The Staccioli Digital Archive: Using Knowledge Graphs to power digital art history catalogues and art exhibitions

Klaus Werner[1], Pietro Liuzzo[1], Alessandro Adamou[1]
[1] Bibliotheca Hertziana – Max Planck Institute for Art History, Italy – *firstname*.*lastname*@biblhertz.it

ABSTRACT (ENGLISH)
This paper presents the Staccioli Digital Archive, a comprehensive catalogue of digitized material and semantic metadata revolving around the life and career of Italian sculptor Mauro Staccioli. Through the shared usage of open standards, conventions and systems, it attests to a fruitful interaction between art history research institutions, individual scholars, and the archives and foundations providing the source material. The Staccioli Archive also powers a persistent virtual exhibition built upon the same data platform as the archive, and supported a physical exhibition that ran in 2023-24. This collaborative development follows an extensible data management model based on FAIR standards such as Linked Data, CIDOC-CRM and IIIF, from inventory to publishing. The workflow relies upon shared conventions designed in such a way that data capture, digitization, IIIF image publishing, and the identity management of archival units could proceed each at its own pace and independently of one another. This model is intended to serve as a blueprint for the creation of future explorable art history datasets with support for discoverability, to potentially pair with virtual exhibitions to come.
Keywords: digital archives, Knowledge Graph; Linked Data; Art History; 20th-century sculpture

ABSTRACT (ITALIANO)
L'Archivio Digitale Staccioli: utilizzare i Knowledge Graph per alimentare cataloghi di storia dell'arte e mostre digitali. Il contributo presenta l'Archivio Digitale Staccioli, un catalogo completo di materiale digitalizzato e metadati semantici che ruotano intorno alla vita e alla carriera dello scultore italiano Mauro Staccioli. Attraverso l'uso condiviso di standard, convenzioni e sistemi aperti, esso è testimone di una proficua interazione tra istituti di ricerca di storia dell'arte, singoli studiosi, ed archivi e fondazioni da cui provengono le fonti materiali. L'Archivio Staccioli alimenta anche una mostra virtuale persistente costruita sulla stessa piattaforma di dati dell'archivio e ha supportato una mostra fisica svoltasi nel 2023-24. Questo sviluppo collaborativo segue un modello di gestione dei dati estensibile basato su standard FAIR come Linked Data, CIDOC-CRM e IIIF, dalla fase di inventario a quella di pubblicazione. Il flusso di lavoro si basa su convenzioni condivise, progettate in modo tale che l'acquisizione dei dati, la digitalizzazione, la pubblicazione di immagini IIIF e la gestione dell'identità delle unità archivistiche possano procedere ciascuna al proprio ritmo e indipendentemente l'una dall'altra. Questo modello progettuale è destinato a fungere da base per la creazione di futuri dataset di storia dell'arte con supporto per l'esplorabilità, da abbinare potenzialmente alle mostre virtuali a venire.
Parole chiave: archivi digitali, Knowledge Graph; Linked Data; storia dell'arte; scultura contemporanea

1. INTRODUCTION
In 2023, curator Marica Antonucci of the exhibition "Mauro Staccioli: Cementing an Artistic Legacy" hand-picked a selection of visual material documenting the defining moments in the career of Italian contemporary sculptor Mauro Staccioli (1937-2018) (Dorfles, Barbero, & Bazzini, 2008). Prior to the exhibition, which was open from October 2023 to January 2024, the archive was only available as a plaintext inventory with image galleries of partially digitized material, mostly limited to photographs of Staccioli's art installations and designs. The physical inventory, however, comprised a wealth of additional documentation, including preliminary drawings, press, correspondence, and exhibition flyers, catalogues and tickets, all of which nurture a greater understanding of Staccioli's artistic production context and were hitherto available in physical form only.
Aside from providing the curator with additional material to include in the digital counterpart to the exhibition, digitizing these resources also had the purpose of adding a dimension of serendipitous discoverability, thus allowing the viewer to branch out of the exhibition and explore other digital, if not curated ad-hoc, resources. A structured and standardized knowledge organization of the data for these documents is key to achieving this goal, for which Linked Data provide an ideal implementation.
Additionally, setting up such a complex environment requires the contribution of departments and teams—the Library, Photo archive, digitization team, curators and scholars carrying out onsite fieldwork—that

operate on greatly differing schedules: while their respective outputs must be integrated seamlessly, the paces at which they produce it must be taken into account without disrupting their workflows.
For the above reasons, the outcome of this effort, the Staccioli Digital Archive,[1] rather than addressing specific research questions in art history, focuses on how to integrate the existing infrastructure, including assets which are usually isolated, such as an archive, a database and an exhibition. The front-facing component of the project is implemented on a fork of the ResearchSpace platform, which sits atop the knowledge graph of the Bibliotheca Hertziana, and on the facilities of its Photo archive for publishing its visual material using the IIIF framework, while the ePIC identifier system provides IDs to archival units. These components operate independently, yet are seamlessly integrated through a shared knowledge of how the material is organized, which is encoded in the knowledge graph using the CIDOC-CRM core model. The proposed set-up has the added benefit of leaving space for further possible layers of reading, narrative and interpretation to be added explicitly or simply formed by users in their experience with the application.

2. BACKGROUND

The physical archive of Mauro Staccioli's oeuvre is located at his landhouse in Mazzolla, Tuscany. The material is stored in 190 boxes sorted per year and/or project; 40 moleskin notebooks containing mostly theoretical, partly already published writings; 10 horizontal folders with plans and sketches; 30 rolled-up plans; a cabinet full with slides used for seminars at the University of Milan; and 15 maquettes of his projects. A project may be intended as the preparatory work for, or documentation of, an art exhibition or installation, or the design of an individual sculpture. The criterion around which they are organized is primarily event-driven and chronological.
A digital inventory of the physical archive contents does exist, however, its entries are limited to listing the amount of units per type contained in each folder[2] in one single HTML page in Italian, despite the website being bilingual. Since a full digitization of the complete archive would have taken two to three years, our digitization campaign hand-picked a selection of the most relevant projects, detailed in Section 4.

3. FIELDWORK AND DATA CAPTURE

To accommodate the schedule of the art historian tasked with scrutinising Staccioli's physical archive in Mazzolla and recording its data, the timescales of a native Linked Data implementation of the data capture workflow were not feasible: a field database was then developed on top of a commercial relational DBMS, MementoDB, which is also available for tabletop systems and could be installed on the scholar's tablet device for the on-site recording of data from the archive's folders. Although this part of the implementation was relational, it was design reusing terms from the CIDOC-CRM ontology to facilitate subsequent transformations. Whereas the database was supposed to only give us a crude overview of the material—box, folder, and item with its relevant project title and timeframe—it quickly evolved to accommodate a more fine-grained representation and, most importantly, linking to external datasets and authority files. The final schema for data capture was thus as follows:

- **Document**(ID, box, folder, item, title, date, conservation, note, Keyword, P2_has_type[Type], P62_depicts[Work], P62_depicts[Person], P67_refers_to[Person], P45_consists_of[Material], P108i_was_produced_by[Person], P70_documents[Event], P67_refers_to_Institution[Institution])
- **Work**(ID, title, date, fate, Wikidata, description, P53_has_former_or_current_location[Place], P94i_was_created_by[Person], P103_was_intended_for[Event], P45_consists_of[Material])
- **Material**(ID, name, AAT)
- **Institution**(ID, name, GND, P69_has_association_with[Event])
- **Person**(ID, name, ULAN, Wikidata)
- **Place**(ID, name, geometry, Wikidata)
- **Event**(ID, title, date, P7_took_place_at[Place], P147i_was_curated_by[Person])
- **Keyword**(ID, name, GND)

Simplifications with regards to CRM include the fact that P94i_was_created_by or P108i_was_produced_by were linked directly to persons, rather than to activities as mandated by CRM, as the original archive does not distinguish individual production activities, therefore the reification of the activities required by the CRM could be deferred to a later stage, where the RDF data for the knowledge graph were produced. Links to the German authority file GND, to the Getty thesauri AAT and ULAN, and to Wikidata, as per the eponymous fields in the schema, were added manually after each fieldwork session.

1 Staccioli Digital Archive: https://staccioli.biblhertz.it/ (cons. 24/01/2025).

2 Original Staccioli inventory: https://maurostaccioli.org/it/inventario-dellarchivio-mauro-staccioli/ (cons. 22/01/2025).

4. DIGITIZATION AND DIGITAL IMAGE STORAGE

As the material was progressively described, the curator selected a subset of it to prioritize for digitization targetting the planned exhibition. Of more than 150 events (exhibitions, showcases etc.) to which Staccioli took part between 1965 and 1988, which had been documented in the database, three were selected, owing to the fact that they were well-documented and covered an important phase in the development of the artist. These are: Staccioli's solo exhibition in Volterra, 1972; the 1978 Biennale in Venice, and "Arte e Critica" at the National Modern Art Gallery in Rome, 1981 (GNAM'81).

The boxes were transferred to the Bibliotheca Hertziana, where the digitization of the papers, sketches, photos and slides was performed. The storage schema closely resembled the physical order, as indicated in the "Document" entity of the database: boxes, folders and items as nested (sub)directories containing scans with 4-figure numbers. The compiled material amounted to over 22,000 scans of ca. 7,500 items.

The digital images produced were stored on a Nexus-POSIX filesystem, which already hosted all the images of the Photo archive of the Bibliotheca Hertziana, and were similarly processed on a high-performance computing (HPC) cluster for transforming them to the JPG2000 format and longterm archival. The JPG2000 derivates are served using the Internet Image Interoperability Framework by a Cantaloupe image server backed by an RDF triple store. For each image a minimal set of triples was produced, which included the absolute path to the Nexus- POSIX storage and the accessibility information (property v:access) needed by the server to check whether it has the rights required for publishing it in open access.

5. SEMANTIC DATA MODELLING

Whereas setting up a relational data model served as a stratagem to maximize the throughput of data entry, achieving compliance to the FAIR principles (findability, accessibility, interoperability, reusability) requires that a Linked Open Data derivative of the data be produced and published.

The data collected revolved around a restricted set of entity classes and relations between them:

- Sculptures and other artworks authored by Staccioli himself or his collaborators.
- Events in which said artworks were featured or discussed.
- The documentation that constitutes the actual digitized material of our collection.
- Places may be event locations, or places where the artworks rest or are conserved.
- People related to artworks or documents as their authors, or as being depicted or mentioned.

It was ascertained that most of the above could be modelled using version 7 of the CIDOC CRM core ontology (Bekiari, Bruseker, Doerr, Ore, Stead, & Velios, 2021), with the Keyword entity in the original relational database being modelled as a SKOS thesaurus, and with minimal usage of other vocabularies like schema.org. External links are modelled as owl:sameAs.

```
event:QFJDK2YycilqZCNEanhhI01Sb3U/collection
  a crm:E78_Curated_Holding, crm:E72_Legal_Object,
    crm:E24_Physical_Human-Made_Thing ;
  rdfs:label "\"Alternativa - II Festival Internacional de Arte Viva\""@it;
  crm:P19i_was_made_for event:QFJDK2YycilqZCNEanhhI01Sb3U ;
  crm:P46i_forms_part_of <http://data.biblhertz.it/curated_holding/archivio_staccioli> .

event:QFJDK2YycilqZCNEanhhI01Sb3U a crm:E7_Activity ;
  crm:P7_took_place_at place:Rio5Tj47dCtYSm87UW1AQVlTajg ;
  crm:P19_was_intended_use_of event:QFJDK2YycilqZCNEanhhI01Sb3U/collection .

place:Rio5Tj47dCtYSm87UW1AQVlTajg a crm:E53_Place ;
  rdfs:label "Almada" ;
  owl:sameAs wd:Q328015 .
```

Listing 1. Example of RDF/Turtle rendering of a collection associated to an event

Listing 1 shows an example of how an event, the "Festival Internacional de Arte Viva" that took place in Almada, Portugal, was modelled in the knowledge graph as an E7_Activity in CRM, which is associated to a collection of materials that is part of the larger Staccioli archive.

The mapping from the initial database to RDF was performed using SPARQL-Anything, a software library for querying data in various formats using the SPARQL language with an extended syntax for query

federation and additional utility functions (Asprino, Daga, Gangemi, & Mulholland, 2022). Because MementoDB exposes its data using a cloud service with a Web API in JSON format, it was possible to query it directly in SPARQL-Anything, use queries of type CONSTRUCT to generate new RDF data in the desired schema based on CIDOC-CRM, and store the result in the knowledge graph infrastructure of the project. A dataset of nearly 250,000 RDF triples was produced for this project.

6. END-USER APPLICATION

Due to its flexible templating mechanism and versatile support for different semantic storage solution, ResearchSpace (Oldman & Tanase, 2018) was adopted as the semantic platform upon which to build the catalogue and digital exhibition. Unlike default ResearchSpace installations, the instance of the Staccioli archive does not have its own RDF store, but is instead fed from the knowledge graph of the Bibliotheca Hertziana, which hosts multiple datasets along with the Staccioli one: therefore, no further data integration was necessary after the re-engineering from the original MementoDB instance. This was also made possible by the fact that the editorial workflow was implemented outside the platform.

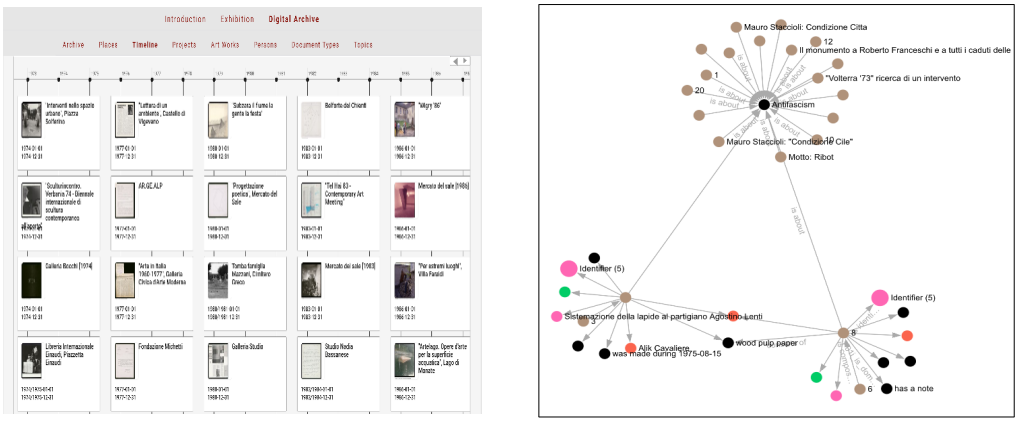

Figure 1. Exploring the Staccioli Digital Archive through a timeline view (left) and the SigmaGraph (right).

Visualizations were implemented using off-the-shelf components of ResearchSpace, such as the timeline view of events and projects, map-based geospatial exploration, and entity-based search. Additionally, the standard graph-based visualization, which displays a fixed set of notes retrieved from a simple SPARQL, was replaced with a third-party component, the Semantic SigmaGraph[3] (Figure 1), which allows dynamic query expansion upon clicking on individual nodes, thereby encouraging the serendipitous exploration of knowledge around the sculptor's career, beyond the curated content of the exhibition (Po, Bikakis, Desimoni, & Papastefanatos, 2020). The Staccioli Digital Archive thus blends in one place both the exhibition contents selected from the archive and the digitized part, in its entirety, of the archive itself. While this is an unusual choice, because a scholarly dataset is typically not something one would want to expose to the visitors of the exhibition, it allows users to choose the level of complexity at which they wish to explore the content (Foo, Theng, Go, & Na, 2009). There were also noticeable benefits in the development stage, as it did not prove necessary to implement one system for the researcher interested in the data, and one for the visitor interested in further documentation to the items on display at the exhibition. In this way the restricted pre-organized view into the material which the exhibition favoured was preserved, but also the possibility was given to enhance one's knowledge by allowing a wider, and at the same time more refined, view into the available documentation.

7. IMAGE COLLECTIONS, RESOURCE IDENTITY, AND SEAMLESS INTEGRATION

A dedicated parallel structure was also built on the basis of the existing material exposed by the image servers, so that collections and manifest documents conform to the IIIF Presentation API could be constructed from the knowledge graph, independently from the ResearchSpace application. This makes it possible for these collections to be usable in any IIIF viewer, including the one set up as a standard for the Photographic collection of the Bibliotheca Hertziana, which hosts the digital archive.
Collections and Manifests are generated from an XQuery module stored in eXist-db. The module is integrated into a declarative structure and queries the knowledge graph directly over HTTP. The SPARQL

3 F. Kräutli. Visualising Networks. The Sigma Graph Component: https://www.sari.uzh.ch/en/Insights/Visualising-Networks.-The-Sigma-Graph-Component.html (cons. 24/01/2025).

queries are tailored each Collection and Manifest. Starting from a main Collection, several subcollections are available, which query the knowledge graph on the fly to list available items. This bore at least two advantages to the various workflows at play in the realization of this project:

- The Staccioli catalogue and exhibition can use the IIIF image viewer of the Photographic collection, without any need to host its own IIIF substructure, yet with interlinking between the two.
- Both the collections on the IIIF platforms, and the image links in the Staccioli catalogue, could be generated altogether, without having to wait for the digitization department to finish its job. Any empty collections or broken image links would automatically be fixed as the corresponding scan was performed and the resulting images processed and saved to longterm storage.

For each archival unit, a persistent identifier (PID) was also generated and stored into the knowledge graph, which replicates the box-folder-unit organization of the archive, upon querying the knowledge graph itself. As the identity provider, the ePIC mechanism[4] was selected due to its flexibility and open source API. For instance, PID http://hdl.handle.net/21.11153/biblhertz_staccioli_box_1_folder_6_unit_19 resolves to the 19th unit catalogued in folder 6 of box 1.

8. CONCLUSION AND OUTLOOK

Not only does the Staccioli Digital Archive offer an instance of co-habiting digital catalogue and exhibition, but also its implementation tested the ability of different departments with different workflows to interact with one another, while preserving their priorities and setups. By relying upon open standards, shared and well-understood conventions, and open source software, this endeavour repurposed existing technologies that are being adopted in other projects, with minimal adaptation. It also proved to be an effective way forward for several other similar projects, and set a precedent for a more integrated implementation of said existing resources in future efforts.

Funding is actively being sought to bring the digitization of the entire archive in Mazzolla to completion. This will be accompanied by further enhancements of the front-facing application, potentially expanding it to address the research questions of scholars in contemporary art history, naming providing alternative visualizations and prototypical queries to the knowledge graph. Due to the evolution of other ResearchSpace-based projects in the meantime, this second iteration aims at implementing the data curation workflow natively as Linked Data, thereby abating the need to perform an extraction process from a relational database.

ACKNOWLEDGEMENTS

The project was developed in collaboration with Associazione Archivio Mauro Staccioli, Fondazione Cassa di Risparmio Volterra, Galleria Il Ponte, Firenze, the Galleria d'Arte Niccoli, Parma, and the Swiss Art Research Infrastructure, Zurich. For detailed acknowledgements visit https://staccioli.biblhertz.it.

REFERENCES

Asprino, L., Daga, E., Gangemi, A. & Mulholland, P. (2022). Knowledge Graph Construction with a façade: a unified method to access heterogeneous data sources on the Web. *ACM Transactions on Internet Technology*. Association for Computing Machinery. https://doi.org/10.1145/3555312.

Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S., & Velios, A. (Eds.). (2021, April). Volume A: Definition of the CIDOC Conceptual Reference Model, Version 7.1.1. CIDOC CRM Special Interest Group. https://doi.org/10.26225/FDZH-X261.

Dorfles, G., Barbero, L.M., Bazzini, M. (Eds.). 2008. *Mauro Staccioli. All'origine del fare*. Mantova: Corraini Edizioni. ISBN: 978-8875701901.

Foo, S., Theng, Y., Goh, D. H., & Na, J. (2009). From Digital Archives to Virtual Exhibitions. *Handbook of Research on Digital Libraries: Design, Development, and Impact*. Y. Theng, S. Foo, D. Goh, & J. Na (Eds.), IGI Global Scientific Publishing. 88-100. https://doi.org/10.4018/978-1-59904-879-6.ch009.

Oldman, D. & Tanase, D. (2018). Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace. *ISWC 2018 - 17th International Semantic Web Conference*, Monterey, CA, USA 8-12 November 2018. Vrandecic, D., Bontcheva, K., Suàrez-Figueroa, M.C., Presutti, V., Celino, I., Sabou, M., Kaffee, L.A. & Simperl, E. (Eds.). 325-340. https://doi.org/10.1007/978-3-030-00668-6_20.

Po, L., Bikakis, N., Desimoni, F., & Papastefanos, G. 2020. *Linked Data Visualization: Techniques, Tools, and Big Data*. Springer Cham. https://doi.org/10.1007/978-3-031-79490-2.

4 European Persistent Identifier Consortium: https://www.pidconsortium.net/ (cons. 24/01/2025).