EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2023)
A. Bucciero, B. Fanini, H. Graf, S. Pescarin, and S. Rizvic (Editors)

DataSpace-ISPC: a semantic platform for Heritage Science

Alberto Bucciero, Alessandra Chirivì, Gerardo Anglada, Emanuel Demetrescu, Bruno Fanini, and Nicolò Paraciani.

Abstract

The contribution presents the DataSpace of the Institute of Heritage Science (ISPC) of the Italian National Council for Research. Cultural Heritage is well known to be a multidisciplinary domain where heterogeneous data, methods and techniques overlap, co-exist and collaborate to create evidence-based wisdom represented with local and (rarely) inter-operable standards. DataSpace aims to overcome the need to harmonise converging research perspectives, allow collaborative research, and debabelize the presentation of data. This horizontal perspective is based on a graph of knowledge, ruled by a semantic reference model, and acts as a glue to connect different digital resources created with different perspectives. The DataSpace, which is built on top of the Arches platform integrated and customised with original tools, has been validated with positive results against two case studies to cover the wider domain of Heritage Science. The contribution explains the activities and technological solutions adopted and goes into detail about the problems and limitations that arose with the implementation of the DataSpace offering a critical yet positive perspective on the use of the DataSpace-ISPC in production scenarios.

CCS Concepts

• Applied computing** → **Arts and humanities; Physical sciences and engineering; • Information systems** → **Digital libraries and archives; Information storage systems; Data management systems; Information retrieval;

1. Introduction

In recent years, there has been an increase in research projects aimed at strengthening the dialogue between the various scientific communities interested in the knowledge, conservation, valorisation, and restoration of Cultural Heritage (CH). The rapidly growing interdisciplinary dialogue covers a vast range of knowledge: from archaeology [Art10] to artificial intelligence, from art history to diagnostics applied to the study of cultural heritage, from immersive 3D visualisation [GCD*16] to the "Heriverse" (Metaverse for CH), etc. The DataSpace project, currently under development, was created to guarantee and strengthen these synergies and encourage the digital transition according to the open science principles [`https://dataspace.ispc.cnr.it`]. The digital platform aims to address the current needs of the Heritage Science community, an interdisciplinary research domain based precisely on the transversal approach to improve knowledge of material and immaterial Cultural Heritage. In the Heritage Science field, the accessing, sharing, manipulating, and long-term conservation of vast and very heterogeneous amounts of data play a crucial role in advancing knowledge and preserving our cultural legacy.

In this paper, we describe the main aspects of the DataSpace project and its implementation process, highlighting the steps taken to achieve the platform's vision and objectives. Additionally, we showcase two interesting case studies where real-world projects have been successfully uploaded into DataSpace. By examining these examples, we aim to illustrate the platform's practical application and its ability to cater to diverse cultural heritage research initiatives. Furthermore, we discuss the design and implementation of data presentation models within DataSpace. This encompasses the visualisation and organisation of data to enable effective exploration and analysis, empowering researchers to gain valuable insights from the information.

While DataSpace has achieved significant milestones in its development, there are ongoing matters that require further attention and refinement. We acknowledge these challenges and provide an overview of the current open issues, seeking to foster a collaborative environment where researchers, developers, and stakeholders can work together to enhance the platform's capabilities and address any limitations that may exist.

It is important to note that DataSpace is currently under development, and the information and structure discussed in this article do not represent the final version of the platform. Nonetheless, it explains the long-term perspective of the authors (around a knowledge-graph-centric approach) to data as well as valuable insights into the key features and capabilities of DataSpace as it stands in its current development stage.

After a wider definition of the DataSpace project in section 2 the methodological and technological motivations for the adoption of Arches as the project's base are explained (sec. 3). In section 4 the components of the platform as well as their basic functionalities are described. The case studies in section 5 represent the application core of the project and demonstrate the capability of the platform to embrace the complexity and the diversity required by two different scenarios (and relative data-sets, Marmora Phrygiae and MOLAB). A key topic of the project is the design of cus-

DOI: 10.2312/gch.20231166

tomised layouts to represent data accordingly with the disciplinary needs of the stakeholders (described in sec. 6). A critical appraisal (sec. 7) that frankly underline the successes and the issues of the project close the contribution aiming to describe the future work of the project as well as the aspects that still remain unclear and should stimulate further methodological discussions with a multidisciplinary approach.

2. DataSpace ISPC project

DataSpace is a powerful data management platform for the sharing, accessing, and long-term data conservation produced by the Institute of Heritage Science. It allows access to innovative digital tools & services to increase knowledge and enhance the conservation of Cultural Heritage from an interdisciplinary perspective.

DataSpace represents the first step towards the design of a shared and more general platform (DIGILAB) within the European Research Infrastructure for Heritage Science (E-RIHS) [NMAP17]. It will consist in the core of the research facilities devoted to the study of Heritage through integrated Heritage Science and Humanities digital data, providing:

- data repositories: dealing with all activities of the data life-cycle like creation, ingestion, preservation, access and reuse and taking into account the heterogeneous nature of born-digital and digitised data, such as texts, numeric results, images, audio, video, 3D and interactive animations and considers internationally recognised standards on these data;
- services and tools related to tangible or intangible assets such as massive data processing, data analysis and visualisation, and data integration;
- a portal: including data and services catalogues, training, and help desk;
- a collaborative workspace: data and services sharing, scientific orchestrator manager able to define and connect services, even provided by different Virtual Research Environments, in a workflow shared between different remote researchers.

DataSpace provides access to research data, initially generated by HS researchers, curating it according to the FAIR data principles and enabling its interoperability to Virtual Research Environments, where they can access, visualise, interrogate, and manipulate such data, to create new Heritage-related knowledge [PCFL20].

Its architecture has been designed starting from results of European-related projects, i.e. E-RIHS PP and IPERION HS, and in dialogue with the E-RIHS- IP project (submitted at the call: HORIZON-INFRA-2021-DEV-02-02). It will also be aligned with EU policies regarding data management (FAIR, Open Research Data) and the European Open Science Cloud strategy (EOSC).

DataSpace can also facilitate the exchange of data and their interoperability, promoting the development of open science in the Heritage Science field. It offers virtual access to tools and data supporting multidisciplinary research (i.e. results of scientific measurement, environment for immersive and interactive visualisation of the data-set, historical documentation and literature sources, etc.). It guarantees data search-ability with advanced tools based on metadata stored in federated repositories. It also grants accessibility of resources through mechanisms of federated identity, delegating access and delivery control to the nodes. Data interoperability is granted by the use of appropriate and shared standards, while re-use is guaranteed by the possibility of data processing through specialised services (i.e. virtual access, scientific visualisation, and simulation for digital heritage, collaborative research in immersive and interactive environments accessible on-site o remotely, data geo-referencing annotation, analysis, etc.) integrated into the E-RIHS ecosystem.

3. Why did we choose Arches?

As previously mentioned, one of the main needs expressed by the HS community is the capacity to fully describe and be able to outline content relating to different typologies and disciplinary areas ranging from immovable property, such as historic buildings, sites, landscapes, environments; to movable ones, as works of art or technology and science. They also include immaterial resources such as music, performances, traditions, practices, representations, expressions, knowledge, and skills, as well as the instruments, objects, and cultural spaces associated with them and recognised as such by a community or group of people. These aspects are intertwined with each other, linking for example a building to its content and its history. To denote each heritage item, the term heritage asset [NFH22] can be used, which may refer to any kind of heritage, either tangible, both movable and immovable, or intangible.

To preserve or establish connections between various types of digital scientific content, ranging from historical analysis to purely diagnostic information, it is essential to acquire the necessary tools or resources; on the one hand with a software platform for data management that can describe these mutual semantic relationships and on the other it is necessary to use a data model that is capable of representing the asset heritage with its dense network of attributes.

A knowledge graph is a data structure that represents information in the form of entities (nodes) and their relationships (edges). It organises knowledge in a way that allows for efficient storage, retrieval, and analysis of complex data. In a knowledge graph, entities can represent various things such as concepts, objects, events, or people, and the relationships between them capture the connections or interactions between those entities.

From a conceptual point of view one of the most successful ways of representing information, easy to read and manipulate is the one called a graph of knowledge. A knowledge graph is typically built using semantic technologies and graph databases. It incorporates principles from graph theory and semantic web technologies to create a rich network of interconnected data. This structure enables powerful querying and reasoning capabilities, as well as the ability to infer new knowledge by traversing the relationships within the graph. More, there are practical reasons to adopt a graph data structure: it avoids redundancy (node elements re-used in connection with several others) and it is basically schema-less: you can ingest both only a minimal information to describe cultural heritage assets or you can extend the properties for a specific asset without loosing the general coherence of the data-set.

In the context of heritage science, which focuses on the study, preservation, and management of cultural heritage, knowledge graphs can play a vital role in organising and analysing diverse information related to artefacts, historical events, conservation practices, and more. By structuring heritage data as a knowledge graph, it becomes possible to connect various elements of heritage knowledge, uncover hidden relationships, and derive meaningful insights.

There are just a few software platforms and frameworks available that utilise knowledge graphs for heritage science applications, one of the most advanced and reliable is Arches [`https://archesproject.org`].

Arches is an open-source platform created by the Getty Conservation Institute and the World Monuments Fund, specifically designed for cultural heritage inventory and management. It uses a knowledge graph to represent heritage data, allowing for complex querying, data visualisation, and collaborative workflows.

Designed for Cultural Heritage: Arches is specifically designed for cultural heritage inventory and management. It takes into account the unique requirements and complexities of heritage data, making it well-suited for representing and organising diverse heritage assets.

- Knowledge Graph Approach [DQJL18]: Arches uses a knowledge graph to represent heritage data. This allows the modelling of complex relationships and connections between different elements of heritage knowledge. By structuring data as a graph, Arches enables powerful querying, analysis, and visualisation capabilities.
- Flexible Data Modelling: Arches provides a flexible data modelling framework that can be customised to suit the specific needs of a heritage project. It allows the creation of custom data types, attributes, and relationships, enabling the representation of various types of heritage assets and their specific characteristics.
- Open-Source and Extensible: Arches is an open-source platform, which means it is freely available and can be customised and extended by the heritage community. This open nature encourages collaboration, sharing of best practices, and continuous platform improvement.
- Multilingual and Multi-script Support: Arches supports multilingual and multi-script data, making it suitable for representing heritage assets from different cultural contexts and languages. It is also possible to have multilingual labels, descriptions, and other textual information.
- Collaborative Workflows: with the scientific world producing data at an explosive rate, workflows have an important part to play in the end to end management of scientific data [Sim08] Arches provides features for collaborative workflows, enabling multiple stakeholders to contribute, validate, and review heritage data. This is particularly useful in heritage projects that involve diverse experts, researchers, and community members.
- Data Visualisation and Mapping: Arches offers interactive data visualisation and mapping capabilities, allowing users to explore and present heritage data in meaningful ways. It supports geospatial data integration and mapping, which is essential for heritage assets that have a spatial component.
- Integration and Interoperability: Arches supports data integration and interoperability with other systems and standards. It can import and export data using various formats, ensuring compatibility with existing heritage data sources and tools.

Overall, Arches provides a robust and customisable platform for representing heritage data assets. Its knowledge graph approach, flexibility, collaborative features, and support for multilingual and multi-script data make it a compelling choice for heritage science projects. Additionally, being an open-source platform, it encourages community involvement and fosters the development of a shared knowledge base for cultural heritage management. For all the previous reasons we have chosen Arches as the funding platform for our DataSpace.

4. Platform implementation

DataSpace users can input, store, and organise large amounts of data in a structured manner, ensuring data integrity and easy accessibility. With support for various data types, including geospatial data, DataSpace empowers users to handle and analyse related information effectively.

One of the key strengths of DataSpace lies in its data discovery capabilities. Users can easily search, filter, and retrieve specific data-sets or subsets based on their desired criteria. This streamlined approach to data retrieval saves valuable time and allows users to focus on gaining insights from the data rather than struggling with manual searches.

DataSpace also offers powerful visualisation tools to enhance data exploration. The platform supports interactive maps, charts, and graphs, enabling users to gain insights from their data and effectively communicate their findings to others (see Figure 1).

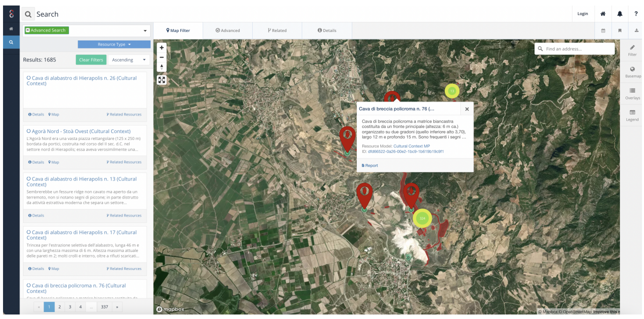

Figure 1: *Geospatial search in the DataSpace.*

DataSpace also has the capability to generate semantic data according to standards, which is significant because it enhances the possibility of creating data that can work well together and be maintained over time. Users can upload and utilise various ontologies, such as the CIDOC Conceptual Reference Model (CRM). Once they create instances of objects, DataSpace generates a node map that illustrates the relationships between resources (see Figure 2).

4.1. DataSpace Software Dependencies

DataSpace is built on the foundation of Arches, an open-source data management tool created by the Getty Conservation Institute and

Proceedings published by Eurographics - The European Association for Computer Graphics.

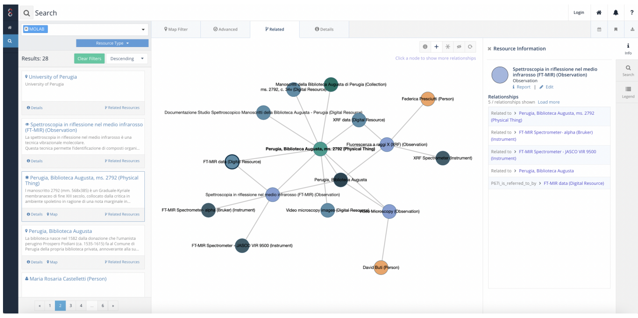

Figure 2: *An example of related resources in the DataSpace.*

the World Monuments Fund. It is powered by Django, a Python-based framework that streamlines the development of websites and web applications. Figure 3 outlines some of the features Arches provides.

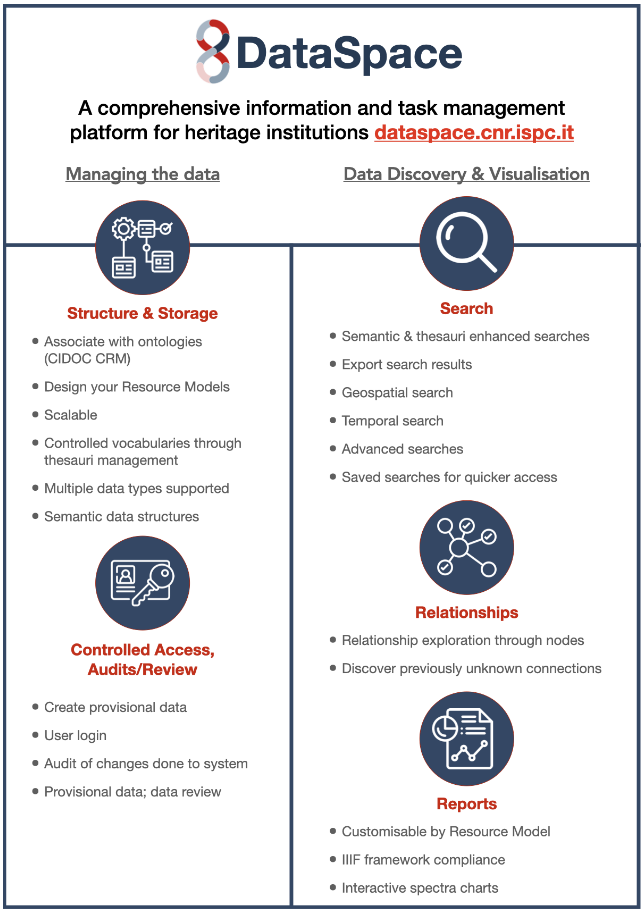

Figure 3: *Current features and capabilities of the DataSpace.*

While the initial configuration files of DataSpace are written in Python, the majority of the platform is developed using HTML, CSS, and JavaScript. The use of KnockoutJS, a JavaScript library, enhances the interactivity and responsiveness of the platform's user interface.

Overall, DataSpace serves as a versatile and comprehensive solution for data management, discovery, visualisation, and project management.

Other dependencies include:

- node.js: a JavaScript runtime environment that allows you to execute JavaScript code outside the browser,
- yarn: a package manager for JavaScript,
- PostgreSQL with PostGIS as its database management system (DBMS), and
- Elasticsearch: a search and analytics engine designed for full-text search, real-time analytics, and scalable data storage.

4.2. Data Model

Arches has a variety of features and capabilities for data management and allows for flexible customisation of the data model. To implement DataSpace, we used different modelling solutions, starting from heterogeneous digital resources managed in a completely different way. Currently, we evaluated the data of two projects and we assumed them like case studies to compare the distinct solutions, their strengths and limits. The relationship between the abstraction level of data models and the need not to lose information about them is a crucial topic for reflection (see below).

Actually, our interest was to reconciliate different disciplinary approaches within the same project using a knowledge-graph-centric to ingest and visualize data from different points of view (see 4).

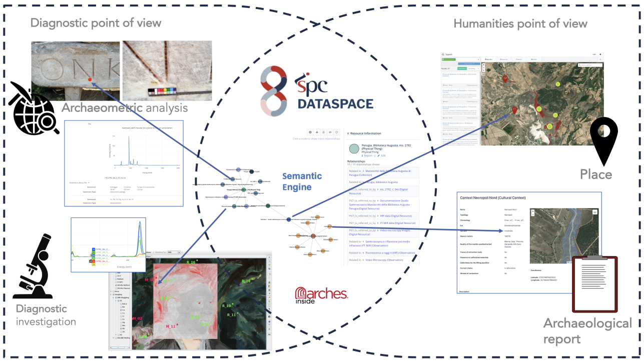

Figure 4: *Different points of view reconciliated in Dataspace.*

To this end, we identified two different use cases that differ in strategies, solutions, research data management, types of assets studied, and the tools and services provided to the community. The first case study is dedicated to the Marmora Phrygiae project and the second to a project of the E-RIHS mobile laboratories (MOLAB)(see below).

For managing the archaeological and archaeometric data produced within the Marmora Phrygiae project we developed new Resource Models for Context, Object, Sample, and DataSpace automatically generating the required relationships and performing data normalisation internally. These Resource Models were created using DataSpace's user interface (recalling that the original data model designed for the implementation of the Marmora Phrygia database is not available anymore).

Unlike Marmora Phrygiae, we used a suite of Resource Models derived from the ongoing development of Arches for Science, a collaborative project within the Arches community, to manage the analytical data from MOLAB's integrated diagnostic investigation techniques. This package includes a range of pre-existing resource models based on the CIDOC Conceptual Reference [Doe03]. However, we have been adding new concepts in the available thesauri and within the Related Resource models to manage all essential digital data.

The different ways used to manage digital resources in DataSpace have an impact on their interoperability and sustainability (see below) and require us to face some issues.

4.3. Interactive 3D Presentation

Regarding 3D data managed by the DataSpace, interactive presentation services are required. To maintain a scalable and modular approach, the open-source ATON framework [FFD*21] is employed for such tasks. The framework is based on Node.js (already a dependency of the DataSpace) and its architecture allows different deployment scenarios:

1. On the same node where DataSpace is deployed,
2. On a dedicated node for 3D service, or
3. Distributed across multiple nodes in a Research Infrastructure.

The REST API provided by the framework allows different integrations with Arches (e.g. using reports, see Figure 5) as well as exploiting built-in front-end components, inspecting a specific 3D resource (like "Hathor", see [FFD*21] - section 3.8) - or custom front-ends, for instance depending on 3D resource type or properties. This also enables different interfaces to present 3D data, including custom functionalities, custom presentation of semantic annotations, and metadata.

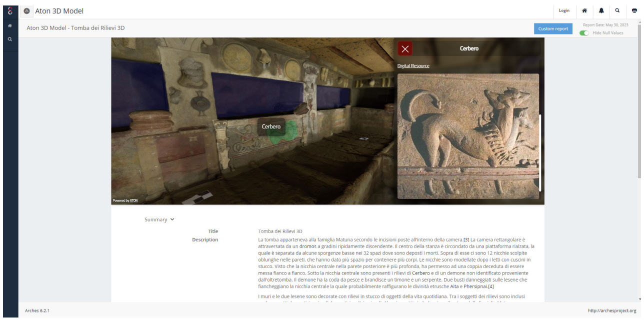

Figure 5: *An example of a report with interactive 3D component.*

The framework is also based on open, robust, extensible, and interoperable 3D standards (like glTF ISO standard - `https://www.khronos.org/gltf/`, and others), thus aligned with FAIR data principles, at the international level.

On client-side, ATON offers a cross-device presentation layer: any frontend based on the framework is capable of automatically adapting the rendering system and interaction model to the consuming device, ranging from mobile (smartphones, tablets), desktop (laptops, workstations) up to XR devices.

Such properties combined with the ongoing active development of the framework, provide a reusable, interoperable, and modular 3D component for universal 3D presentation in the DataSpace.

5. Case studies

After engaging in discussions with researchers and research teams within the Institute of Heritage Science, we identified some projects and initiatives to prioritise in this initial phase of activities. In our selection, we considered the Institute's diverse sectors, numerous themes, and various lines of research, ensuring that the chosen projects originated from different disciplinary fields to address the needs and requirements of the relevant scientific communities. This approach allowed us to leverage diverse resources, encompassing disciplines such as archaeology, the history of illuminated manuscripts, architecture, and diagnostics for knowledge, conservation, and restoration of cultural heritage.

Simultaneously, the inclusion of projects from different disciplinary fields enabled us to understand specific researchers' needs, reconstruct workflows and operational processes followed by various stakeholders, and address specific conceptual models, as well as different types of data and resource management methods. Meeting these diverse requirements while designing the DataSpace and choosing appropriate technological solutions to harmonise the needs with the opportunities provided by Arches posed significant challenges throughout the project.

5.1. Marmora Phrygiae Project

The first of these projects is Marmora Phrygiae [IS16]. The project (2013-2016) dealt with the study of ancient marble, alabaster, and breccia quarries in the territory of Hierapolis of Phrygia (Denizli, Turkey) with a multidisciplinary approach and the reconstruction of the procurement strategies of stone materials for the construction of buildings urban. For the characterisation of the materials and the determination of the origin, complementary diagnostic investigations were carried out: mineralogical-petrographic, cathodoluminescence analysis, analysis of the stable isotopes of carbon and oxygen, and XRD (X-Ray Diffraction) analysis. These searches have therefore produced a large amount of archaeological and archaeometric data: alphanumeric data resulting from quantitative investigations, textual data, macro photographs of thin sections, XRD spectra, etc. For management, exchange, and sharing of digital resources generated by several scientific teams involved and located in Italy and abroad implemented a geodatabase which unfortunately has not been accessible for years.

This problem concerns an important aspect of managing digital assets, namely their long-term conservation. This refers to the process of ensuring that digital information remains accessible, usable, and authentic over time [`https://dpconline.org/handbook`]. It is a critical task as digital information can quickly

Proceedings published by Eurographics - The European Association for Computer Graphics.

become obsolete or unusable due to various factors. Some challenges associated with long-term data preservation are technological obsolescence, media decay, data corruption, and changing standards and formats. Technological obsolescence refers to the rapid pace of technological change that can make it difficult to access old digital information. Media decay is the natural degradation of storage media over time, which can render stored data unreadable. Data corruption can occur due to errors in storage or transmission, which can lead to the loss of important information. Changing standards and formats can make it difficult to access and use old digital information as technology evolves. Addressing these challenges requires a combination of strategies such as migration to new formats, creating multiple copies of data, and regular testing to ensure data integrity.

Lack of attention toward following best practices referring to long-term data preservation led the server to break down and without documentation nor the people who initially set up the database, we were given the task of extracting the information directly from the hard disk.

The old Marmora Phrygiae platform made it possible to manage large quantities of geo-referenced data relating to the context under study (Cultural Context), to the artefacts examined (Object), and to the analyses performed (Sample & Analysis), and to create new resources. For some years, it was a valuable tool for scholars and members of the scientific community interested in studying the quarries and construction sites of Hierapolis, Phrygia. The open data policy ensured that this information was available for free to those who wished to use the results for their needs.

The database was originally designed with three main tables, as previously mentioned: Cultural Context, Object, and Sample. These tables were characterised by a multitude of fields containing arrays of pre-set values. To ensure normalisation, a separate table was created within the PostgreSQL database for each of these multi-value fields [IS16]. Each table comprised two columns: one with an index (ID) and the other with the textual value of the array (thesaurus). This normalisation process was undertaken by the archaeologists during the system's development. Consequently, the database structure consisted of a series of linked tables associated with each of the main tables.

Once we successfully extracted all the necessary information, we proceeded to create new models using DataSpace's user interface (see Figure 6). DataSpace, being based on Django, a Python framework for web development, offers an Object Relational Mapper that enables users to insert data and create tables without the need to write specific SQL queries [https://pythonistaplanet.com/advantages-of-django]. This functionality facilitated the creation of "Resource Models" within DataSpace. We developed new Resource models for Context, Object, and Sample, and DataSpace automatically generated the required relationships and performed data normalisation internally.

5.2. MOLAB Project

In implementing DataSpace, we decided to dedicate a case study to an illuminated manuscript analysed by the MOLAB (MObile LABoratory) of E-RIHS [BMR*16]. This work is part of the broader

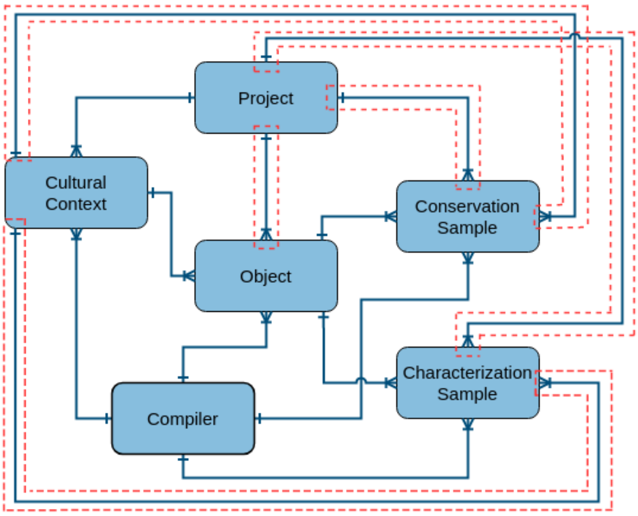

Figure 6: *ER Diagram of Marmora Phrygiae's database, reverse-engineered from its original implementation. Outlined in dotted, red boxes are redundant relations which we removed within DataSpace.*

and more articulated digital space dedicated to data provided by the MOLAB teams. The manuscript we have focused on is codex 2792 held in the Augusta Library. Non-invasive integrated analyses were carried out to increase the knowledge relating to the characterisation of the materials, the executive techniques, and the assessment of the state of conservation. The techniques used are imaging techniques (visible photography, infrared reflectography, fluorescence, and false colour infrared imaging) and punctual and non-invasive spectroscopic techniques (X-ray fluorescence, infrared spectroscopy, UV-Vis spectroscopy, and X-ray diffraction) (see Figure 7).

Figure 7: *Investigated points FT-MIR, and imaging analysis from c. 34v., in the DataSpace*

To manage this data, unlike Marmora Phrygiae, we used a suite of Resource Models derived from the ongoing development of Arches for Science. Data Space allows one to view all the heterogeneous related resources associated with the manuscript, the tab of which contains information about the object (description, size, collection, etc.). Related resources are viewed at the bottom of the sheet or via the map. The manuscript object is therefore accompanied by a series of contents of different types, which together provide information on the work in its material and immaterial complexity. Additionally, among various customisations to the Arches platform, we extended the core functionality that allowed users to upload their diagnostic data files in a .txt format only if they wanted their charts rendered, to allow for more file formats. This way, particular proprietary formats (e.g. ".mca", ".dpt") are now recognised and plotted accordingly. Users can use analytics tools to interpret their raw data obtained thanks to integrated analytical investigations such as the comparative evaluation of different spectra (see Figure 8). The available thesauri have been enriched by putting specific concepts of the diagnostic investigations for Cultural Heritage.

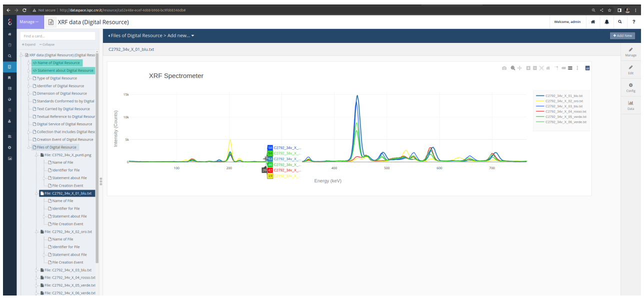

Figure 8: *Collected spectra compared for data analysis, in the DataSpace.*

6. Design and implementation of some data presentation models

The Arches platform provides a default frontend for the presentation of structured data into reports, which are implemented as views in the Django framework. Reports are the main way for users to visualise and consult information related to resources. Since the default presentation was considered not appropriate for DataSpace content, a custom design was developed (see Fig. 9 and 10). The information related to a CH asset goes far over its description: it includes its cultural context, the relations with other assets as well as the formal way it is supposed to be presented, according to the standards of evidence and wisdom of each discipline (archaeology, conservation, architectural studies, etc..). In other words, the representation goes with the data. In the Dataspace-ISPC, we decided to connect data and its visualisation paying special attention to finding a formal way to store and retrieve this information, by means of data layouts resulting in "custom reports".

It is possible to define these "custom reports" within Arches by "inheriting" the default view and implementing the required modifications using both Jinja (extended HTML) templates and CSS and JavaScript.

Therefore, we initially adopted this approach to implement the design for DataSpace reports. However, this proved to be difficult for two main reasons: 1) the underlying CSS had to be modified extensively by overriding most rules in the project's stylesheet; 2) the team was not experienced with Knockout.js, the main frontend JavaScript library that Arches uses to manage data in views dynamically.

Since the platform provides RESTful API endpoints that can be queried to retrieve resource data in JSON format, we decided to

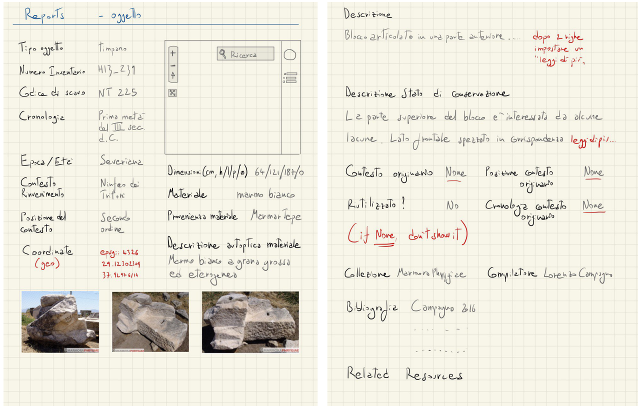

Figure 9: *Layout for the data presentation.*

leverage them to build a small custom frontend. This has been developed as a completely client-side application using HTML, native JavaScript and CSS with the following open-source libraries to include specific functionality:

- Spectre: a minimal, responsive CSS framework that provides many utilities and features;
- Spotlight: a native JavaScript solution to implement responsive image galleries;
- Leaflet: a JavaScript library to embed Web maps showing georeferenced data;
- Plotly: another JavaScript library to create plots as SVG images from text data.

The frontend works as a standalone application that retrieves JSON data from Arches by making HTTP requests to its REST endpoints via the Fetch API. This is a native JavaScript API that all modern browsers implement and so does not require additional dependencies. The JSON document is then processed to extract information about the resource, including its type, and the corresponding report is generated dynamically by inserting HTML elements into the page's skeleton. To do this, the frontend implements a rudimentary "templating" system, which consists of structured map objects that define the fields and the related data to be extracted for each supported resource model.

Currently, only resource models related to the Marmora Phrygiae project are supported - including analysis data presented in the "Sample" model. However, the system - which could be considered a proof of concept at this stage - tries to implement modularity to be easily extensible or replaceable.

This custom frontend can be deployed independently from the platform since it works entirely on the client side and it only requires an active internet connection to retrieve Arches JSON data. A button has been included in DataSpace reports within Arches to allow users to switch to the custom view for the resource (based on its UUID).

The code for this frontend is itself open source, released under the AGPL 3.0 licence (like Arches).

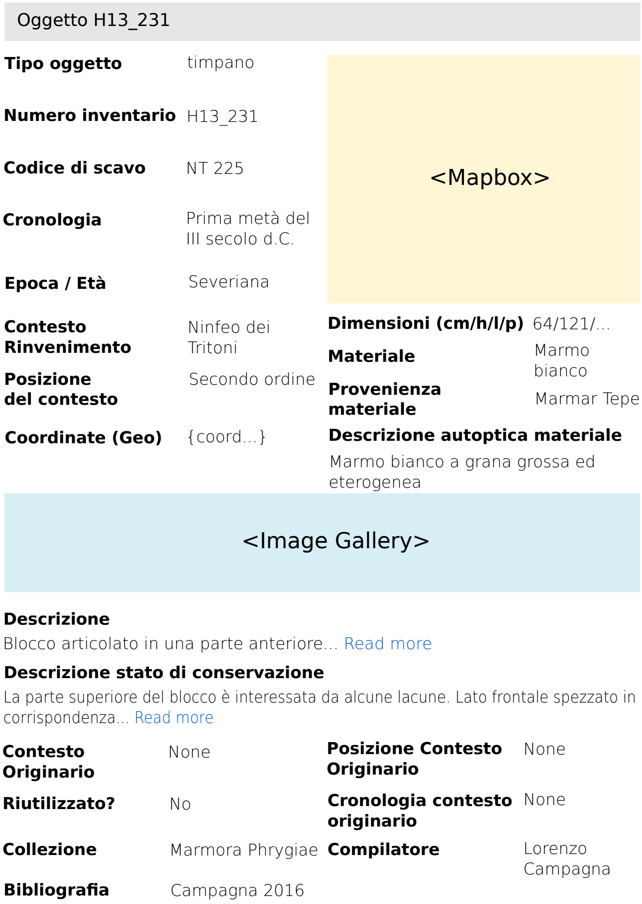

Figure 10: *SVG version of the layout shown in Fig. 9.*

7. Open Issues

Even if the Arches platform has proved to be powerful beyond expectations, flexible and capable of modelling, albeit not completely, generic data ingestion workflows, the usability aspect still remains a weak point. In fact, the possibilities for customising the flow of activities that make up the scientific data production process are very limited and this is a potentially disorienting factor for the researcher who risks losing his or her points of reference.

At the same time, the limited ability to customise reports confuses users who are used to expecting a non traditional graphic layout closer to their disciplinary needs. However, although these considerations on the usability of the platform have been made, it has not yet been possible to conduct an appropriate verification campaign towards end users. Despite this lack, which is well known, we already started the alpha test phase of the DataSpace platform within the Institute of Heritage Science internal user community, about 200 researchers and technicians, to measure the degree of ease of use of the various features offered and to collect feedback in order to improve its overall performance.

Another great challenges we currently face with the Arches platform (DataSpace) is managing permissions. Specifically, the platform is optimised to manage items or objects under a single "manager", which means that all of the different collections/objects can be overseen by a single user. The platform is not designed for different groups to work on the same database and restrict access or use of some objects to one group or another.

This can become particularly problematic when one group of people would like to hide their instances from another group. Currently, we have two solutions to this issue. The first is creating a "Resource Model" that is specific to the type of instance the group would like to hide. Arches allows us to create permissions at the Resource Model level, which means that we can "hide" all of the objects from a particular Resource Model from any users or group of users that the manager of that Resource Model decides. However, there is a caveat here - creating a Resource Model for every group can be time-consuming, and we may lose the semantic relation between the digital resources. The second solution is specifying the permissions one by one on each instance of an object that the group would like to hide. However, this can also be time-consuming, particularly if we eventually want to make all of the objects "public". In that case, we would have to go back one by one to each of the objects to revert or update the permissions, which can be a laborious task.

In addition to managing permissions, we've also encountered a second challenge with the Arches platform: providing a designated space for manual image analysis. After in-depth research, we have agreed to include in DataSpace the IIIF model. IIIF is a globally-adopted open source standardised model for delivering many types of image-based resources on the web in many different formats so that audiences can interact with them [`https://training.iiif.io/iiif-online-workshop/day-one/whatisiiif.html`]. This framework is now widely used by prestigious institutions operating in the cultural heritage field. This means that there are plenty of resources available to help us create and expand a viewing component within Arches. Currently, the IIIF component of the platform allows for this image analysis only in Edit Mode, which can be problematic. Most likely, a user who wants to analyse an image doesn't want to edit the information anymore, so entering edit mode to use these tools isn't intuitive. Fortunately, there are ways to work around this issue. For example, we could create a designated space within the platform where users can conduct manual image analysis without having to enter Edit Mode. This would make the process more intuitive and user-friendly, while also ensuring that users have access to the tools they need to properly analyse and manage cultural heritage data, and we're confident the IIIF Model is certainly one we are keen to implement.

The implementation of workflows is the third and final (at the time this was written) open issue that we face while working on DataSpace. Workflows provide a richer experience for the user to create and update objects within the platform. Currently, we are exploring options to implement workflows by using an existing package called "Arches for Science", which was created by the Arches community and is currently in the development phase. Our team is engaged in discussions around whether to adapt our needs to fit within the Arches for Science package, despite the presence of bugs due to its ongoing development. Alternatively, we are also considering creating our workflows from scratch, which would require us to develop the necessary plugins on our own.

This decision is not an easy one, as both options present their challenges. Adapting our needs to fit within the Arches for Science package may be more efficient in terms of development time, but we would need to be willing to work around its bugs, which could impact the functionality of our workflows. On the other hand, creating our workflows from scratch would allow us to have full control over the development process, but it would also be a more time-consuming and resource-intensive endeavour.

Regardless of which option we choose, our team is committed to finding a solution that will allow us to fully leverage the power of workflows within Arches. We believe that this will ultimately lead to a better user experience and improved data management capabilities, which will benefit our organisation as a whole.

Despite these challenges, Arches remains a powerful open-source software platform and we are continuously working to implement Arches in the best way to manage our community's needs.

8. Acknowledgements

The research leading to these results has received funding from the National Ordinary Fund for Research Institutions (FOE) "Innovazione Tecnologica per la protezione, valorizzazione e sicurezza del Patrimonio Culturale" and PON IR 2014-2020 SHINE Project grant PIR01-00016.

The authors are especially grateful to Dr. Costanza Miliani, director of the Istitute of Heritage Science (CNR-ISPC), for her constant and tangible support.

References

[Art10] ARTIOLI G.: *Scientific Methods and Cultural Heritage: An introduction to the application of materials science to archaeometry and conservation science*. Oxford University Press, 07 2010. doi:10.1093/acprof:oso/9780199548262.001.0001. 1

[BMR*16] BRUNETTI B., MILIANI C., ROSI F., DOHERTY B., MONICO L., ROMANI A., SGAMELLOTTI A.: Non-invasive investigations of paintings by portable instrumentation: The molab experience. *Topics in current chemistry 374* (02 2016). doi:10.1007/s41061-015-0008-9. 6

[Doe03] DOERR M.: The cidoc conceptual reference module: An ontological approach to semantic interoperability of metadata. *AI Magazine 24*, 3 (Sep. 2003), 75. doi:10.1609/aimag.v24i3.1720. 5

[DQJL18] DOU J., QIN J., JIN Z., LI Z.: Knowledge graph based on domain ontology and natural language processing technology for chinese intangible cultural heritage. *Journal of Visual Languages & Computing 48* (2018), 19–28. doi:10.1016/j.jvlc.2018.06.005. 3

[FFD*21] FANINI B., FERDANI D., DEMETRESCU E., BERTO S., D'ANNIBALE E.: Aton: An open-source framework for creating immersive, collaborative and liquid web-apps for cultural heritage. *Applied Sciences 11*, 22 (2021), 11062. 5

[GCD*16] GALEAZZI F., CALLIERI M., DELLEPIANE M., CHARNO M., RICHARDS J., SCOPIGNO R.: Web-based visualization for 3d data in archaeology: The ads 3d viewer. *Journal of Archaeological Science: Reports 9* (2016), 1–11. doi:10.1016/j.jasrep.2016.06.045. 1

[IS16] ISMAELLI T., SCARDOZZI G.: *ANCIENT QUARRIES AND BUILDING SITES IN ASIA MINOR. Research on Hierapolis in Phrygia and other cities in south-western Anatolia: archaeology, archeometry, conservation*. Edipuglia, 2016. 5, 6

[NFH22] NICCOLUCCI F., FELICETTI A., HERMON S.: Populating the data space for cultural heritage with heritage digital twins. *Data 7*, 8 (2022). doi:10.3390/data7080105. 2

[NMAP17] NICCOLUCCI F., MEGHINI C., ACHILLE F., PEZZATI L.: Condivisione dei dati sui beni culturali: Digilab, l'esperienza di ariadne e di e-rihs. *CONFERENZA GARR*, 1 (2017), 1,3. doi:10.26314/GARR-Conf17-proceedings-24. 2

[PCFL20] PIERGROSSI A., CARAVALE A., FROMAGEOT-LANIEPCE V.: Note e recensioni, what is f.a.i.r.? *Archeologia Calcolatori 31*, 1 (2020), 253–264. doi:10.19282/ac.31.1.2020.12. 2

[Sim08] SIMMHAN Y.: End-to-end scientific data management using workflows. In *2008 IEEE Congress on Services - Part I* (July 2008), pp. 472–473. doi:10.1109/SERVICES-1.2008.22. 3