Journal of Cultural Heritage xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Cultural Heritage

journal homepage: www.elsevier.com/locate/culher

SI: Notre-Dame de Paris

A digital platform for the centralization and long-term preservation of multidisciplinary scientific data belonging to the Notre Dame de Paris scientific action

Ariane Néroulidis[a,*], Thomas Pouyet[b], Sarah Tournon[c], Miled Rousset[d], Marco Callieri[e], Adeline Manuel[a], Violette Abergel[a], Olivier Malavergne[f], Isabelle Cao[a], Roxane Roussel[a], Xavier Granier[c], Xavier Rodier[b], Livio De Luca[a]

[a] Modèles et simulations pour l'Architecture et le Patrimoine - UMR 3495 MAP, Campus du CNRS (Bâtiment US) - 31, chemin Joseph Aiguier, 13402 Marseille, Cedex 20, France
[b] UMR 7324 - CITERES MSH Villes et Territoires, 33, Allée Ferdinand de Lesseps, 37204 Tours, Cedex 03, France
[c] Archeovision 3D SHS - UMS CNRS, Archéopôle d'Aquitaine, Esplanade des Antilles, Université Bordeaux Montaign, 33607 Pessac, Cedex, France
[d] Maison de l'Orient et de la Méditerranée Jean Pouilloux (MOM) - FR 3747, 7 Rue Raulin, 69007 Lyon, France
[e] CNR-ISTI, Istituto di Scienza e Tecnologie dell'Informazione, Via Giuseppe Moruzzi, 1, 56127 Pisa, Italy
[f] Laboratoire de Recherche des Monuments Historiques, 29 rue de Paris, 77420 Champs-sur-Marne, France

ARTICLE INFO

Article history:
Received 11 September 2023
Accepted 28 September 2023
Available online xxx

Keywords:
Heritage science
Digital platform
Collaborative studies
Computational modelling
Digitisation
Information systems

ABSTRACT

This article explores the design, development and deployment of a digital platform for scholarly work at Notre Dame Cathedral and demonstrates the transformative impact of digital technology on heritage disciplines. By merging technology and human expertise, the platform facilitates the creation, integration, sharing, and analysis of extensive scientific data on the multidisciplinary post-fire study of the cathedral. This multi-layered approach includes community building for collaborative efforts, digital tools tailored to different stakeholders, data structuring approaches for managing multidimensional features, and experience-based workflows for documenting, categorising and semantically enriching scientific and restoration data. The overall goal is to introduce an integrated solution for collaborative studies and to promote a digital memory of the collective initiative in accordance with the principles of FAIR for scientific heritage data. This initiative not only supports the research and restoration of Notre Dame, but also serves as a paradigm for future conservation and documentation efforts in the field of cultural heritage.

Introduction

Digital tools for supporting multidisciplinary knowledge

In recent years, digital technology has revolutionised the way cultural heritage-related disciplines work. The introduction of new 'digital objects' has proved useful for studying forms and materials (e.g. reality capture sensors), searching and sharing data (e.g. databases, information systems, ontologies) or disseminating scientific results (e.g. multimodal virtual environments), but few experiments have explored the full path from data production to knowledge production as key to embedding human knowledge in n-dimensional digital objects, especially in the context of co-production of multidisciplinary knowledge around the same cultural heritage object. Beyond the development of an innovative technological framework in the service of collaborative studies, the introduction of a digital ecosystem (with actors, methods, protocols and tools that produce data, information and knowledge) raises new scientific questions in the context of heritage sciences by moving the cursor from the analysis of the features of the cultural object to the analysis of the knowledge mobilised to understand it. The multidisciplinary nature of a cultural object is the ideal context in which to explore this topic, and the Notre Dame Restoration Project (involving hundreds of scholars from diverse backgrounds) presents an unprecedented opportunity. This paper presents the design and ongoing implementation of the centrepiece of this ecosystem, a digital platform to support the study and restoration of the cathedral by integrating data, information and knowledge from the many stakeholders. The project is part of the official Digital Data Working Group of the Notre Dame Scientific Action of the CNRS and the Ministry of Culture [1,2], organised around four complementary axes: the collection and integration of existing data, the production of new data, their sharing and

* Corresponding author.
E-mail address: ariane.neroulidis@map.cnrs.fr (A. Néroulidis).

https://doi.org/10.1016/j.culher.2023.09.016
1296-2074/

archiving, and their structuring and semantic enrichment. The design of this digital platform takes place at the intersection of these themes, aiming to collect, correlate and analyse masses of digitally generated scientific data on the cathedral's architecture, its changes over time, its structural and acoustic behaviour, its materials, as well as the damage caused by the fire in April 2019. By introducing an innovative collaborative framework, our platform also includes the implicit goal of creating a kind of "digital memory of a collective endeavour", whose content would be accessible and reusable under the FAIR principles as an emblematic example of scientific heritage data.

Collaborative efforts and technological integration

The Digital Data Working Group [3], progressively established during the first year of Notre Dame's scientific action, brings together complementary competences for the creation/documentation of scientific data central to the restoration project [LRMH (CRC) and C2RMF (NEW AGLAE), in collaboration with the SRA of DRAC Ile-de-France], for the acquisition, merging, and reconstruction of spatial data at different scales (Archeovision, MAP, MIS, LASTIG, Centre Chastel), to databases and information systems for archaeological research or heritage conservation (MSH VdL, MOM, MAP, ETIS), and to knowledge of the history of Notre Dame through the cultural, social and material dimension (Centre Chastel). Our platform also brings together in a single technological environment specific software tools from research units whose work already interacts closely with the IR* Huma-Num Very large research infrastructure through the MASA and 3D for humanities consortia. Finally, although outside our working group, other thematic working groups of the general scientific action, the lead architects and the public body in charge of the restoration work (EPRNDP) form a second and natural ring of partners linked to this ambitious project.

Integration with current scientific paradigms

Cathedrals, a prime example of the complexity of cultural heritage, offer a relevant case study for experimenting with systemic approaches in digital documentation. The Notre Dame scientific action is a paradigm shift, bringing together representatives of scientific and methodological issues in the field of cultural heritage and digitisation: Platform, Software, Data, Metadata. Dynamic processes, organisations and actors thus occupy a central place in this approach, like a digital twin, as opposed to GIS or BIM models [4]. The principle of interoperability of digital assets is central, as is the question of the life cycle of digital data. Issues related to the acquisition, structuring and preservation of data (iconographic, textual, 3D, ...) in the context of the digital representation of buildings draw directly on the results of French and European initiatives to implement FAIR data ecosystems, taking into account the conceptual model CIDOC CRM [5] - and its extensions [6]. Finally, at the technological level, isolated work on reality-based 3D annotation [7], signal processing for iconographic or handwritten data [8,9] and approaches from the field of knowledge modelling [10] can be mobilised and integrated to manage datasets of different types, production types [11] and processing stages along their relationships. This underlines the potential of linking production and collaborative analysis of multidisciplinary resources at the interface between human analysis and computational support.

Methodological foundations

Our platform represents the digital dimension of a socio-technical ecosystem designed to foster the semantic enrichment of data from restoration and scientific worksites involving a large and multidisciplinary team. A collaborative toolkit for the human-driven description and analysis of digital objects (texts, images, 3D models, sounds, videos, ...) is combined with a correlation engine (currently under construction) responsible for extending the semantic enrichment of the corpus based on spatial, temporal and morphological criteria. The organisation of the methodological approach follows the data flow from in-situ (or in-lab) production of data to their long-term archiving through the different stages of human-driven and computer-assisted semantic enrichment.

- **In situ production of enrichment-enabled data.** This involves the development of documentation processing methods to support daily data production on the restoration site [12], particularly in the operations of sorting, cataloguing and analysing remains. Starting from the indexing of hundreds of thousands of images, we focus on the implementation of innovative methods for the daily monitoring of the site. The aim is to coherently combine the production of scientific data on site with its natural uptake into the digital platform.
- **Spatial and temporal distribution of data streams.** The aim is to define new methods for integrating and referencing datasets from the construction site. Starting from existing methods [13] to integrate data with spatial reference (3D point clouds, photogrammetry, multispectral images, ...), we propose to extend the same approaches to data types that do not yet have spatial and temporal reference (technical drawings, sound recordings, materials and structures...).
- **From data to disciplinary knowledge.** The focus is on defining a protocol for storing, indexing and contextualising data, based on a detailed study of the specificities of the different disciplines mobilised on the site and on the FAIR (Findable, Accessible, Interoperable and Reusable) principles. These principles are implemented by assigning unique identifiers, using controlled vocabularies and structures, building metadata records [14] that fully exploit the potential of the languages of the semantic web, or adapting to standards such as CIDOC CRM and its extensions.
- **Enriching process storage and long-term archiving of data.** This issue goes through the entire data flow to store the key moments of its semantic enrichment (actor, technical or intellectual activity, ...) and contextualise it within the totality of activities carried out by the community of actors involved in the scientific site. This work also includes the alignment of data, metadata and paradata with existing standards, especially in relation to the challenges of long-term preservation of 3D-based cultural content [15].

The methodological issues described above form the core strategy for building our digital platform. Each of these facets underlines our ambition to seamlessly combine technology and human expertise to jointly produce rigorous and relevant scientific documentation in the field of conservation. The following sections of this article address specific areas of our efforts. Section 4 addresses our efforts to create a coherent collaborative space that brings together a variety of skills and expertise. Section 5 describes the technical dimension of our system and highlights how technology supports and extends our interdisciplinary collaboration. Section 6 presents typical examples of our data curation processes, providing an overview of our approach. In the conclusion, we summarise the key findings of our methodology and provide insight into the upcoming developments and future perspectives of our platform.

Building the collaborative framework

Digital community integration

One of the main concerns of the Digital Data Working Group (WG) is to build a community around the digital framework, made

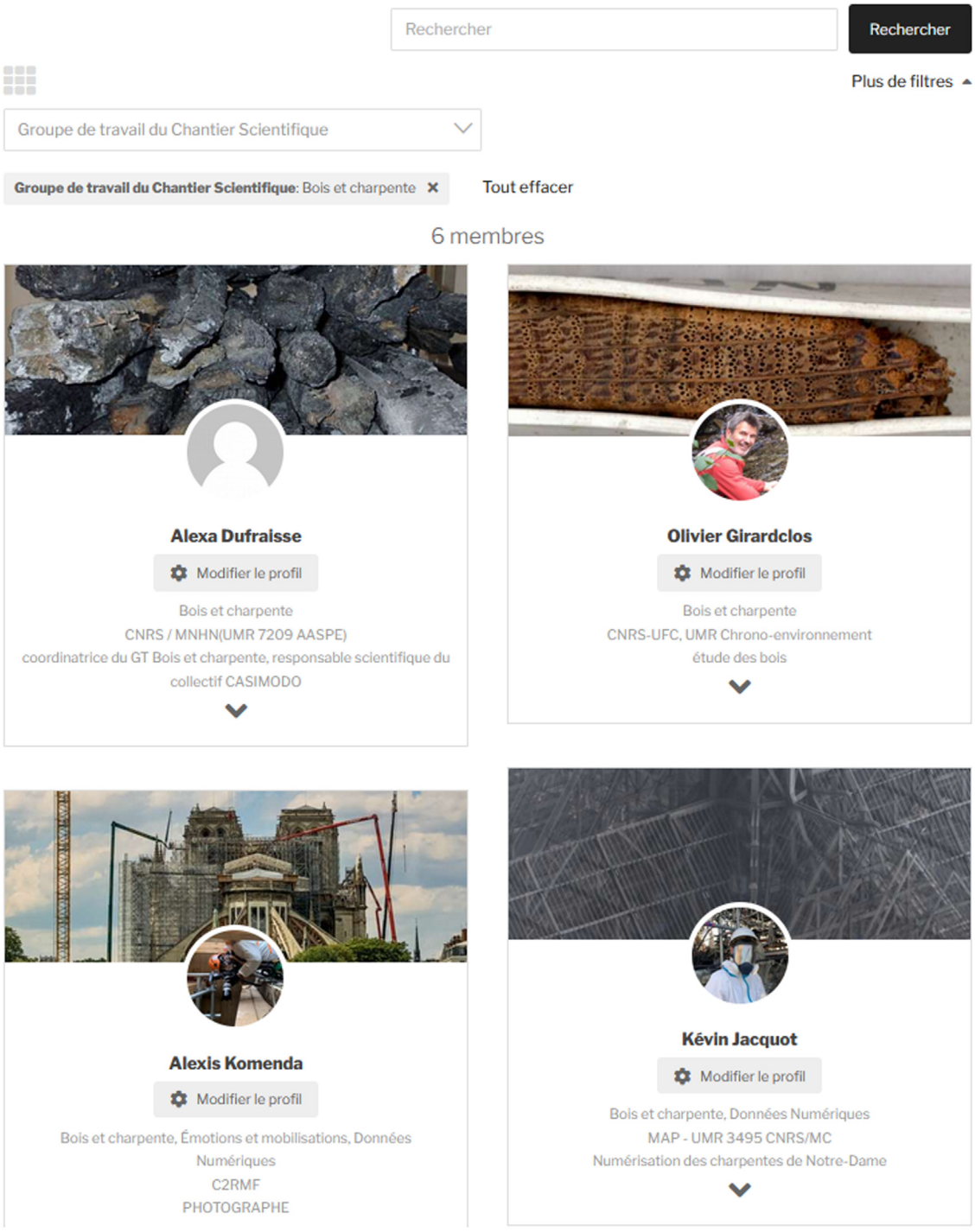

Fig. 1. Some users belonging to Wood WG are shown in the member directory of the collaborative website.

up of those involved in the Notre Dame scientific project and the cathedral conservators. For this integration, a synergistic platform for the Notre Dame scientific action has been set up, complementing the official portal of the project [16] and anchored on a common infrastructure with unique authentication via Huma-Num identifier. Consequently, the platform is connected to the Huma-Num directory via the Lightweight Directory Access Protocol (LDAP) to manage user interactions.

The collaborative portal allows users to access tools curated by Digital Data WG. Support options have also been integrated in the form of user guides, demonstration videos and a support forum. An interactive forum serves a dual purpose: to support users and to promote discourse amongst participants. The membership directory highlights the areas of expertise of project participants, leading to interdisciplinary engagement (see Fig. 1).

Virtual events such as practical workshops are organised, attended by different participants of the working group. These events are crucial for community building and interaction between the working groups. In addition, topic-specific mailing lists are set up to promote scientific discourse on research topics. These communication channels serve to chronicle the discourses that develop during the academic endeavour and summarise the questions and

insights brought forward by researchers during the course of the initiative.

Data paradigms and collection

Following the devastating fire at Notre Dame Cathedral, numerous photographs have appeared in a wide variety of formats. This compilation includes a rich selection of photographic images, architectural drawings, iconographic archives, 3D lasergrammetric point clouds, videography and audio recordings.

Even before the fire, remarkable datasets contributed significantly to our understanding. Andrew Tallon's lasergrammetric datasets of the entire cathedral from 2010, Didier Groux's extensive photographic portfolio and Rémi Fromont's and Cédric Trentesaux's architectural investigations of the roof are outstanding examples. Thanks to the support of the French Ministry of Culture and the RNDP, historical records were collected a few months after the fire and made available to the researchers and professionals involved in the restoration project. Beyond the historical documentation, a new chapter of documentation is being opened, characterised above all by the compilation of 3D digitisation, photographic surveys and technical data on the state of conservation of the cathedral after the fire, produced by private companies (such as Art Graphique et Patrimoine, GEA, Life3D, …) and research laboratories belonging to the scientific action. Since 2019, this digital mosaic has been further enriched by the ongoing efforts of the digital data WG and other thematic WGs, contributing to a cumulative dataset of 30 terabytes. This dynamic collection now includes more than 180,000 photographs, 5000 3D point clouds, hundreds of engineering drawings, and numerous structured 3D models that offer insights into the cathedral before and after the fire. In addition, an archive of 5000 documentary sources, including archives, bibliographies and iconography, offers insights into the history of Notre Dame.

Building the technological framework

Our digital platform includes a set of tools designed for (or adapted to) stakeholders involved in the Notre Dame science project. This toolkit supports functions such as data ingestion (Esmeralda), content indexing (ArcheoGRID), thesaurus structuring (Opentheso), interactive 3D exploration of archaeological remains (3DHOP), image-based 3D annotation of architectural scenes (Aioli) and multimodal exploration of the cathedral-scale data corpus (NDP 3D Viewer). The following sections describe in detail the functions of these tools and how they are linked.

Esmeralda: a web interface for data ingestion

Esmeralda provides a dedicated web interface tailored to Notre Dame research initiative stakeholders, enabling efficient data storage on the Huma-Num Box server, a system that operates exclusively via Secure File Transfer Protocol (SFTP). This tool seamlessly integrates and automates several key functions:

- normalised structuring of metadata,
- generation of thumbnails for visualisation in ArcheoGRID,
- monitoring of data transfer via the integrative tools of the digital platform.

Basically, the Huma-Num Box acts as a central repository containing both archived (cold) and frequently accessed (warm) datasets, strategically distributed across a synergistic server network. The data, supplemented by associated metadata, are organised in directories delineated by the contributing institution, and access modalities are determined by the authoritative Digital Data WG.

When launching the Esmeralda interface (see Fig. 2), users are confronted with the task of finding a metadata schema that matches the typology of their dataset. These schemas cover a wide range: from basic frameworks such as Dublin Core to complex schemas tailored to specific data attributes (created in collaboration with other thematic working groups). After selecting the schema, users are guided through the data upload sequence, specifying the most relevant metadata for optimal storage on the Huma-Num Box server. Following this, the data and metadata are merged and fine-tuned. At this point, integrating controlled lexicons from Opentheso and adding entries to the thesaurus (by a semi-automatic process) becomes plausible. After rigorous validation, users can recover a data package prepared for secure storage in the Huma-Num Box infrastructure.

ArcheoGRID: a comprehensive database for collaborative documentation of digital multimedia assets

ArcheoGRID [18] is a robust, web-based environment for the categorisation, documentation and dissemination of data. A notable feature that ensures its longevity is the ability to tag data with a distinctive identifier (DOI). The synergistic link between ArcheoGRID and the National 3D Data Repository [19], which emerged from the Huma-Num "3D for Humanities", facilitates the publication of 3D datasets and associated documentation originating from both ArcheoGRID and the Notre Dame initiative.

A variety of data and metadata schemas are housed within the ArcheoGRID ecosystem, facilitating the seamless assimilation of the heterogeneous data spectrum of the Notre Dame project. An automated batch process controls the ingestion of data and metadata (from Esmeralda) into a wide-ranging technical system. After assimilation and subsequent enrichment in ArcheoGRID, the data is easily accessible via dedicated APIs. One of the most important APIs developed in the Notre Dame initiative is responsible for extracting comprehensive resource details within ArcheoGRID and passes them to the NDP 3D Viewer (see Section 5.6). In parallel, ongoing efforts are being made to map metadata congruently with CIDOC–CRM to improve data interoperability. In addition, the rescind integration of an IIIF image server enhances the platform's capabilities, particularly in facilitating the visualisation and the annotation of high resolution images.

In the ArcheoGRID paradigm, the notion of 'object' — whether tangible (like a voussoir) or digital (like a virtual simulation) — serves as a nexus that connects disparate documents or metadata. This hodgepodge of data is conceived as a comprehensive digital passport of the object and made accessible via a QR code. Within the framework of an experimental approach, some selected voussoirs analysed by the Stone WG of the main scientific action, have been equipped with this digital passport. The corresponding QR codes have been materialised and attached to the physical voussoirs so that an instant scan can display a range of data: from the findings curated by the Stone WG to visual material from different teams, 3D object simulations and other pertinent information (as shown in Fig. 3).

Opentheso: a comprehensive thesaurus management system

Opentheso [20], an open source software, is developed on MOM [21] in collaboration with the FRANTIQ network [22], the MASA consortium [23] and the Huma-Num infrastructure [24]. It is primarily designed as a web-centric thesaurus management apparatus whose main functionalities include vocabulary management and structured lexicon creation.

Within the Notre Dame initiative, the thematic working groups work with specific lexicons relevant to their respective scientific fields. The challenge was to develop a streamlined solution to unify

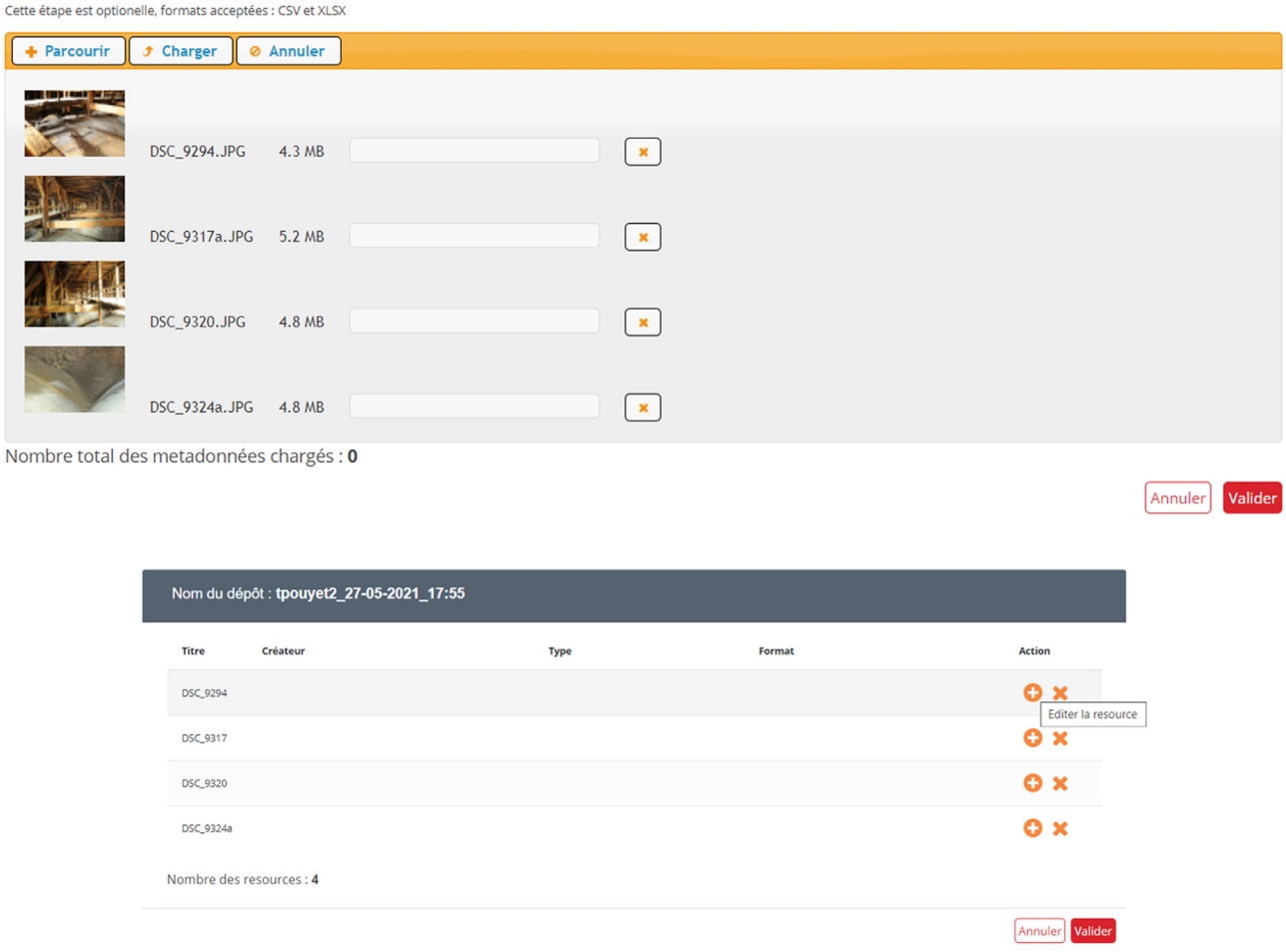

Fig. 2. Esmeralda web service interface for data ingestion.

and synchronise these different vocabularies in a single vocabulary management system. A collaborative project was initiated to formulate this reservoir of important controlled vocabulary using the following features:

- **Integrated matching module:** This facilitates the synchronisation of thesauri with freely available vocabularies, including but not limited to Wikidata [25], Geonames [26], IdRef [27], Agrovoc [28] and Getty AAT [29].
- **Terminology candidate management:** This module enables collaborative improvement of thesauri. A unique discussion and voting system is used to curate potential terms.
- **Change tracking:** The system keeps an accurate record of changes made to the thesaurus, which are attributed to the relevant contributor.
- **Automatic term completion:** This feature, which helps users select terms, is based on generic widgets associated with Opentheso.
- **Permanent identifier:** Each concept in the thesaurus is assigned a unique identifier via the system ARK [30] (Archival Resource Key). This identifier is crucial for linking various software components of Notre Dame's digital platform, including Aioli [31], ArcheoGRID [18], NDP 3D Viewer and Esmeralda [17].

Opentheso is the nerve centre of the digital platform and enables intricate cross-referencing between the different tools, as shown in Fig. 4.

A dedicated 3D viewer for studying the voussoirs of archs

A voussoirs viewer has been developed using 3DHOP (3D Heritage Online Presenter) [32], an open-source framework for 3D visualisation on the web platform targeted at cultural heritage applications. 3DHOP uses the WebGL component of HTML5, which enables native 3D rendering in browsers on all platforms. Thanks to the multiresolution encoding of the 3D models, which is compressed and streaming-friendly, it is possible to use the high-resolution geometry of the voussoirs directly, ensuring maximum consistency with the measured data. 3DHOP can be easily customised to develop a visualisation web application tailored to a specific project.

The goal in developing this viewer was to provide as many options/tools for navigation, visualisation and measurement as possible, creating a sandbox environment that meets the diverse needs of the experts involved in the project.

The viewer offers a simple two-axis navigation (turntable): this paradigm is easy to use, even for non-experts in 3D, but still allows to reach all parts of the voussoir. The user can switch between two different object poses: "flat", as the object is currently stored, and "upright", which is more like the original positioning. Predefined views allow the user to easily jump to any of the numbered pages. A small gizmo shows the current pose/orientation of the object to help the user stay orientated.

It is possible to control the lighting and change the direction of the light (to allow for grazing light inspection), adjust the surface transparency and gloss, and turn the colour on and off. A real-time

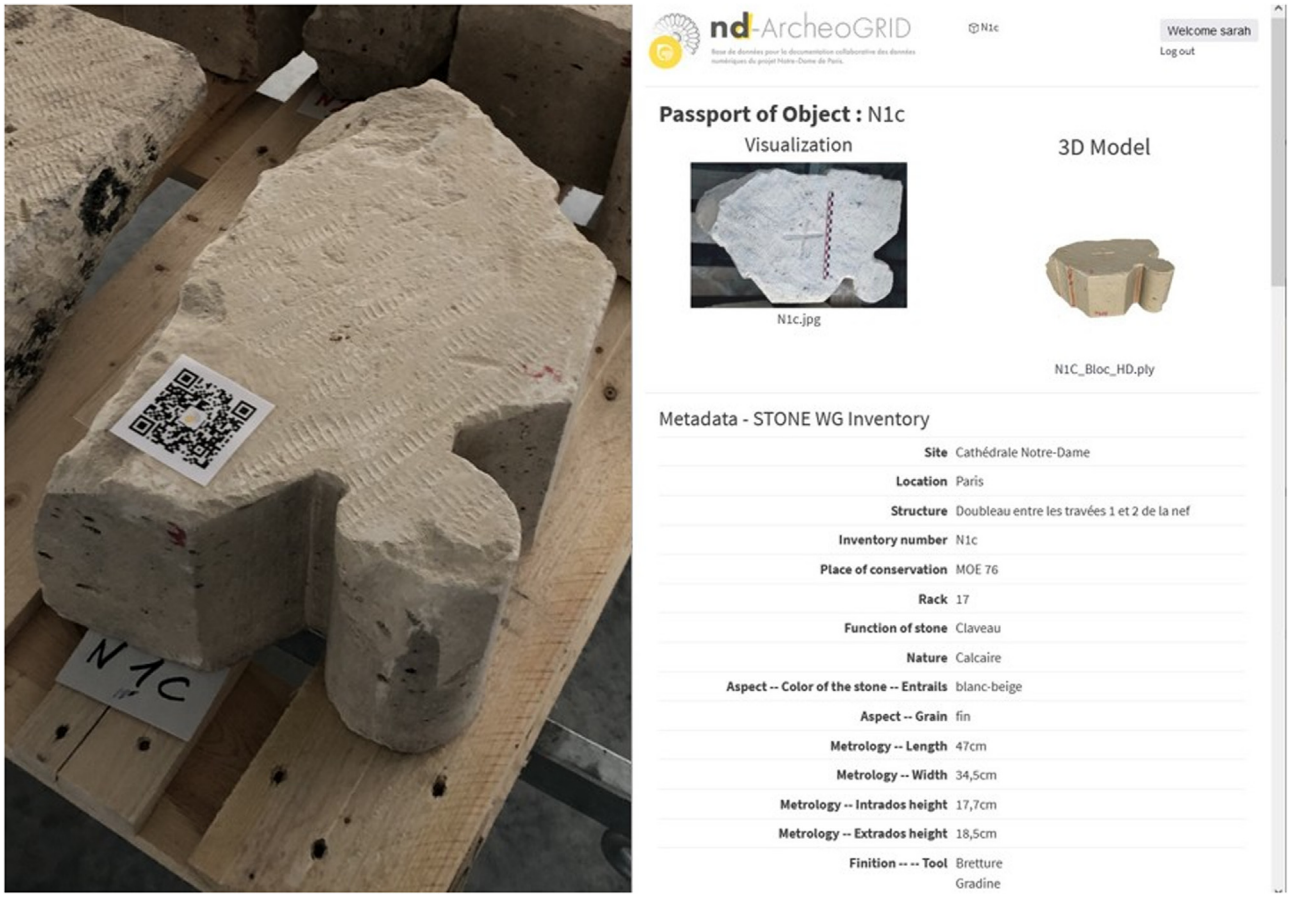

Fig. 3. QR Code of a voussoir and its digital passport in the ArcheoGRID database.

section can be used to cut the voussoir interactively along its main axes.

Interactive tools allow you to measure point-to-point distances and angles directly on the 3D surface. Metric grids are available to estimate the size of object details.

To support the process of distributed collaborative exploration of the voussoirs, the user can add annotations to the 3D object, place annotated points on the surface, and save annotated views over the voussoirs; the views also record the status of the viewer (i.e. light direction, rendering options) (Fig. 5). These annotations can then be saved locally and shared with the other experts working on the voussoirs. Screenshots can also be saved directly via the web interface. On-going works are focusing on expanding the possible annotation types by introducing polyline and area-based annotations.

Aioli: an integrated framework for collaborative 3D annotation

Aioli [31] has been developed to manage data & information related to multiple aspects of material heritage asset, encompassing geometry, visual features and semantics. This integrated approach to documentation ensures a seamless transition from reality-based data capture to analysis and interpretation, in a collaborative environment. Aioli pioneers the linking of information between material objects and their digital counterparts and has several integrated features [33]:

- **Incremental image-based 3D spatiality**: this process facilitates the geometric fusion of photographs originating from different authors at different points in time.
- **2D/3D annotation system**: users can delineate, visualise and register relevant surface areas using 2D images that are spatially synchronised with evolving 3D representations.
- **Multi-layered morphological data structuring**: this model provides a granular description of tangible objects and captures their intricate geometric nuances while incorporating observations from an interdisciplinary perspective.

Within Aioli, annotations can be enriched using customisable description templates. These allow users to define their own data architecture by attaching different fields: text content, dates, URLs, concepts derived from the thesaurus and even additional documents (such as images, PDFs, audio files or videos). Fields based on thesaurus concepts can be linked to Opentheso, and data entry is simplified by an autocomplete function that is limited to terms within the associated thesaurus. Significantly, Aioli's data structuring paradigm is consistent with the standards of CIDOC–CRM. To demonstrate Aioli's efficiency in monitoring cathedrals, comprehensive diagnostic reports (and related scientific data) were transcribed and integrated into the Aioli system (see Section 6.3).

NDP 3D viewer: a visualisation environment for the exploration of heterogeneous digital corpuses

The diversity of the digital corpus associated with the cathedral - characterised by different file formats, geometric nature, temporal states and focal regions - necessitated the development of a dedicated 3D web viewer. The NDP 3D Viewer, mainly based on the PotreeJS library [34], was developed to spatially centralise and align all digital assets related to the cathedral, regardless of their production date in relation to the fire event and the restoration

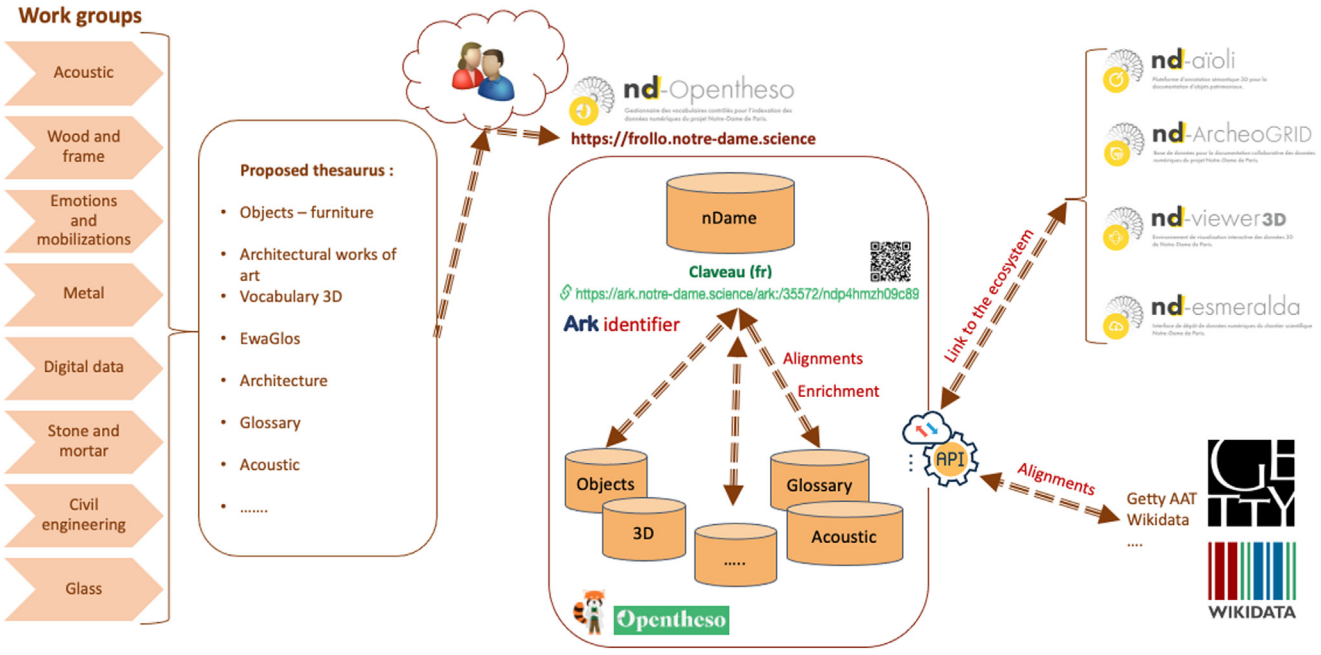

Fig. 4. Management of the Notre Dame thesaurus in Opentheso.

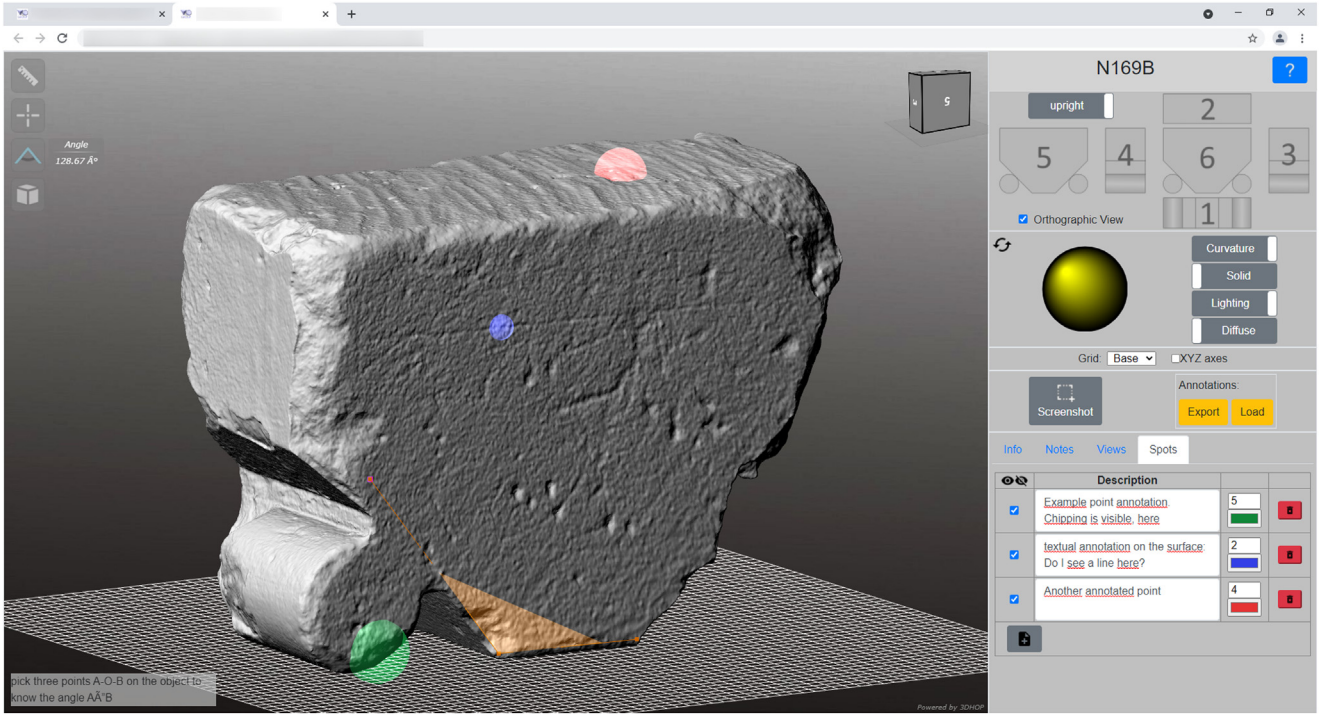

Fig. 5. The 3D viewer of the voussoirs, showing the geometry of an element with grazing light and annotated points.

process. This compilation of digital assets requires a preliminary registration process for each data item. The overall objective is to facilitate the integrated visualisation of assets obtained or created by each working group associated with the Notre Dame scientific project. This aggregation is intended to promote the discovery of potential correlations between different analyses and observations.

Any dataset can be integrated into the viewer, provided that its spatialisation parameters can be determined. This integration establishes a permanent link to the ArcheoGRID database, so that each component is permanently linked to the corresponding metadata. In this way, the lexicon used for documentation remains accessible, promoting efficient data use and ensuring traceability. To improve user navigation, the spatial domain of the cathedral has been divided into different volumes that correspond to the restoration project's spatial nomenclature defined in Opentheso. For example, the user can choose to visualise components specific to the choir, limit themselves to the stands of the apse, or further limit themselves to a specific area. This spatially orientated categorisation serves as an initial filter for the 3D visualisation. A similar principle could be applied to metadata to facilitate the creation of filters based on documentary or semantic affiliations.

Fig. 6. Virtual anastylosis of the voussoirs of the collapsed arch in the nave of the cathedral, displayed in the NDP 3D Viewer.

The dynamic nature of the NDP 3D Viewer allows for iterative improvements to meet the evolving needs of both restoration and comprehensive scientific research. Collaboration with each thematic working group allows for the incorporation of new features tailored to their specific needs. For example, collaboration with the Stone Working Group has led to the development of a dedicated tool for the virtual anastylosis of demolished vaulted arches in the nave of the cathedral (see Fig. 6). This tool allows users to quickly evaluate different localisation hypotheses. It allows instant positioning, rotation, colour changes to indicate degrees of uncertainty and optimised export of hypothesis aggregates.

Examples of data curation workflows within the digital platform

Systematic categorisation of photographs for the chronological documentation of remains recovery

The need to meticulously document the condition of the cathedral after the fire began with a rigorous recording protocol. Immediately after the fire, the recovery of the remains followed a predetermined protocol that facilitated the identification and enumeration of the recovered components of the cathedral. Through this systematic approach, these elements were sorted by container type and then stored on pallets.

During each stage of this sorting process, photographs were carefully taken from different angles by a number of field workers (see Fig. 7).

Subsequently, the focus was on structuring a hierarchical taxonomy in anticipation of comprehensive data use. This included first and foremost the establishment of uniform file naming conventions, followed by a logical segmentation of the images into specific sub-collections. At the same time, a differentiated classification scheme was created.

To facilitate the efficient allocation of the photos to the directories of the scheme, a methodical approach was chosen that relied on the EXIF timestamps of the images. This timestamp quickly became the linchpin of the entire image management system. The classification scheme was essentially divided into two segments: one for sorting the images and the other for all remaining location photos.

The documentary enhancement of these images was then carried out through the use of controlled vocabularies and a complex indexing mechanism. These lexicons were used to create a customised mini-thesaurus. Images were then indexed using these hierarchical terms in conjunction with systematically named files.

To complete this process, the unique metadata of each image was extracted to create a comprehensive output directory to accompany the images in their designated storage area. Currently, this documentation process includes about 60,000 images.

Harmonisation of multidisciplinary lexicons

To achieve synchronised alignment of multidisciplinary lexicons, the Digital Data Working Group (WG) has pioneered the development of a dedicated thesaurus in Opentheso. This approach facilitates the uniform indexing of documents based on terminologies jointly selected by the expert community. The goals are multifaceted and include the creation of a unified semantic environment for the Notre Dame scientific initiative, the merging of different terminologies and standardised description methods for optimal data retrieval.

The Notre Dame thesaurus [35] is a collaborative lexicon composed of micro-thesauri created by the respective WG. It currently includes about 4000 terms covering various areas such as classifications of actors, architectural concepts, scientific processes, types of materials, construction methods and restoration techniques. It also contains a nomenclature curated by the Digital Data WG to precisely locate the individual components of the cathedral (see Fig. 8).

The Notre Dame Thesaurus is organised by a grid of interrelated concepts and relies on hierarchical and associative relationships to structure its content. The concept aggregates are stratified both thematically and by WG, allowing for greater accessibility and fluid navigation within the thesaurus while clarifying disciplinary affiliations.

In its quest to create a comprehensive lexicon, the Digital Data WG has also integrated existing thesauri for adaptive reuse. These

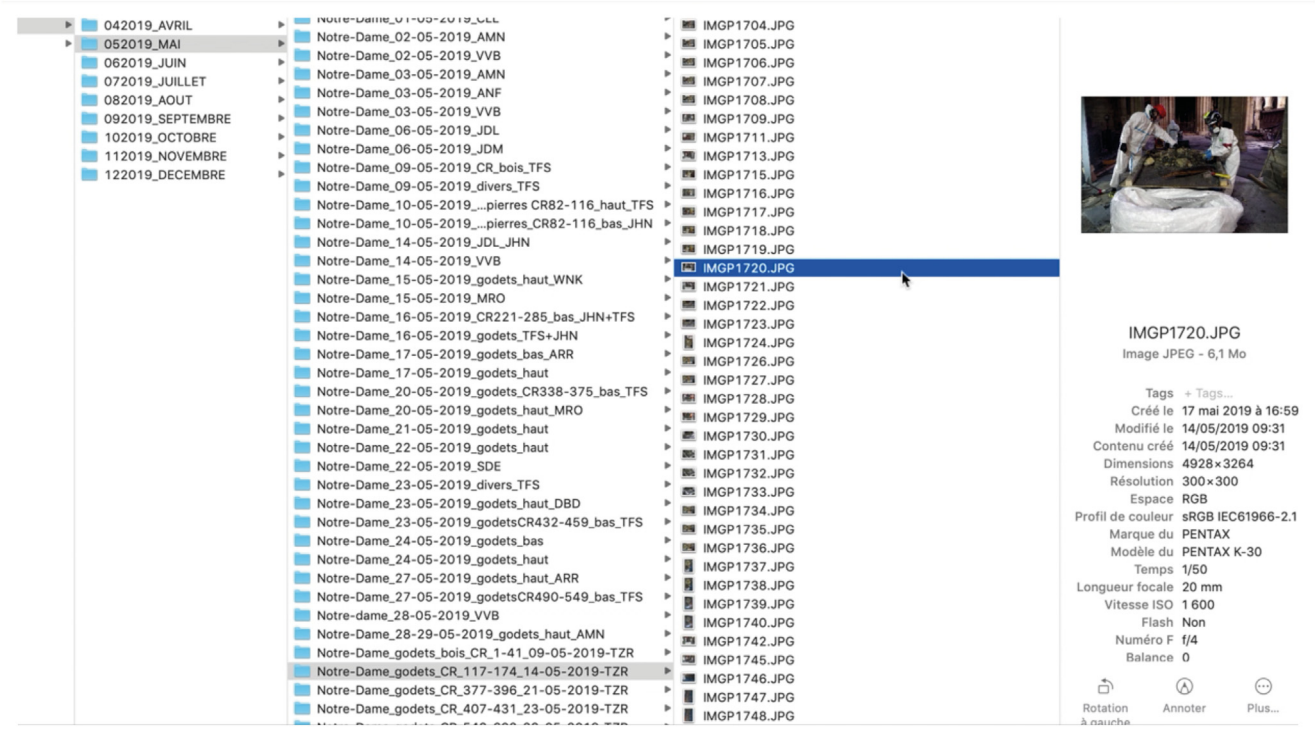

Fig. 7. Chronological photographic documentation of the remains recovery operations.

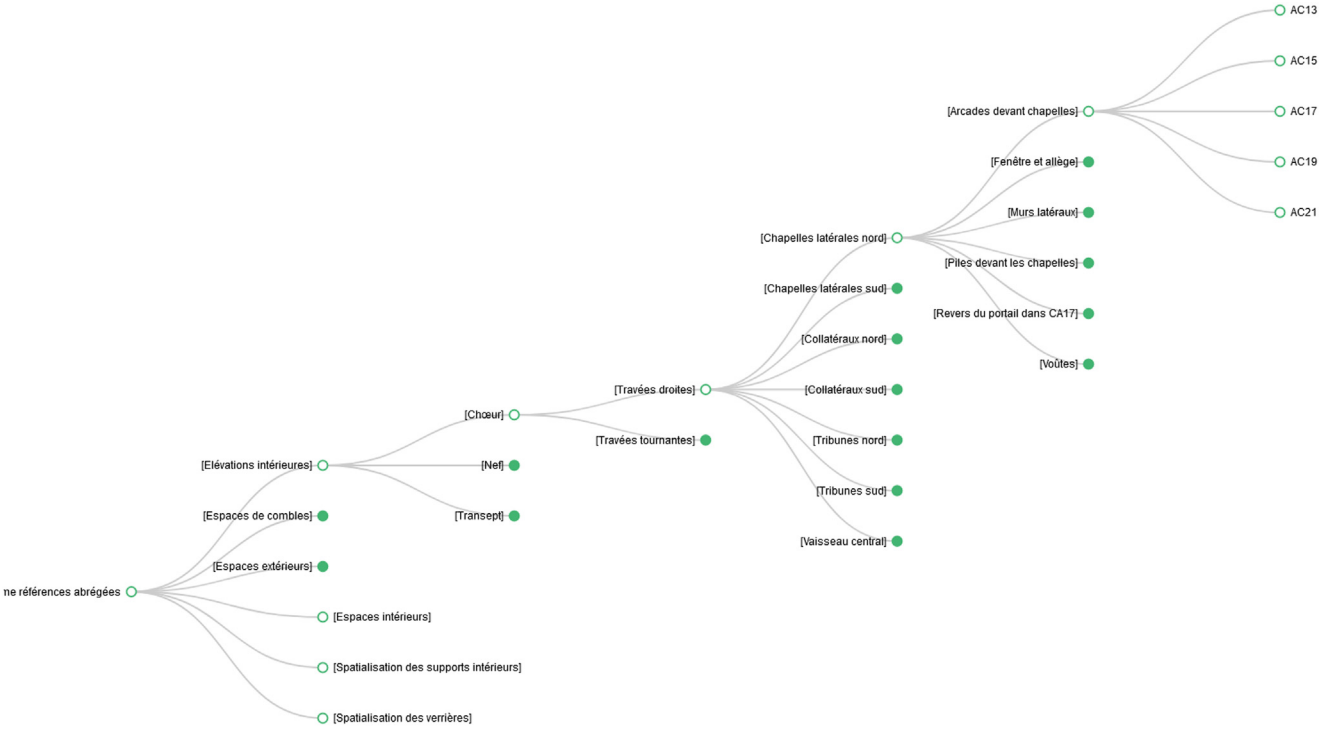

Fig. 8. Graph-based representation of a nomenclature branch in Opentheso.

include authoritative vocabularies from institutions such as the french Ministry of Culture [36] and analytical lexicons of technology from projects such as PARCOURS [37]. This drive to integrate multidisciplinary references also extends to specialised glossaries, notably the architectural vocabulary of Jean-Marie Pérouse de Montclos [38].

To ensure contemporary adaptability, these assimilated glossaries have been transformed and aligned with the Semantic Web for Interoperability (SKOS) model [39]. Notable examples include the 3D Vocabulary Lexicon for the Humanities and Social Sciences by the French consortium 3D SHS [40] and the illustrative glossary on the decay of stones by ICOMOS-ISCS [41].

2D/3D Annotation framework and cathedral diagnosis

Using the Aioli platform, a comprehensive digital condition report of the cathedral was created. Initial tests were carried out involving the acquisition of several sets of photogrammetric ac-

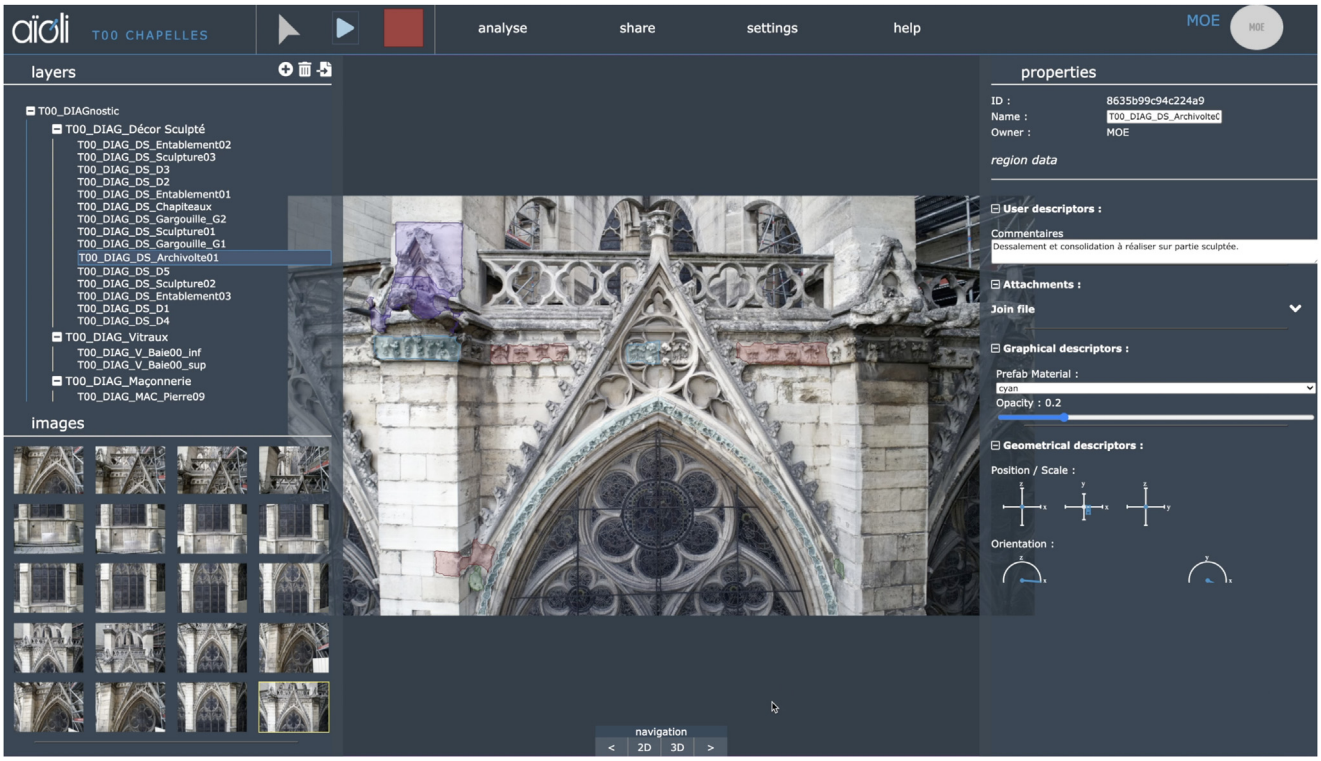

Fig. 9. 2D/3D annotation process in the Aioli platform for the cathedral condition report.

quisitions (including intricate structures such as the vaults of the nave, the choir and the remains of the collapsed wooden framework scattered over the vaults) and intricate details (such as the stained glass). These projects were systematically documented by using the 2D/3D annotation approach, based on observational analysis of their current condition's state. The preliminary analysis confirmed the platform's ability to effectively manage several hundred annotations, including custom descriptors and additional attachment files with detailed observations (see Fig. 9).

After this initial assessment, the extensive process of organising and categorising the 50,000 exterior photographs of the cathedral began. This involved dividing the cathedral spatially and hierarchically, resulting in the formulation of hundreds of individual projects that together represent the entire exterior and interior structure of the cathedral. Each project was meticulously annotated to reflect the official diagnostic report, including associated colour schemes, captions and additional descriptions. Detailed photographs have been included as appendices, while user descriptions provide additional insight into the state of preservation and historical context of specific segments. Each project's unique identification number facilitated its integration into the NDP 3D Viewer, allowing users to either view individual projects or group them together for a holistic diagnostic overview (and multi-lenses documentation) of the cathedral.

Conclusion

The digital platform we have presented in this article is characterised by the fusion of two directions of innovation. First, it goes beyond the usual metadata documentation by capturing the scientific journey from raw data to informed interpretation. Second, it uses a complex data correlation approach rooted in the digital twin of the cathedral to reveal intertwined regions of enquiry that are closely linked spatially, temporally or semantically across multiple disciplinary profiles of a single cultural monument. This approach not only underlines the multidisciplinary nature of the research activities, but also provides fertile ground for spontaneous semantic connections that foster advanced data analysis. Bringing together approaches from the humanities and social sciences (HSS) and information and communication technology (ICT), our approach aims to create a comprehensive Notre Dame-centred dataset that encompasses both data production and semantically enhanced data curation. With its reproducible methods and adaptable technology, the project paves the way for similar undertakings on other historic properties. It embodies the possibility of moving from individual practises to a collective framework that places heritage data at the centre of interdisciplinary scientific research.

Acknowledgements

The Digital Data Working Group of the Notre Dame Scientific Action is supported by several institutions and projects. These include the Centre National de la Recherche Scientifique (CNRS), the French Ministry of Culture, the Établissement public chargé de la conservation et de la restauration de Notre-Dame de Paris (EPRNDP), the Fondation des Sciences du Patrimoine (FSP) and the European Research Council (ERC) Advanced Grant nDame_Heritage project funded by the European Union.

The authors thank Jonathan Truillet (EPRNDP), Philippe Villeneuve, Pascal Prunet, Rémi Fromont (ACMH) for access to data from the cathedral restoration site and Joël Marchand (CNRS) for technical assistance with Huma-Num services.

References

[1] P. Ball, The huge scientific effort to study Notre-Dame's ashes, Nature 577 (2020) 153–154, doi:10.1038/d41586-020-00008-5.
[2] Le chantier scientifique Notre-Dame de Paris. Ministère de la Culture. Available: https://www.culture.gouv.fr/Thematiques/Sciences-du-patrimoine/Thematiques-de-recherche/Le-chantier-de-Notre-Dame-de-Paris. Last accessed on July 6, 2023

[3] Members of the digital data woring group of the Notre Dame scientific action. Available: https://www.notre-dame.science/membres-gt-donnes-numeriques/. Last accessed on July 6, 2023.
[4] C. Boje, A. Guerriero, S. Kubicki, Y. Rezgui, Towards a semantic construction digital twin: directions for future research, Automat. Construct. 114 (2020) 103179, doi:10.1016/j.autcon.2020.103179.
[5] M. Doerr, Definition of the CIDOC conceptual reference model, ICOM/CIDOC, 2010.
[6] I. Bannour, C. Marinica, L. Bouiller, O. Malavergne, et al., CRMcr - a CIDOC-CRM extension for supporting semantic interoperability in the conservation and restoration domain, 2018. HAL: hal-01872164.
[7] F. Ponchio, M. Callieri, M. Dellepiane, R. Scopigno, Effective annotations over 3D models, Comput. Graphics Forum 39 (2020) 89–105, doi:10.1111/cgf.13664.
[8] F. Maiwald, T. Vietze, D. Schneider, F. Henze, et al., Photogrammetric analysis of historical image repositories for virtual, reconstruction in the field of digital humanities, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Copernicus GmbH, 2017, pp. 447–452, doi:10.5194/isprs-archives-XLII-2-W3-447-2017.
[9] G. Muehlberger, L. Seaward, M. Terras, S.Ares Oliveira, et al., Transforming scholarship in the archives through handwritten text recognition: transkribus as a case study, JD 75 (2019) 954–976, doi:10.1108/JD-07-2018-0114.
[10] D. Oldman, D. Tanase, Reshaping the knowledge graph by connecting researchers, data and practices in researchspace, in: The Semantic Web - ISWC 2018, Springer, Cham, 2018, pp. 325–340, doi:10.1007/978-3-030-00668-6_20.
[11] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, et al., Building Rome in a day, Commun ACM 54 (2011) 105–112, doi:10.1145/2001269.2001293.
[12] C. Niang, C. Marinica, E. Leboucher, O. Malavergne, et al., Supporting semantic interoperability in conservation-restoration domain: the parcours project, J. Comput. Cult. Heritage 10 (2017) 16, doi:10.1145/3097571.
[13] A. Pamart, F. Morlet, L. De Luca, P. Veron, A robust and versatile pipeline for automatic photogrammetric-based registration of multimodal cultural heritage documentation, Remote Sens. 12 (2020) 2051, doi:10.3390/rs12122051.
[14] X. Rodier, O. Marlet, Digital ecosystem for the French archaeological community, (2019). DOI: 10.34894/BDC22F.
[15] A. Alaoui M'Darhri, V. Baillet, B. Bourineau, A. Calantropio, G. Carpentiero, et al., Share - Publish - store - preserve. methodologies, tools and challenges for 3d use in social sciences and humanities, parthenos and consortium 3D-SHS and LIA MAP-IS.
[16] Chantier scientifique Notre-Dame de Paris. Available: https://www.notre-dame.science. Last accessed on July 6, 2023.
[17] Esmeralda, interface de dépôt de données numériques du chantier scientifique Notre-Dame de Paris. Available: https://esmeralda.notre-dame.science. Last accessed on July 6, 2023.
[18] ArcheoGRID, Archeovision. Available: https://www.archeogrid.fr Last accessed on July 6, 2023.
[19] Conservatoire National des Données 3D, Consortium 3D SHS. Available https://3d.humanities.science Last accessed on July 6, 2023.
[20] Carnet Hypothèses sur Opentheso. Available: https://opentheso.hypotheses.org Last accessed on July 6, 2023.
[21] Maison de l'Orient et de la Méditerranée. Available: https://www.mom.fr Last accessed on July 6, 2023.
[22] Fédération et Ressources sur l'Antiquité. Available: https://www.frantiq.fr Last accessed on July 6, 2023.
[23] Carnet de recherche sur Hypothèses sur le Consortium MASA. Available: https://masa.hypotheses.org Last accessed on July 6, 2023.
[24] TGIR Huma-Num. Available: https://www.huma-num.fr Last accessed on July 6, 2023.
[25] Wikidata. Available: https://www.wikidata.org Last accessed on July 6, 2023.
[26] GeoNames geographical database. Available: https://www.geonames.org Last accessed on July 6, 2023.
[27] Identifiants et Référentiels pour l'enseignement supérieur et la recherche. Available: https://www.idref.fr Last accessed on July 6, 2023.
[28] AGROVOC multilingual thesaurus. Available: https://www.fao.org/agrovoc/fr/search Last accessed on July 6, 2023.
[29] Art & architecture thesaurus. Available: https://www.getty.edu/research/tools/vocabularies/aat Last accessed on July 6, 2023.
[30] ARK alliance, home of the archival resource key. Available: https://arks.org Last accessed on July 6, 2023.
[31] Aioli, a reality-based 3D annotation platform for the collaborative documentation of heritage artefacts. Available: http://www.aioli.cloud Last accessed on July 6, 2023.
[32] M. Potenziani, M. Callieri, M. Dellepiane, M. Corsini, F. Ponchio, R. Scopigno, 3DHOP: 3D heritage online presenter, Comput. Graph. 52 (2015) 129–141 November 2015ISSN 0097-8493.
[33] A. Manuel, A. Alaoui M'Darhri, V. Abergel, F. Rozar, L. De Luca, A semi-automatic 2D/3D annotation framework for the geometric analysis of heritage artefacts, in: 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), 2018, pp. 1–7, doi:10.1109/DigitalHeritage.2018.8810114.
[34] Potree library on GitHub. Available: https://github.com/potree/potree Last accessed on July 6, 2023.
[35] Notre Dame thesaurus in Opentheso. Available: https://frollo.notre-dame.science Last accessed on July 6, 2023.
[36] Ministère de la Culture, SKOS vocabularies. Available: http://data.culture.fr/thesaurus/ Last accessed on July 6, 2023.
[37] C. Niang et al.: Supporting semantic interoperability in conservation-restoration domain: the PARCOURS project.
[38] Jean-Marie Pérouse de Montclos, 1972. Architecture: méthode et vocabulaire. Ministère des affaires culturelles, Inventaire général des monuments et des richesses artistiques de la France.
[39] W3C, SKOS Primer. Available: https://www.w3.org/TR/skos-primer/ Last accessed on July 6, 2023.
[40] X. Granier et al.: Vocabulaire 3D - Lexique pour les Sciences Humaines et Sociales. Available: https://hal.archives-ouvertes.fr/hal-02297326 Last accessed on July 6, 2023.
[41] ICOMOS-ISCS: Illustrated glossary on stone deterioration patterns. Available: https://www.icomos.org/publications/monuments_and_sites/15/pdf/Monuments_and_Sites_15_ISCS_Glossary_Stone.pdf Last accessed on July 6, 2023.