

Bringing Corato's historical records together

Abstract

This article presents a proposal for the integrated indexing and matching of historical records from Corato, Puglia, including parish registers, the 1754 *Catasto onciario* and civil registration records. The project follows a two-phase approach: first, the systematic transcription (indexing) of these records and, then, an algorithmic matching system designed to identify and consolidate references to the same individuals across some or several of them. The method emphasizes key genealogical criteria such as filiation, chronological consistency and relational context, with manual verification for ambiguous matches. A practical case study of the Mastrolillo family demonstrates how this approach reconstructs multi-generational family histories by linking fiscal, ecclesiastical and civil sources. The proposed system is locality-agnostic and scalable to other towns with similar sources. Moreover, it is designed to integrate data from external partners such as FamilySearch, thus facilitating transnational migration research and supporting institutional record management. This model positions Corato as a prototype for developing an international database on emigration from Puglia, while also serving as a tool for academic collaboration in the digital humanities and migration studies.

Riassunto

In questo articolo viene presentata una proposta per l'indicizzazione e l'abbinamento integrati dei documenti storici di Corato, Puglia, tra cui registri parrocchiali, il catasto onciario del 1754 e registri dello stato civile. Il progetto segue un approccio in due fasi: prima, la trascrizione sistematica (indicizzazione) di questi registri e, poi, un sistema di corrispondenza algoritmica progettato per identificare e consolidare i riferimenti agli stessi individui in alcuni o più di essi. Il metodo enfatizza criteri genealogici fondamentali quali filiazione, coerenza cronologica e contesto relazionale, con verifica manuale delle corrispondenze ambigue. Un caso di studio pratico della famiglia Mastrolillo dimostra come questo approccio ricostruisca le storie familiari multigenerazionali collegando fonti fiscali, ecclesiastiche e civili. Il sistema proposto è indipendente dalla località ed è scalabile ad altri comuni con fonti simili. Inoltre, è progettato per integrare dati provenienti da *partner* esterni come FamilySearch, facilitando così la ricerca sulle migrazioni transnazionali e assistendo nella gestione dei registri istituzionali. Questo modello posiziona Corato come prototipo per lo sviluppo di una base di dati internazionale sull'emigrazione pugliese, fungendo al contempo da strumento di collaborazione accademica nell'ambito delle discipline umanistiche digitali e degli studi sulle migrazioni.

Introduction

Corato has access to an extraordinary collection of historical documents dating back over four centuries. These include baptismal and marriage registers beginning in 1582, the eighteenth-century household census known as the *Catasto onciario*, and civil-registry books established during the Napoleonic era and maintained continuously up to the present day. The civil registry, in particular, contains comprehensive records of births, marriages, and deaths.

These historical resources form an invaluable network of information, allowing descendants of *coratino* emigrants and residents alike to trace family histories across generations and countries. Such documents are critical for historians studying the patterns and impacts of migration from Puglia to destinations including France, the Americas, and beyond. Additionally, these resources support municipal officers in the *Stato civile*, who regularly fulfill requests for official documentation and proof of lineage.

However, despite their immense value, these collections currently remain fragmented. While many records are scanned and available online, others, like the *Catasto onciario* and parish registers, have been digitized but are not yet publicly accessible online due to institutional and ecclesiastical restrictions. The existing digital images alone, without a comprehensive indexing system, often prove challenging to navigate. This fragmented accessibility complicates the task for descendants who typically start their searches online, expecting an efficient process. Clerks at the *Stato civile*, similarly, face unnecessary difficulties in quickly locating records due to incomplete or missing indexes.

The immediate challenge is therefore to connect these separate archives through an integrated indexing and matching system. This will ensure that the rich historical heritage of Corato is not only preserved but becomes more easily accessible, efficiently serving the needs of descendants, historians, and municipal officers alike.

A practical solution

The proposed solution involves creating a comprehensive, user-friendly online index that would integrate historical records from various collections into one centralized resource. This online index would function much like a detailed directory, providing essential details such as names, event dates, and basic family relationships including the identities of parents and spouses, for each historical document.

When the same individual appears across different types of documents – for example, appearing both in civil registry records and parish registers – the index would seamlessly link these records, clearly indicating that they refer to the same person. This approach significantly simplifies the

process for users who are seeking information about their ancestors or needing official documentation.

The integrated index would provide guided searches accessible both to the general public and to official staff, allowing users to enter a name and quickly receive an organized list of matching records. Depending on access permissions, users would either see the scanned image of the original document or just the indexed details extracted from it.

Images that are already publicly available through the *Archivio di Stato*'s servers could eventually be directly integrated into the index. Additionally, future partnerships, potentially involving institutions like FamilySearch or academic collaborators, could facilitate the publication of further document images, depending on mutual agreements and compliance with privacy and archival regulations.

How indexing works

An index serves as a comprehensive electronic guide to historical documents, making the process of finding specific records quicker and easier. Rather than manually examining each page, users can perform an electronic search for essential details like names, dates, and relationships that have been transcribed from the original documents. This structured approach significantly simplifies accessing historical records, allowing users to easily locate the information they need without needing to search through the images one by one or, even worse, physically handle the fragile original sources.

The work of indexing is carried out by trained volunteers, who may include local history enthusiasts, descendants of emigrants, and students from partner universities that join the project. These participants view high-resolution images of the historical documents and transcribe the essential information into an online form, following clear guidelines to ensure accuracy and consistency.

The basic cycle of a volunteer in the indexing process

1. Opens the next image.
2. Reads the record (a baptism, marriage, passport application etc.).
3. Types the names, dates, and other specified details relevant to that particular type of record into clearly labeled boxes.
4. Saves the entry.
5. The system queues the next image.

Every record is indexed twice by different volunteers; the software flags any discrepancies for a supervisor to resolve.

Which details are recorded?

Doc. type	Indexable data (fields transcribed)
Baptism (parish)	Child's full name; date of birth; date of baptism; father's name; mother's name; paternal grandfather's name; maternal grandfather's name; godparents' names; indication of whether grandfathers were alive; whether the child was a foundling (<i>trovatello</i>); whether the child was a twin; whether the child was posthumous (born after father's death); parish reference (volume, folio)
Marriage (parish)	Names of bride and groom; father's name for each; indication whether the father was alive; whether the bride/groom is widowed; date of marriage; parish reference
Burial (parish)	Name of the deceased; names of parents or spouse (if recorded); indication whether parents or spouse were alive; date of burial; parish reference
Birth (Stato civile)	Child's full name; names of parents; names of grandfathers; indication whether grandparents were alive; parents' ages; parents' places of birth; date of birth; date of baptism (if recorded); whether the child was stillborn; whether the child was a foundling (<i>trovatello</i>); whether the child was a twin; whether the child was posthumous (born after father's death); date of registration; act number
Marriage (Stato civile)	Names of bride and groom; dates of birth (if recorded); places of birth; ages; names of parents; indication whether parents were alive; dates and places of death of parents (if applicable and recorded); dates and places of death of paternal grandfathers (if applicable and recorded); date of civil marriage; date of religious marriage (if recorded); date of registration; act number
Death (Stato civile)	Name of the deceased; place of birth; age; names of parents; indication whether parents were alive; ages of parents (if recorded); names of spouse(s); indication whether spouse(s) were alive; date of death; date of registration; act number
Catasto onciario	Full name of the head of household; name of the father of the head of household and whether he was alive (if recorded); names of all other inhabitants of the household; their relationship to the head; their ages; marital status (if recorded); <i>fuoco</i> (household) number
Passport application	Passport number; applicant's full name; name of father; date and place of birth; address (if recorded); occupation; intended destination; date of issuance; remarks (if recorded)

Only these concise facts are entered. Full images and any sensitive notes remain secure and are shown only when authorized.

Case study: The Mastrolillo family

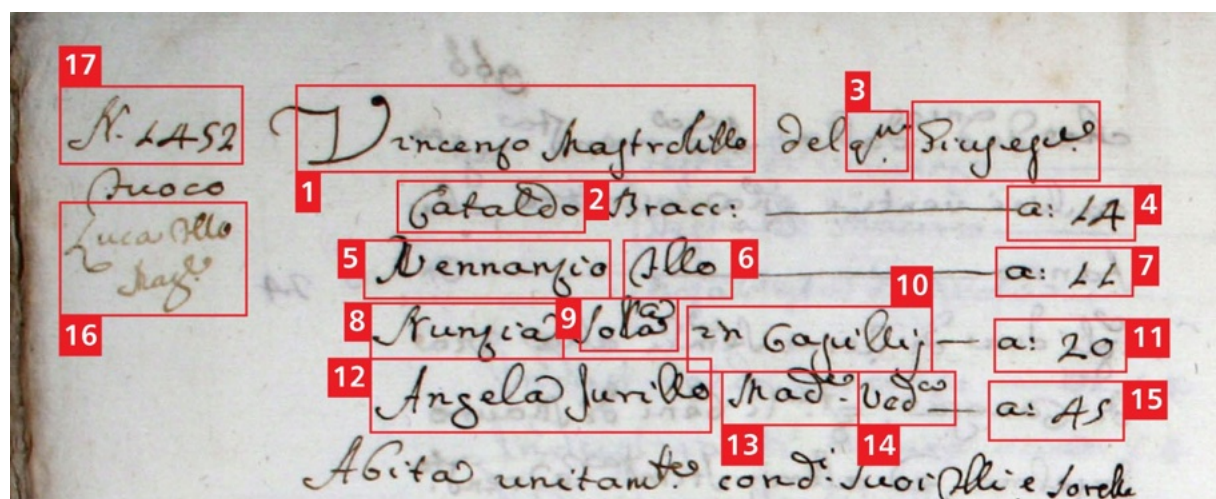
To demonstrate in concrete terms how the integration of primary records can clarify and reconstruct family histories, we present a real case from the 18th and 19th centuries involving the Mastrolillo family of Corato. By connecting fiscal, ecclesiastical, and civil sources, we can follow the trajectory of this family across generations, illustrating the strength of an integrated indexing project.

The documents below show how various types of records, each preserving distinct information, complement one another to create a coherent genealogical narrative.

Catasto onciario household No. 1452 (1754)

Type: Fiscal census record

In 1754, Vincenzo Mastrolillo, a 14-year-old boy, is listed as the *capofuoco* (head of household) in the *Catasto onciario*. Although legally a minor, he is registered as the family's fiscal representative, likely due to the death of his father, Giuseppe Cataldo Mastrolillo. His widowed mother, Angela Iurillo (45), and his siblings Vennanzio (11) and Nunzia (20) are listed in the same household. The *Catasto* record also mentions an older brother, Luca Mastrolillo, living separately.



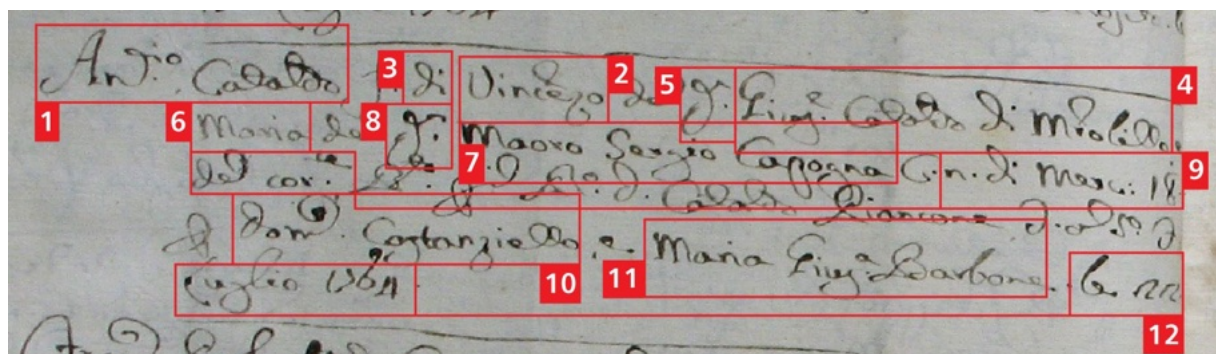
1. **Vincenzo Mastrolillo (person A)**: *capofuoco* (head of household).
2. **Giuseppe Cataldo (B)**: The *capofuoco*'s father.
3. **q[uonda]m**: The *capofuoco*'s father was already dead at that time.
4. **a[nni] 14**: Vincenzo was 14 years old.
5. **Vennanzio (C)**: Another resident of the household.

6. *f[rate]llo*: A brother of the *capofuoco*.
7. *a[nni]* 11: Vennanzio was 11 years old.
8. **Nunzia (D)**: Another resident of the household.
9. *so[re]lla*: A sister of the *capofuoco*.
10. *in capillis*: She was single, of marriageable age.
11. *a[nni]* 20: Nunzia was 20 years old.
12. **Angela Iurillo (E)**: Another resident of the household.
13. *mad[r]e*: The *capofuoco*'s mother.
14. *ved[ov]a*: She was a widow.
15. *a[nni]* 45: Angela was 45 years old.
16. **Luca f[rate]llo mag[gior]e (F)**: An older brother of the *capofuoco* living elsewhere, probably married.
17. **N. 1452**: Household number for identification purposes.

Baptism record (1764)

Type: Parish baptism record

On 22 July 1764, Antonio Cataldo Mastrolillo is baptized by d. Cataldo Piancone at the parish of Corato. He had been born a few days earlier, on 18 July 1764. His father is Vincenzo Mastrolillo and his mother is Maria Capogna. The record names both paternal and maternal grandfathers, Giuseppe Cataldo Mastrolillo and Maoro Sergio Capogna, both noted as deceased. Antonio's godparents are Domenico Costanziello and Maria Giuseppa Barbone. This baptism entry confirms the continuity of the Mastrolillo family and provides additional generational links.



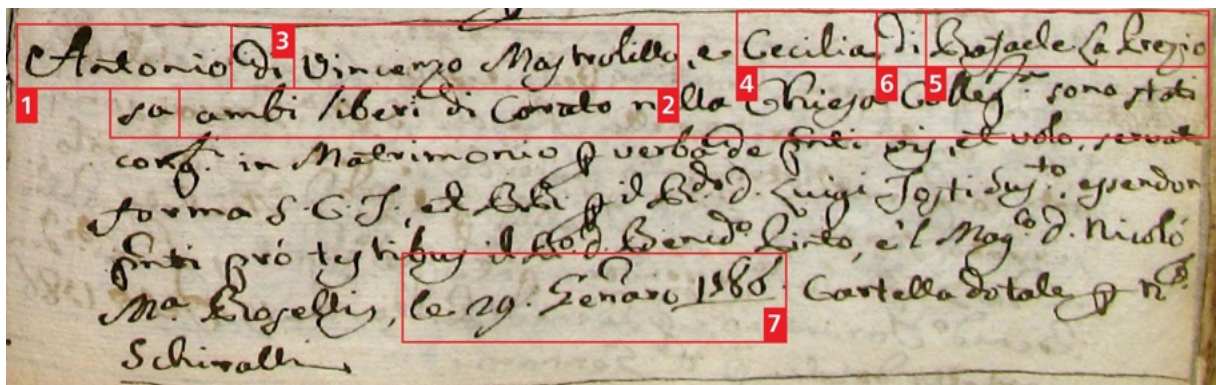
1. **Ant[oni]o Cataldo (G)**: Newborn being baptized.
2. **Vince[n]zo (A)**: The newborn's father.
3. **di**: The newborn's father was alive.

4. **Gius[epp]e Cataldo di M[ast]rolillo (B)**: The newborn's paternal grandfather.
5. **q[uonda]m**: The newborn's paternal grandfather was dead.
6. **Maria (H)**: The newborn's mother.
7. **Maoro Sergio Capogna (I)**: The newborn's maternal grandfather.
8. **q[uonda]m**: The newborn's maternal grandfather was dead.
9. **n[ato] di merc[oledi] 18 del cor[ren]te**: Born on Wednesday, 18th of the current month.
10. **Dom[eni]co Costanziello (J)**: The newborn's godfather.
11. **Maria Gius[epp]a Barbone (K)**: The newborn's godmother.
12. **le 22 Luglio 1764**: Baptized on 22 July 1764.

Marriage record (1786)

Type: Parish marriage record from the Chiesa Collegiata

On 29 January 1786, Antonio Cataldo Mastrolillo marries Cecilia La Preziosa. The marriage is celebrated by d. Luigi Tosti, with d. Benedetto Pinto and *Magnifico* d. Nicolò Maria Rosellis acting as witnesses. Antonio is explicitly identified as the son of Vincenzo Mastrolillo.



1. **Antonio (G)**: The groom.
2. **Vincenzo Mastrolillo (A)**: The groom's father.
3. **di**: The groom's father was alive.
4. **Cecilia (L)**: The bride.
5. **Rafaele La Preziosa (M)**: The bride's father.
6. **di**: The bride's father was alive.
7. **le 29 Gen[n]aro 1768**: Married on 29 January 1768.

Death mention (death record No. 23, 1823)

Type: Civil death record (Stato civile)

In 1823, Cecilia La Preziosa, recorded as the widow of Antonio Cataldo Mastrolillo, dies. Although Antonio's own death record has not been found, Cecilia's civil death record confirms that Antonio had predeceased her, and likely died before 1809, the year from which civil registration officially begun in the Kingdom of Naples. This final piece allows us to close the cycle on this branch of the Mastrolillo family.

ATTO DI MORTE.

N. d' Ordine *Venticinque*

11, Anno milleottocento *ventitree*, il dì *quindici*
del mese di *Gennaio* alle ore *quindici*
avanti di noi ** Don Nicola Spasini Sindaco*
ed ufficiale dello stato civile del Comune di *Corato*
Distretto di *Barletta*, Provincia di Terra di Bari,
sono comparsi *Francesco Bone* di
anni *venti* di professione *bracciante* regnicolo
domiciliato *Stradale Mierza*, e *Luigi*
Sarvagallo di anni *quarantacinque*
di professione *bracciante* regnicolo domiciliato *Strada*
Antona i quali han dichiarato, che nel giorno
soprascritto del mese di *Gennaio* anno *corrente*
alle ore nove è morta nella propria casa
1 *L. Cecilia la Preziosa* *di anni cinquanta* **2** nata in *Corato*
Concedina domiciliato *3* *Abate Antonio Totoli* **4** figlia
dr. fu Raffaele di professione **5**
6 *del fu* **7** *Maria Giuseppe Anala*
domiciliata *Verona di Antonio Mastrolillo* **8**
10 Per esecuzione della **9** legge ci siamo trasferiti insieme co'
detti testimonj presso la persona defunta, e ne abbiamo rico-

1. *giorno soprascritto* del mese di Gennaio anno corrente**: Death occurred on the same day the record was made (indicated by an asterisk: 15 January 1823).
2. **Cecilia la Preziosa (L)**: The deceased.
3. *di anni cinquanta*: Of (approximately) 50 years of age.
4. *nata in Corato*: The deceased was born in Corato.
5. **Raffaele (M)**: The deceased's father.
6. *del fu*: The deceased's father was dead.

7. **Maria Giuseppa Ferola (N)**: The deceased's mother.
8. **della fu**: The deceased's mother was dead.
9. **Antonio Mastrolillo (G)**: The deceased's spouse.
10. **vedova di**: The deceased's spouse was dead.
11. **N. d'Ordine Ventitre**: Act number.

The strength of an integrated indexing system

This case study of the Mastrolillo family clearly illustrates the strength of an integrated indexing system. A researcher searching for **Antonio Cataldo Mastrolillo (G)** in a simple, isolated manner might find only a single record – perhaps his marriage, or a mention in a death certificate. Yet, when records from different sources are indexed together programmatically, each small piece builds a bigger picture.

Starting with his baptism, we confirm Antonio's parentage: **Vincenzo Mastrolillo (A)** and **Maria Capogna (H)**. The baptismal record further reveals the names of his grandfathers: **Giuseppe Cataldo Mastrolillo (B)** and **Maoro Sergio Capogna (I)**. The names of his godparents, **Domenico Costanziello (J)** and **Maria Giuseppa Barbone (K)**, show the family's network within the parish.

The marriage record connects Antonio (G) to his wife **Cecilia La Preziosa (L)** and his father-in-law **Raffaele La Preziosa (M)**. The marriage record also confirms that Antonio's father, Vincenzo (A), was still alive in 1786.

Through the *Catasto onciario* of 1754, we trace Antonio's paternal family during Vincenzo's (A) youth, discovering his grandmother **Angela Iurillo (E)** and his paternal uncles and aunts, **Vennanzio (C)** and **Nunzia (D)**. This early record, though distant from Antonio's own lifetime, becomes critical in establishing the family's social standing and economic situation at mid-century.

Finally, the *Stato civile* death record of **Cecilia La Preziosa (L)** shows us that her mother's name was **Maria Giuseppa Ferola (N)** and indirectly confirms Antonio's (G) death prior to 1823, allowing us to place the probable date of his passing even though no individual death record for him was found.

Thus, a search beginning with a single name (Antonio Cataldo Mastrolillo) could, thanks to cross-indexing, immediately reveal connections to at least thirteen other individuals (A to N) across four different types of sources, spanning nearly seventy years. It would enable a user, whether a descendant, a scholar, or a municipal officer, to reconstruct the full outline of a family story that would otherwise be lost among thousands of isolated entries.

This example shows how, through careful and unified indexing, Corato's historical records can cease to be a fragmented archive and become instead a living memory, accessible to all who seek to understand their roots.

But how is this done?, or: How matching works

After all the historical documents have been carefully indexed, a second step begins. This step is where technology plays its role: using the indexed information, a matching system is able to find connections between different records that speak about the same person or family, even when those records were created years or even decades apart.

The system does not simply look for identical names. It works more carefully: it compares names, dates, relationships, and other details to determine when two different documents are, in fact, referring to the same individual. For example, if a baptismal record names a child Antonio Mastrolillo, son of Vincenzo, born in 1764, and a marriage record twenty years later shows an Antonio Mastrolillo, also son of Vincenzo, marrying in Corato, the system recognizes the strong likelihood that these two records refer to the same person. It connects them together in a process that is invisible to the user.

From the perspective of a researcher or descendant searching for their ancestor, it appears simple: by typing one name into the system, they can instantly see all the records that relate to that person, from baptism to marriage to burial, and even property listings or migration documents. Instead of having to search manually through thousands of pages, each from a different source, the connections are already made, ready to be explored.

This method of matching allows the town's memory, spread across many books and collections, to be reunited electronically, making it possible to tell complete family stories that would otherwise remain hidden in separate, silent archives.

The Mastrolillo family, part II: How the matching happens in practice

Once all the documents have been indexed, the system begins a second, crucial phase: the matching of information across different sources. This process is carried out in two distinct stages: first, the internal work of matching and verification; then, the experience offered to the public user.

First stage: Matching and review

The matching system compares all indexed entries systematically. When two records match perfectly on key points, such as name, parents' names, and relevant dates, the system automatically establishes a link between them with a high degree of certainty.

In cases where small discrepancies are found (for example, a slight variation in the spelling of a name or an approximate date*), the system proposes a match but does not confirm it immediately. Instead, these suggested matches are flagged for human review. A volunteer or researcher, possibly but not necessarily the same person who contributed to the indexing effort, examines the proposed connection.

The reviewer checks whether the surrounding information, such as family relationships, occupation, residence, or associated persons, supports the proposed match. Only after this manual verification is the link confirmed within the system. This careful two-step process ensures that users will navigate a database where family connections are reconstructed with both precision and prudence.

Second stage: What the user sees

When a user (for example, a descendant of Antonio Cataldo Mastrolillo) accesses the system, they are presented with a complete and coherent view of the information already reviewed and confirmed.

The user searches for “Antonio Cataldo Mastrolillo”. Instantly, without any need to perform complex research or to examine records individually, the system displays a series of results:

- The baptism record of Antonio, showing his parents, Vincenzo Mastrolillo and Maria Capogna, and naming his grandparents and godparents.
- The marriage record where Antonio marries Cecilia La Preziosa, with references to their parents and witnesses.
- The *Catasto onciario* record from Vincenzo’s youth, providing a view of Antonio’s family background a generation earlier.
- The death record of Cecilia La Preziosa, confirming the death of her husband Antonio.

Each of these records is already connected by the system based on verified links. The user does not need to guess whether these pieces belong to the same family: the platform presents them as parts of a single, unified story.

In this way, technology and human expertise work hand in hand. The user benefits from the patient and careful work already done behind the scenes, accessing in moments what once would have required months or even years of manual research.

* This can occur even when the document is faithfully indexed because not all recorded data reflect strict factual accuracy. Ages declared at death, for example, or the ages of parents reported in birth records, are often approximations or estimates and can deviate significantly from reality. Furthermore, different scribes had different views on how a name should be spelled.

Family history search — Results for: Antonio Cataldo Mastrolillo

Name: Antonio Cataldo Mastrolillo
Born: 18 July 1764 – Corato, Bari, Puglia, Italy
Father: Vincenzo Mastrolillo
Mother: Maria Capogna
Spouse: Cecilia La Preziosa

Related records:

Catasto onciario – Vincenzo Mastrolillo (father) – 1754

Baptism – Antonio Cataldo Mastrolillo – 22 July 1764

Marriage – Antonio Cataldo Mastrolillo & Cecilia La Preziosa – 29 January 1786

Death – Cecilia La Preziosa (spouse) – 15 January 1823

[Back to search](#)

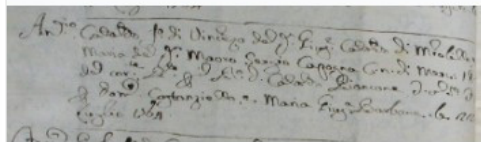
Illustrative mockup of the user interface.

Baptism – Antonio Cataldo Mastrolillo – 22 July 1764

Indexed data

Reference: Vol. B 1763-1777, fol. 16
Full name: Antonio Cataldo Mastrolillo
Foundling? No
Father: Vincenzo Mastrolillo
Was the father alive? Yes
Mother: Maria Capogna
Paternal grandfather: Giuseppe Cataldo di Mastrolillo
Was the paternal grandfather alive? No
Maternal grandfather: Maoro Sergio Capogna
Was the maternal grandfather alive? No
Godfather: Domenico Costanziello
Godmother: Maria Giuseppa Barbone
Date of birth: 18 July 1764
Date of baptism: 22 July 1764

[View image \(click to enlarge\)](#)



Record view (indexed data and corresponding image).

Bridging a larger gap

While reconstructing family stories within Corato already demands the careful combination of parish, fiscal, and civil documents, an even greater challenge arises when the story extends beyond local borders. Beginning in the late 19th century, thousands of *coratini* emigrated – first to elsewhere in Italy, then to France, to the Americas, and to even more distant places.

In these cases, it is not enough to connect different types of records from within Corato. The person we seek may have left few traces in local archives after emigration, and yet their life continued elsewhere, recorded in foreign documents, censuses, naturalization files, passport applications, and burial registries in distant lands.

It is here that larger partnerships become essential. Institutions such as FamilySearch, as well as national archives and academic centers abroad, have gathered extensive collections of digitized records relating to Italian emigrants. Some of these collections have already been indexed, allowing searches by name and place; others exist only in digital images, awaiting structured access.

By integrating data from these external sources into the matching system, it becomes possible to bridge the gap between the native *coratino* and the emigrant living in France, the United States, or beyond.

For instance, a person found in Corato's *Stato civile* registers as a child might appear decades later in a French residence permit, a Brazilian immigration file, a U.S. census... or maybe even in all of them. Without unified indexing, these fragments would remain scattered across continents. With it, descendants can follow a single continuous path across space and generations, reconnecting with the family's complete history.

Thus, while building a solid foundation with Corato's own resources may be the first step, the true potential of the project lies in creating bridges outward – using international partnerships and shared databases to rebuild the migratory journeys that have shaped so many lives around the globe.

The need for a strong academic partner

The vision we propose – reconstructing individual life stories, linking generations, and bridging migrations across continents – is ambitious by nature. It reflects the richness and complexity of the human experience. Yet to bring such a project fully to life, technical means alone are not enough. It requires the dedicated support of an institution capable of understanding both the historical significance and the methodological challenges involved.

An academic partner, particularly a university engaged in migration studies, can offer the necessary foundation to realize this vision. Beyond technical assistance, universities bring a

tradition of structured research, critical validation, and a capacity to train future generations of scholars who will carry this work forward. Their participation ensures that the project is not only built with rigor, but also studied, expanded, and kept alive as a living resource for both academic and public audiences.

Furthermore, an academic institution can act as a bridge between different worlds: between local archives and international repositories; between traditional historical research and modern digital humanities; between the descendants seeking their roots and the broader narratives of migration and cultural transformation.

Because the method of indexing and matching we propose is built on universally structured genealogical principles – names, dates, family relationships, places – it remains agnostic to specific localities. Once established, the same system can be applied without modification to any other community that possesses similar sources: parish registers, *Stato civile* acts, fiscal censuses such as the *Catasto onciario*. The only limits are the availability of digitized material and the dedicated effort of trained volunteers to carry out indexing and verification of matches.

In this sense, Corato would serve not merely as an isolated case, but as a prototype for developing an international database on emigration from Puglia. Through careful expansion, the project could later include other towns whose histories mirror Corato's.

This should be read as an invitation: the opportunity to contribute to a project that connects personal stories to global history, that serves both the citizens of Corato and Puglia today and the wider community of scholars and descendants across the world.

With the right partnership, Corato's archives will not simply be preserved. They will be given new life, and new meaning for generations to come.

May 2025; rev. July 2025

Rafael Caterina (São Paulo, Brazil)

For the Atelier Généalogique