



HAL
open science

Combining FAIR principles and long-term archival of 3D data

Matthieu Quantin, Sarah Tournon, Valentin Grimaud, Florent Laroche,
Xavier Granier

► To cite this version:

Matthieu Quantin, Sarah Tournon, Valentin Grimaud, Florent Laroche, Xavier Granier. Combining FAIR principles and long-term archival of 3D data. ACM Web3D conference, Oct 2023, San Sebastian, Spain. pp.16:1-6, 10.1145/3611314.3615913. hal-04199555

HAL Id: hal-04199555

<https://hal.science/hal-04199555v1>

Submitted on 7 Sep 2023

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 4.0 International License

Combining FAIR principles and long-term archival of 3D data

Matthieu Quantin*

matthieu.quantin@ls2n.fr
Digital laboratory LS2N, Ecole
Centrale de Nantes, Nantes University
Nantes, France

Sarah Tournon

sarah.tournon@cnrs.fr
CNRS, Archéosciences Bordeaux
UMR 6034
Bordeaux, France

Valentin Grimaud

valentin.grimaud@univ-nantes.fr
CReAAH, LARA laboratory, Nantes
University
Nantes, France

Florent Laroche

florent.laroche@ec-nantes.fr
Digital laboratory LS2N, Ecole
Centrale de Nantes, Nantes University
Nantes, France

Xavier Granier

xavier.granier@insitutoptique.fr
Institut d'Optique Graduate School,
LP2N
Bordeaux, France

ABSTRACT

This paper presents an update of the French solution for long-term archiving combined with online publication of 3D research data in the humanities. The choice of data organisation, metadata, standards and infrastructure is in line with the FAIR principles of the semantic web. Since 2017, our consortium 3D for Humanities offers this service through the French National 3D Data Repository for Humanities. This was the start of our collaboration with CINES (Centre Informatique National de l'Enseignement Supérieur), the standard Open Archival Information System (OAIS) infrastructure for research data in France. In 2021, with more than one thousand (1000) projects registered, driven by laboratories from all over the country, open questions were posed to the community. In response, we have developed a new metadata schema that allows a more precise description of the research content, greater openness to other non-archaeological humanities fields, and better FAIR compliance. This metadata schema is aligned with standard vocabularies and mapped to the Europeana Data Model (EDM). Among other online features, a 3D viewer has been implemented to meet the needs of researchers and public communication. As designed, aLTAG3D, the desktop UI software we developed to help research labs create Submission Information Packages (SIPs), has adapted itself to the new schema through the XSD content.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives**; • **Computing methodologies** → **3D imaging**; • **Applied computing** → **Arts and humanities**; **Document management**.

KEYWORDS

3D data, heritage, FAIR, aLTAG3D

ACM Reference Format:

Matthieu Quantin, Sarah Tournon, Valentin Grimaud, Florent Laroche, and Xavier Granier. 2023. Combining FAIR principles and long-term archival of 3D data. In *Proceedings of ACM Web3D conference, (Web3D 2023)*. ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

1.1 Domain specific issues

In recent decades, 3D recordings, such as 3D digitization, modeling or reconstruction have played a growing role in the humanities. In this broad field of research, the volume of 3D data has increased. The importance of this data and the trust users place in it has grown. For the majority of rescue archaeology projects, records contain scientific evidence and traces of elements that may be lost forever. This work raises many expectations in the French community. Long-term archiving of completed projects is currently mainly carried out locally by the laboratories, using the - sometimes poor - IT resources available.

The Huma-Num consortium “3D for humanities” aims at federating the community of 3D producers and users in the context of Human and Social Sciences. The consortium is based on the network of collaborators coming from more than 20 laboratories and research team in France. It involves national infrastructures such as Huma-Num and C.I.N.E.S. (Centre Informatique National de l'Enseignement Supérieur). It is providing recommendation from the creation, up to the publication and archiving of such data. It is developing the national 3D data repository for Human Sciences together with the required metadata scheme and a companion software to simplify the documentation process. It aims at implementing FAIR principles [Wilkinson et al. 2016].

1.2 Proposal

Since 2017, a fully integrated solution for the long-term archiving and online publication of 3D data in the humanities has been available to researchers in France. The service is divided into a web interface called “The French National 3D Data Repository for Humanities” (CND3D in French – <https://3d.humanities.science>) for publication purposes and a desktop software called a Long Term Archive Generator for 3D (aLTAG3D – <https://altag3d.huma-num.fr>) for archive package creation. During 5 years of successful service, the provenance of the 3D data has gradually broadened, initially focusing on archaeological issues, then opening up to cultural heritage elements in general, and more recently to all areas of the humanities: psychology, theatre, literature, etc. To accommodate these new types of data and their description, domain-specific metadata, data production processes and data types have been updated. The absence of any physical source object creates a major conceptual gap

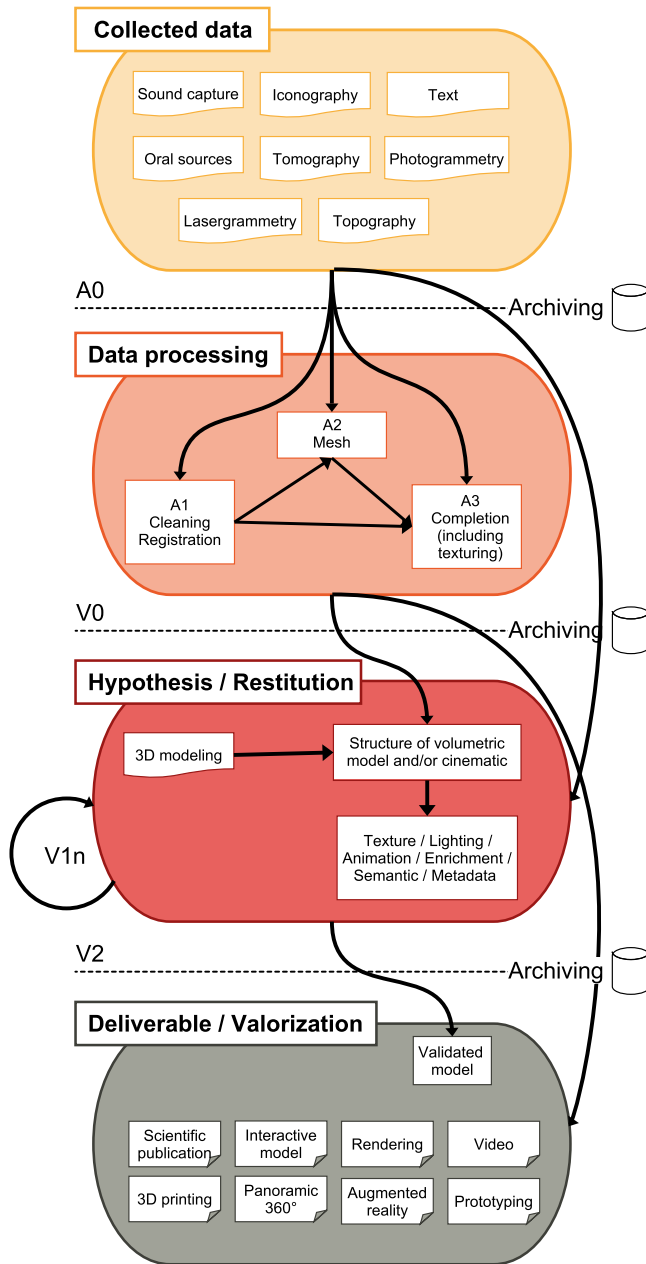


Figure 1: Sequential graph of 3D heritage data production.

between heritage fields and these new disciplines. Figure 1 describes the macro process that is nowadays used by all the French communities when digitizing, modeling, manipulating virtual heritage objects.

In this article, we present a new metadata schema, a more complete and exhaustive documentation of this schema, a mapping of this schema to the Europeana Data Model (EDM), and the integration of these new descriptors into our two existing tools: CND3D, our online platform, and aLTAG3D, our desktop tool.

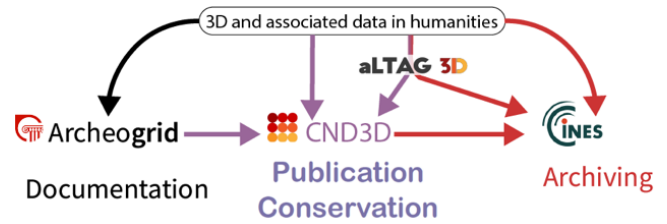


Figure 2: Data process and software associated.

1.3 Scope of application

Our proposal applies only to completed and closed projects. The archive can be built up over time with other tools (like ArcheoGrid for example), but we only process data once its content has stabilized. Figure 2 shows the presented tools aLTAG3D and CND3D in the data lifecycle. Documentation of unfinished 3D models is possible, but it is advisable to limit long-term preservation and documentation to completed models in order to limit storage space to scientifically validated data.

2 RELATED WORKS

2.1 Metadata schema

Since the 2000's, numerous vocabularies have been developed, sometimes at a high level, sometimes more specific to the fields of archaeology or cultural heritage. CIDOC-CRM [Doerr 2005] is the reference for the description of data in these fields. Vocabularies dedicated to cultural heritage or other disciplinary fields (psychology, theatre, etc.) complement our work. These vocabularies provide a precise description of the content in terms of concepts and their relationships between these concepts. These descriptors and their associated semantics are not included in our proposal. However, CIDOC-CRM, which is very complete, covers our entire schema on a top level. A mapping to CIDOC-CRM and DublinCore ensures strong interoperability with our records. Specialized schemas for the description of 3D data have also been published. At the crossroads of the humanities and 3D, CARARE [D'Andrea and Fernie 2013b], from the 3D ICONS project [D'Andrea and Fernie 2013a], offers a very flexible and lightweight metadata schema, widely recognized at European level. CARARE is a bundle of other standards (EDM, MIDAS, LIDO, DublinCore). This schema is too light for our archiving constraints based on the file content. Nevertheless, inspired by CARARE, the high-level and administrative parts of our schema overlap with its requirements. Thus, and similar to [Nicolucci et al. 2022], our schema remains interoperable through a mapping. In addition, the Europeana Data Model (EDM) has inspired the architecture of our schema with core and contextual classes.

Mapping to reference schemes for disciplines other than archaeology and heritage will be defined as needs arise and data is ingested, depending on the needs, uses and skills of the members of the emerging discipline of our consortium. Other cross-discipline schemas have been developed to describe the data production process [Belhajjame et al. 2013; Dudek et al. 2015; Homburg et al. 2021]. These descriptors are not included in the proposed version, but we are currently working on integrating this information explicitly

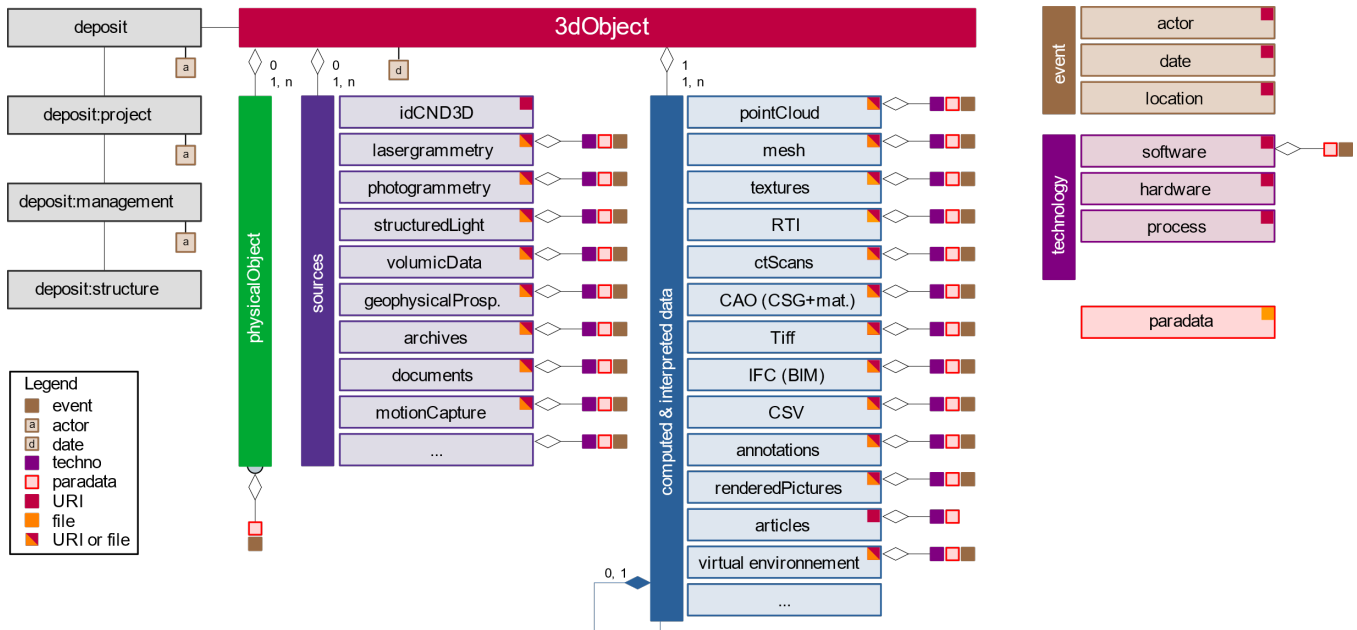


Figure 3: The architecture of the metadata schema.

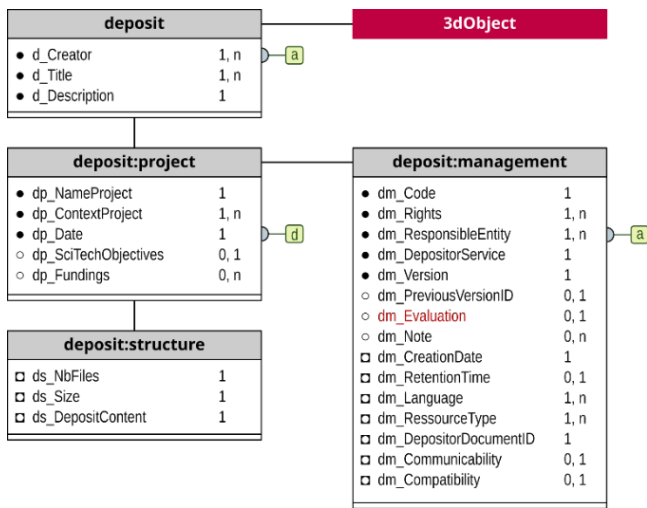


Figure 4: Detailed deposit part of metadata schema.

into our model. Finally, we are required to fulfil some basic descriptors for the creation of the SIP at CINES [CINES, 2013]. These elements are therefore included in our schema.

2.2 Tools and workflows

The Community Owned digital Preservation Tool Registry (COPTR – <https://coptr.digipres.org>) initiative of COST Action SEADDA (CA18128) combined with the Digital Research Infrastructure for the Arts and Humanities SSH Open MarketPlace (DARIAH SSHOM – <https://marketplace.sshopencloud.eu>) helped us to position our tools and digital preservation workflows among other European

initiatives. Of the 594 archiving tools listed by the COPTR project and the 61 tools dedicated to archiving, only our 2 tools, aLTAG3D and CND3D, can create packages for archiving 3D data. Besides this core functionality, our aLTAG3D tool appears to be the only XML visual programming editor available.

3 THE PACKAGE

3.1 Architecture of the metadata schema

The metadata schema we propose is organized into 5 core classes and 3 contextual ones (see Figure 3). It is explicitly formalized in an XSD file. The 5 core classes are the administrative parts, the 3D object, the physical object, the sources, and the computed outputs. The schema is centred on the 3D object. This object may use input sources: photogrammetry, image archive, etc. and must be composed of outputs "computed and interpreted data", containing at least one 3D file. The object may represent a physical object. Shared contextual classes (complex types) "actor", "date", and "software" (and their subtypes) are used across the schema for different entries.

With the new architecture of the schema, the expression of actors, locations and dates have been deported in separate classes (Actor, Location, Date) that can be grouped together in the "event" class. These classes are mapping more easily with the EDM contextual Classes Agent, Place and TimeSpan. The new physicalObject class helps mapping data with the EDM ProvidedCHO classe with separate metadata related to the physical object. To map the EDM WebResource class, we align metadata from our 3dObject class and other information from deposit part of the schema. Concept information is provided by the use of interoperable thesaurus, PACTOLS. At last, most of the information of the Aggregation EDM class are computed from information provided by the deposit management

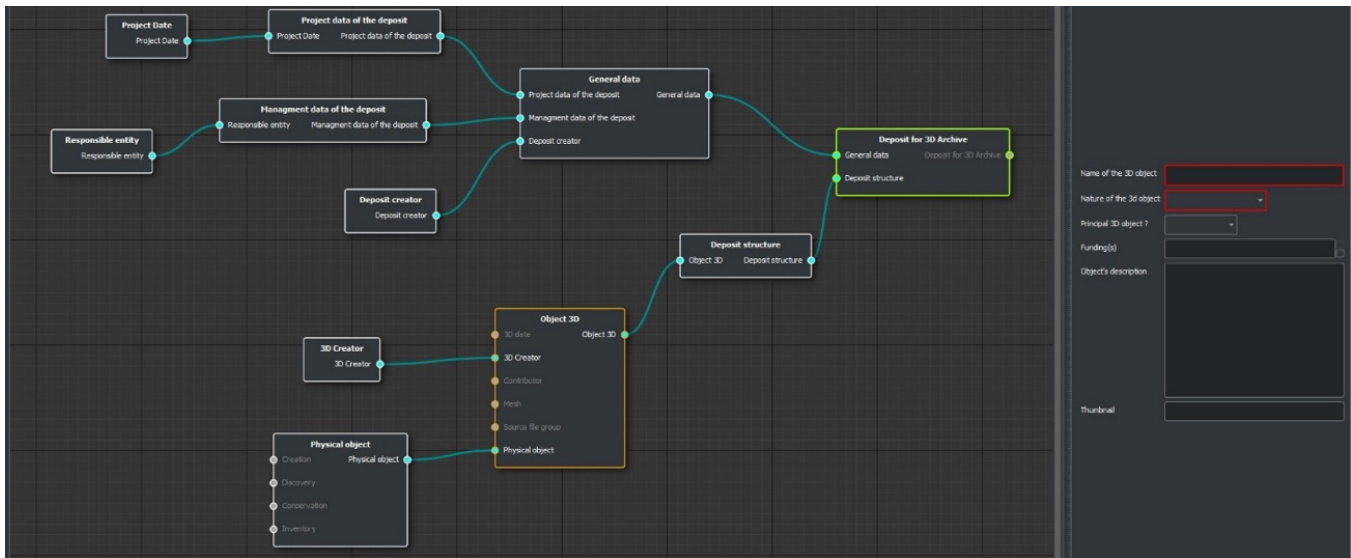


Figure 5: aLTAG3D [Dutailly et al. 2023] screenshot.

class (dm_Rights, ...) as see in Figure 4. The final mapping document will soon be available online, as soon as Europeana has validated it.

3.2 Archive content

"Sources" and "computed and interpreted data" may refer to files included in the package or to external URIs maintained by other institutions. The package focuses on files (local or URI) associated with users input description. The local files in the archive package must pass the CINES eligibility test <https://facile.cines.fr>, and they can only be of three kinds:

- (1) On the one hand the raw sources: scientific evidence and sensor data (may not be 3D)
- (2) On the other hand, the final outputs of the project: articles, 3D reconstructions, etc. (at least one of which must be 3D). No intermediate products are kept.
- (3) A third kind of file is the paradata: any contextual related documentation file: contract, scanning procedure, etc.

In the metadata, to leverage the potential of interoperability, free text user input fields are replaced where possible by either standard datatypes (e.g. xsd:date), closed lists (e.g. ISO639-2, custom one) or thesaurus standards (eg. PACTOLS (1)).

3.3 Documentation and tool integration

aLTAG3D [Dutailly et al. 2023] reads an XSD file and provides a visual programming interface to create and organize XML elements. An XSD validation of the user input content is processed in real time. This flexibility allows the user interface to be easily updated for the new schema with very few changes on the software side. Many automated process reduces the amount of information that the researcher will have to fill in manually. In the Figure 5 red boxes indicates mandatory data in the updated metadata schema. CND3D publishes part of the package in an online format. The update of

the metadata schema will affect the online platform by providing more details when available. The CND3D creates a DOI for the package and a web page with basic public information about it. It exposes the metadata via an OAI-PMH endpoint. Depending on the depositor's choice, it may display some of the package's 3D files online via its 3DHOP [M. et al. 2015] or Potree [Schütz 2016] viewer instance. CND3D stores the packages on secure servers for cold data (Huma-Num Box) and is the sole data provider for the CINES. Based on the XSD file, an exhaustive documentation of the schema has been created (not yet online in production), so to give everyone the possibility to create an SIP by their own, contribute, and create new tools or plugins for aLTAG3D.

4 APPLICATION

The whole process, from 3D creation to long term preservation and publication has been successfully experimented for a few years now, thanks to aLTAG3D software and CND3D platform. A schema in Figure 6 explain the complete workflow. Once integrated in the CND3D platform, documented 3D data can be harvested by other portals like OpenArchaeo <http://openarchaeo.huma-num.fr/explorateur/home> in France. Thanks to linked references in the new metadata schema, data can be cross referenced in CND3D and other portals of online publication. An example is exposed in Figure 7.

5 CONCLUSION AND PERSPECTIVES

Dedicated to the 3D data produced by the humanities in France, this new metadata schema and its ecosystem of tools meets the challenge of a greater openness to other non-archaeological humanities fields, a more detailed and unambiguous definition of research files, a better alignment in a semantic web FAIR perspective, a clearer documentation, and a better dissemination via machine-readable files (OAI-PMH endpoint plus EDM mapping) and a general public online 3D viewer. The next challenge is to port the existing records from the first schema to the new one. The advantage of opening

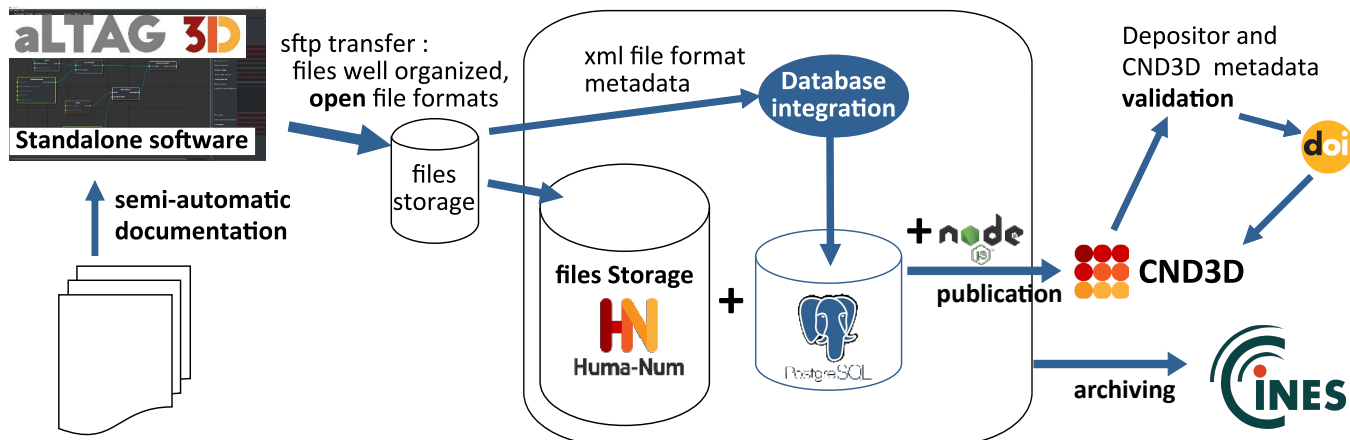


Figure 6: 3D data preservation workflow

to URI instead of files in the package is that in a semantic web paradigm it removes some useless duplicates. On the other hand, in a long-term archiving paradigm, this may result in incomplete archives, depending on the continuity of the URI and the readability of their content. Downloading and including content from external URIs during the SIP generation process for long-term archiving purposes is a discussed option. Environmental and financial concerns may drive the decision.

REFERENCES

- Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. 2013. *PROV-O: The PROV Ontology*. Technical Report. W3C. <https://www.w3.org/TR/prov-o/>
- Andrea D'Andrea and Kate Fernie. 2013a. 3D-ICONS Metadata Schema for 3D Objects. *Newsletter Archeologia CISA 4* (2013), 159–181. <https://api.semanticscholar.org/CorpusID:65382106>
- Andrea D'Andrea and Kate Fernie. 2013b. CARARE 2.0: A metadata schema for 3D cultural objects. In *2013 Digital Heritage International Congress (DigitalHeritage)*, Vol. 2. IEEE, Marseille, France, 137–143. <https://doi.org/10.1109/DigitalHeritage.2013.6744745>
- Martin Doerr. 2005. The CIDOC CRM, an Ontological Approach to Schema Heterogeneity. In *Semantic Interoperability and Integration (Dagstuhl Seminar Proceedings (DagSemProc), Vol. 4391)*, Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, and M. Uschold (Eds.), Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 1–5. <https://doi.org/10.4230/DagSemProc.04391.22>
- Iwona Dudek, Jean-Yves Blaise, Livio De Luca, Laurent Bergerot, and Noémie Renaudin. 2015. How was this done? An attempt at formalising and memorising a digital asset's making-of. In *2015 Digital Heritage*, Vol. 2. IEEE, Granada, Spain, 343–346. <https://doi.org/10.1109/DigitalHeritage.2015.7419519>
- Bruno Dutailly, Sarah Tournon, Mehdi Chayani, Valentin Grimaud, and Xavier Granier. 2023. aLTag3D: A User-Friendly Metadata Documentation Software. DARIAH Annual Event 2023: Cultural Heritage Data as Humanities Research Data?. <https://hal.science/hal-04115897> Poster.
- Timo Homburg, Anja Cramer, Laura Raddatz, and Hubert Mara. 2021. Metadata schema and ontology for capturing and processing of 3D cultural heritage objects. *Heritage Science* 9, 1 (July 2021), 91. <https://doi.org/10.1186/s40494-021-00561-w>
- Potenziani M., Callieri M., Dellepiane M., Corsini M., Ponchio F., and Scopigno R. 2015. 3DHOP: 3D heritage online presenter. *Computers & graphics* 52 (2015), 129–141. <https://doi.org/10.1016/j.cag.2015.07.001>
- Franco Niccolucci, Achille Felicetti, and Sorin Hermon. 2022. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* 7, 8 (2022), 105. <https://doi.org/10.3390/data7080105>
- Markus Schütz. 2016. *Potree: Rendering Large Point Clouds in Web Browsers*. Master's thesis. Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/E193-02, A-1040 Vienna, Austria. <https://github.com/potree/potree>.
- Mark D. Wilkinson, Michel Dumontier, I. Jbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da

Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (Mar 2016), 160018.

The screenshot shows the HAL portal interface. At the top, there is a navigation bar with the HAL logo and a language dropdown set to 'FR'. Below this, a dark blue button says 'Télécharger pour visualiser'. The main content area is divided into several sections:

- Dates et versions:** Shows 'hal-02045447, version 1 (22-02-2019)'.
- Identifiants:** Shows 'HAL Id: hal-02045447, version 1'.
- Citer:** Provides a citation for Robert Vergnieux's work on the AtonNum exhibition.
- Exporter:** Offers options like BibTeX, TEI, Dublin Core, DC Terms, EndNote, and Datacite.
- Données associées:** Lists the DOI '10.34969/CND3D/474506.d.2011'.
- Publication details:** 'Autre Publication Scientifique', 'Année: 2016', 'ATON-NUM AKHENATON & NEFERTITI', 'Robert Vergnieux (1)', and '1 Archeovision CNRS'.
- Keywords (Mots clés):** Karnak, Aton, 3D, Amarna.
- Domains:** Archéologie et Préhistoire, Sciences de l'Homme et Société.
- Buttons:** 'Afficher plus de détails' and 'Liste complète des métadonnées'.



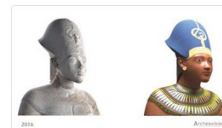
A propos /
Publication

Conservatoire National des Données 3D

fr en

Buste d'AKHENATON conservé au Louvre

Nature de la ressource	dépôt
Date du dépôt	2011-01
Créateur du dépôt	Sarah Tournon-Valiente
Entité déposante	Archeovision
Nb objets virtuels 3D	1
Nombre de fichiers	0
Nom du site (texte libre)	Paris
Propriétaire du site	Egypte
Propriétaire de l'objet	Musée du Louvre
Description archéologique	Buste d'Akhenaton
Date projet	2008,2011,2015
Date archéologique min.	-1360
Date archéologique max.	-1360
Période archéologique (libre)	/Nouvel Empire XVIIIème Dynastie/Amenhotep IV-Aménophis IV-Neferkheperure-Akhenaton
Commanditaire	Archeovision
Objectifs scientifiques et techniques	Restauration et restitution numérique du buste d'Amenhotep IV - Akhenaton
Description des fichiers (type, nombre)	1 fichier 3D (PLY)
Programme de recherche	"Aton 3D" (ANR-08-BLAN-0202-01), "Retrocolor 3D" Projet Région Nouvelle-Aquitaine
Lieu de découverte (texte libre)	El-Amarna
Numéro d'inventaire	E 11076
Référence publication HAL ou autre	hal-02045447
Date du dépôt	2011-01-01
Nature du dépôt	ACQUISITION_RESTITUTION
PeriodO	Antiquité



Localisation



Lieu de découverte (GPS)



Mot-clé

buste

Citation: Pascal Mora, François Daniel, Loïc Espinasse, Robert Vergnieux (2011). Buste d'AKHENATON conservé au Louvre, Archeovision. [10.34969/CND3D/474506.d.2011](https://doi.org/10.34969/CND3D/474506.d.2011)

Figure 7: CND3D assigned DOI and cross reference in HAL portal (<https://hal.science/>).