



**HAL**  
open science

## Des données numériques aux connaissances pluridisciplinaires : enjeux de documentation et de formalisation

Violette Abergel, Pierre Arese, Florent Comte, Iwona Dudek, Anaïs Guillem, Antoine Gros, Kévin Réby

### ► To cite this version:

Violette Abergel, Pierre Arese, Florent Comte, Iwona Dudek, Anaïs Guillem, et al.. Des données numériques aux connaissances pluridisciplinaires : enjeux de documentation et de formalisation. Naisance et renaissance d'une cathédrale - Notre-Dame de Paris sous l'oeil des scientifiques, Apr 2024, Paris, France. hal-04576767

**HAL Id: hal-04576767**

**<https://hal.science/hal-04576767v1>**

Submitted on 17 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Des données numériques aux connaissances pluridisciplinaires: enjeux de documentation et de formalisation

Abergel V.<sup>1</sup>, Arese P.<sup>1</sup>, Comte F.<sup>1</sup>, Dudek I.<sup>1</sup>, Guillem A.<sup>1,2</sup>, Gros A.<sup>1,3,4</sup>, Reby K.<sup>1</sup>

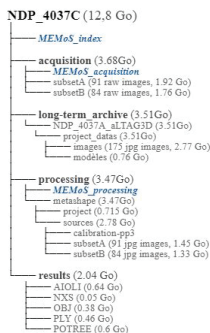
- 1 - CNRS/MAP/Marseille
- 2 - Ludwig Maximilian University of Munich/München
- 3 - CNRS/Laboratoire de Mécanique et de Génie Civil/Montpellier
- 4 - LISPEN EA 7515/Laboratoire d'Ingénierie des Systèmes Physiques et Numériques/Aix-en-Provence



La gestion massive des données multidimensionnelles produites dans le contexte du chantier de Notre-Dame soulève aujourd'hui de nouvelles questions et défis, particulièrement en termes de traçabilité, d'analyse, d'interrogation, ou encore de visualisation. La valeur de ces données numériques réside surtout dans toutes les connaissances qu'elles sont susceptibles de véhiculer pour une communauté d'acteurs à travers des regards orientés. L'enrichissement sémantique permet une connexion durable entre la connaissance humaine et le traitement informatique : notre travail vise à pérenniser toutes ces ressources précieuses et explorer le chemin qui mène de la génération de données à la production de connaissances.



Figure 1. Description d'un processus d'acquisition de données photogrammétriques et enregistrement de métadonnées relatives à un résultat de recherche (research output), 2023. URL: <http://memoria-dev.gamsau.archi.fr/enter.php>



En suivant ce lien, vous pouvez télécharger la publication d'origine :

F. Comte, A. Pamart, K. Réby, L. De Luca. *Strategies and Experiments for Massive 3D Digitalization of the Remains After the Notre-Dame de Paris' Fire*. 10th International Workshop 3D-ARCH «3D Virtual Reconstruction and Visualization of Complex Architectures», Feb 2024, Sienna, Italy.



En suivant ce lien, vous pouvez télécharger la publication d'origine :

I. Dudek, J.-Y. Blaise, M. Rabefandroana. *MEMORIA SI en ligne pour décrire des protocoles de recherche*. *Revue des Nouvelles Technologies de l'Information, 2022, Extraction et Gestion des Connaissances 2022, RNTI-E-38*, pp.453-460.

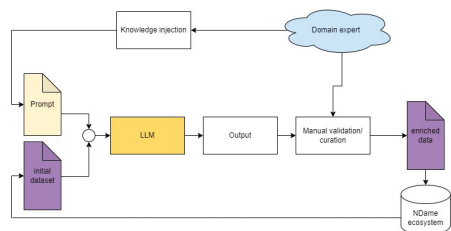
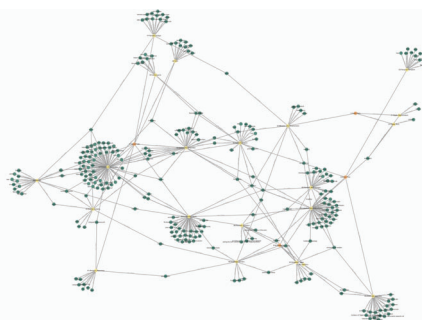


Figure 2. Schéma générique de l'enrichissement de données en utilisant les LLM par la création de prompts, la préparation de jeux de données et l'évaluation des résultats et visualisation des données extraites sous la forme d'un graphe.



## Pérenniser des données multidimensionnelles

La complexité et l'hétérogénéité des données du chantier scientifique de Notre-Dame impose une réflexion sur leur condition de production et leur cycle de vie. En particulier, le volume des données d'acquisition 3D a nécessité de proposer une structure permettant l'enregistrement de chaque étape du travail de numérisation, tout en répondant aux besoins inhérents à leur pérennité à court et long terme. Cet effort s'appuie sur: 1) une automatisation des processus de numérisation, 2) une documentation systématique par la production de métadonnées et parodonnées selon une structure hiérarchique en amont de leur production (MEMoS), 3) Une préparation des données pour leur archivage à long-terme grâce à Altag3D (logiciel de préparation des données).

De plus, avec l'ambition de mémoriser et analyser non seulement les données, mais aussi les processus et résultats des activités de recherche afférentes, nous avons proposé une approche de description à travers le système d'information expérimental MEMORIA.

## Extraire des informations sémantiques

En plus de stratégies d'enrichissement sémantique basées sur l'annotation, permettant d'intégrer des connaissances pluridisciplinaires au sein de représentations numériques 2D et 3D de la cathédrale, nous proposons des approches automatisées permettant d'extraire les éléments clés de notre corpus de documentation scientifique. Pour les textes, nous nous appuyons sur les LLMs (Large Language Models). Ceux-ci nous permettent une curation de contenus efficace et contrôlée, tant dans les actions de sélection, que de formalisation et de partage.

## Des données hétérogènes aux graphes de connaissances

L'ensemble de ces approches offre un cadre méthodologique permettant de construire et caractériser des relations entre les données et les connaissances pluridisciplinaires du chantier scientifique, en exploitant à la fois leurs attributs spatiaux, géométriques, temporels et sémantiques. Il s'agit d'exploiter les données multidimensionnelles de Notre-Dame pour construire un modèle de connaissances capable de supporter de nouvelles stratégies d'analyse, basées notamment sur de la détection de motifs récurrents ou de la classification. L'utilisation conjointe d'approches par graphes et de méthodes de traitement géométrique permet alors de définir de nouvelles modalités d'exploration et d'interrogation de corpus complexes en s'appuyant sur la navigation dans un espace nD.

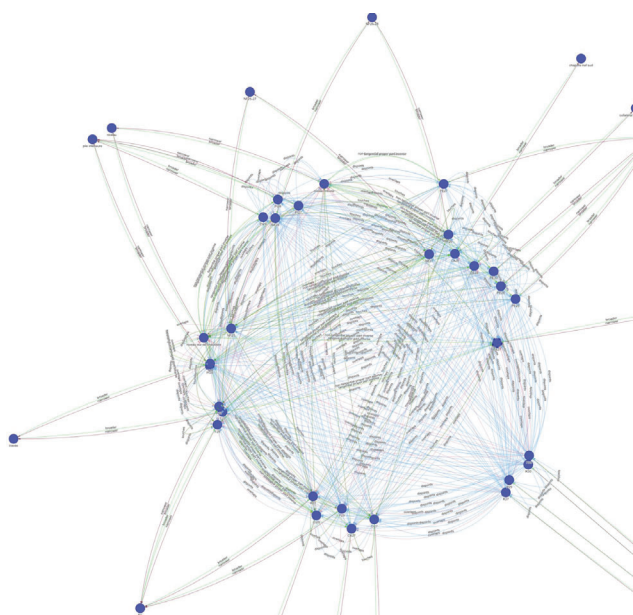
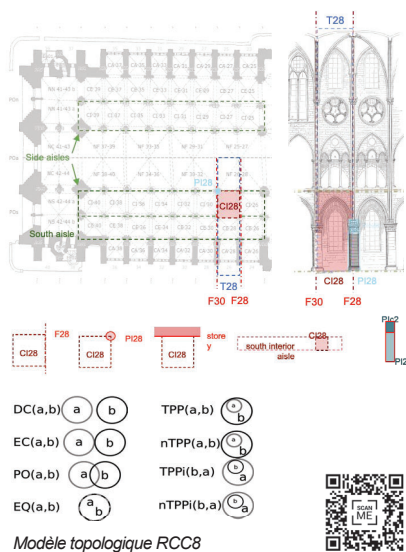


Figure 3. Représentation des relations spatiales des espaces et éléments de Notre-Dame de Paris. Elles sont exprimées avec le modèle topologique du RCC8 sous la forme d'un graphe de données.



Modèle topologique RCC8

En suivant ce lien, vous pouvez télécharger la publication d'origine :

A. Guillem, A. Gros, K. Réby, V. Abergel, L. De Luca. *RCC8 for CIDOC CRM: Semantic Modeling of Mereological and Topological Spatial Relations in Notre-Dame de Paris*. SWOCH'23: International Workshop on Semantic Web and Ontology Design for Cultural Heritage, Nov 2023, Athènes, Greece.