraction phase and the verification 'offline' using a shared template and then choose one person for the ingestion phase?

The team preferred the second option for two main reasons:

The short and detailed summaries were first reviewed and corrected, if necessary, by the project's heads of units and subsequently standardized and verified against the extracted data by two editors. Given the time-consuming nature of this process, the team concluded that data ingestion could only be carried out efficiently and accurately once this phase was fully completed;

When a single individual is responsible for data ingestion, they are more likely to identify inconsistencies, missing standardizations, or unresolved ambiguities in the data. To address such issues, the postdoctoral researcher selected for the ingestion part constantly consulted one of the units' heads responsible for standardization to resolve any outstanding questions. Indeed, during the ingestion phase, the designated researcher identified missing or conflicting information which led to the addition of several fields in the template to clarify ambiguous identifications, relationships between individuals, and gender attributions.[24]

To ensure consistency in data extraction, a working template was created using standardized vocabulary and provided to the four postdoctoral researchers. This template was designed to align with the structure of the Omeka S template for letters, thereby facilitating the subsequent ingestion process **[fig. 9]**.

24 For a discussion of this issue, see Gabriella Desideri, *infra*.

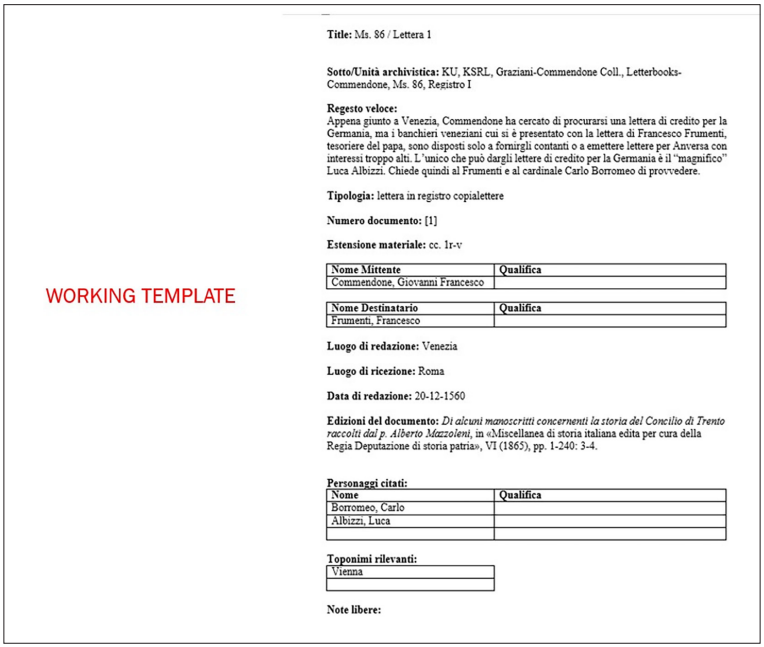


Figure 9 Working template for extracting data from the letters

The team recognized that, in certain cases, disambiguation was necessary for specific individuals or locations. Additionally, ensuring consistency in name standardization was particularly challenging given the involvement of four researchers, necessitating a structured approach to team coordination in order to maintain controlled data usage.

To maximize efficiency and accuracy, a detailed workflow was developed. As explained above, one of the project's unit heads was designated as the standardization editor, and two shared Word files were created in Drive – one for individual names and the other for locations. Each postdoctoral researcher was responsible for entering standardized names in alphabetical order, using an assigned identification colour. The editor would then review the entries, verify them, and either approve the standardized names by marking them in black or flag them for further investigation by highlighting them in yellow **[fig. 10]**. Each researcher was subsequently responsible for updating the working template with the verified standardized names.

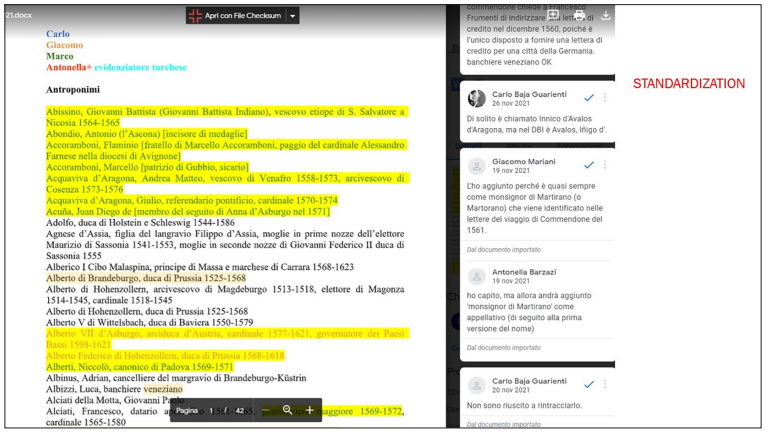


Figure 10 Shared file for individual names' standardization

The outcome of this labour-intensive process was a relatively smooth and efficient ingestion phase, which, in turn, facilitated the rapid uploading of images. Each image had been pre-named according to a standardized format, incorporating the manuscript shelf mark, the letter number, and a sequential number corresponding to its position within the letter unit (e.g., if a letter spanned three pages, the images would be numbered 1, 2, 3, etc.).

The preparation of the *Lettere* section proved to be the most complex and time-consuming phase of the project. In parallel, work began on the *Inventari* section; however, its ingestion had to be postponed until the completion of both the *Lettere* and *Persone* sections, as many individuals mentioned in the inventories were referenced in both sections. The preparatory phase for the inventories ingestion required an initial selection of individuals named in the inventories but not mentioned in the letters, the standardization of their names, research into their titles, occupations, and relationships, and only then the ingestion of this information into the Graziani Archives portal.

7 From Backend to Frontend

A word of caution is necessary for this section. My aim is not to explain the distinction between backend and frontend from an application perspective, nor to provide a technical overview of the programming languages involved. Rather, this is a straightforward humanistic reflection on how humanities scholars may navigate this complex phase and effectively communicate their preferences to the custom web designers responsible for implementation. This account

chronicles the experience of the Graziani Archives team, with the hope that it may assist other teams in transforming their ingested data into a user-centric and user-friendly platform.

Upon completing the data ingestion phase, it became necessary to shift our focus toward the publication phase. As humanists, we are typically well acquainted with the process of print publication; many of us have even learned to format texts using various software programs. However, web publication, and above all – custom web design – presents an entirely different challenge, requiring a distinct approach.

What, then, is the fundamental difference between backend and frontend? In our case, given that we were using Omeka S, the infrastructure was already in place – it merely needed to be adapted to our data architecture based on the predefined sections (as previously discussed). The backend serves two primary functions: first, to enable researchers to ingest data in a structured and retrievable manner; and second, to store this data and facilitate user access. In Omeka S, these functions are performed through a set of modules, each containing multiple fields that the team created according to its needs, along with the possibility to establish relationships between different fields or modules.

From a graphical perspective, each module is designed in the backend in a vertical continuous layout, which can vary in length depending on the number of fields included. Researchers must use the mouse wheel to scroll up and down to access all fields. When establishing a relationship between data elements, a new sidebar is displayed on the right side of the screen, introducing all existing items and allowing for further selection by category to locate the desired item. The team found this design intuitive and worked comfortably with it throughout the ingestion phase.

However, when we were asked to conceptualize a frontend representation of the data we had ingested, we found ourselves at a loss. We were unsure of what was expected of us and unfamiliar with the distinction between backend and frontend. It was at this moment that we realized our primary focus had been on the ingestion process, with little consideration given beforehand to the user experience.

At this stage, we had the support of three professionals: the graphic designer, who had previously designed the project's web pages for the University of Parma (the project's principal investigator), as well as custom web designers from Libnamic Digital Humanities, who implemented the proposed graphic design into a web customization using Figma, a collaborative design tool.[25]

25 https://www.figma.com/.

Before delving into this process, it is useful to take a step back. In many projects, teams first create a webpage or website to communicate the project's objectives, ongoing progress, events, and publications. The design of such a webpage is typically entrusted to a graphic designer and is often influenced by whether it is developed within an existing platform (e.g., a university platform) and whether the platform's administrators have already established fixed layouts, graphic rules, and colour schemes. Fortunately, in our case, the University of Parma permitted the creation of a website while allowing the team full discretion over its layout.

The graphic designer initially proposed a set of colours and fonts for both the project's website and its event brochures. Two years later, when discussing the graphic design of the Graziani Archives portal, she recommended using a different colour scheme to distinguish the portal from the project's main website. We found this rationale compelling and readily accepted the suggestion. Consequently, the team collaborated with her to design the various sections of the portal. The final proposal was then submitted to Libnamic for the web customization phase.

To illustrate the final outcome of the customization phase, I would like to summarize key points that, while seemingly straightforward, may assist other scholars when customizing the frontend of their database. Given that we are not web designers, I wish to highlight several considerations that a humanities team should take into account when approaching the web design and customization phase:

1. Adopting a user-centric approach – It is essential to anticipate the needs of future users, categorizing them – if possible – into groups with distinct objectives and predicting their behaviour. In short, the goal is to effectively communicate what is happening on the screen and to enhance the usability of the portal by minimizing the number of clicks required to access the desired content;
2. Considering screen orientation – Unlike traditional A4 paper formats, where the longer dimension is vertical, computer, laptop, and tablet screens typically have a horizontal orientation. While scrolling allows users to access additional content, the customization process should ensure that the horizontal screen layout immediately conveys the overall content structure of the webpage and enables users to navigate efficiently;
3. Designing the initial portal webpage with distinction – The initial portal webpage may differ slightly from the other pages within the portal. In the case of the Graziani Archives, the goal was to create a visual link between the project's webpage on the University of Parma's website and the portal itself. To achieve this, a distinctive graphic element – a bluish

Figures 11-12
Above the NSA Parma University's project website. Below the Graziani archives initial website page

image of the Graziani manuscripts with the portal's name superimposed in white – was used as a visual bridge between the two websites. This graphic element appears only on the initial portal page, as it occupies significant space that would otherwise be allocated to content **[figs 11-12]**;

4. Signage is essential – Users must always be aware of the portal's name, the menu, and their current location within the portal's sections. This requires the portal header to remain consistent across all pages, displaying the logo and portal name at the top of the screen, with the menu positioned either directly below or on the left or right side. Visual hierarchy is equally important: "Use different font sizes, colours, highlighting effects, etc. to help users distinguish categories of menu items or levels of importance" (Toulson 2021);
5. Menu bar positioning – There are two primary approaches to menu positioning. If the goal is to maximize the space below the logo and dedicate it to the display of content, a left-side menu is preferable. However, if the layout is divided into two or more horizontally aligned columns – as was the case for the Graziani Archives portal – a top menu bar placed directly below the portal logo is a more effective solution. In any case, the header should not occupy more than 25-30% of the initially visible screen space. Users should be able to immediately recognize the presence and nature of the webpage's content;
6. Providing a concise summary of portal content on the homepage – Displaying a brief overview of the portal's content on the initial screen is crucial. Users should be able to grasp the essential information about the portal and its sections at a glance, without needing to navigate between sections to understand their purpose. More detailed descriptions of each section can be provided on their respective landing pages **[fig. 13]**;
7. Dividing the screen into left and right columns – Displaying long passages of text on a horizontally oriented screen may not be the most effective solution, as extended line lengths and high text density can hinder readability. A more user-friendly approach is a split screen: dividing the content into left and right columns. Additionally, if two different types of information need to be presented on the same page, organizing them into columns is preferable to requiring users to scroll down to locate the secondary content. A relevant example is the *Lettere* section of the Graziani Archives portal. In this section, the left-hand column provides a detailed explanation of the section and the materials available, while the right-hand column presents a list of selected manuscripts with two navigation options. Users can either click on the + symbol (located to the left of the manuscript number) to

Figure 13
The Graziani archives initial website page with titling, menu bar, a short explanation of the portal's contents and the short text of the first section

reveal a dropdown displaying the manuscript title and editorial notes, or they can click directly on the manuscript number to access its main page **[fig. 14]**. On this page, the left-hand column displays the list of letters, while the right-hand column contains the full manuscript description. Furthermore, on the manuscript page, at the top of the right-hand column, a navigation widget allows users to browse through the selected manuscripts directly, without the need to click in and out of individual records **[fig. 15]**. This same navigation logic applies to each specific letter's page as well as to all other sections of the Graziani Archives portal, where users can seamlessly move back and forth between the items **[fig. 16]**;

Figure 14 *Lettere* section with a manuscript title and editorial note displayed in the right-side drop-down column

Figure 15 Manuscript page in the *Lettere* section: on the left column the list of letters; on the right column the full manuscript description and on the upper side, the possibility to shift to the next manuscript (in blue)

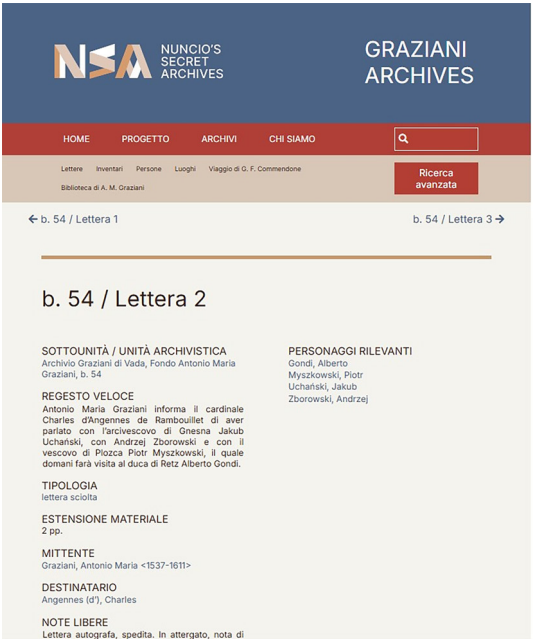


Figure 16 Letter page in the *Lettere* section displaying the possibility to shift to the previous or next letter (top right and left columns – in blue)

8. The Use of Colour – As previously noted, visual hierarchy can significantly enhance user experience. In our case, the graphic design aimed to distinguish the project's website from the Graziani Archives portal by implementing a structured colour scheme. Four distinct colours were used for the Graziani Archives portal: dark electric blue as the background for the logo and portal name; medium carmine red for the top bar, which contains the primary menu presenting the project, its objectives, and the people involved; and pale silver for the secondary menu, which provides access to the content-specific sections of the portal. The fourth colour, a pale grey-yellow (*Isabelline*), was selected as the background for the text. It is crucial to choose a 'neutral' background colour that contrasts well with the text colour, ensuring readability without causing visual strain. Additionally, to help users navigate the relationships between different items, we used dark electric blue to indicate clickable elements – such as persons, places, letter types, hosting institutions, collections, manuscript numbers, and letters – allowing users to access further information about each entity.

Finally, a note on the customization of the interactive map.[26] As previously explained, the objective of this section was to visualize the diplomatic journey undertaken by Cardinal Commendone between 1560 and 1562, during which he travelled across northwestern Europe to inform bishops and princes of the convening of the Council of Trent. This was achieved by geolocating and marking the sites on a Google map he visited and, in selected cases, attaching descriptions of the individuals he met, and, where possible, images of his interlocutors or contemporary maps of the locations.

The primary challenge was to geolocate 229 sites on the map while maintaining a coherent representation of the entire itinerary – covering all of Europe – without requiring a zoomed-in view. Additionally, the map needed to clearly indicate the direction of travel and categorize locations into six distinct geographical areas.

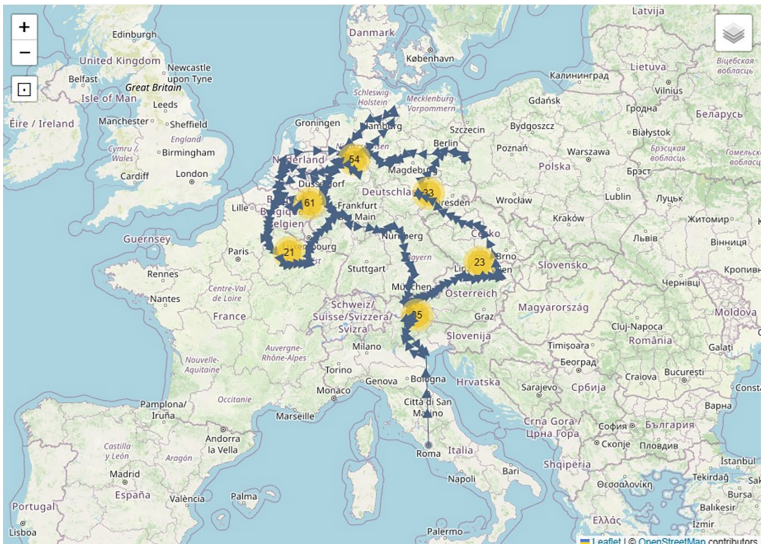


Figure 17 Interactive map with Commendone's diplomatic journey in Europe in 1560-62

The Graziani Archives team requested that when zooming in or using the scrolling function, users would be able to distinguish between locations marked on the map without additional information – represented by small light grey dots – and those containing content, symbolized by blue map markers. When a blue map marker is selected, it turns red, and simultaneously, an expandable sidebar opens on the

26 See Omeka S User Manual: `https://omeka.org/s/docs/user-manual/modules/mapping/`.

left side. It displays either Ruggieri's or Commendone's description of the event, a contemporary image or map of the location, and biographical information about the individuals mentioned in the text. Additionally, the timeline below serves as a chronological search tool, allowing users to locate specific sites along the itinerary.

Figure 18 Interactive map displaying locations, further information on the selected location and a timeline

8 Long-Term Preservation

"Double, double toil and trouble," recite the three witches in *Macbeth* (Shakespeare, *Macbeth*, IV, I). Indeed, after the arduous task of constructing a portal comes a double challenge (and trouble): determining where to store the data and ensure long-term preservation. As has long been highlighted in various scientific forums, "the European investments for digital preservation in the last decade have been large and persistent but not able to support, at the moment, an accepted common vision, general services and adequate infrastructures" (Guercio 2013, 467).

In the absence of a national or European solution, the *Nuncio's Secret Archives* project team first had to determine where to host the

Omeka S backend they were about to configure. The cost-free solution was chosen: a virtual machine running the Linux distribution UBUNTU Core,[27] provided by one of the universities involved in the project.

Researchers in the humanities are not typically aware that long-term data preservation involves "policy questions, institutional roles and relationships, legal issues, intellectual property rights, and metadata" (Thibodeau 2002). The primary issue is not merely data preservation, as this does not necessarily require maintaining all of a dataset's digital attributes – data can easily be stored in CSV format on Zenodo, the Internet Archive, or other repositories. Rather, the challenge lies in preserving relational data and the query functionalities that a portal typically provides to users for interrogating the dataset but also links to external web pages, style sheets, graphic images, JavaScript elements, data extracted from databases, and other related components (Thibodeau 2013, 16). Simply storing data in CSV or another tabular format is insufficient, as it does not support user interrogation based on the relationships established between properties.

As Thibodeau noted in 2002, four criteria must be followed when considering long-term preservation:

1. Feasibility – which requires hardware and software capable of implementing the method;
2. Sustainability – which entails avoiding technological obsolescence and ensuring interoperability with other methods, such as those for discovery and delivery;
3. Practicality – which demands reasonable limits in terms of difficulty and cost;
4. Appropriateness – which depends on the types of objects to be preserved and the specific objectives of preservation.

I will begin with the final criterion, which also encompasses one of the core pillars of the FAIR principles: the fair use of data on the web. We obtained permission from the hosting libraries of the selected manuscripts from the Commendone-Graziani correspondence to publish them on the portal for the benefit of users. These users may view or download each image and utilize them for research or publication purposes, but not for commercial use. The images were processed using Mirador,[28] an image viewer optimized for displaying resources compliant with the International Image Interoperability Framework (IIIF, or Triple I-F), a standardized method for describing images on the web.[29] Mirador adheres to the Web

27 https://ubuntu.com/core.

28 https://projectmirador.org/.

29 https://iiif.io/api/.

Content Accessibility Guidelines (WCAG) 2.1 AA, which are legally adopted in the United States, Canada, and Europe.[30]

Additionally, we incorporated for some individuals or locations some images downloaded from Wikipedia, explicitly noting that they are classified as public domain.[31] When working with data that has already been published online or when publishing their own data and images, researchers must remain aware of both their responsibilities toward content created by others and their rights in disseminating their own work. It is crucial to consider in advance the type of licenses they wish to grant to other users.[32]

Among these four criteria, technological obsolescence is particularly problematic: a website can only remain viable if it is dynamic and evolves in response to the "environment of ongoing, open-ended and multidimensional change in which digital information exists" (Thibodeau 2013, 16-17). However, researchers are generally not in a position to anticipate technological changes or determine how best to address them, whether by locating the latest software release, replacing outdated formats, or adapting to shifts in standards.

The current state of academia leaves researchers in a precarious position.[33] They are required to submit a Data Management Plan before even beginning to develop a portal, yet many of them are encountering the online digital landscape for the first time – a sort of 'far west' in which each researcher or team must independently seek out a computer scientist or a storage provider and rely on their expertise, often without fully understanding the technical terminology or the implications of each decision. How can a humanities researcher be certain that the proposed software can support the data architecture they have designed? How should one mitigate the risks of rapid technological obsolescence? As Ramalho et al. observe,

> Databases are complex digital objects that contain heterogeneous information, often accompanied by structural definitions and even documentation. Their complexity makes it difficult to preserve this kind of object whilst maintaining all its significant properties. (2020, 109)

30 https://www.w3.org/WAI/standards-guidelines/wcag/.

31 See an example of the woodcut of Nuremberg in Hartmann Schedel's Nuremberg Chronicle in https://it.wikipedia.org/wiki/Cronache_di_Norimberga#/media/File:Nuremberg_chronicles_-_Nuremberga.png.

32 See the Creative Commons licenses list in https://creativecommons.org/licenses/list.en.

33 Thibodeau (2013, 19) refers to "a plethora of choices for preservation".

Paradoxically, funding programs typically require only a short-term preservation guarantee after a project's conclusion (usually three to five years). Furthermore, no repository exists at the European, national, or even university level that is specifically designed to host digital portals. This issue was raised years ago by several scholars: "the importance of digital preservation research has been growing for the past ten-fifteen years. Many papers and books describing problems, tools and techniques for digital preservation have been written, and standards providing preservation models have been published. However, most of this work focuses on preservation of file-based digital objects like documents, images, and web pages. Much less work has focused on the preservation of databases and scientific data, where there is a recognized need to preserve scientific data" (Stefanova, Risch 2013). No further steps were taken.

As long as we were ingesting data into the Omeka S backend configured by our team, we remained confident that the UBUNTU virtual machine allocated to us by one of the project's partner universities was a suitable solution. However, when the team reached the custom web design stage and began developing the frontend, it became evident that this solution had limitations. Moreover, the university could not guarantee ongoing maintenance or protection against technological obsolescence.

Thus, the project's research team found itself searching for a suitable service provider capable of ensuring the long-term preservation of our Omeka S-based portal. We were unable to locate such a provider in Italy. Fortunately, Libnamic – the company that had already provided us with web design and customization services – also offered long-term preservation solutions for Omeka S-based projects.[34] Initially, we had legal concerns regarding potential litigation or jurisdictional disputes (Duranti 2013, 28), as Libnamic is based in Spain. However, the matter was resolved to the satisfaction of all parties. The migration process took just over an hour, after which the web design and implementation proceeded on Libnamic's servers.

Nevertheless, the broader situation remains uncertain, as was already highlighted by Luciana Duranti (2013, 28):

> what we need is an internationally agreed upon legal framework that will support the development of integrated and consistent local, national and international networks of policies, procedures, regulations, standards and legislation concerning digital records, to ensure public trust grounded on evidence of good governance.

34 Libnamic Hosting. https://hosting.libnamic.com/.

The European Open Science Cloud (EOSC) is designed to

> provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes.[35]

It remains far from being able to guarantee long-term preservation for portals developed through funded projects. Such projects frequently involve the creation of relational databases with query functionalities, archival source images, tagging, and other complex digital elements. Only a fully integrated and effective digital research ecosystem at a European level can ensure that research outputs from funded projects remain searchable and usable over the long term.[36]

Bibliography

Agosti, M.; Ferro, N.; Silvello, G. (2011). "Handling Hierarchically Structured Resources. Addressing Interoperability Issues in Digital Libraries". Biba, M.; Xhafa, F. (eds), *Learning Structure and Schemas from Documents*. Berlin; Heidelberg: Springer-Verlag, 17-49.

Aliaga, D.G.; Bertino, E.; Valtolina, S. (2011). "DECHO-A Framework for the Digital Exploration of Cultural Heritage Objects". *Journal on Computing and Cultural Heritage* 3(3), 12, 1-21. https://doi.org/10.1145/1921614.1921619

Al Kalak, M. (2024). "Due secoli e tre continenti. Il carteggio tra Girolamo Lagomarsini e Lodovico Antonio Muratori". Boutier, J.; Forner, F.; Paoli, M.P.; Tinti, P.; Viola, C. (a cura di), *Le stagioni dell'erudizione e le generazioni degli eruditi. Una storia europea (secoli XV-XIX)*. Bologna: CLUEB, 675-91.

Berti, P. (1864). *Catalogo delle pergamene e manoscritti già spettanti alla famiglia Graziani di Città di Castello*. Firenze: M. Cellini.

Blaney, J.; Milligan, S.; Steer, M.; Winters, J. (2021). *Doing Digital History. A Beginner's Guide to Working with Text as Data*. Manchester: Manchester University Press.

Bonora, E. (2019). "Comprendre et décrire un autre monde. Le voyage d'un nonce dans l'Europe des confessions et du pluralisme religieux (1560-62)". Ferrant, J.; Guillabert-Madinier, T. (éds), *Le Langage et la Foi dans l'Europe des Réformes XVIe siècle*. Paris Éditions Classiques Garnier, 215-24.

Bonora, E. (2023). "Gli archivi segreti del nunzio". *Riforma e movimenti religiosi*, 13, 177-84.

35 https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en and https://open-science-cloud.ec.europa.eu/.

36 These were the conclusions of the EOSC Symposium held in Berlin on 21-23 October 2024. https://open-science-cloud.ec.europa.eu/news/eosc-eu-node-spotlight-highlights-and-milestones-2024-eosc-symposium.

Bradley, K.; Lei, J.; Blackall, C. (2007). Towards an Open Source Repository and Preservation System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development. https://unesdoc.unesco.org/ark:/48223/pf0000154761

Cappellatto, L. (2015-16). *Studio e realizzazione di una piattaforma di archiviazione di contenuti digitali per l'Università degli Studi di Padova* [Tesi di Laurea]. Venezia: Università Ca' Foscari Venezia.

Corsini, M. (2000). "La Biblioteca e l'archivio Graziani di Vada". *Rara volumina* 7, 127-40.

Dallas, C. (2004). "Policies and Strategies for Digital Preservation". Tola, V.; Castellani, C. (eds), *The Future of Digital Memory and Cultural Heritage*. Roma: ICCU, 281-7.

Dunn, S.; Hedges, M. (2012). "Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research". London: Arts and Humanities Research Council. http://crowds.cerch.kcl.ac.uk/

Duranti, L. (2013). "Trust and Conflicting Rights in the Digital Environment". Duranti, Shaffer 2013, 24-30.

Duranti, L.; Shaffer, E. (eds) (2013). *Proceedings of the Memory of the World in the Digital Age: Digitization and Preservation. An International Conference on Permanent Access to Digital Documentary Heritage* (26-28 September 2012). Vancouver: UNESCO.

Ganguly, R.; Budroni, P.; Sánchez Solís, B. (2017). "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud". Chan, L.; Loizides, F. (eds), *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*. Limassol: Cyprus ELPUB, 203-10.

Gilliland, A.J.; McKermmish, S.; Lau, A.J. (2017). *Research in the Archival Multiverse*. Clayton, Victoria: Monash University Publishing.

Guercio, M. (2013). "Digital Preservation in Europe. Strategic Plans, Research Outputs and Future Implementation. The Weak Role of the Archival Institutions". Duranti, Shaffer 2013, 467-81.

Hardesty, J.L. (2014). "Exhibiting Library Collections Online: Omeka in Context". *New Library World*, 115(3-4), 75-86. http://dx.doi.org/10.1108/NLW-01-2014-0013

Hawkins, A. (2022). "Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-digital Archives Via the Semantic Web". *Archival Science*, 22, 319-44. https://doi.org/10.1007/s10502-021-09381-0

Jaitner, K. (2004). "Miszelle das Archivio Graziani in Vada". *Quellen und Forschungen aus italienischen Archiven und Bibliotheken,* 84, 509-12.

Jaitner, K. (2021). *Instruktionen und Relationen für die Nuntien und Legaten an den europäischen Fürstenhöfen von Sixtus V. bis Innozenz IX. (1585-1591)*. Freiburg; Basel; Wien: Herder.

Kamal, A.M.; Golub, K. (2025). "Subject Matters: Metadata Standards and Subject Access for Library and Museum Catalogues". Hanssen, J.-M.; Furuseth, S. (eds), *The Hermeneutics of Bibliographic Data and Cultural Metadata*. Oslo: National Library of Norway, 204-39.

Lewis, S.; Hayes, L.; Stangeland, E.; Shepherd, K.; Jones, R.; Roos, M. (2010). "DSpace Under the Hood: How DSpace Works". *The 5th International Conference on Open Repositories (OR2010)* (Madrid, Spain, 6-9 July 2010). https://doi.org/10.2390/biecoll-OR2010-52

Mariani, G. (2022). "Sisto V e Antonio Maria Graziani. Documenti del pontificato sistino. Dall'archivio privato della famiglia Graziani". *Sisto v e la comunità camerte = Conference proceedings* (Camerino, 12 December 2022).

Maron, D.; Feinberg, M. (2018). "What Does It Mean to Adopt a Metadata Standard? A Case Study of Omeka and the Dublin Core". *Journal of Documentation* 74(4), 674-91. https://doi.org/10.1108/JD-06-2017-0095

Marsili, M. (2002). s.v. "Graziani, Antonio Maria". *Dizionario biografico degli Italiani*, vol. 58, 801-4.

Mazzatinti, G. (1904). *Archivi della storia d'Italia*, IV. Rocca S. Casciano: Cappelli.

Michetti, G. (2013). "Archives Are Not Trees. Hierarchical Representations in Digital Environment". Duranti, Shaffer 2013, 1002-10.

Miksa, T.; Sánchez Solís, B.; Schrauf, C. (2016). "Data Management Plans. Fortbildungsseminar für Forschungsdaten und e-Infrastrukturen". http://phaidra.univie.ac.at/o:441308

Montanaro, D. (2016-17). *Digital Ecosystems for Open Science. Il caso Università di Vienna, l'esempio di Phaidra* [Tesi di laurea]. Venezia: Università Ca' Foscari. https://unitesi.unive.it/retrieve/fc08629a-5ddd-4fc3-a0be-10c31b239a40/855470-1213774.pdf

Moretti, M. (2012). *Lettere di Pieter de Witte. Pietro Candido nei carteggi di Antonio Maria Graziani, edizione critica*. Roma: De Luca.

Moretti, M. (2015). "Committenti, intermediari e pittori tra Roma e Venezia attorno al 1600. I ritratti di Domenico Tintoretto per il nunzio Graziani e una perduta 'pentecoste' di Palma il Giovane per Fabio Biondi". *Storia dell'Arte,* 41, 21-43.

Moretti, M. (2018). "Antonio Maria Graziani e le fatiche della carriera. L'altare di famiglia a Sansepolcro e la commissione dell''Assunta' a Palma il Giovane". *Storia dell'Arte*, 15(2), 18-67.

Moretti, M. (2021). "L'altare Graziani da Raffaello a Palma il Giovane. Una copia della 'Madonna' Canossa e una 'sentenza' sfavorevole a Giovanni De' Vecchi". *Storia dell'Arte* 155-6(1-2), 61-87.

Moretti, M. (2023). "L'ozio religioso in villa. Un ritiro 'alla Cappuccina' per Carlo Graziani, maestro di casa del cardinale Francesco Barberini". Porzio, G.; Primarosa, Y. (a cura di), *Orazio Gentileschi e l'immagine di san Francesco. La nascita del caravaggismo a Roma*. Milano: Officina Libraria, 112-23.

Morton, A. (2011). "Digital Tools: Zotero and Omeka". *The Journal of America History,* 98(3), 952-3. https://doi.org/10.1093/jahist/jar520

Opgenhaffen, L. (2022). "Archives in Action. The Impact of Digital Technology on Archaeological Recording Strategies and Ensuing Open Research Archives". *Digital Applications in Archaeology and Cultural Heritage*, 27. https://doi.org/10.1016/j.daach.2022.e00231

Raines, D. (2024). "Un archivio diffuso. La storia dell'archivio di Antonio Maria Graziani". *La Chiesa di Roma e l'Europa multiconfessionale nella prima età moderna: attori, politiche, esperienze* (Parma, 17-19 aprile 2024).

Ramalho, J.C.; Ferreira, B.; Faria, L.; Ferreira, M. (2020). "Beyond Relational Databases: Preserving the Data". *New Review of Information Networking* 25(2), 107-18.

Sprout, B.; Romkey, S. (2013). "A Persistent Digital Collections Strategy for UBC Library". Duranti, Shaffer 2013, 257-68.

Stefanova, S.; Risch, T. (2013). "Scalable Long-Term Preservation of Relational Data through SPARQL queries". *Semantic Web journal*. https://www.semantic-web-journal.net/content/scalable-long-term-preservation-relational-data-through-sparql-queries-1

Thibodeau, K. (2002). "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years". *The State of Digital Preservation: An International Perspective*. Washington D.C.: Council on Library and Information Resources. https://www.clir.org/pubs/reports/pub107/thibodeau/

Thibodeau, K. (2013). "Wrestling with Shape-Shifters. Perspectives on Preserving Memory in the Digital Age". Duranti, Shaffer 2013, 15-23.

Toulson, S. (2021). "Four Header Styles That Will Optimize Your Service Portal Navigation and Look Good Doing It". *Portalorem*, 7 April 2021. https://portalorem.com/blog/five-header-styles-that-will-optimize-your-service-portal-navigation)

Trace, C.B. (2022). "Archives, Information Infrastructure, and Maintenance Work". *Digital Humanities Quarterly,* 16(1). https://www.digitalhumanities.org/dhq/vol/16/1/000603/000603.html

Tritch Roman, G. (2018). "Scalar and Omeka". *Journal of the Society of Architectural Historians*, 77(1), 122-3.