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

Metapolis: Spatializing Histories Through Archival Sources

Lukas Klic
I Tatti - The Harvard University Center for Italian Renaissance Studies, Italy

Abstract *Metapolis* endeavours to fill a gap within digital infrastructures tailored for the Humanities, particularly those designed to encompass the entire research continuum of geospatial historical research. While numerous digital tools exist to cater to the research requisites of scholars engaged in spatio-temporal inquiries at various levels, these tools often operate in isolation from one another, resulting in a fragmented research process. *Metapolis*, in contrast, amalgamates a suite of tools and provides essential support for humanities scholarship, thereby empowering researchers in their endeavour to reconstruct locations across temporal dimensions within a geospatial framework. Functioning as an interactive map-centric publication platform, it affords users the capacity to intermingle archival, bibliographic, and multimedia resources with interpretative research, thereby enabling their correlation and visual representation on a geographical map. Leveraging ResearchSpace, an open-source Semantic Web research environment, *Metapolis* facilitates the reuse and dissemination of Linked Open Data. Additionally, an extensive array of functionalities facilitates the enhancement of data through external knowledge repositories like VIAF, WikiData, Worldcat, and the Getty vocabularies. Conceived both as a research instrument and a publication medium, this software enables cohorts of scholars spanning a wide spectrum of humanistic disciplines to harmonize their analyses and bolster each other's discoveries through the overlaying of historical maps, interlinking them with sources to empower users in constructing a deeper understanding of the world and its historical evolution.

Keywords Metapolis. Linked Open Data. Research infrastructures. ResearchSpace. Geospatial data.

Summary 1 Introduction. – 2 Background. – 3 System Architecture. – 4 Spatializing Histories. –5 Data Sustainability. – 6 Conclusion.

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 2023-10-03 | Accepted 2024-01-18 | Published 2025-05-22

DOI 10.30687/978-88-6969-919-1//004

1 Introduction

The *Metapolis* project[1] is an initiative to develop a digital research infrastructure to support scholarship in the humanities that focuses on identifying and articulating the complex network of relationships that surround places throughout time. The architecture of the platform is designed both as a research and publication tool and allows groups of scholars from a wide range of humanistic disciplines to connect their research and interlink and augment each other's findings. While advancements brought about by the Open-Source Geospatial Foundation and other initiatives have been instrumental in advancing geospatial research, real-world applications and tooling for the humanities are still lacking. Furthermore, the various tools available to scholars that could support an end-to-end research workflow are fragmented across various pieces of software that do not communicate with one another (Waters 2023). In order to support novel and more innovative forms of inquiry into the history of urban spaces and culture, the field is in need of a map-based research infrastructure geared toward non-technical users, empowering them with the necessary set of digital tools to support innovative historical scholarship (Knowles 2008).

Metapolis builds on the advancements of the ResearchSpace initiative,[2] a collaborative research environment developed by the British Museum (Oldman, Tanase 2018; Oldman, Tanase, Santschi 2019). Various projects have demonstrated how the platform can be used as an interpretive research tool that powers the full life cycle of digital scholarly research and publication.[3] *Metapolis* can be used to map people, an idea, an art form, a philosophical movement, or any concept that can be georeferenced on a map. One could just as easily track the birth and proliferation of Jazz in New Orleans, the forcible relocation of the Cherokee Nation along the Trail of Tears or map the impact of COVID-19 on given communities. Historical census data, deeds, account books, wills, and inventories could be uploaded and georeferenced. Layers of digital media can then be interwoven to create a depth of information that is impossible to achieve with conventional tools. Most software that supports geospatial research

Metapolis is generously supported by a Level III Digital Humanities Advancement grant from the National Endowment for The Humanities. The project is a highly collaborative effort between Researchers in the Digital Humanities Lab at the Harvard Center for Italian Renaissance Studies: Ludovica Galeazzo, Remo Grillo, Lukas Klic, and Gianmarco Spinaci. This project would not be possible without this team of highly dedicated scholars.

1 https://dh.itatti.harvard.edu/#Metapolis.

2 http://researchspace.org/.

3 https://hokusai-great-picture-book-everything.researchspace.org/; https://florentinedrawings.itatti.harvard.edu/.

focuses on visualizing a limited set of data that is decoupled from these historical sources. *Metapolis* enables the publishing of these sources, together with scholarly narratives or articles alongside a map and offers innovative methods of interaction, search, and visualization. By georeferencing media such as photographs or oral histories, the historical record can be expanded to provide novel views of culture that could otherwise remain undiscovered.

ResearchSpace is an open-source collaborative Semantic Web environment designed to use and build knowledge about the world and its history. The platform is already rich with features: data extraction and integration, linking and enrichment, storage, querying, inferencing, search, visualization, and authoring, all using well-established open standards of the Semantic Web that enable reuse and repurposing in Linked Open Data environments. As an open-source system, the platform allows for enhancements from the community, which in turn can be contributed back to the source code, allowing for dissemination to a global community. The software architecture is data-centric, allowing user interfaces to be built and centered around any kind of research and collection data (Klic 2019, 31-2).

In order to accommodate complex historical reconstructions, a series of necessary customizations and upgrades to geospatial functionality have been implemented in *Metapolis*. Traditional Geographic Information Systems (GIS) such as ArcGIS[4] and QGIS[5] allow scholars to analyze and superimpose historical maps on modern Cartesian coordinates (Gregory, Geddes 2014). While these systems can be used to create visualizations of transforming urban spaces, they generally lack the ability to publish these findings in a web-based infrastructure that supports search and further interpretation by end-users (Huang, Harrie 2020). GIS systems typically employ functionality where map 'layers' serve to represent segments of time, a component that often places limitations on scholars' ability to express a more complex and dynamic urban landscape. Additionally, these systems limit scholars to representing an interpretive analysis that is the result of years of research without the ability to link back to original sources. As a result, archival documents, visual representations, and secondary literature that serve as the foundation of research become secondary and are often buried in footnotes. *Metapolis* seeks to overcome these limitations by extending a rich set of features in ResearchSpace, with functionality commonly found in GIS systems. It aims to create a platform where users can seamlessly navigate between historical sources and spatial data in ways that foster new opportunities for interpretation and research.

4 https://www.arcgis.com/index.html.

5 https://qgis.org/it/site/.

Institutions and scholars globally will have a research infrastructure capable of hosting their own geospatial projects. Interactive publications and narratives can be coupled with annotations and assertions on maps, historical documents, and iconographic sources. These semantically enriched and structured data would in turn contribute to a culture of open scholarship and collaboration among researchers, disrupting barriers posed by proprietary databases and publishing platforms where information is kept in silos. The platform would also serve as a pedagogical instrument that could be incorporated into the curriculum of courses worldwide, both in-person and in Massive Open Online Course (MOOC) environments. The nature of this machine-readable data lends itself well to playful and serendipitous discoveries, making it an attractive and engaging tool for both undergraduates and seasoned scholars alike. The project aims to advance the global paradigm shift in publishing models, away from an inward-looking, closed, and costly model towards an open and inclusive one based on collaboration and open access (Klic 2019, 24-5).

2 Background

The number of humanities projects that employ digital geospatial methodologies and tools is vast since geographic coordinates, together with dates, form part of a limited set of historical data that are computationally actionable at a larger scale. International consortiums such as the Open Source Geospatial Foundation (OSGeo)[6] have done exceptional work on this front, allowing for unprecedented interdisciplinary collaboration in the field. Groundbreaking open-source tools developed by this consortium, such as QGIS,[7] a desktop Geographical Information System, and OpenLayers,[8] a set of open-source libraries for visualizing and publishing geospatial data on the web have enabled numerous digital humanities projects to be brought to fruition. Other tools, like Neatline,[9] have been integrated into the open-source digital publishing software such as Omeka,[10] enabling for the visualization of humanities data to be published in web environments in productive ways. These have even resulted in digital publications such as *The Chinese Deathscape*,[11] a digital volume from Stanford University Press that explores grave relocation and burial reform in

6 https://www.osgeo.org/.
7 https://www.osgeo.org/projects/qgis/.
8 https://openlayers.org/.
9 http://www.neatline.org/about/.
10 https://omeka.org/.
11 https://chinesedeathscape.org/.

China. Projects such as *Digimap*[12] have allowed for the publication of stratified historical and contemporary maps to facilitate scholarly analysis. Research projects like *DECIMA*[13] (Terpstra, Rose 2016), *Visualizing Venice/Visualizing Cities*[14] (Huffman, Giordano, Bruzelius 2018) and the *Venice Time Machine*[15] (Abbott 2017) have sought to reconstruct the history of urban landscapes at scale.

All these research and infrastructure projects have used desktop GIS systems, either ArcGIS or QGIS, to perform the complex historical analysis of these spaces. By georeferencing historical maps over modern-day ones and cross-referencing historical documents, researchers were able to determine the history of political boundaries, the transformation of urban spaces, and the movement of objects, actors, or goods over space and time in novel ways that are often impossible in the analog form. Notwithstanding the gamut of digital tools at the disposal of scholars, the output of this often-painstaking research has generally been a print publication that does not provide access to the source data. Publications in the digital form have not allowed users to search for historical data and sources within a particular space and time or visualize the results on a map in meaningful ways that can facilitate further interpretation (Gaehtgens 2013). Additionally, these publications often employ flat data structures that do not do justice to the rich scholarship behind the complex network of relationships that compose the publication.

Historical reconstructions require the ability to visualize and search for historical features, generally in the form of polygons and lines, that can take the form of buildings, outdoor spaces or political boundaries. Although Omeka does offer tools for data input and visualization on a map, it is not able to author or visualize anything other than individual points or allow for geospatial search. Tools such as GeoNetwork[16] can manage complex geospatial features but are built on data management tools that do not adapt well to humanities research and are often used for managing sets of geospatial data rather than publishing them in an intuitive user interface. GeoBlacklight[17] allows for the publishing of geospatial datasets and has limited visualization options but is more geared towards providing raw data that the user can download and reuse in other contexts, rather than

12 https://digimap.edina.ac.uk/.

13 https://decima-map.net/.

14 https://www.beniculturali.unipd.it/www/ricerca/linee-di-ricerca/visualizing-venice-exploring-the-citys-past/.

15 https://www.epfl.ch/research/domains/venice-time-machine/.

16 https://geonetwork-opensource.org/.

17 https://geoblacklight.org/.

serving as a tool for scholars to author and interact with data. Carto[18] offers powerful visualization tools that can be pre-configured by the author to visualize a single idea but does not allow for subsequent reconfiguration and interpretation by end-users, nor has any data management functionality. The *Metapolis* project has evolved out of years of experimentation with these existing tools and hopes to fill these gaps by extending existing work in the field and building on top of it rather than in parallel.

3 System Architecture

The ability to represent artistic production, historical documentation and urban change over time is instrumental for research projects that tell the complex histories of the urban, architectural, cultural, and socio-economic dynamics of early modern cities, as well as of the actors and objects pertaining to these spaces. The *Metapolis* project is rooted in an iterative expansion of the ResearchSpace infrastructure in order to be able to represent and publish these histories.

After several years of successfully working within the ResearchSpace ecosystem, researchers in the Digital Humanities Lab at I Tatti, The Harvard University Center for Italian Renaissance Studies, began planning for a more ambitious set of tools to incorporate geospatial mapping functionality. This work has largely been informed by a collaboration with the ERC project *Venice's Nissology – VeNiss*, led by PI Ludovica Galeazzo (2022),[19] and her previous project *Archipelago* (2021).[20] This research aims to digitally reconstruct the history of a cluster of sixty islands scattered across the Venetian lagoon from the late fifteenth century onwards. Both these projects originally employed a relational database, FileMaker Pro, to collect and store all historical data and sources, alongside a geographic information system, ArcGIS or QGIS, for mapping and geocoding urban data. While geodatabases store spatial data, including digital reconstructions of the ancient landscapes represented through points, lines, and polygons and provided mapping functionality, they are never directly connected to relational databases through any integration. The research questions behind the *VeNiss* project made working with this software configuration exceedingly

18 https://carto.com/.

19 The project *Venice's Nissology. Reframing the Lagoon City as an Archipelago: A Model for Spatial and Temporal Urban Analysis (16th-21st centuries)* has been funded with a five-year grant by the European Research Council (ERC-2021-StG, n. 101040474). https://www.unipd.it/en/sites/en.unipd.it/files/Galeazzo_VeNISS.pdf.

20 https://dh.itatti.harvard.edu/#Archipelago.

cumbersome and insufficient. One could not, for example, visualize textual or iconographic sources describing historical urban places with digital representations of their ancient configuration, or connect certain geographical modifications to their related urban events and agents.

The software architecture of *VeNiss* echoes the requirements of many other projects in the field: a necessity to integrate geospatial data with historical sources in the form of structured data, and to provide a research and publishing platform that seamlessly integrates these two in a web-based environment made available to non-technical users. Following this extensive analysis of existing software, researchers at the Digital Humanities Lab together with numerous other stakeholders determined that it would be impractical to attempt to recreate a single web-based tool that could match the robust functionality of desktop software such as ArcGIS or QGIS. They decided that the most effective path forward would be to extend the rich data management and publishing features of ResearchSpace with functionality to support geospatial data, search, annotation, and analysis. A seamless software integration with QGIS would then allow scholars to perform the complex work of geocoding and drawing in the more powerful desktop software. Beginning with the research questions, the team then started to map out the architecture of the platform. Before working on any software development, the group spent three months designing extensive mock-ups that laid out the full user interaction and design of the platform, in order to ensure a user experience that was intuitive and productive for non-technical users **[figs 1-2]**.

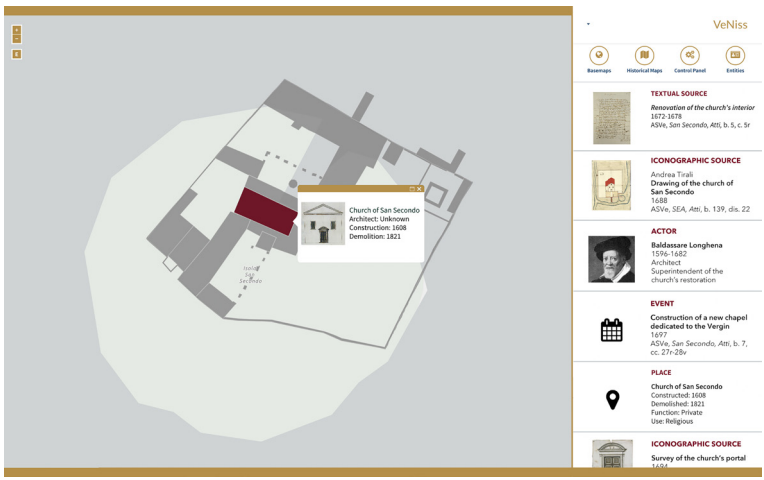

Figure 1 Feature popover opens when clicked

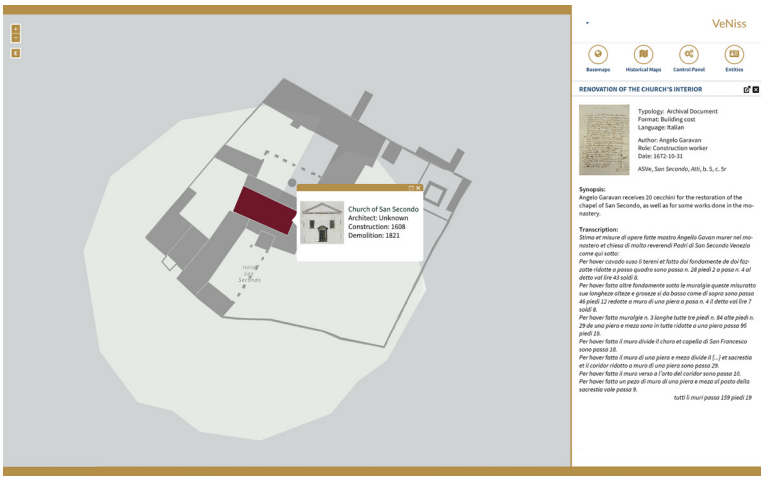

Figure 2 Navigator overlay with specific information of the selected entity

In order to ensure the overall technical feasibility of the project architecture, the team also began developing one critical software component that would allow for the communication between the relational geospatial database back end of QGIS, with the graph database back end of ResearchSpace (Blazegraph).[21] Following months of software development, this critical first step was successfully completed in November of 2021, forming the backbone for integrating GIS-based systems with a Linked Open Data research platform capable of connecting, enriching, and publishing archival data in a geospatial research environment.

4 Spatializing Histories

An extensible architecture for the geospatial mapping components is critical to support a range of research questions into the history of a given space. These components allow for interaction between the map and existing data management and publishing features of ResearchSpace.

Functionality to be able to change the basemap tiles and mapping providers allows scholars to choose a map background that best serves the needs of their individual research. This will permit them to use a wide range of map tile providers: Mapbox, MapTiler, ArcGis, or Open Street Maps. Georeferenced images can also be uploaded, superimposed, and used as a reference to draw geospatial features

21 https://blazegraph.com/.

on the maps in the form of points and polygons, as well as annotation functionality that enables researchers to connect these features to data in the knowledge graph. Visualization tools include the ability to add any number of historical maps as separate layers **[fig. 3]**.

Figure 3 Georeferenced historical maps

Functionality to set transparency levels on these layers that allow researchers to compare and contrast them, together with 'swipe'[22] and 'eyeglass'[23] functionality, allows these layers to be compared and contrasted at both micro and macro levels. The styling (through colors and text labels) of geospatial features based on type, allows users to visualize their historical functions and use. They also allow for the visualization of the movement of objects or goods across space-time, enabling researchers to map their displacement. Most importantly, a timeline slider allows users to limit their visualization of geospatial features and archival sources to a particular point in time. Non-technical users are able to upload images of archival documents to the platform and publish them through an International Image Interoperability Framework (IIIF)[24] server embedded in ResearchSpace. Web forms allow researchers to transcribe data from these sources, publishing them as Linked Open Data. These data can be properly attributed to multiple researchers and can be coupled with textual narratives that provide additional

22 https://openlayers.org/en/latest/examples/layer-swipe.html.

23 https://openlayers.org/en/latest/examples/layer-spy.html.

24 https://iiif.io/.

levels of interpretation. A seamless integration with external knowledge bases allows users to query the title, author, ISBN, or DOI of a particular book or article and seamlessly cite secondary literature.

Figure 4 Dynamic feature styling and labeling

Robust search and faceting functionality permit any user to fine-tune the visualization of archival sources and geospatial features based on their own research interests **[fig. 4]**. The ability to annotate existing entities and text within a web-based editor, allows for the cross-pollination of geospatial data, textual narratives, and data from archival sources to enable a cohesive visualization of these histories.

5 Data Sustainability

Libraries and archives have come to accept the fact that while the underlying software that supports a digital project is inherently ephemeral, the data and research that go into these projects need to be portable from one system to another. SQL structures where the metadata model is built into the database or software are difficult to interpret without the underlying software. By employing graph structures, data is stored in the Resource Description Framework (RDF) model,[25] ensuring greater portability to other systems in the future. Because the ontology that describes both the meaning and context of the data is built into the model itself, it can easily be ported to any other graph data

25 https://www.w3.org/TR/rdf-schema/.

structure that supports RDF, and a user interface can be developed on top. Well-documented ontologies like the Conceptual Reference Model (CIDOC CRM)[26] allow for an extremely detailed description of data, its meaning, and context. This approach is extremely attractive to libraries and archives as the raw data could be retrieved hundreds of years into the future and one would be able to understand and decipher its meaning without any kind of specialized software.

6 Conclusion

The *Metapolis* project, through its integration of advanced geospatial capabilities with the open-source ResearchSpace infrastructure, marks a significant leap forward in the field of digital humanities. Our work has addressed existing gaps in the availability of comprehensive, user-friendly digital tools that enable in-depth research and interactive publishing in humanities disciplines. *Metapolis* is designed to move beyond the limitations of traditional Geographic Information Systems, empowering scholars to seamlessly explore the interplay between historical data and spatial phenomena. It brings the capacity for novel historical inquiry to the forefront by incorporating archival documents, visual representations, and secondary literature directly into its spatial framework, rather than relegating these crucial sources to footnotes or separate databases.

Building on the established functionalities of ResearchSpace, *Metapolis* delivers a user-centric interface that enables not only the visualization but also the dynamic interpretation of spatial-temporal data. By allowing users to integrate georeferenced media and link these to the underlying sources, *Metapolis* expands the historical record and provides fresh perspectives on cultural patterns and developments that might otherwise remain hidden.

The project source code is open-source and freely available on GitHub, allowing for contributions from a global community.

Finally, *Metapolis* aims to contribute to the global shift in publishing models by promoting a collaborative and open-access approach. This departure from traditional proprietary databases and publishing platforms allows for a more transparent and integrated scholarship, paving new paths for scholarship working with historical reconstructions.

The project encapsulates an ambitious vision for the future of digital humanities research and education. By providing a familiar Google Maps-like interface to users with a time slider and direct access to archival sources, *Metapolis* allows new paths of scholarly

26 https://www.cidoc-crm.org/.

inquiry to be tested and defined. As we continue to refine and expand this platform, we anticipate that it will stimulate innovative scholarship, foster global collaboration, and reshape our understanding of the world's cultural and historical landscapes.

Bibliography

Abbott, A. (2017). "The 'Time Machine' Reconstructing Ancient Venice's Social Networks". *Nature News*, 546(7658), 341.
https://doi.org/10.1038/546341a

Gaehtgens, Th.W. (2013). "Thoughts on the Digital Future of the Humanities and Art History". *Visual Resources*, 29(1-2), 22-5.
https://doi.org/10.1080/01973762.2013.761110

Galeazzo, L. (2021). "Autorità ecclesiastica e civile nell'iconografia dell'arcipelago veneziano tra XVI e XVII secolo". *In_Bo*: *Dominio del Sacro. Immagine, cartografia, conoscenza della città dopo il Concilio di Trento*, 12(16), 186-97.

Galeazzo, L. (2022). "Analysing Urban Dynamics in Historic Settlements Using a Geo-spatial Infrastructure. The Venice's Nissology Project". *Journal of Art Historiography*, 27, 1-13.

Gregory, I.N.; Geddes, A. (eds) (2014). *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington: Indiana University Press.

Huang, W.; Harrie, L. (2020). "Towards Knowledge-Based Geovisualisation Using Semantic Web Technologies: A Knowledge Representation". *International Journal of Digital Earth*, 13(9), 976-97.
https://doi.org/10.1080/17538947.2019.1604835

Huffman, K.L.; Giordano, A.; Bruzelius, C. (eds) (2018). *Visualizing Venice: Mapping and Modeling Time and Change in a City*. London: Routledge. Routledge Research in Digital Humanities.

Klic, L. (2019). *Digital Publishing and Research Infrastructure for Cultural Heritage: an Institutional Roadmap* [PhD dissertation]. Venice: Ca' Foscari University of Venice.
http://hdl.handle.net/10579/15587

Knowles, A.K. (ed.) (2008). *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Redlands (CA): ESRI Press.
https://doi.org/10.14714/cp63.160

Oldman, D.; Tanase, D. (2018). "Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace". Vrandečić, D. et al. (eds), *The Semantic Web – ISWC 2018*. Cham: Springer, 325-40. Lecture Notes in Computer Science.
https://doi.org/10.1007/978-3-030-00668-6_20

Oldman, D.; Tanase, D.; Santschi, S. (2019). "The Problem of Distance. A ResearchSpace Case Study on Sequencing Hokusai Print Impressions to Form a Human Curated Network of Knowledge". *International Journal for Digital Art History*, 4, 5.29-5.45.

Terpstra, N.; Rose, C. (eds) (2016). *Mapping Space, Sense, and Movement in Florence GIS and the Early Modern City*. London: Routledge.

Waters, D.J. (2023). "The Emerging Digital Infrastructure for Research in the Humanities". *International Journal on Digital Libraries*, 24(2), 87-102.
https://doi.org/10.1007/s00799-022-00332-3