Narrating through data: life and relationships of a XIVth-century's Tuscan merchant

Emiliano Degl'Innocenti
ORCID: 0000-0002-3839-9024
Contact: emiliano.deglinnocenti@cnr.it

Alessia Spadi
ORCID: 0000-0002-9670-3426
Contact: spadi@ovi.cnr.it

Federica Spinelli
ORCID: 0000-0002-2195-3930
Contact: spinelli@ovi.cnr.it
Consiglio Nazionale delle Ricerche, Istituto Opera del Vocabolario Italiano, Firenze, Italy

This contribution will present the work carried out by the DARIAH[1] research team at the OVI-CNR[2] in close collaboration with research communities, domain experts, and citizen scientists, to shape the development of an interoperable digital ecosystem for research in the Humanities and Cultural Heritage, populated with data, tools, and services.

In the context of sharing knowledge, narratives have long played an important role, especially since the advent of the Semantic Web, when large amounts of structured and interconnected data – related to various kinds of resources in different scientific domains – have become freely accessible. Cultural heritage institutions started producing a great deal of data, together with relations, from their resources, enabling connections in a linked open data[3] context. However, data without context cannot tell any story and yield results that are neither understandable nor interesting; in fact, both theoretical and domain knowledge of the resources, plus advanced data analysis techniques, are required to create a compelling and profound narrative.

The workflow presented in this contribution serves to explain the approach followed by our team to manage (i.e. ingest, map, model, enrich, publish) data coming from multiple disciplinary

[1] Digital Research Infrastructure for the Arts and Humanities: http://dariah.cnr.it/.
[2] Istituto Opera del Vocabolario Italiano of the National Research Council of Italy: http://www.ovi.cnr.it/.
[3] Linked Open Data are the sources from which data scientists can extract the relevant knowledge to engineer stories and create visual representations of copious amounts of information to be presented to the public.

contexts and populate a highly integrated research environment.[4] The workflow is meant to manage the whole digital lifecycle of the resources provided by the project partners, starting from the ingestion (direct upload to the platform, harvesting, etc.), encompassing the mapping (syntactic, conceptual) and data normalization phases – managed with custom scripts – and ending with the data modeling (from an internal custom schema, to the CIDOC-CRM subset we used) and data publication. For this last step a dedicated publication framework (totally independent from the data) was developed in order to: a) represent the various needs of the project's stakeholders (Archival view, Works of Art view, Philological View); and b) represent the resources within their spatial and temporal development (provenance, age determination, etc.), using timelines and geographic metaphors. During the implementation of the RESTORE project, the DARIAH.it team aimed to both develop best practices and contents supporting the innovative use of historical documentation in a multidisciplinary environment, promoting understanding and encouraging its re-use by the community. The data management methods implemented make it possible to enable collections of heterogeneous data to be integrated and used for storytelling.[5]

To demonstrate the methodology used, this paper will focus on the resources within the RESTORE platform,[6] initially developed as pilot project within the SSHOC Cluster, now available also on the SSHOC Open Marketplace[7] and further developed as a set of tools and technologies to support data collection and integration in the Humanities and Heritage Studies within the IPERION-HS project.[8] The RESTORE dataspace is populated with objects related to Francesco Datini and his entourage's lives and commercial activities, in particular: letters and other documents held by the Archivio di Stato in Prato (digital images, metadata, and archival descriptions in EAD format); transcriptions of the letters and documents' contents, with philological and lexicographic annotations, provided by the CNR-OVI (with custom encoding formats and TEI headers); digital representations of works of art commissioned or owned by Datini and/or his family, preserved by the Museo di Palazzo Pretorio in Prato (images, metadata, and descriptions encoded with Italian standards: ICCD-OA xml files).

The project's main purpose was the recovery, integration, and accessibility of data and digital objects collected by partners in order to build a knowledge base made of information regarding the history of the city of Prato, its civic institutions, the development of its economic and entrepreneurial system, the role of women in the development of a welfare state and network. The datasets collected for the RESTORE project were analyzed, treated, and integrated for the sake of shedding light on the

[4] The workflow implemented for the semantization of the datasets, realized with open access tools, allow for transformation from different file formats, in order to foster data integration, going from the acquisition of materials organized by each domain standards (EAD - EAC, ICCD, TEI) to their transformation in the CIDOC-CRM ontology.
[5] In particular, the platform implemented provides open access to a unique data lake, where both the original resources and their triples (RDF) can be navigated, visualized, and, if needed, downloaded.
[6] http://restore.ovi.cnr.it/
[7] https://marketplace.sshopencloud.eu/
[8] https://www.iperionhs.eu/

world of Francesco di Marco Datini – a very influential merchant from Prato who lived in the XIVth century – his social network and the relationships established within his entourage. Among the most promising results of RESTORE is the possibility to track over time and space the activities of a wide network of people related to the same professional milieu (merchants), allowing researchers to start with a prosopographic focus (e.g. Italian merchants in southern Europe and northern Africa) but to narrow or restrict their research to specific aspects (e.g. welfare and devotion in XIVth-century Prato), accessing several datasets (images, transcriptions) with different levels of detail (e.g. onomastics of people and places, specialized lexica, etc.), providing different points of view (descriptions) to describe the same objects, people, phenomena, and supporting various research scenarios (philology, conservation, archival and library science, history of ideas, people and places, citizen science).

Starting from a local history approach (i.e. the Datini family and the city of Prato), it is nonetheless possible to broaden the focus from the local dimension to reconstruct a significant part of the commercial and economical history of European and Mediterranean cities of the XIVth century.

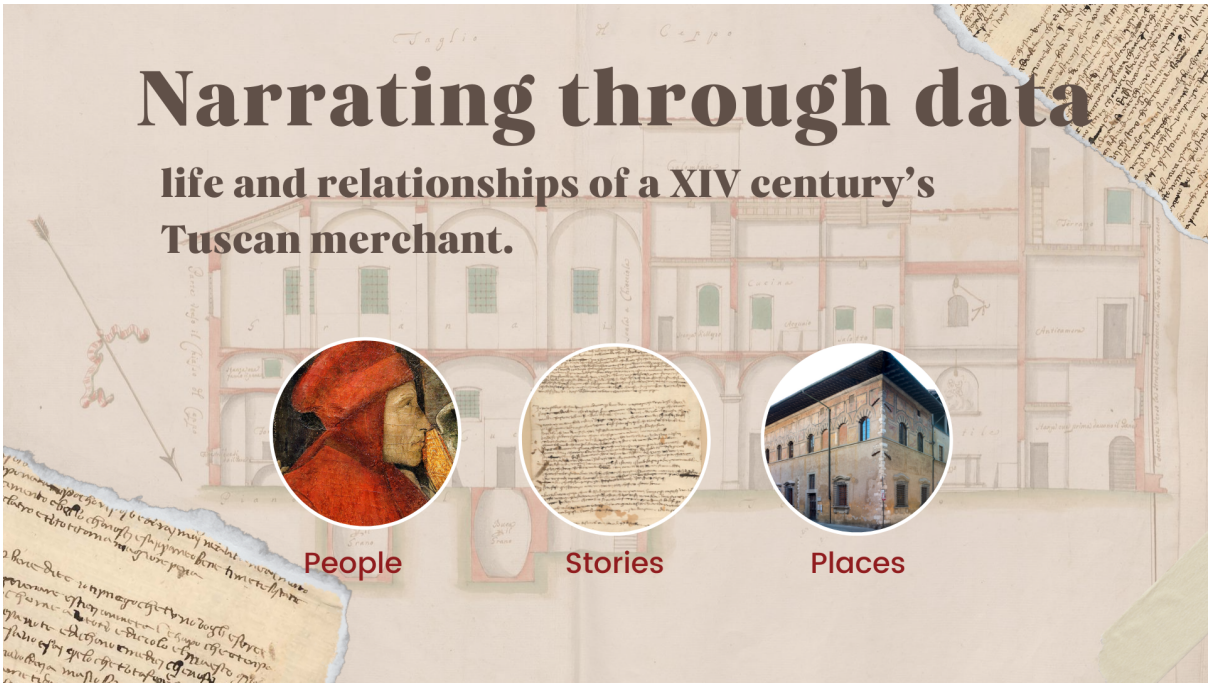

Fig. 1