DARIAH.it: Data Integration Strategies and Solutions for Digital-Resources Management and Research in the Arts and Humanities

Emiliano Degl'Innocenti, Carmen Di Meo, Alessia Spadi

> "[...] the astounding growth that our resources have undergone in terms of their precision and adaptability will in the near future confront us with very radical changes indeed in the ancient industry of the beautiful. In all arts there is a physical component that cannot continue to be considered and treated in the same way as before; no longer can it escape the effects of modern knowledge and modern practice. [...] We must be prepared for such profound changes to alter the entire technological aspects of the arts…"
>
> Paul Valéry, *Pièces sur l'art*[1] (1934) (Benjamin, Foreword, 1)

Introduction

The role of researchers has evolved in tandem with the availability of huge data sources, necessitating increased interoperability as research domains expanded. This shift mirrors the emergence of data-intensive research, initially rooted in pure sciences and now permeating traditional humanistic research paradigms. The evolution of data analysis from the 1960s to contemporary big-data analysis exemplifies a transition from mere data aggregation to knowledge extraction. The advent of the internet in the late twentieth century facilitated an unprecedented scale of information production by individuals, leading to challenges in validating authoritative information in the digital deluge. This prompted a paradigm shift toward data-driven methodologies and the emergence of data science as a significant field (Hilbert 2012).

In this context, research infrastructures[2] emerged as essential mecha-

---

[1] Benjamin, Walter. *The Work of Art in the Age of Mechanical Reproduction*. Translated by J. A. Underwood, Penguin Adult, 2008.

[2] The definition of *research infrastructure* was given in Article 2 (6) of the Regulation (EU) No. 1291/2013 of December 11, 2013. *Establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014- 2020)*: "[r]esearch infrastructures mean facilities, resourc-

nisms to consolidate resources within the digital humanities (DH) ecosystems, fostering interoperability and mitigating technological and knowledge islands. However, sustainability remains a critical concern, requiring robust governance models and financial support to guarantee and maintain accessibility with adherence to standards. Infrastructures are central to the digital transition in research methodologies, becoming integral to researchers' work. The vast amount of available digital resources as well as the use of advanced technological tools and methodologies – including ontologies, triplestores, and AI techniques such as deep learning and machine learning – contributed to realizing the digital transition in the SSH domain. Big-data analysis, although mainly associated with voluminous and rapid data, holds significance in DH for its potential to analyze comprehensive datasets, avoiding reliance on statistical samples.

As a consequence, research infrastructures (RIs) have transformed significantly over time, extending far beyond the traditional scientific realms to encompass a multiple range of objects, platforms, and services. In the landscape of DH, these infrastructures play a pivotal role in supporting multifaceted research, facilitating data-driven methodologies, and enabling the transition from data collection to knowledge inference.

RIs have undergone a profound evolution in recent years, expanding their scope beyond the borders of traditional scientific domains and they have emerged as advanced digital environments supporting researchers in various research endeavors. The characterization of infrastructures is based on: (1) accessibility, (2) interoperability, and (3) facilitating rather than conducting research activities. These infrastructures facilitate access to resources, support data-science environments, and accommodate diverse research domains and workflows. In this context, the Digital Research Infrastructure for the Arts and Humanities (DARIAH) emerges as an ESFRI Landmark for the Social and Cultural Innovation (S&CI) sector. DARIAH supports researchers in the diverse fields of arts and humanities. This network, composed of human expertise, knowledge bases, content, methodologies, and tools, empowers researchers to engage with digital or

es and services that are used by the research communities to conduct research and foster innovation in their fields. Where relevant, they may be used beyond research, for example for education or public services. They include major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives of scientific data; e-infrastructures such as data and computing systems and communication networks; and any other infrastructure of a unique nature essential to achieving excellence in research and innovation. Such infrastructures may be single-sited, virtual or distributed."

hybrid resources through processes of building, analysis, and interpretation. Consequently, DARIAH fosters the sustainable development of research and teaching endeavors that leverage digital technologies through a network that aggregates the combined expertise and resources of its member countries.

This article will focus on the research and implementation work done by the group of researchers gathered around the DARIAH-IT node, based in Florence at the Istituto Opera del Vocabolario Italiano of the National Research Council of Italy (OVI-CNR) that contributed to several national- and EU-funded projects, such as the SSHOC EOSC thematic cluster (Social Sciences & Humanities Open Cloud) and the IPERION-HS (Integrated Platform for the European Research Infrastructure on Heritage Science) Horizon 2020 project. This research group has developed in the context of the Italian Roadmap for the development of the national node of the DARIAH-ERIC pan-European infrastructure, which received a national funding within the framework called the National Operational Program (PON) on Research and Innovation 2014-2020 for the DARIAH.it project (that is, *Developing National and Regional Infrastructural Nodes of DARIAH in Italy*). Very recently, this national infrastructure has been funded by the Italian National Recovery and Resilience Plan (for instance, the Humanities and Heritage Italian Open Science Cloud – H2IOSC federated infrastructure) to create a network in which Italian nodes of four research infrastructures in the ESFRI domain of Social and Cultural Innovation (S&CI) will participate: DARIAH.it, CLARIN.it, OPERAS.it, E-RIHS.it.

Recently, DARIAH.it focused on fostering interoperability and access provision to digital resources related to tangible (objects) and intangible (intellectual and conceptual items) aspects of cultural heritage held by cultural and/or memory institutions (such as galleries, libraries, archives, museums, and research institutes) based in Italy, as well as on promoting training and capacity-building activities (such as doctoral courses, data and metadata stewardship, or curation tasks). Working along with established research communities, the DARIAH.it team at OVI-CNR has collaborated with domain experts and citizen scientists to shape the development of an interoperable digital ecosystem for the research in the humanities and cultural heritage, populated with data, tools and services.

Landscaping Core Research Services and Facilities in the SSH Sector in Italy: The DARIAH.it Experience

DARIAH was conceived as a distributed research infrastructure supporting digitally enabled research and teaching for arts and humanities schol-

ars throughout EU countries. Entered into the ESFRI Roadmap in 2006, DARIAH was established as a European research infrastructure consortium (ERIC) in 2014, and was awarded landmark status in 2016 as a pan-European research infrastructure of scientific excellence. DARIAH-EU recently includes twenty-two member states, among which Italy is one of the founding members, and in 2017, DARIAH.it headquarters were established at the OVI Institute of the Italian National Research Council (Istituto Opera del Vocabolario Italiano – Consiglio Nazionale delle Ricerche) in Florence.

DARIAH-EU is, first and foremost, a network of services and researchers from different member countries. One of the most important missions informing by the DARIAH-EU legal statutes is ensuring that best practices, and methodological and technical standards, are followed by each country and each targeted participating research community. The last point is highlighted here, for it stands right at the heart of the approach in designing research projects and contributions realized by the Italian node of the research infrastructure, in the development and implementation of the digital solutions provided to the main RI and other research communities.

As a concrete result of this approach, the multiple initiatives that will be presented here are characterized by a variety of ventures and by cohesion and adhesion to the practices, policy, and standards (either technical, cataloguing, conceptual) shared in the ERA (European Research Area) and developed by means of a common effort with other EU researchers and RPO and RFO (research performing/funding organizations). With regard to the objective of favoring a possible boost to the availability and use of digital solutions and facilities in the SSH research sector, one of the first and most important initiatives programmed and then financed by the Italian Government (MUR - Ministero dell'Università e della Ricerca) as a main promoter the DARIAH.it node, was the PON (Piano Operativo Nazionale – National Operational Program on Research and Innovation 2014-2020), named *Developing nAtional and Regional Infrastructural nodes of dAriaH in ITaly (DARIAH.it)*.[3]

This project, coordinated by the OVI-CNR Institute in the person of Dr. Emiliano Degl'Innocenti, who is also the DARIAH.it national coordinator, was awarded funding of 13,469,000 euros, with the mission of (1)

---

[3] PIR01_00022. Ministero dell'istruzione, dell'università e della ricerca, decreto 9 maggio 2019 Ammissione del progetto "DARIAH.it - Developing National and Regional Infrastructural nodes of DARIAH in Italy" al finanziamento previsto dal decreto direttoriale 28 febbraio 2018. (Decreto n. 900/2019). (19A04611) (GU Serie Generale n.165 del 16-07-2019).

mapping the needs of the SSH research communities in Italy in terms of digital and technological advance and innovation and (2) building a network of data centers across Italy, with six regional nodes (data centers) in Toscana, Lazio, Puglia, Campania, and Sicilia. The new arrangement was seen as a necessary step to support the transition to open-science and data-driven science, setting the stage for the implementation and use of HPC and HTC, big-data storage, computation on knowledge graphs, data visualization, simulation and game programming, 3D rendering, language analysis, DL, ML, and AI.

The six nodes have been located as follows: Two main nodes, performing HPC, big-data management, computation on knowledge graphs, data visualization, simulation and game programming, are based in Lecce and Naples; a second level, thematic node, was set up in Catania with three mirror nodes and supporting functionalities in Pisa, Rome, and Florence.

The National Operational Program funding, from which the DARIAH.it node benefited, was the instrument established by Italy to contribute to the improvement of the quality of higher education and the enhancement of research, technological development, and innovation achieving the objectives fixed by the EU cohesion policy and strategy for member countries' disadvantaged areas, the European Regional Development Fund (ERDF), and the European Social Fund (ESF). Managed by the Ministry of Universities and Research (MUR), formerly Ministry of Education, Universities, and Research (MIUR), the program focused especially on transitional regions (Abruzzo, Molise, and Sardinia) and less developed regions (Basilicata, Calabria, Campania, Puglia, and Sicilia), with a comprehensive budget of 1.189 million euro. The PON on Research and Innovation is developing in line with the strategic objectives of the EU Horizon 2020 work programs, in synergy with the National Smart Specialization Strategy (NSSS) and the Regional Operational Program (POR).[4]

Mapping Needs in Digital Research Communities: Some Premises

According to the ESFRI Roadmap 2021, innovation in the social sciences and humanities domain must bring, in addition to the long-lasting scientific impact expected from the creation and access to services and facilities provided by research infrastructures to data communities of researchers, a core societal impact – created by scholars while contributing to the shaping

---

[4] For further information on the PON program, visit http://www.ponricerca.gov.it/pon-ricerca/programme/.

of public policies, and to the realization of public engagement activities, fostering citizens' understanding of culture and cultural heritage in societies, by creating resources and tools for "handling citizens' needs, and [...] creating collaboration among multiple stakeholders."

> The ability to do high value SSH research increases exponentially if such information and data is more easily available from a single point or at least a smaller number of points even if they arise from different sources, different countries and different points in time." (ESFRI Roadmap 2021, 106-107.)

This statement directly poses the issue of those identified as "emerging drivers" in the SSH research field, such as the necessity of being able to manage as smoothly as possible an increasing scale of digitized material, that is, big-data management, which is also crucial in SSH institutions and repositories (Wilkinson et al. 2016). In the authors' view, this very important theme can be linked to the discussion in a scientific contribution devoted to analyzing this new landscape entitled *Cultural Heritage Infrastructures in Digital Humanities* (Benardou 2018). According to observations made in this book, the point is to get from quantitative research to qualitative research, which must be data-driven and community-specific. Technologies, tools, digital services, and devices have to be developed in line with the contribution of the partner institutions and the direct involvement of domain experts and scholars.

With regard to the societal impact mentioned above, SSH research and, in our case, arts and humanities objects and their digitization, can be seen as valuable means of:

1. Strengthening the sense of community, by reinforcing the knowledge of the history of cultural city centers and their communities;
2. Strengthening and fostering digital literacy;
3. Supporting scientific careers and employment;
4. Bridging the gap through fields of study and making learning in SSH a competitive research ground, in line with natural sciences;
5. Promoting computationally engaged research, teaching, and publishing;
6. Sustaining valorization of the linguistic element and variety across culture and epochs;
7. Contributing to building more sustainable cities, in their physical environments and ecosystems, in which data centers and storage are shared by data communities.

The data needs we identified, in line with our own research experience, are as follows:

1. Making use of open-source applications;
2. Gathering together data across disciplines;
3. Distinguishing between the tangible and the intangible aspects of a cultural object;
4. Mapping and modeling of data;
5. Representing data semantically;
6. Managing data visualization and analysis;
7. Processing automatization in text/image visualization and analysis;
8. Sharing the acquired knowledge;
9. Planning for sustainability;
10. Maintaining and finding data repositories that meet the technical or legal requirements of owned research data;
11. Working with interoperable frameworks and workflows;
12. Developing not-too-narrow tools (and licensing);
13. Setting an endpoint to other researchers and research institutions;
14. Offering services to communities of scholars;
15. Being represented in marketplaces;
16. Practicing research data management, data ownership, and data stewardship, according to shared standards and best practices.

Data modeling and the need to represent semantically the information to be conveyed impose a deep reflection on the consistency of the information to be represented digitally. With these reflections and approaches in mind, the OVI-CNR research team developed an interest and a scope to reach out to and to engage with the extensive GLAM communities of researchers, through the design and development of a heterogeneous data-integration suite, so as to foster interoperability between data produced by either historians, art historians, archivists, librarians, and conservators, scientists, and analysts, to create a "common language" for semantic data integration.

Case Study: The RESTORE Project

Given all the previous considerations, the DARIAH.it team worked on a data pilot, the RESTORE (smaRt accESs TO digital heRitage and mEmory) project, co-financed by the Regione Toscana and started in June 2020. The project, available at http://restore.ovi.cnr.it/, had as its main

purpose the recovery, integration, and accessibility of data and of digital objects produced by partner institutions, the Archivio di Stato (Archivio di Stato di Prato, n.d.) and the Museo di Palazzo Pretorio (Museo di Palazzo Pretorio, n.d.) of the city of Prato, the OVI - Istituto Opera del Vocabolario Italiano of the Italian Research Council that coordinated the project consortium, the Archival and Bibliographic Superintendency of Tuscany (Soprintendenza Archivistica e Bibliografica della Toscana, n.d.) and the *SPACE SpA* software company.

The Archivio di Stato di Prato holds an extensive collection of sources from various institutions including civil, religious, and charity organizations. When combined with archives from other civic entities like the Museum of Palazzo Pretorio, it offers a comprehensive view of the urban community, territory, and society from the fourteenth century to the present. Notably, the Fondo Datini[5] and Ospedale Misericordia e Dolce[6] collections have been digitally cataloged, featuring archival descriptions, digital document images, text transcriptions, and a lemmatized text corpus courtesy of the OVI-CNR Institute. These resources, created through disparate projects and activities, lack the added value of integrated information from various actors working on the same cultural heritage reference.

All of these custodian actors play a key role in preserving a significant part of Prato city's cultural and artistic heritage. The most consistent part of data provided comes from the Ospedale Misericordia e Dolce of Prato and regards cultural objects created expressly for or acquired by Prato families or religious bodies, subsequently merged into various collection of documents, works of art, and textual corpora. In collaboration with the partners, the team recovered data that could have been lost, because of an obsolescent cataloging format, and converted them to a new standard to make accessible the information they conveyed: Data records have been FAIRified (GoFair, n.d.), made semantically interoperable for human readability, and integrated with information collected from diverse contexts. This resulted in a knowledge base concerning the history of

---

[5] The archive of administrative documents and correspondence of the merchant Francesco di Marco Datini (1335-1410) bears witness, through his extensive activities in the industrial, commercial, and banking fields, to the economy and social life throughout the entire Mediterranean region.

[6] Charitable institution that since the thirteenth century has been taking care of travelers, the poor, and abandoned children. The digital resources related to this collection can be accessed on the archive's Web site: http://www.archiviodistato.prato.it/accedi-e-con-sulta/aspoSt005/tree.

the city of Prato and of its institutions, through which is possible also to reconstruct the development of its economic and entrepreneurial ecosystem, starting from the figure of the merchant Francesco di Marco Datini, and the central role played by women in the management and creation of business exchanges and networks and in promoting the city itself.

The critical approach implemented to address data modeling and presentation challenges for cultural-heritage objects addresses inconsistent data formats and lack of interoperability between datasets from different institutions. The RESTORE project adopted a multi-faceted approach to data management and integration. First, they ensured consistency by converting data from disparate sources into a common format based on the CIDOC CRM ontology. Furthermore, vocabularies and thesauri were employed to establish a unified language for describing objects and their characteristics. This facilitated the linking of related information across datasets and eliminated redundancy through a process of aligning information from diverse sources. Finally, a critical review process was applied, particularly to provenance data associated with artworks. This critical analysis enabled differentiation between objects directly commissioned by an institution and those simply housed within its collections.

The project's architecture is modular, employing customized or existing components supporting domain standards (Wisser, 2011).[7] To achieve a shared data model, the project followed a workflow involving data transformation into semantic triples,[8] mapping processes, and alignment of information across different datasets. The CIDOC CRM ontology (CIDOC CRM, 2024) served as the basis, formalizing knowledge and making it in a machine-readable format, aided by vocabularies and thesauri for data integration. Vocabularies[9] helped link common entities across datasets, enabling integration based on shared references.

In particular, the dataset provided by the Museum of Palazzo Pretorio concerns strongly historicized artworks. The bibliography has considered a large number of artworks, currently in the municipal collections since the second half of the nineteenth century, as "coming from the hospital." In fact, these artworks left the hospital premises to be properly exhibited in

---

[7] Domain standards covered by the project are EAD, EAC, XML-TEI for libraries/archives, ICCD/OA for museums

[8] For a definition of semantic triple and to delve deeper into the context of use, see W3C, 2014.

[9] Toponyms, anthroponyms, and other entities were linked through vocabularies like Getty ULAN, TGN and AAT, VIAF, and Iconclass system for artwork subject description.

the municipal gallery, inaugurated in 1858, as documented in the metadata accompanying the artworks. Research conducted in recent years, including a review of archival documentation, has been able to clarify certain aspects and consequently divide this large group of artworks generically referred to as "coming from the hospital" into at least three groups: Works commissioned by the hospital and retained by it since the fourteenth century; works that reached the hospital through donations or bequests; works that only passed through the hospital for a certain period of time before entering the art collections of the municipality of Prato and being displayed to the public – whose ownership by the hospital is therefore not proven. Thanks to RESTORE, it was possible to make these divisions clear, meaning it is possible to look for works commissioned by the hospital and those of the so-called Galleria Martini, which came to the hospital as part of Giovanni Martini's legacy.

The cataloging standard used within the superintendencies responsible for the preservation, protection, and enhancement of artistic, archeological, and architectural assets in Italy is structured by the ICCD (Central Institute for Cataloging and Documentation), established under the Ministry of Culture (MIC). This standard comprises a set of rules and cataloging principles, converging into various cataloging models for the specific types of inventoried, documented, and preserved assets forming part of the national cultural heritage (ICAR 2024). It includes regulations, supportive tools, and controls (vocabularies, lists of terms).

The documents used for mapping an artwork's data in the project mainly include:
1. Cataloging regulations, data model descriptions, and authority files;
2. Support and control tools: thesauri and terminological tools developed to uniformly acquire data and create a "common and shared language," essential for correct information utilization during searches and for data interoperability in cultural information;
3. Cataloging schemas: descriptive models collecting information in a structured manner, following a "knowledge path."

Vocabularies and thesauri are also important tools to outline a common language of reference during the mapping of resources. In fact, resources coming from different sources can refer to the same vocabulary to describe its content and/or characteristic, effectively creating a connection with other resources sharing the same meaning. By referencing the vocabularies, toponyms and anthroponyms were identified and indexed as contact

points across datasets even as common entities across different datasets have to share links to the reference vocabulary. When entities from various datasets refer to the same URI, once integrated into the semantic Web environment, they automatically connect through the reference vocabulary's URI.

Finally, mapping refers to the process of aligning parameters from one database to another (in this case from datasets into triples). It serves as the initial step in facilitating data migration, integration, and other data management activities. For data integration, common entities across multiple datasets must reference the same URIs, serving as the contact point in the semantic web.

The resulting components can be summarized as follows:

Data Ingestion Tool Suite	Tool suite to support the recovery, integration, and accessibility of data and digital objects collected
Data Management Tool Suite	Tool suite to support the indexing and management of resources within the data lake
Data Transformation Tool Suite	Tool suite to support the creation of a knowledge base integrating information and metadata about the collected resources and digital objects
Data Exposition Tool Suite	Set of APIs to exploit the data stored into the data lake
Data Visualization Tool Suite	Tool suite to support the creation of a digital environment for data integration focused on datasets coming from the domains of digital humanities and GLAMs, interoperable with solutions and platforms developed by other projects
Atlante	Tools for geo-referencing historical data, to be applied to lexical or grammatical atlases
Chronos	Tool to compare timelines coming from different datasets

As evidence of its scalable and buildable nature, the RESTORE project has been included and scaled up to a research data pilot project within the context of the SSHOC (*Social Sciences and Humanities Open Cloud*, H2020-INFRAEOSC-04-2018) Project, as the so-called Datini use case, a collaboration that has brought to the publication of RESTORE as a new

service for the Digital Arts and Humanities Research[10] as the RAISE tool. RAISE,[11] the Restore dAta Integration SuitE, extended the functionalities developed for the RESTORE project, because it consists of a toolbox that can be used to collect, align, merge, and transform heterogeneous datasets from different GLAM research fields, with the aim of producing integrated and interoperable data and representing different aspects of the tangible and intangible dimensions of cultural heritage. The code for several of its components is available at the DARIAH.it git service at http://dev.restore. ovi.cnr.it:3000/explore/repos.

In this context, the DARIAH.it team has also been developing and validating internationally the selected options of implementation of the tools and software components customized for the ingested datasets. The choice of the ontology of reference (CIDOC CRM), for instance, has been made and commented on with the acknowledgment of the data communities active in the project. With references to the workflows and tools developed by the data community within SSHOC, the implementation of a linked open-data environment helped domain experts in heritage science to evaluate the efficiency of the data-integration strategy adopted: Checking the information had to be visible and communicated to both academic and general end users, such as information about physical supports, relevant for users interested in the tangible characteristics of cultural heritage. The conceptual reference model and ontology revealed itself as key to foster open access and FAIRification of the partners' data, without preferring specific and too narrow domain-related logics.

In this context, the DARIAH.it team further developed the mapping and modeling techniques for the OA (Opera d'Arte) – ICCD (ICCD 2018) dataset, leading to its conversion to the CIDOC CRM model, as a strategic implementation of the SSHOC data-integration suite, along with the perspective of further enabling the addition of data referencing the diagnostic operations done on cultural objects. The work on material provided by the Museum of Palazzo Pretorio for the RESTORE project, related to the iconography of Datini and to the Ospedale della Misericordia e Dolce, thanks to the research work done within the RESTORE project, gave visibility to fundamental research work still in process, distinguishing between the nucleus of works of art directly commissioned by the Institute

---

[10] For more information, please visit the SSHOC Web site: https://sshopencloud.eu/restore, at the RESTORE factsheet page; see also Degl'Innocenti et al. 2021.
[11] RAISE - Restore dAta Integration SuitE. Developmental version. Retrieved July 19, 2023, from https://marketplace.sshopencloud.eu/tool-or-service/pIyy0d.

from the fourteenth century on, and the large amount of works simply stored in it, belonging to the municipal collections (Galleria Martini).

The technological solutions developed into RAISE, can be extended to other projects and reused in similar contexts by other institutions within the GLAMs, the digital humanities and heritage science research communities. Furthermore, the project's development benefits from the collaboration with DARIAH-ERIC (ESFRI Landmark for the humanities and social sciences) and E-RIHS[12] (ESFRI project for heritage science), as well as the European Open Science Cloud (EOSC).

In this respect, the collaboration with the IPERION-HS[13] presented a valid stage to expand the range of information connected to artwork and cultural heritage resources. The information already mapped on the premise of the SSHOC project were therefore integrated with datasets coming from mobile labs such as those commonly in use in use in conservation labs and institutes internationally within the framework of the open access Web tools for data storage and data fruition created at the OVI, developing a cross-disciplinary information integration system. From this point of view, digitized materials have been collected and stored, and information related to the resources (covering both the tangible and intangible aspects of cultural heritage artifacts) were then mapped and aligned using a set of algorithms and vocabularies and, eventually, modeled using the CIDOC CRM ontology.

Continuing with the approach undertaken in SSHOC with respect to art assets, the same workflow implemented for the artworks in the RESTORE project, and refined in the RAISE tool suite, was used to create the Manuscript working demo for IPERION-HS with some customized elements. This study allowed the DARIAH.it team to expand the scope of the data-processing techniques implemented for the artwork with information provided by scientific analyses carried out by heritage science researchers. The new dataset[14] included data collected from measurement

[12] E-RIHS.it is the Italian node of the European research infrastructure on heritage science. For more information, please see https://www.e-rihs.it/.
[13] The IPERION HS "Integrated Platform for the European Research Infrastructure on Heritage Science" project (GA 871034) integrates a range of national facilities, recognized for their expertise in the study of heritage science, to create an open infrastructure of institutions and specialist researchers, engaged in collaborative research and providing services to the wider field of heritage science.
[14] The dataset was provided by the Istituto di Scienze del Patrimonio Culturale (CNR-ISPC), specifically by the XRAYLab of ISPC-Catania.

campaigns: (1) EDF[15] data (and text "scanning parameters") representing the results of the measurement process (spectra); (2) XRF images (TIFF format) representing pictorial pigments and their distribution on the pictorial surface; (3) binary files with readable header and spectral data in binary format; and (4) information (metadata) about the object analyzed.

The information provided was processed and ingested in the data lake to be mapped and modeled according to the workflow implemented. To include these data in the LOD created for the artworks provided by the Museo di Palazzo Pretorio, it was necessary to study the techniques employed and the standard used.

This work resulted in the creation of a working demo, the so-called Manuscript demo, featuring the *Marlay Cutting (Presentation in the Temple)*, which displays the following information: (1) information about the illumination; (2) information about the manuscript; and (3) results of the measurement campaign. For each piece of this documentation, the textual information, metadata and images were provided. Moreover, anthroponyms and toponyms were also extracted and mapped on digital vocabularies. Last, all the text provided on the resource was analyzed to search for terms to connect to the vocabularies and thesauri of reference. When an instance was found, it was mapped and connected to the corresponding URI. Having done this, the user is able to navigate the resources to check if the same information is reported for other resources.

The result was a knowledge base featuring integrated information coming from different domains blended on a semantic ontology to allow cross-domain exploration of data. The successful case described in the previous paragraphs opens up a more detailed exploration of research domains and their resources to reach the widest possible audience.

Conclusions

The workflow's innovation lies in its approach to data integration, employing semantic modeling and standards such as CIDOC-CRM ontology to ensure not just interoperability but also semantic consistency across disparate datasets. The project's success in fostering collaboration across GLAM institutions and disciplines offers a valuable model for future endeavors. RAISE, the resultant tool suite, emerges as a testament to RESTORE's legacy, offering a versatile toolbox for collecting, aligning, and trans-

---

[15] European Data Format (EDF). For more information on this format please see the official website at https://www.edfplus.info/.

forming heterogeneous datasets within the GLAM research fields. All the activities undertaken show the advantages of thinking in a sustainable perspective regarding projects. The infrastructure is therefore able to provide researchers an always upgradable and deconstructable set of tools to exploit data to help them in their endeavors.

Works Cited

Archivio di Stato di Prato. n.d. "Patrimonio." Accessed April 18, 2024. https://archiviodistatoprato.cultura.gov.it/home.

Benardou, Agiatis, Erik Champion, Costis Dallas, and Lorna Hughes. 2018. *Cultural Heritage Infrastructures in Digital Humanities*. London and New York: Routledge.

Bekiari, Chryssoula, George Bruseker, Martin Doerr, Christian-Emi Ore, Stephen Stead, and Athanasios Velios. 2021. *Definition of the CIDOC Conceptual Reference Model v7.1.1 (Version v7.1.1). The CIDOC Conceptual Reference Model Special Interest Group*. https://doi.org/10.26225/FDZH-X26.

Degl'Innocenti, Emiliano, Carmen Di Meo, Francesco Coradeschi, Maurizio Sanesi, Elisa Brunoni, Athina Kritsotaki, and Eleni Tsoulouha. 2021. *MS49 Heritage Science and Humanities Pilot alpha release (v1.0)*. Zenodo. https://doi.org/10.5281/zenodo.5217113.

Consiglio dell'Unione Europea. 2021. "Raccomandazione (UE) 2021/2022 del Consiglio del 26 novembre 2021 su un patto per la ricerca e l'innovazione in Europa." *Gazzetta Ufficiale dell'Unione Europea*. https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32021H2122.

European Strategy Forum on Research Infrastructures (ESFRI). 2021. *Roadmap 2021. Strategy report on research infrastructures*. https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf.

GoFair, n.d. "FAIR principles." Accessed February 21, 2024. https://www.go-fair.org/fair-principles/.

Hilbert, Martin. 2012. "How Much Information Is There in the 'Information Society'?" *Significance* 9: 8–12. doi:10.1111/j.1740-9713.2012.00584.x.

Istituto Centrale per gli Archivi - ICAR. 2024. "Standard internazionali." Accessed February 22, 2024. https://icar.cultura.gov.it/standard/standard-internazionali.

Ministero dei beni e delle attività culturali e del turismo. Istituto per il Catalogo centrale e la documentazione (ICCD). 2018. *Normativa OA – Opere e oggetti d'arte, Versione 3.00. Struttura dei dati delle schede di catalogo ICCD*. http://www.iccd.beniculturali.it/getFile.php?id=7508.

Ministero dell'Istruzione, dell'Università e della Ricerca. 2019. Decreto 9 maggio 2019 "Ammissione del progetto «DARIAH.it - Developing nAtional and Regional Infrastructural nodes of dAriaH in ITaly al finanziamento previsto dal decreto

direttoriale 28 febbraio 2018." (Decreto n. 900/2019). (19A04611) (GU Serie Generale n.165 del 16-07-2019).

Museo di Palazzo Pretorio. n.d. "Il museo," "Le opere". Accessed April 27, 2024. http://www.palazzopretorio.prato.it/.

Il Parlamento Europeo e il Consiglio dell'Unione Europea. 2013. "Article 2 (6) of the Regulation (EU) No 1291/2013 of 11 December 2013: "Establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)." *Gazzetta Ufficiale dell'Unione Europea*. https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32013R1291.

Soprintendenza Archivistica e Bibliografica della Toscana. n.d. "Soprintendenza Archivistica e Bibliografica della Toscana." Accessed February 22, 2024. https://sab-toscana.cultura.gov.it/home.

W3C. 2014. "RDF 1.1 Primer." Accessed February 22, 2024. https://www.w3.org/TR/rdf11-primer/.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3, no. 160018. doi: https://doi.org/10.1038/sdata.2016.18.

Wisser, Katherine M. 2011. "Describing Entities and Identities: The Development and Structure of Encoded Archival Context—Corporate Bodies, Persons, and Families." *Journal of Library Metadata* 11 nn. 3–4: 166–75. doi: https://doi.org/10.1080/19386389.2011.629960.