Early Modern Digital Itineraries: The Italian Core Set

Rachel Midura
Virginia Polytechnic Institute and State University (Virginia Tech)

Small, cheaply published itinerary books written by professional travelers indicated precisely which routes to utilize, where to stay, which sites to see, and even provided tools for navigating foreign customs, language, and currency in the early modern period. Early Modern Digital Itineraries (EMDigIt) transforms itinerary books printed from the sixteenth through eighteenth centuries into a unique dataset for the study of historical mobility. This piece describes the workflow for semi-automatic transcription, data tidying, and linking historical place names to modern geographic data. It discusses use cases of the Italian core set, which consists of eight itinerary books published in the Italian language and featuring more than four thousand waypoints. This data is freely available for textual, spatial, and network digital exploration and analysis.

I f you teach history; read, or write historical fiction; play, or design historical games; you have likely faced a similar set of logistical questions. The process of imagining yourself in the well-worn shoes of an early modern traveler raises questions such as: where could I travel, and how? How much would it cost me? How long would it take? And, trickiest of all, how would I know any of this?

The initial answer is often an unsatisfactory «it depends», or «it's complicated». Travel memoirs and letters are vague on these nuts-and-bolts details, while official legislation and secondary sources range from the overwhelmingly encyclopedic to the narrowly parochial. Writing about historical mobility requires hours tracking sources to answer a single question for an individual, route, or time period. Recent years have brought excellent new work in mobility studies, as scholars broaden the cast of known travelers, or reconstruct the sensory experience of the travel. Yet the editors of one such volume call for increased attention to the «quotidian mechanics», in other words, «how mobility worked (or did not) in a very concrete sense»[1]. Much like reimagining a travel experience, digital tools for spatial analysis and visualization demand greater clarity than a traditional prose account.

This article introduces the Early Modern Digital Itinerary Project (EMDigIt) which brings sources uniquely suited to answering these questions together with the digital capacity

---

1   Rosa Salzberg — Paul Nelles, *Movement and Mobility in the Early Modern World: An Introduction*, in: *Connected mobilities in the Early Modern world: the practice and experience of movement*, ed. by P. Nelles, R. Salzberg, Amsterdam: Amsterdam University Press, 2023, p. 8.

for multi-scalar analysis and exploration. Small, cheaply published itinerary books written by professional travelers indicated precisely which routes to utilize, where to stay, which sites to see, and even provided tools for navigating foreign customs, language and currency. Itinerary books structured European conceptualization and navigation of space from antiquity through the eighteenth century. They featured tables of routes detailing how to proceed from one city to another as lists of intermediary stops, often glossed with historical sites, postal stations, or border crossings[2].

EMDigIt pilots a semi-automated transcription, data tidying, and visualization workflow, transforming the itineraries into datasets appropriate for textual, spatial, and network analysis. An EMDigIt web platform will facilitate individual inquiries (e.g. «how could I get from Milan to Paris in 1575?») as well as more complex stories about how routes grew, shrank, and shifted over time. This project joins studies of European road networks such as Itiner-e and Viabundus[3], with the additional benefit of selecting for routes that were widely republished, often over long time periods, and consulted by a variety of readers[4]. Merchants, diplomats, pilgrims, tourists, soldiers, and couriers traveled by these arteries of a pan-European route network.

This article accompanies a first major data release, which we call the Italian Core Itineraries. Eight itinerary books published in Italian from 1562-1720 provide 4,500 individual geo-referenced waypoints utilized by early modern travelers. These eight titles include:

– SAB1562: Anonymous, *Le Poste, Necessarie A Corrieri & Viandanti, Per L'Italia Francia, Spagna, & Alemagna Con le Fiere che si fanno per il Mondo,* Brescia: 1562.
– GH1563: Giovanni dell'Herba, Cherubino Stella, *Itinerario delle poste per diverse parti del mondo*, Roma: 1563.
– OC1608: Ottavio Codogno, *Nuouo itinerario delle poste per tutto il mondo*, Milano: 1608.
– FS1610B: Franciscus Schottus, *Itinerario, overo nova descrittione de' viaggi principali d'Italia*, Padova: 1610.
– OC1623: Ottavio Codogno, *Compendio delle poste*, Milano: 1623.
– SASD: Anonymous, *Poste Diverse d'Italia, Alemagna, Spagna, e Francia*, Milano: SD.
– GM1684: Giuseppe Miselli, *Il burattino veridico, overo, Instruzione generale per chi viaggia*, Roma: 1684.
– GV1720: Giovan Maria Vidari, *Il viaggio in pratica*, Napoli: 1720.

---

2   The outcomes of the 2023-2025 NEH Digital Humanities Advancement Grant are available in the project white paper: Rachel Midura, *Early Modern Digital Itineraries: Workshops for Data-Driven Approaches to Premodern Travel White Paper*, 2025 <https://apps.neh.gov/publicquery/AwardDetail.aspx?gn=HAA-293210-23>, and <https://github.com/rmidura/EMDigIt/blob/63b5fef0f9085c6c4cbd465dbaeb990a29477396/Early%20Modern%20Digital%20Itineraries%20White%20Paper.pdf>
3   Bart Holterman et al., *Viabundus Pre-modern Street Map 1.3*, <http://www.viabundus.eu/>.
4   Rachel Midura, *Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700,* «Journal of Social History» 54 (2021), n. 4, p. 1023–1063.

1. Project Background

The EmDigIt Project began during my time as a senior graduate research fellow at the Stanford Center for Spatial and Textual Analysis (CESTA). I was at work on my dissertation, now adapted into my first book, on the advent of Europe's early modern postal networks[5]. Ottavio Codogno, the postmaster lieutenant of Milan, published two such itineraries: the *Nuovo itinerario* in 1608 (and reproduced in Venice in 1611 and 1616, see Fig. 1), and the *Compendio delle poste* in 1623. These books are well known for their chapters describing the history and the contemporary operations of the Habsburg postal systems run by the Tassis family of postmaster and postmistress generals. However, most of the book consists of semi-structured route tables (Fig. 1) drawn from Codogno's professional knowledge of pan-European transit and commerce. Codogno assured his readers from the front page that these routes were reproduced «not just for secretaries, but for clerics and merchants»[6].

Figure 1. *Page from a Venetian edition of Ottavio Codogno's Nuovo itinerario (1611). Route tables frequently featured observations on borders, scenery, and historical or religious sites of interest. Bayerische Staatsbibliothek (BSB), München, Res/Geo.u. 87*

---

[5] Rachel Midura, *Postal Intelligence: The Tassis Family and Communications Revolution in Early Modern Europe,* Ithaca (NY): Cornell University Press, 2025.

[6] O. Codogno, *Nuovo itinerario,* cit.

Codogno's book is perhaps the best-known example of a much broader genre. I assembled a bibliography of itineraries and a database of route headers that could be found across itinerary books printed in many different languages and countries[7]. This first iteration of the EMDigIt database supported an article, *Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700*, in which I argue that network methods reveal the only semi-spatial indexing of space. Proximity was one method among many at work in drawing associations and directionalities among routes. These served as mnemonics for space in a largely pre-cartographic world[8].

In summer 2023 I received a Level 1 Digital Humanities Advancement Grant from the United States National Endowment for the Humanities (NEH) to support a series of workshops advancing data-driven approaches to the history of travel. Our goal was to establish a professional community of researchers on premodern, digital, spatial history and explore how geographic information found in primary sources such as letters and journals could be extracted and mapped with EMDigIt to trace the movement of people and goods. Participants explored shared research questions over the course of the year, at the same time as student research assistants helped to extract and refine data from the corpus of early modern itineraries. By doing so, we aimed to bring desired audiences as stakeholders in the early stages of project design[9], presenting the results at a one-day workshop in conjunction with the Digital Humanities Organization conference in Arlington, Virginia in August 2024.

The year-long series of meetings began with only two itinerary books that had been processed to the point of basic geo-reconciliation (*Compendio delle poste* and *Itinerario delle poste*). From this, we built out an adaptable workflow (Fig. 2) for other itineraries, beginning from those with the shared publication language of Italian. Work with student collaborators significantly improved the accuracy and granularity of the data, including extracting new qualitative features and descriptors, such as borders, monasteries, and warnings[10]. Over 2024, we complete the workflow for an additional six Italian-language itinerary books, chosen in consultation with participating scholars for their omnipresence in libraries and private collections across Europe.

---

[7] Armando Serra, *"Monopolio naturale": di autori postali nella produzione di guide italiane d'Europa, fonti storico-postali tra cinque e ottocento*, «Archivio per la storia postale: comunicazioni e società», (2023), n. 14–15, p. 40–51. Individual books have been reproduced in edited volumes and datasets: Robert Hibberd — Jack B. Owens, *Before Highway Maps: Creating a Digital Research Infrastructure Based on Sixteenth-Century Iberian Places and Roads*, «Bulletin for Spanish and Portuguese Historical Studies», 40 (2015), n. 1; Clemente Fedele — Armando Serra — Marco Gerosa, *Europa Postale*, Bergamo: Museo dei Tasso e della Storia Postale, 2014.

[8] R. Midura, *Itinerating Europe*, cit.

[9] For the project participants see: <https://emdigit.org/>.

[10] With thanks to the Virginia Tech Computational Modeling and Data Analytics Capstone instructors and students.
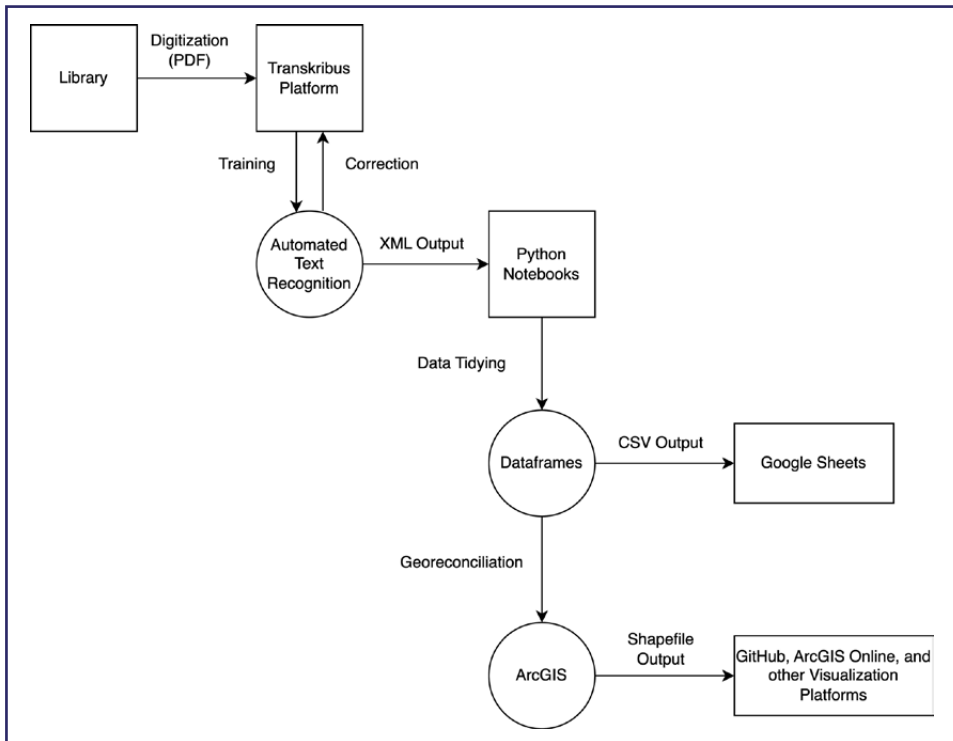
Figure 2. *Generalized model of the EMDigIt Workflow. The Transkribus (READ-COOP)[11] platform provides auto-mated text recognition and can be refined through further training on completed itineraries. We then tidy the XML output into data-frames using a set of Python Notebooks. The data can then be exported to Google Sheets as relational tables or imported into ArcGIS for transformation into shapefiles for use in visualization*

We begin from PDFs or archival photographs of the itineraries, uploading them to the Transkribus servers. We then run a neural-network model for text recognition trained on the hand-corrected transcription of *Compendio delle poste* and *Itinerario delle poste*. We use Python to re-segment the data, building and reconciling against a growing project gazetteer. We link out to the GeoNames gazetteer where possible, and calculate ap-proximated coordinates where not based upon journey bearing and distance from prior locations. Our data in turn will be made available through the World Historical Gazetteer (University of Pittsburgh)[12].

Our goal is to release data in stages, with each core set reflecting the most widely repub-lished and utilized itineraries in a given language, as informed by the domain knowledge of project affiliates. With the help of Molly Taylor-Poleskey of Harvard Map Collec-tion, Eva Chodějovská of Masaryk University and a pilot project between Innodata[13] and Harvard Libraries, work is well underway on a set of German-language itineraries.

---

[11] https://app.transkribus.org/.

[12] Ruth Mostern — Karl Grossner et al, *World Historical Gazetteer*, <https://whgazetteer.org/>.

[13] https://innodata.com/.

The semi-automatic extraction, segmentation, and georeconciliation process will require some refitting to each language, as it relies upon on identifying certain prepositions or key terms. The periodic revision of the workflow, however, also provides staged opportunities to revisit other aspects and to bring in new scholars with pertinent domain knowledge as stakeholders in the project.
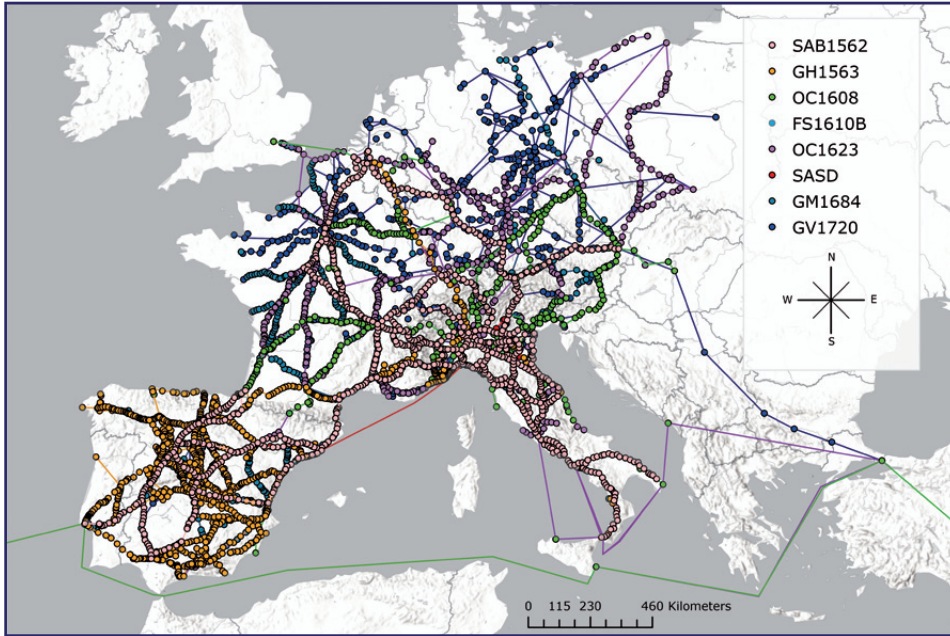

Figure 3. *Map representing all eight itineraries featured in the Italian Core Set of EMDigIt. Each point is an origin, waypoint, or destination that occurs along one or more routes. The itineraries are arranged chronologically, inferring a likely date of publication for the anonymous and undated* Poste Diverse d'Italia, Alemagna, Spagna, e Francia. *Points or routes that occur in multiple itineraries are shown with their earliest title*

The EMDigIt data (Fig. 3) will remain a digital translation of a pre-existing source base of reference literature. While we will continue to publish our workflow and code for scholars at work in other geographic and temporal contexts, we do not intend to open the platform in the style of crowd-sourcing projects. This does mean that EMDigIt will replicate many of the limitations and biases of the sources it transforms. These include (but are not limited to) Eurocentrism, emphasis upon land routes over sea routes, and clear geographic "holes" in the minds of itinerary creators – see, for example, the notable absences in southern Italy, or the Baltic. EMDigIt is not a perfect window on to the experience of travel but rather sheds light on the information that an early modern traveler might have had at hand.
The dynamic nature of our data and intended web platform most closely resembles Orbis: The Stanford Network Model of the Roman World (Stanford University)[14].

---

[14] Walter Scheidel — Elijah Meeks, *Orbis*, <https://orbis.stanford.edu/>.

This web platform combines cartographic space with experiential factors such as seasonality, elevation, and mode of transport to simulate cost-benefit analysis with contemporary conditions. Users can select points and modes to map their journeys. The site enjoys use by scholars and a wider public for measuring premodern travel, even outside the temporal bounds of the ancient Roman world. The applicability of Orbis for the early modern world is nonetheless limited, as new linguistic, political, and confessional boundaries shaped traveler routes. Orbis has inspired several efforts to produce an «Orbis-in-a-Box», separating methodology and content[15]. We pick up the torch with an important innovation: restricting our data to a delimited and exportable database. Furthermore, EMDigIt will continue to develop in ways that draw from and support projects such as Beyond the Horizon (Moravian Library)[16], Viabundus (Universität Göttingen)[17], and Itiner-e[18].

2. The Italian Core Set

The EmDigIt data consists of four primary objects of analysis: works, editions, routes, edges, and locations. Here I will briefly consider each object in turn as well as their potential for inquiry. Franciscus Schottus' *Itinerario, overo nova descrittione de' viaggi principali d'Italia* serves as a representative example. This work had at least 43 known editions republished across Europe and in many languages. The 1610 edition features 34 routes. These routes include 195 edges, meaning connections among 284 unique locations. While these numbers may seem straightforward, they tell a more complex story.

2.1 Works

First, how do we define a *work,* and an *edition*? There is no need to reinvent the wheel, given the ample scholarship in the history of the book. EMDigIt follow a typical pattern by distinguishing editions by indications that they were printed at a different time, in a different location, or by different publishers. For example, we use FS1610B to refer to the Italian edition published in Venice by Francesco Bolzetta. FS1610A refers to another edition that appeared in Latin in Vicenza the same year[19]. Route tables, however, only appeared in the Italian editions of the work. They had been added by the entrepreneurial Francesco Bolzetta, who continued to publish new editions for several decades. In fact, we see from the digital transformation that publishers regularly intervened in route publication in this way, often copying and

---

[15] Maxim Romanov — Masoumeh Seydi — James Baillie — Karl Grossner — Rainer Simon — Marfa Vargha, *Orbis-in-a-Box (OIB): Modeling Historical Geographical Networks in ADHO DH Conference at Utrecht University*, 2019, <https://staticweb.hum.uu.nl/dh2019/dh2019.adho.org/programme/book-of-abstracts/index.html>.

[16] Eva Chodějovská et al., *Beyond the Horizon,* <https://beyondthehorizon.mzk.cz/>.

[17] B. Holterman et al. *Viabundus*, cit.

[18] Tom Brughmans — Joseph Guitart — Santiago Muxach — Pau de Soto, *Itiner-e,* <https://itinere.iec.cat/>.

[19] Franciscus Schottus, *Itinerarium Nobiliorum Italiae Regionum, Urbium, Oppodorum, et Locorum*, Vicenza: Petrum Bertellium, 1610.

incorporating routes from other publications or making changes where information had presumable become outdated[20].

2.2 Routes

Second, the itineraries are lists of *routes*, with connections between locations either explicitly stated (e.g. «From Rome to Bologna») or simply implied by the sequential logic of a list. We use the Transkribus tool to transform images into XML which preserves important information about the pixel coordinates of text on the original page. As a reference material, itineraries are formulaic by nature: they use prepositions, symbols (such as numbers, asterisks, or printed manicules), and position on the page to signify the purpose of a given text. We can use regular expressions in a Python workflow to distinguish text as a route header, a location name, or a page number.

Symbolic logics do change among titles. Using a dictionary helps to adapt the workflow to a given book but requires periodic human attention. That attention can be productive in its own right; for example, EMDigIt project participants noted that itinerary authors occasionally changed the method by which they measured distance or formatted their text based upon the region in question. This led to new observations regarding where Roman system of measurements predominated over more localized systems, such as the French mile. Furthermore, the digital approach facilitates moving between a qualitative and quantitative consideration. Once location points are geo-referenced, we can measure the modern kilometer distance between them and compare to the original sixteenth or seventeenth century estimation. This in turns provides a dictionary with highly granular data as to what one mile likely meant when used in Provence, Lazio, or Bavaria (Table 1).

Region	Early Modern Italian Mile	Average Distance in Modern Km
Emilia-Romagna	1	1.58
Lazio	1	1.95
Liguria	1	1.48
Lombardy	1	1.84
Piedmont	1	1.82
Veneto	1	1.60
Italy (All Regions)	**1**	**1.84**

Table 1. *While itinerary creators often referred to a unified Italian mile, comparison by modern regions shows the variance from region to region*

---

[20] It will therefore be valuable to add more editions in future releases. This will support comparison within a given title among editions, and offer greater granularity of change across time to the overall picture of route networks.

2.3 Locations

What, then, are the nodes of this network? These are third objects, or the *locations*. The use of a gazetteer, or dictionary of places, is crucial for linking a text object to spatial characteristics such as a set of coordinates and information about region or state. The Geonames Gazetteer (Unxos GmbH)[21] offered a freely accessible starting point, however it is a modern gazetteer that is ill-equipped to help with identifying the location of the «inn of the two sisters», or a location simply designated as «borgo». To this end, we built and published a project gazetteer.

Name variants from itineraries	Standardized name	GeoNames ID
aborquecque; alberquech; alborqueque; albroqueque; albucheche; alburqueque|berquech; borquech	Alburquerque	2522183
betzanon; bressanon; bressenon; brexanon; brixanon; brixen; brixien; presenon	Bressanone/Brixen	6535887
lengiera; lenguiera; lenguieve	Laneuville au Bois	2793410
mentinsuol; mettinbol; mettinfol; mintecoalt; mirencoalt; mitencoalt; mittewald; montisol	Mezzaselva	3173492
mersulsog; merzhofen; merzuelalag; merzuslalag	Mürzzuschlag	2705
nava de roa; la nava; naue de roa; nava; nava de rona; nava de rova; nava di rova; novederosa; novedrosa	Nava de Roa	3115689

Table 2. *Sample of reconciled place names. The column on the left includes some of the variant names that appear. We can reconcile these with high confidence to a standardized modern place name and authority file (represented by the GeoNames id) based upon their repeated appearances and similar roles across the itineraries*

Reconciliation, deduplication, and reambiguation each play an important role. Our overriding goal is *reconciliation* of each name variant that appears to an authority file, represented here by a GeoNames id and a row in our project gazetteer (Table 2). The *georeconciliation* refers to our ability to fetch spatial data (notably, the latitude and longitude values). To do so, we must often *disambiguate* by drawing connections among name variants. At its simplest, this might mean assuming the Padova that appears in FS1610B is the same as Padova that appears in GH1563. At its most complex, it can mean combining fuzzy matching, experimental mapping, and further research.

---

[21] https://www.geonames.org/.

We also must often *reambiguate* locations that have been incorrectly collapsed. Examples include the many locations named Villanova, La Venta, or Neumarkt, each a toponym which occurs frustratingly often. This may mean adding a row for which we are not yet able to make an authority file match, but we can still thereby distinguish that a given attestation is *not* the same location as another, matched value.

The eight itineraries feature over 8,800 unique location names. Using a combination of manual review, the fuzzywuzzy package in Python[22], and a custom-built reconciliation tool in Python Bottle[23], we were able to whittle these down to just over 4,500 unique locations. As shown in table 2, many of the name variants strayed far from the modern equivalents. Reading across the itineraries gave important clues for the early modern game of telephone in which Nava de Roa became «Novedrosa», for example. In doing so, we believe we have produced an additional resource for scholars interested in toponymic etymology or linguistics. Variants likely reflect how an Italian author transcribed a name as spoken by locals or other travelers.

In cases where we were not able to match data to a modern gazetteer, the nature of the itinerary still permits well-supported inference using route information about bearing and distance from prior and following identified points. We can provide approximate coordinates using this method, accompanied by a confidence metric. Because many of these locations may no longer exist or go by the same names, the EMDigIt gazetteer also contributes to the long-standing effort to add temporal bounds to geographic data.

The final data object, or *edge*, requires the most inferential approach. We use the term edge here to reflect that each route is a directed network of connected locations (nodes). A route consists of many edges, and edges might be repeated across many different routes. At first glance, the proliferation of waypoints may seem to have solved the teleportation problem of only knowing a start and end point for a given journey. Zooming in, however, reveals that the straight lines between waypoints are equally misleading, zipping imaginary travelers straight across mountaintops or through the middle of lakes.

In network terms, the itineraries give priority to the *nodes* rather than the *edge* connections, which are often left implicit. When the itineraries noted terrain hazards or travel modes, they did so in the same list as other locations. We can use ArcGIS to make more reasonable inferences that consider elevation, terrain, and contemporary borders as edge characteristics (Fig. 4). This is a slow process that often requires the consultation of additional sources, straying farther from the original source material. While this process has only been completed for modern Italy thus far, we hope to offer it as an option for other locations, as it helps to calculate more accurate distance, elevation, and slope attributes for use in determining optimal paths.

---

[22] https://pypi.org/project/fuzzywuzzy/.
[23] <https://bottlepy.org/docs/dev/>. We took inspiration from the tool designed and published by Ruth and Sebastian Ahnert at <https://github.com/tudor-networks-of-power/code>.

Figure 4. *A map represents the route network published in Franciscus Schottus' Itinerario, overo nova descrittione as published by Francesco Bolzetta in 1610. EMDigIt's data construction facilitates moving from the consideration of a full route network to considering single works or editions*

Figure 5. *All eight of the Italian Core Itineraries have been used to create this route network among locations in modern Italy. Instead of straight lines between nodes, the route lines have been further modified to reflect elevation and terrain as well as ancient and modern road networks*

3. Insights

In summary, we have introduced several potential use cases for the EMDigIt data:

–   The organizing logics of space: as a spatial tradition that predated and cross-pollinated with cartography, the itineraries reflect a "mental map" that is not otherwise easily accessible. The EMDigIt data surfaces organizing the continuities and disruptions of organizing logics; these include the choice to use one system of distance measure or another, designate a place by a given toponym, or reorganize locations into new routes, thereby elevating new locations as origins and destinations.

–   The contextualization of individual works or editions: this history of the book approach facilitates the rapid visual comparison of dense reference materials, revealing how authors and publishers have borrowed or modified material over time.

–   The enrichment of individual journeys: users seeking answers to logistical questions for a given time, place, or traveler can find contemporary travel guidance to infer where a journey may have gone and experiential details such as its timing, sights and hazards, or modes of travel.

Many of the test cases piloted by project participants involved mutually enriching the EMDigIt data and additional sources, such as *album amicorum* (signature albums), *avvisi* (early newsletters), or postal timetables. EMDigIt remains, at its core, a reference material to provide a comparison and sense of divergence from a norm: as any modern traveler knows, journeys rarely follow their Platonic ideal. EMDigIt benefits from this comparison in turn, as we can refine our calculations of an average pace for different travel modes and determining what factors beyond simple distance went into determining an optimal path. EMDigIt data can help to determine where a traveler might have encountered hazards such as brigands or rough terrain, but it is the traveler accounts that confirm these estimations.

As EMDigIt proceeds through the language core sets, we also look forward to bringing its findings to new audiences. This includes the support of a web platform, but also exploration of its potential use in other creative modes, such as storytelling, games, and augmented reality apps. We take inspiration from projects like HistoryCity (formerly Hidden Cities)[24] which uses spatial data to link travel experiences across the centuries within urban space. We will continue to involve student researchers and collaboration with partners in the U.S. and abroad, as the promise of EMDigIt remains as much in the journey as the destination.

---

[24] Fabrizio Nevola — David Rosenthal et al., *History City* Apps, <https://historycityapps.org/about/>.

Nella prima età moderna, agili ed economiche guide di viaggio realizzate da viaggiatori professionisti indicavano ai viandanti quali percorsi utilizzare, dove alloggiare, quali siti visitare, fornendo anche strumenti per orientarsi tra usanze, lingue e valute straniere. Il progetto Early Modern Digital Itineraries (EMDigIt) *ha lo scopo di trasferire le informazioni contenute nei libri di viaggio stampati dal XVI al XVIII secolo in un dataset, per consentire lo studio della mobilità in questo periodo storico. L'articolo descrive il flusso di lavoro utilizzato per la trascrizione semiautomatica, la pulizia dei dati e il collegamento dei nomi dei luoghi storici ai siti geografici moderni. Sono analizzati come caso d'uso otto libri di itinerari pubblicati in lingua italiana, il set dedicato alle guide italiane, contenenti più di quattromila punti di riferimento. Questi dati sono resi disponibili gratuitamente per l'esplorazione e l'analisi testuale, spaziale e di rete.*

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.