

The Data Crew—Archive with ARCHE and Enrich with OpenAtlas

Martina Trognitz ^{1,*}, Massimiliano Carloni ¹ and Bernhard Koschicek-Krombholz ¹

¹Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, 1010 Vienna, Austria; E-Mails: Martina.Trognitz/Massimiliano.Carloni/Bernhard.Koschicek-Krombholz@oeaw.ac.at

* Corresponding author

Abstract

In digital humanities research, different tools and services need to be integrated to allow for a seamless and streamlined workflow. With the example of project INDIGO, the tasks of archiving, annotating, enriching, and disseminating data with the digital archive ARCHE, the spatial database OpenAtlas, and ACDH-CH's Vocab Repository are described in detail. Each of the three services is presented along with its core tasks and how it is used within the exemplary workflow of project INDIGO. New functionality that was necessary for the integration of these services within INDIGO was implemented and is also described.

Keywords

Controlled vocabulary; Data enrichment; Long-term preservation; Metadata; Spatial database

1. Introduction

Research with digital methods often requires the interplay of different tools and services responding to diverse needs. Examples include archiving, annotating, and enriching data, as well as disseminating data and results to a wider audience. Since each task may be the domain of one or more separate pieces of software, a robust and streamlined workflow between the different components is essential to ensure the sustainability of the research project's data and its adaptability to different scenarios.

The graffiti along Vienna's Donaukanal (Eng. Danube Canal), which are documented and disseminated within project INDIGO for future reusability and further analysis (Verhoeven et al., 2022), pose an example of the use of several tools and services. From their initial acquisition until their dissemination, the data undergo an extensive workflow across five so-called 'research pillars': Acquisition, processing, management, dissemination, and analysis (Verhoeven et al., 2022, p. 514). All these phases involve the collaboration of various institutional partners.

We will focus on the research pillars 'management' and 'dissemination' and explore in more detail the workflow being developed to enable information exchange between the digital archive ARCHE, the spatial database OpenAtlas, and ACDH-CH's Vocab Repository (Figure 1). These three services are developed at, and hosted by, the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) of the Austrian Academy of Sciences. Within the context of project INDIGO, ARCHE, OpenAtlas, and the Vocab Repository are used for archiving, enriching, and disseminating the data created and collected during the project.

After a brief presentation of each service and its main functionalities, the contribution will focus on how these services are joined and which tasks are solved with each service in the overall workflow. New functionality implemented to streamline the workflow is also presented along the way.

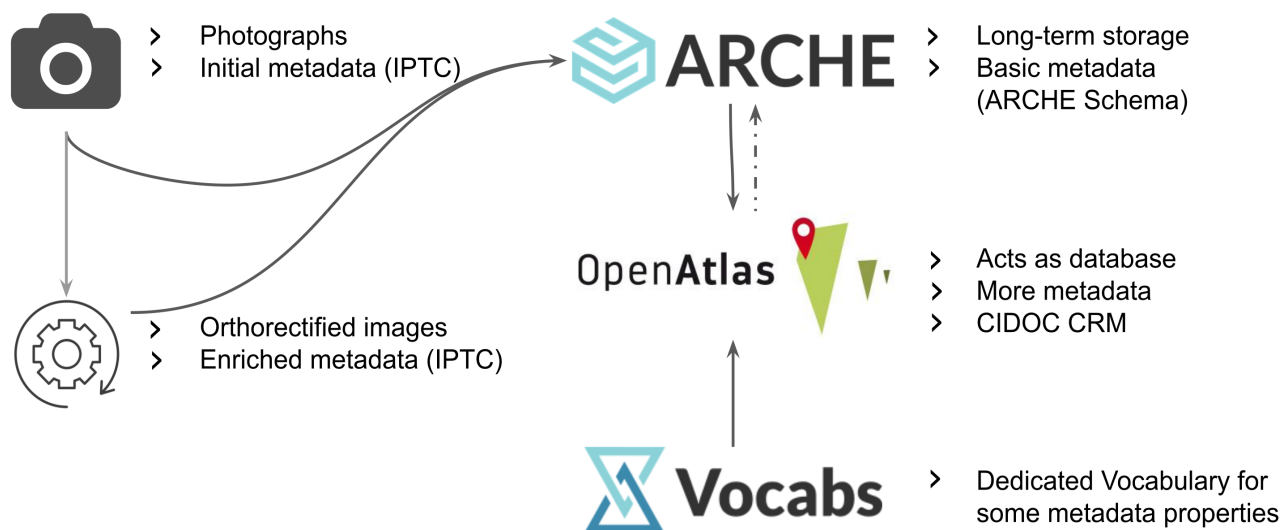


Figure 1. Schematic of the workflow enabling the information exchange between the digital archive ARCHE, the spatial database OpenAtlas, and ACDH-CH's Vocabs Repository. Source: Martina Trognitz.

2. ARCHE

ARCHE, A Resource Centre for the HumanitiEs (ACDH-CH, n.d.–c), is a digital archive established in 2017 and certified with the CoreTrustSeal (CoreTrustSeal, n.d.), a core-level certification for trustworthy data repositories. It offers long-term digital preservation of humanities research data, either related to or originating from Austria. ARCHE does not only aim at the long-term preservation of data, but has a broader perspective that also includes data dissemination to enable data publication, permanent referenceability, sustainable reusability, and reproducibility of research results (Trognitz et al., 2022, p. 231). Preservation and dissemination of the data underlying research results is an important cornerstone of conducting research conforming to general rules of good scientific practice (ALLEA, 2023), including principles such as Open Science (FOSTER, n.d.; Open Knowledge Foundation, n.d.) and FAIR data (Wilkinson et al., 2016).

ARCHE is specialised in preserving research data, which is something few repositories in Austria do (Trognitz, 2021). Supported data types include images, texts, structured and tabular data, audio recordings, videos, 3D models, geographic information, and many more. Individual files are grouped within collections, which can be nested. The technical infrastructure, underlying policies, and design

principles behind ARCHE fully embrace and support the FAIR principles for Findable, Accessible, Interoperable, and Reusable data and metadata.

Findability is ensured by the assignment of persistent identifiers to all resources and collections as well as by extensive accompanying metadata. A custom metadata schema was developed for ARCHE to allow for the description of heterogeneous data from a wide range of humanities disciplines (Trognitz & Ďurčo, 2018). Findability is further improved by the fact that resources deposited in ARCHE are harvested by a wide range of aggregators, which include domain-specific portals, such as the CLARIN Virtual Language Observatory (CLARIN, 2023) for linguistic resources, and general-purpose databases about research data, like OpenAIRE (OpenAIRE, 2023) or Europeana (Europeana, n.d.).

Accessing data and metadata is not only possible via a GUI but also via an ever-growing range of dissemination services that adhere to Semantic Web and Linked Open Data (LOD) (Berners-Lee, 2006) standards (Žóltak et al., 2022). Dissemination services are applications capable of presenting specific data types online or delivering them in different formats, thus further enhancing data

interoperability. Examples of dissemination services include a IIIF (International Image Interoperability Framework)-based image viewer and endpoint, and the ExifTool service. The IIIF service allows viewing images online and can disseminate images in various formats and sizes. The ExifTool service is based on the eponymous command-line tool by Phil Harvey (Harvey, 2023) and can read embedded metadata like Exif (Exchangeable Image File Format) and IPTC (IPTC Photo Metadata Standard), and delivers them in JSON (JavaScript Object Notation) format.

Similarly, metadata stored in ARCHE can be delivered in various metadata formats with an OAI-PMH templating system (Żółtak et al., 2022), such as the Europeana Data Model (EDM), which is harvested by the Austrian aggregator for Europeana.

Reusability of data and metadata is further strengthened by the provision of an Application Programming Interface (API), which offers a way to search and retrieve data programmatically. This allows (re-)users to integrate data stored in ARCHE directly within their custom web applications. By allowing the storage of extensive metadata and thus contextualising the data, e.g. with information about applied methods and used software, reusability is further improved.

Within the context of project INDIGO, ARCHE serves as the place where all project data are preserved for the long term.

3. OpenAtlas

OpenAtlas (OpenAtlas team, 2023) is an open-source database project, developed to acquire, edit, and manage geospatial data and information from humanities disciplines, including history, prosopography, archaeology, and heritage science, as well as related data coming from the natural sciences (Richards et al., 2023). OpenAtlas' data model is mapped to the CIDOC Conceptual Reference Model (CRM) developed by the International Council of Museums (ICOM, n.d.) in the background of the application, allowing users to easily create data conforming to CIDOC CRM without the necessity to be familiar with all the complexities of the ontology. OpenAtlas embraces the principles of FAIR data (Wilkinson et al., 2016), adheres to the standards of

LOD (Berners-Lee, 2006), and whenever possible aims for openness in all its aspects (Richards et al., 2023).

OpenAtlas allows to record and display relations between different entities, like actors, events, or places, and can be customised with a bespoke set of categories and classification options. The software allows linking data to external online and offline sources. These include authority files like the Getty Arts and Architecture Thesaurus (AAT), GeoNames, or a Museums' inventory number. Data in OpenAtlas can be visualised, e.g. as network graphs (Watzinger & OpenAtlas team, n.d.-b). Machine-readability of all data is provided via an API, which serves data in standard formats such as JSON-LD (Sporny et al., 2020). The API allows information to be fetched for reuse in bespoke web applications or local software for further dissemination, processing, and analysis.

The development, expansion, and overall improvement of OpenAtlas are often carried out in collaboration with research projects. In the context of each project, consideration is also given to extending the data model or adding further functionality to the software (Filzwieser & Eichert, 2020). Usually, the expansions are then added to the core code of OpenAtlas, thus allowing future projects to benefit from past projects. Functionalities introduced during project INDIGO are detailed in Section 5.

Within project INDIGO, OpenAtlas, with its powerful API, serves as a middleware between the data archived in ARCHE and the separate frontend platform. Furthermore, OpenAtlas can be used to record and edit all metadata related to an individual graffiti, such as the creator(s), location, creation time, style, colour(s), and dimensions.

4. ACDH-CH's Vocab Repository

The Vocab Repository of ACDH-CH (ACDH-CH, n.d.-a) is a service based on Skosmos (Suominen et al., 2015) to provide hosting, browsing, and visualisation of controlled vocabularies modelled in Simple Knowledge Organization System (SKOS) (Isaac & Summers, 2009). Skosmos implements Linked Data principles and provides an API that allows fetching the vocabulary data and presenting them in custom project applications (Andorfer & Illmayer, 2023). Controlled vocabularies are essential to ensure consistent

use of values in a database (Harpring, 2013, p. 13). In the realm of Semantic Web and LOD, controlled vocabularies are important tools to achieve semantic interoperability across multiple data sources (Zaytseva & Ďurčo, 2020).

The repository does not solely host controlled vocabularies from and for ACDH-CH, but has a much wider scope and is part of the services for the Austrian national CLARIAH-AT consortium, for the Thesaurus Maintenance working group in DARIAH-EU (Tsoulouha & Goulis, n.d.), for WP6 – Services and tools in PARTHENOS (Sbarbati, 2015), and, in a separate instance, for the Social Sciences & Humanities Open Cloud (SSHOC) (Vocabs SSHOC, n.d.).

The Vocabs Repository is used to host the INDIGO Graffiti Thesaurus (Schlegel, Carloni, et al., 2023; Schlegel, Wogrin, et al., 2023), which will also be used to populate controlled lists in OpenAtlas. In addition, the Vocabs Repository hosts a controlled vocabulary used in the GRAPHIS application (as described in Section 4.2 of the contribution by Verhoeven, Wieser, & Carloni in this volume).

5. Joining the services for INDIGO

Before reaching any of the presented services, INDIGO data will already have undergone an extensive acquisition and processing workflow (see left side of Figure 1) described in (Verhoeven et al., 2022; Verhoeven et al., 2023; Wild et al., 2023), with various file types as output: The raw photographs, masked and colour-corrected photographs, orthophotographs generated from the colour-corrected images, 2D polygon files, 3D mesh files, and metadata included in sidecar files or directly embedded in the image files. File formats in the submitted data package include Nikon's raw image format NEF (Nikon Electronic Format), XMP (Extensible Metadata Platform) sidecar files with metadata, processed TIFF (Tag Image File Format) and JPEG (Joint Photographic Experts Group) images, spatial information in GeoJSON (Geo JavaScript Object Notation) and GeoTIFF format, 3D models in PLY (Polygon File Format, Stanford Triangle Format) with JPEG texture files, and documentation in TXT (plain text) and DOCX (Microsoft's Office Open XML) files. Among the processed image files, there are so-called orthophotos in GeoTIFF format, which

are merged from multiple individual colour-corrected raw images, georeferenced, and orthorectified. Each of them represents an individual graffiti (Molada-Tebar & Verhoeven, 2023; Wild et al., 2022).

The first contact point of INDIGO data with the software and services offered by ACDH-CH is the digital archive ARCHE, which will take care of the long-term preservation and publication of the data. The data and metadata undergo an extensive curation process, guided by a curator from ARCHE, that comprises several checks in order to ensure the quality, consistency, and usability of the data. After an initial automatic validation of file formats, reporting of file names, and identification of possible duplicates with the tool repo-file-checker (Czirjak et al., 2023), the metadata are checked for compliance with ARCHE's minimum requirements. Based on these results, a curation strategy is devised. It includes the necessary steps and changes to be applied to the submitted dataset to transform it into the Archival Information Package (AIP) to be ingested. The AIP is an information package type defined by the Open Archival Information System (OAIS) reference model (Consultative Committee on Space Data Systems, 2012), which ARCHE implements (Trognitz & Ďurčo, 2018). Data from INDIGO are expected to come with extensive metadata following the IPTC standard, in XMP sidecar files, and also included within the image files. The curation strategy will cater to this, and a mapping to the ARCHE metadata schema will be prepared, together with an automated workflow for reading and ingesting the metadata into the archive. The final number of raw photographs is estimated at around 150,000. It is expected that raw photographs will not be needed to be served one-by-one to other services and that requests for individual images by users of the archive are unlikely. Therefore, to avoid any performance issues due to the large number of files, batches of raw images comprising one graffiti will be grouped into single BagIt packages (Kunze et al., 2018).

After successful curation of the data and ingestion of the AIP, the data are publicly available on ARCHE with a Persistent identifier (PID) assigned to each resource. PIDs are based on the Handle system (Corporation for National Research Initiatives, 2023) for persistent identification and stable

reference. The data and metadata can now be accessed and reused either manually via ARCHE's web portal, or automatically via the API or the various dissemination services. This is where the next service, OpenAtlas, comes into play.

The import of data from ARCHE into OpenAtlas is performed through a bespoke interface developed for project INDIGO (Watzinger & OpenAtlas team, n.d.-a). The activation of this interface requires specific parameters in a configuration file, including the identifier of the collection that needs to be fetched and the base URI of the ARCHE instance which must be queried. The latter is especially useful in testing environments, when data are still temporarily stored on the staging instance with data collections still undergoing curation, and not yet on the production instance of ARCHE with datasets actually ingested into the archive.

Within the import interface, users with respective rights can check which data available in the collection in ARCHE have not yet been imported into OpenAtlas. These data

are displayed in a table (based on the 'Artifact' class), as can be seen in Figure 2, and can be selected and imported, respectively.

In the case of project INDIGO, not all data stored in ARCHE are needed in OpenAtlas. For example, only merged orthophotographs representing one graffito, and not the individual source photographs, are selected for import. Not only data, but also metadata are imported into OpenAtlas. These include metadata pertaining to the image files, which are modelled according to the ARCHE metadata schema, as well as metadata directly embedded within an image file. The former are exposed and fetched via ARCHE's core API and the latter are fetched through the ExifTool service.

The mapping of metadata coming from ARCHE to OpenAtlas' data model is visualised in Figure 3. This mapping is specific to project INDIGO and assumes that each image represents one artefact, i.e. a graffito, which is an instance of the class E22 Human-Made Object in CIDOC CRM. The import automatically creates instances of relevant classes,



Figure 2. Tabular view of data archived in ARCHE that can be imported into an OpenAtlas project. Source: OpenAtlas.

like 'Artifact', 'Agent', or 'Place'. Most of the metadata to be mapped are fetched from the data served via the ExifTool service and in Figure 3 are indicated with the namespaces 'EXIF', 'IPTC', and 'XMP'.

While the actual graffito is represented as an artefact in OpenAtlas, its digital counterpart, i.e. the orthophoto coming from ARCHE, is treated as a related file (which is modelled

as an instance of E31 Document in CIDOC CRM). Thus, the graffito and the image of the graffito are described with different properties in OpenAtlas, as Figure 3 exemplifies.

Once data and metadata are imported, it is possible to enrich them through the OpenAtlas web interface. In the data model of OpenAtlas, entities are described with types, i.e. attributes or properties, which correspond to the class E55

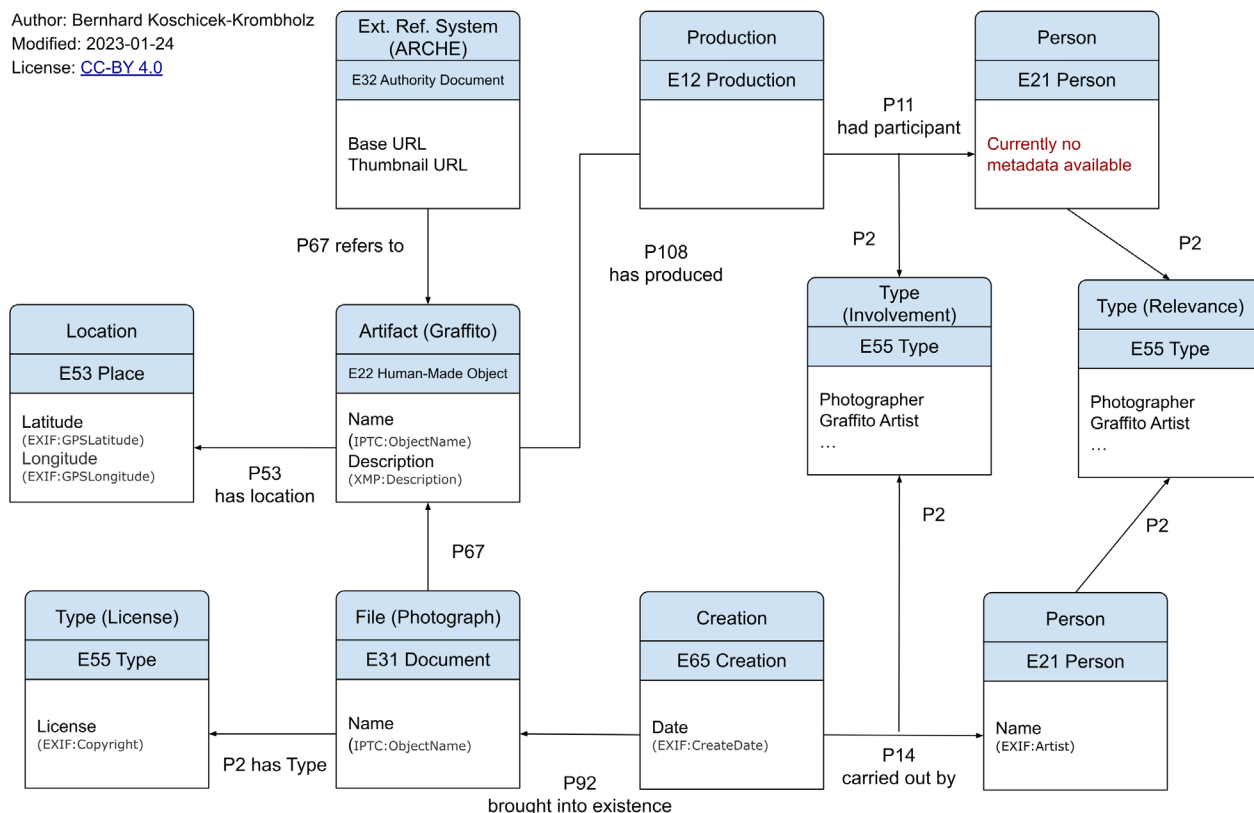


Figure 3. Mapping of ARCHE metadata to the OpenAtlas data model. The namespaces EXIF, IPTC, and XMP indicate metadata fetched by ARCHE's ExifTool service. Source: Bernhard Koschicek-Krombholz via (Watzinger & OpenAtlas team, n.d.–a), [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).

in CIDOC CRM. Hierarchical trees of types can be defined or imported, where the types at a higher level are more general, and those at a lower level, i.e. the subtypes, are more specific. OpenAtlas already ships with some standard types, but each project can define types tailored to their research purposes. In the example of project INDIGO, a standard type is 'Artifact'. Project INDIGO added further custom subtypes to the tree 'Artifact', such as 'visual works', 'graffiti', and 'pieces (graffiti)'. In terms of a database or metadata schema, a type like 'Artifact' can be thought of as the equivalent of a property, while subtypes, like 'visual work', can be considered values for that property. By using types and assigning one or even multiple subtypes to an artefact, like a graffiti, a user can describe it in a more specific and expressive way, thus enabling interesting insights and cross-queries.

For project INDIGO, new functionality was implemented in OpenAtlas to allow controlled vocabularies from ACDH-CH's Vocab Repository to be imported into OpenAtlas types. This was implemented by using the API of the Vocab Repository, which is based on the Skosmos software (ACDH-CH, n.d.–b). More specifically, the new functionality enables to access and retrieve concepts present in a controlled vocabulary, their relevant metadata properties, such as preferred and alternative labels, and their mutual hierarchical relationships, which can be transposed into the hierarchical organisation of type trees in OpenAtlas. For example, the concept 'pieces (graffiti)' (<https://vocabs.acdh.oeaw.ac.at/indigo/piecesGraffiti>) is narrower than the concept 'graffiti' (<https://vocabs.acdh.oeaw.ac.at/indigo/graffiti>); both concepts would be imported as subtypes into an OpenAtlas

type and ‘pieces (graffiti)’ would then be a type subordinate to the type ‘graffiti’. In addition, the integration between ACDH-CH’s Vocab Repository and OpenAtlas also allows access to concepts that are arranged into collections, i.e. groupings of concepts that are related and can be identified through a common label. One example presented in Figure 4 is the Vocab Repository collection ‘sign and element’ (<https://vocabs.acdh.oeaw.ac.at/indigo/signAndElement>), which includes the concepts ‘@ signs (motif)’ (<https://vocabs.acdh.oeaw.ac.at/indigo/atSignsMotif>) and ‘anarchy symbols’ (<https://vocabs.acdh.oeaw.ac.at/indigo/anarchySymbols>). In OpenAtlas, the collection ‘sign and element’ will be imported as a type, of which the two concepts ‘@ signs (motif)’ and ‘anarchy symbols’ will be subtypes.

By leveraging the capabilities of both ACDH-CH’s Vocab Repository and OpenAtlas together, it is possible to arrange and describe necessary type trees in an external knowledge organisation system, providing hierarchical and associative relations between the types and subtypes as well as rich metadata about them, including definitions and sources for each of them. This also allows for wide accessibility and

reusability of the hierarchical type trees across different systems. In project INDIGO, the Graffiti Thesaurus (Schlegel, Wogrin, et al., 2023) has been published on the Vocab Repository and will be used to populate specific predefined and custom types in OpenAtlas.

6. Conclusions

The necessary functionality for the integration of the digital archive ARCHE, the spatial database OpenAtlas, and ACDH-CH’s Vocab Repository is already implemented and has so far been tested with exemplary test data. The main implemented features include ARCHE’s ExifTool service for the dissemination of embedded metadata of files stored within ARCHE, OpenAtlas’ ARCHE module to allow fetching data and metadata from ARCHE, and OpenAtlas’ feature to import concepts and collections from ACDH-CH’s Vocab Repository.

Currently, the OpenAtlas’ ARCHE module is tailored for importing INDIGO data from ARCHE into an OpenAtlas instance for project INDIGO. Thus, the mapping of ARCHE’s and the images’ metadata to the data model of OpenAtlas

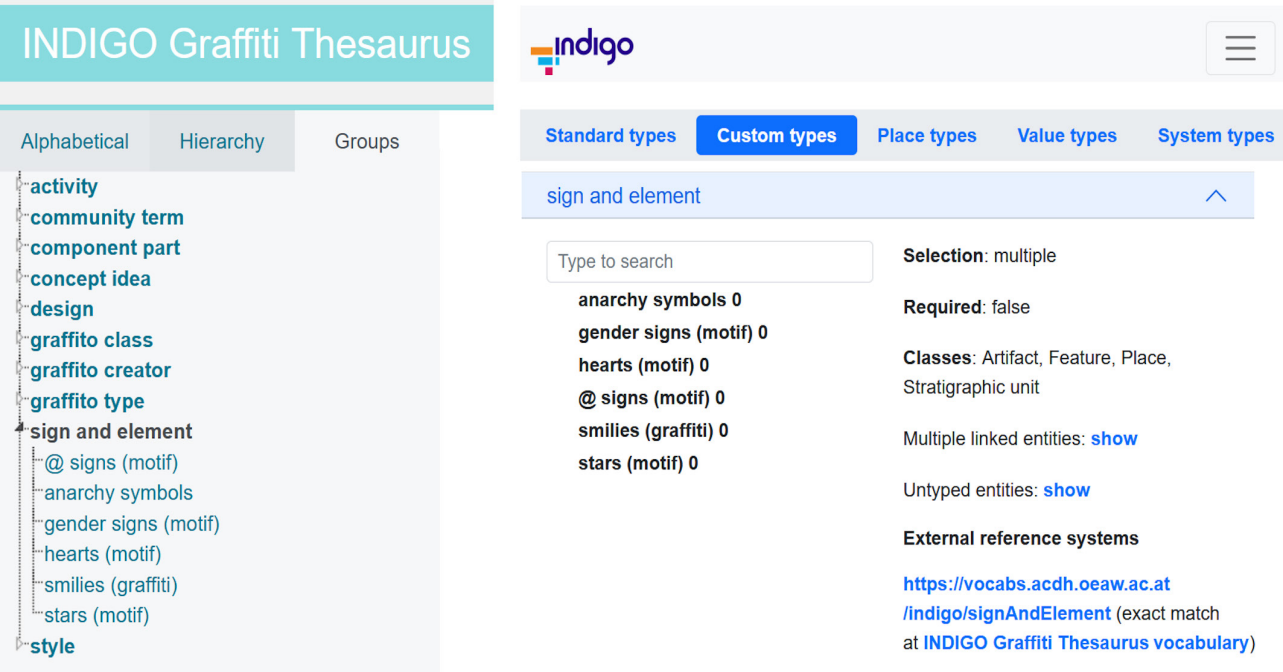


Figure 4. The collection ‘sign and element’ in ACDH-CH’s Vocab Repository and as a type in OpenAtlas. Source: Martina Trognitz via ACDH-CH’s Vocab Repository and OpenAtlas.

is highly customised and can not be reused for any other file structure or project setup. Further work on this module is needed to implement a more flexible or generalised mapping. This would then allow other OpenAtlas projects to use it.

The next steps will include a full ingest of all INDIGO data into ARCHE, which will take some time due to the large amount of files, about 430,000, which include around 150,000 RAW photographs, and accumulate to a total estimated size of 20 TB for the whole dataset. Once ingested, the data can then be imported into OpenAtlas and further enriched. The enrichment of data after they have been archived poses another question to be solved: How can this additional information be fed back into ARCHE to allow the data and metadata to be preserved without depending on OpenAtlas?

As of June 2024, ARCHE has so far archived one dataset originating from an OpenAtlas project: Mapping Medieval Conflict (MEDCON) (Preiser-Kapeller, 2020). The data were archived as a database dump in SQL format with accompanying documentation files. For some other OpenAtlas projects, archiving will be performed in a similar way. Project INDIGO poses a new case because archiving of project data is done first and then an instance of OpenAtlas is set up that fetches data and metadata from ARCHE for further enrichment. While implementing a solution to feed back the data enriched in OpenAtlas to ARCHE is still open, such a solution would help integrate data management tasks related to long-term preservation into the research workflow. This would ensure the correct preparation of data and a smooth transition into the archive from the early stages of the project.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

INDIGO was funded by the Heritage Science Austria programme of the Austrian Academy of Sciences (ÖAW).

References

- ACDH-CH. (n.d.-a). *ACDH-CH/DARIAH Vocabularies*. Retrieved July 9, 2023, from <https://vocabs.acdh.oeaw.ac.at/en>
- ACDH-CH. (n.d.-b). *ACDH-CH/DARIAH Vocabularies - API*. Retrieved August 16, 2023, from <https://vocabs.acdh.oeaw.ac.at/en/about#api>
- ACDH-CH. (n.d.-c). *ARCHE - A Resource Centre for Humanities Related Research in Austria*. Retrieved June 9, 2023, from <https://arche.acdh.oeaw.ac.at>
- ALLEA. (2023). *The European Code of Conduct for Research Integrity*. <https://allea.org/code-of-conduct>
- Andorfer, P., & Illmayer, K. (2023). *Vocabs Editor and vocabulary management at OEAW ACDH-CH*. <https://doi.org/10.5281/zenodo.7963135>
- Berners-Lee, T. (2006). *Linked Data*. <https://www.w3.org/DesignIssues/LinkedData.html>
- CLARIN. (2023). *CLARIN Virtual Language Observatory*. <https://vlo.clarin.eu>
- Consultative Committee on Space Data Systems. (June 2012). *Reference Model for an Open Archival Information System (OAIS)* (Recommended Practice CCSDC 650.0-M-2).
- CoreTrustSeal. (n.d.). *CoreTrustSeal*. Retrieved November 27, 2023, from <https://www.coretrustseal.org>
- Corporation for National Research Initiatives. (2023). *Handle.Net Registry*. <https://www.handle.net>
- Czirjak, N., Žóltak, M., Trognitz, M., & Ďurčo, M. (2023). *acdh-oeaw/repo-file-checker* (Version 3.1.2) [Computer software]. Zenodo. <https://zenodo.org/records/8232570>
- Europeana. (n.d.). *Discover Europe's digital cultural heritage*. Retrieved September 12, 2023, from <https://www.europeana.eu/en>

Filzwieser, R., & Eichert, S. (2020). Towards an Online Database for Archaeological Landscapes. Using the Web Based, Open Source Software OpenAtlas for the Acquisition, Analysis and Dissemination of Archaeological and Historical Data on a Landscape Basis. *Heritage*, 3(4), 1385–1401. <https://doi.org/10.3390/heritage3040077>

FOSTER. (n.d.). *Open Science Definition*. Retrieved February 10, 2023, from <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

Harpring, P. (2013). *Introduction to controlled vocabularies: Terminology for art, architecture, and other cultural works* (Updated edition). Getty Research Institute.

Harvey, P. (2023). *Exiftool - Read, Write and Edit Meta Information!* <https://exiftool.org>

ICOM. (n.d.). *The CIDOC CRM*. Retrieved August 16, 2016, from <http://www.cidoc-crm.org>

Isaac, A., & Summers, E. (2009, September 18). *SKOS Simple Knowledge Organization System Primer*. W3C Working Group Note. <https://www.w3.org/TR/skos-primer>

Kunze, J., Littman, J., Madden, E., Scancella, J., & Adams, C. (2018). *The BagIt File Packaging Format (V1.0)* (Request for Comments No. 8493). RFC Editor. <https://doi.org/10.17487/RFC8493>

Molada-Tebar, A., & Verhoeven, G. J. (2023). Towards colour-accurate documentation of anonymous expressions. In G. J. Verhoeven, J. Schlegel, B. Wild, S. Wogrin, & M. Carloni (Eds.), *Document | archive | disseminate graffiti-scapes. Proceedings of the goINDIGO2022 international graffiti symposium* (pp. 86–130). Urban Creativity. <https://doi.org/10.48619/indigo.v0i0.704>

Open Knowledge Foundation. (n.d.). *Open Definition 2.1 - Defining Open in Open Data, Open Content and Open Knowledge*. Retrieved February 10, 2023, from <https://opendefinition.org/od/2.1/en>

OpenAIRE. (2023). *OpenAIRE*. <https://www.openaire.eu>

OpenAtlas team. (2023). *OpenAtlas*. <https://openatlas.eu>

Preiser-Kapeller, J. (2020). *Mapping Medieval Conflict (MEDCON) [Data set]*. <https://hdl.handle.net/21.11115/0000-000C-D99B-1>

Richards, N., Eichert, S., & Watzinger, A. (2023). One Ontology to Rule Them All—CIDOC CRM in the Humanities and Its Use in OpenAtlas. In G. J. Verhoeven, J. Schlegel, B. Wild, S. Wogrin, & M. Carloni (Eds.), *Document | archive | disseminate graffiti-scapes. Proceedings of the goINDIGO2022 international graffiti symposium* (pp. 220–230). Urban Creativity. <https://doi.org/10.48619/indigo.v0i0.711>

Sbarbati, S. (2015). *Activities and Workflow - PARTHENOS Project*. <https://www.parthenos-project.eu/activities-and-wps>

Schlegel, J., Carloni, M., Wogrin, S., Graff, A. M., & Verhoeven, G. J. (2023). Making a mark—Towards a graffiti thesaurus. In G. J. Verhoeven, J. Schlegel, B. Wild, S. Wogrin, & M. Carloni (Eds.), *Document | archive | disseminate graffiti-scapes. Proceedings of the goINDIGO2022 international graffiti symposium* (pp. 203–219). Urban Creativity. <https://doi.org/10.48619/indigo.v0i0.710>

Schlegel, J., Wogrin, S., Carloni, M., & Verhoeven, G. J. (2023). *ACDH-CH/DARIAH Vocabularies: INDIGO Graffiti Thesaurus v1.0.0*. <https://vocabs.acdh.oeaw.ac.at/indigo/Thesaurus>

Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., Champin, P.-A., & Lindström, N. (2020). *JSON-LD 1.1: A JSON-based Serialization for Linked Data*. W3C Recommendation. <https://www.w3.org/TR/json-ld>

Suominen, O., Ylikotila, H., Pessala, S., Lappalainen, M., Frosterus, M., Tuominen, J., Baker, T., Caracciolo, C., & Retterath, A. (2015). *Publishing SKOS vocabularies with Skosmos*. <https://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf>

- Trognitz, M. (2021). Saving us from the Digital Dark Age: The Austrian perspective. *Internet Archaeology*, 58. <https://doi.org/10.11141/ia.58.2>
- Trognitz, M., & Ďurčo, M. (2018). One Schema to Rule them All. The Inner Workings of the Digital Archive ARCHE. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 71(1), 217–231. <https://doi.org/10.31263/voebm.v71i1.1979>
- Trognitz, M., Ďurčo, M., & Mörth, K. (2022). Text Technology for the Digital Humanities. In A. Witt (Series Ed.) & D. Fišer & A. Witt (Vol. Eds.), *Digital Linguistics: Vol. 1. CLARIN: The infrastructure for language resources* (pp. 223–247). De Gruyter. <https://doi.org/10.1515/9783110767377-009>
- Tsoulouha, E., & Goulis, H. (n.d.). *Thesaurus Maintenance | DARIAH*. Retrieved June 5, 2023, from <https://www.dariah.eu/activities/working-groups/thesaurus-maintenance>
- Verhoeven, G. J., Wild, B., Schlegel, J., Wieser, M., Pfeifer, N., Wogrin, S., Eysn, L., Carloni, M., Koschiček-Krombholz, B., Molada-Tebar, A., Otepka-Schremmer, J., Ressler, C., Trognitz, M., & Watzinger, A. (2022). Project INDIGO – document, disseminate & analyse a graffiti-scape. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022, 513–520. <https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-513-2022>
- Verhoeven, G. J., Wogrin, S., Schlegel, J., Wieser, M., & Wild, B. (2023). Facing a chameleon—How project INDIGO discovers and records new graffiti. In G. J. Verhoeven, J. Schlegel, B. Wild, S. Wogrin, & M. Carloni (Eds.), *Document | archive | disseminate graffiti-scapes. Proceedings of the goINDIGO2022 international graffiti symposium* (pp. 63–85). Urban Creativity. <https://doi.org/10.48619/indigo.v0i0.703>
- Vocabs SSHOC. (n.d.). *SSH Vocabulary Commons*. Retrieved June 5, 2023, from <https://vocabs.sshopencloud.eu/browse/en>
- Watzinger, A., & OpenAtlas team. (n.d.–a). *ARCHE – OpenAtlas 7.16.0 documentation*. Retrieved October 7, 2023, from <https://manual.openatlas.eu/admin/arche.html>
- Watzinger, A., & OpenAtlas team. (n.d.–b). *Network visualization – OpenAtlas 7.16.0 documentation*. Retrieved October 7, 2023, from <https://manual.openatlas.eu/tools/network.html>
- Wild, B., Verhoeven, G. J., Wieser, M., Ressler, C., Schlegel, J., Wogrin, S., Otepka-Schremmer, J., & Pfeifer, N. (2022). AUTOGRAF—Automated Orthorectification of GRAffiti Photos. *Heritage*, 5(4), 2987–3009. <https://doi.org/10.3390/heritage5040155>
- Wild, B., Verhoeven, G. J., Wogrin, S., Wieser, M., Ressler, C., Otepka-Schremmer, J., & Pfeifer, N. (2023). Urban creativity meets engineering. Automated graffiti mapping along Vienna's Donaukanal. In G. J. Verhoeven, J. Schlegel, B. Wild, S. Wogrin, & M. Carloni (Eds.), *Document | archive | disseminate graffiti-scapes. Proceedings of the goINDIGO2022 international graffiti symposium* (pp. 131–145). Urban Creativity. <https://doi.org/10.48619/indigo.v0i0.705>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zaytseva, K., & Ďurčo, M. (2020). *Controlled Vocabularies and SKOS: Version 1.1.0*. DARIAH-Campus. <https://campus.dariah.eu/id/D8d6OrLdpLIGRqBSQDVNO>
- Żóttak, M., Trognitz, M., & Ďurčo, M. (2022). ARCHE Suite: A Flexible Approach to Repository Metadata Management. In M. Monachini & M. Eskevich (Eds.), *Linköping Electronic Conference Proceedings: Vol. 189, Selected Papers from the CLARIN Annual Conference 2021* (pp. 190–199). Linköping University Electronic Press. <https://doi.org/10.3384/p18917>