

The Data Crew - Archive with ARCHE and Enrich with OpenAtlas

goINDIGO 2023

June 14th, 2023 - Massimiliano Carloni, Bernhard Koschiček-Krombholz &
Martina Trognitz, ACDH-CH ÖAW

Services you will get to know



- Austrian Centre for Digital Humanities and Cultural Heritage
- Founded in 2015, restructured in 2019 (+ CH)
 - Language, music, history, archaeology
- International involvement: DARIAH, CLARIN, SSHOC (EOSC)

«Foster the humanities by applying digital methods and tools to a wide range of academic fields»



<https://www.oeaw.ac.at/acdh/>

What is it?



digital archive RDF, OWL



controlled vocabularies SKOS

OpenAtlas

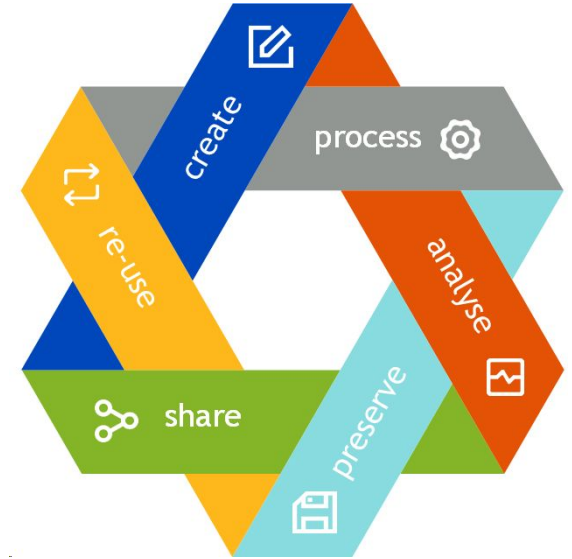


database

CIDOC CRM

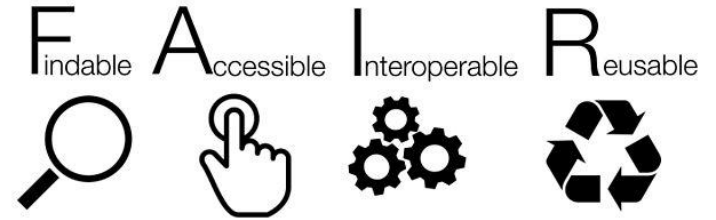
Guiding principles

- Keep the Data Life Cycle going
- Good scientific practice



Guiding principles

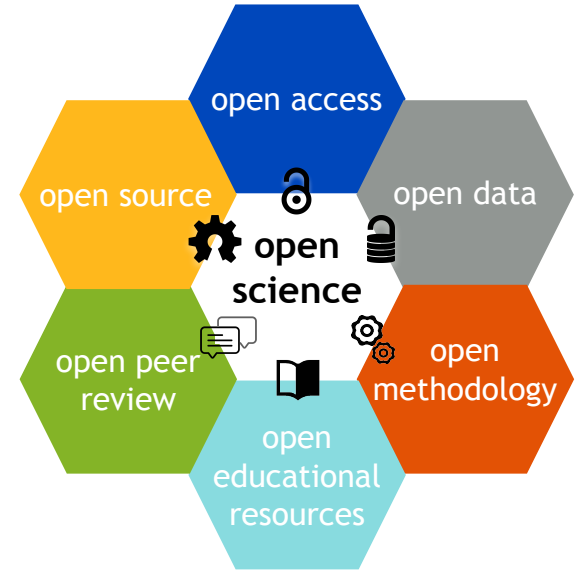
- Keep the Data Life Cycle going
- Good scientific practice
- FAIR data and metadata



[SangyaPundir CC BY-SA 4.0]

Guiding principles

- Keep the Data Life Cycle going
- Good scientific practice
- FAIR data and metadata
- Open Science



Guiding principles

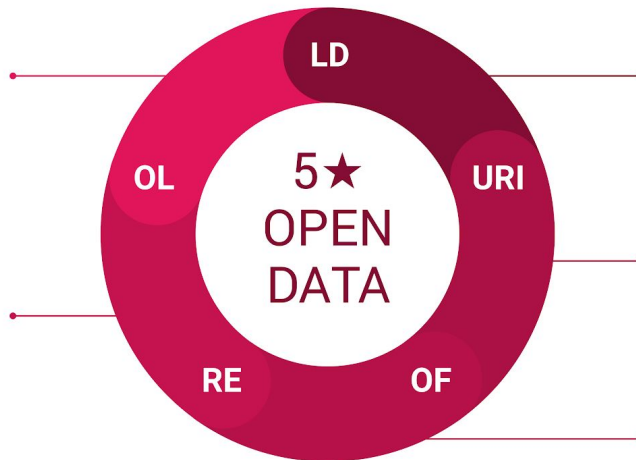
- Keep the Data Life Cycle going
- Good scientific practice
- FAIR data and metadata
- Open Science
- Linked Open Data

★ open licence [PDF]

available on the web (whatever format) but with an open licence, to be Open Data.

★★ machine readable [XLS]

available as machine-readable structured data.



★★★★★ Linked Open Data

★★★★ plus: link your data to other people's data to provide context.

★★★★ use URIs [RDF]

★★★ plus: use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff.

★★★ open format [CSV]

★★ plus non-proprietary format.

[Florian Thiery CC BY 4.0]

Standards to be used

- W3C Semantic Web standards
 - SKOS, RDF etc.
- JSON(-LD)
- Metadata
 - CIDOC CRM
 - IPTC
 - EXIF
 - XMP



ARCHE

- A Resource Centre for the Humanities established in 2017
- A digital archive for humanities data: long-term preservation
- Acquired [CoreTrustSeal](#) in 2017/2018, renewed in 2021

www.arche.acdh.oeaw.ac.at/browser/



ARCHE - Data

- ARCHE preserves
 - Plain text, text documents, PDF documents, images, spreadsheets, databases, audio, video, 3d data, GIS and much more
- ... as long as files are provided in formats suitable for long-term preservation
 - ☞ [IT-Empfehlungen IANUS](#) & [Guides to Good Practice ADS & tDAR](#)
- ... file names comply with the ARCHE naming rules
 - ☞ a-Z, 0-9, -, _
- ... and sufficient Metadata is provided
 - ☞ ARCHE has a [dedicated metadata schema](#)

[www arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata](http://www.arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata)

ARCHE - Dissemination Services & API

- One file: many ways
- Each collection and each public resource gets a PID
- Machine access (APIs)
 - ARCHE main API
 - OAI-PMH
 - IIIF
 - RDF
 - JSON
 - BibLaTeX
 - OpenRefine
 - ... and many more

www.arche.acdh.oeaw.ac.at/browser/api-access

ARCHE - Dissemination Services & API



- One file: many ways
- Each collection and each public resource gets a PID
- Machine access (APIs)

- ARCHE main API
- OAI-PMH
- IIIF
- RDF
- JSON
- BibLaTeX
- OpenRefine
- ... and many more

Dissemination Services



```
JSON Rohdaten Kopfzeilen
Speichern Kopieren Alle einklappen Alle ausklappen JSON durchsuchen
SourceFile: "/tmp/exif_cache_83j07NM"
ExifToolVersion: 12.18
FileName: "exif_cache_83j07NM"
Directory: "/"
FileSize: "77 KiB"
FileModifyDate: "2023:06:13 10:15:30+00:00"
FileAccessDate: "2023:06:13 10:15:29+00:00"
FileInodeChangeDate: "2023:06:13 10:15:30+00:00"
FilePermissions: "rw-r-----"
FileType: "TIFF"
FileTypeExtension: ".tif"
MIMEType: "image/tiff"
ExifByteOrder: "little-endian (Intel, II)"
ImageWidth: 8543
ImageHeight: 2671
BitsPerSample: "8 8 8 8"
Compression: "LZW"
PhotometricInterpretation: "sRGB"
SamplesPerPixel: 4
PlanarConfiguration: "Chunky"
SubFileType: "Reduced-resolution image"
TitleWidth: 256
TitleLength: 256
TitleByteOrder: 00000771
```



OpenAtlas




www.arche.acdh.oeaw.ac.at/browser/api-access

Overview

[↔ Copy Resource Link](#)
[☰ Switch to Expert-View](#)

INDIGO

 Type(s): [acdh:TopCollection](#)

 Contact(s): [Geert J. Verhoeven](#)
 Available Date: 13 Jun 2023

 Number of Items: 157

 Binary Size: 3.36 GB

 Licensor: [Geert J. Verhoeven](#)
 Owner: [Geert J. Verhoeven](#)
 Identifier(s): <https://arche-curation.acdh-dev.oeaw.ac.at/api/1491014>,
<https://id.acdh.oeaw.ac.at/indigo>


Summary

 Subject(s): graffiti


 Description: This is a test collection for project INDIGO. It was created to test the interaction between ARCHE and OpenAtlas, as well as to develop the curation and ingestion workflow for the data coming from project INDIGO.

Cite Resource

[APA 6TH](#)
[HARVARD](#)
[VANCOUVER](#)
[BIBLATEX](#)

 INDIGO. (2023). [Data set]. <https://arche-curation.acdh-dev.oeaw.ac.at/api/1491014>
[Copy](#)

Child Resource(s)

 [Switch to List-View](#)
 [1_RAWs](#)
 [2_JPEGs](#)
 [3_Polygons](#)
 [4_Orthophotos](#)
 [5_MeshAndTextures](#)

Overview

[Copy Resource Link](#)
[Switch to Expert-View](#)

INDIGO

Type(s):	acdh:TopCollection	
Contact(s):	Geert J. Verhoeven	
Available Date:	13 Jun 2023	
Number of Items:	157	
Binary Size:	3.36 GB	
Licensor:	Geert J. Verhoeven	
Owner:	Geert J. Verhoeven	
Identifier(s):	https://arche-curation.acdh-dev.oeaw.ac.at/api/1491014 , https://id.acdh.oeaw.ac.at/indigo	

Summary

Subject(s): graffiti

Description: This is a test collection for project INDIGO. It was created to test the interaction between ARCHE and OpenAtlas, as well as to develop the curation and ingestion workflow for the data coming from project INDIGO.

Cite Resource

APA 6TH HARVARD VANCOUVER BIBLATEX

INDIGO. (2023). [Data set]. <https://arche-curation.acdh-dev.oeaw.ac.at/api/1491014>

Copy

Child Resource(s)

[Switch to List-View](#)

- 1_RAWs
- 2_JPEGs
- 3_Polygons
- 4_Orthophotos
- 5_MeshAndTextures

[INDIGO /4_Orthophotos /INDIGO_2023-03-27_G112 /](#)

Overview

[Copy Resource Link](#)
[Switch to Expert-View](#)

INDIGO_2023-03-27_G112_Ortho_10mm.png

Type(s): [acdh:Resource](#)



Available Date: 13 Jun 2023

Binary Size: 0.45 MB

Category: [Image](#)

Licensor: [Geert J. Verhoeven](#)

License: [In Copyright](#)

Owner: [Geert J. Verhoeven](#)

Access Restriction: [public](#)

Identifier(s): <https://arche-curation.acdh-dev.oeaw.ac.at/api/1491109>,
https://id.acdh.oeaw.ac.at/indigo/4_Orthophotos/INDIGO_2023-03-27_G112
[/INDIGO_2023-03-27_G112_Ortho_10mm.png](https://id.acdh.oeaw.ac.at/indigo/4_Orthophotos/INDIGO_2023-03-27_G112_Ortho_10mm.png)

Part of: [INDIGO_2023-03-27_G112](#)

OpenAtlas

- An open source database software
- Initiated about 10 years ago by Stefan Eichert
- Developed especially to acquire and manage research data
- Used for various fields of humanities like history and archaeology
- Record complex relations as simple as possible
- Uses the international standard of the CIDOC CRM



<https://openatlas.eu/>



074_lowRes



Type **graffito**

Associated concept **greeting highlights (graffito) message shadows (graffito)**

Owned by **unknown**

Class **E22**

Description

text: "Die Bitch wird 20"



INDIGO_2022-07-22_Z7II-A_0076

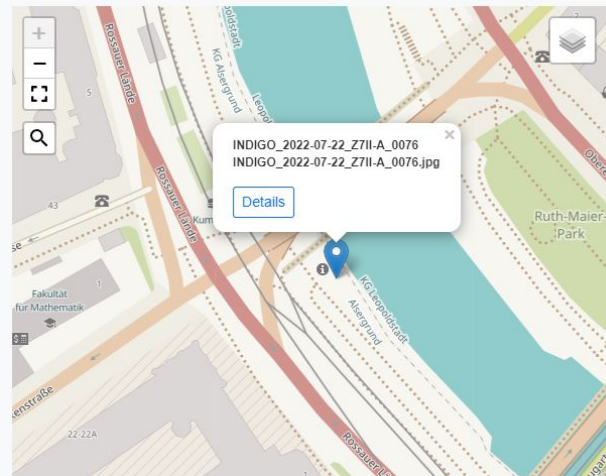


Class **E22**

Created 2023-03-20

External reference systems

https://arche-curation.acdh-dev.oeaw.ac.at/browser/oeaw_detail/1390163 (exact match at ARCHE)



Alphabetical Hierarchy Groups

- Activities <facet>
- Agents <facet>
- Associated Concepts <facet>
- Associated Concepts <hierarchy name>
 - attributes and properties by specific type <guide term>
 - concepts in the arts and humanities <guide term>
 - graffiti concepts
 - backgrounds (graffiti)
 - burners (graffiti)
 - commissioned works (graffiti)
 - concept walls (graffiti)
 - dedications (graffiti)
 - fadings (graffiti)
 - fill ins (graffiti)
 - greetings (graffiti)
 - handstyles (graffiti)
 - memorials (graffiti)
 - messages (graffiti)
 - mirrored (graffiti)
 - productions (graffiti)
 - social science concepts
 - technology and related concepts <guide term>
- Objects <facet>
- Physical Attributes <facet>
- Styles and Periods <facet>

Vocabulary information

TITLE	INDIGO Thesaurus
DESCRIPTION	Placeholder for description.
CREATOR	Jona Schlegel Stefan Wogrin
LANGUAGE	English German
RELATION	https://projectindigo.eu
LAST MODIFIED	2023-02-15
FORMAT	application/x-turtle
TYPE	http://www.w3.org/2004/02/skos/core#ConceptScheme
URI	https://vocabs.acdh.oeaw.ac.at/indigo/Thesaurus

Resource counts by type

Type	Count
Concept	179
• Deprecated concept	0
Collection	11

Term counts by language

Language	Preferred terms	Alternate terms	Hidden terms
English	179	164	0

Associated Concepts <facet> > Associated Concepts
<hierarchy name> > concepts in the arts and humanities
<guide term> > graffiti concepts > burners (graffiti)


PREFERRED TERM **burners (graffiti)** 

BROADER CONCEPT [graffiti concepts](#)

ENTRY TERMS
burner
burner (graffiti)
burners

BELONGS TO GROUP [community term](#)

EDITORIAL NOTE
I do not see this like the hierarchical situation in the Getty THIS is a concept, quite objective - a really good - the best - graffito - this is an aspect that also depends on the judge and audience - it is kind of like a title and appreciation of the graffito - my personal question: how did this concept changed nowadays as the graffitiists have more time to finish their work?

URI <https://vocabs.acdh.oeaw.ac.at/indigo/burnersGraffiti> 

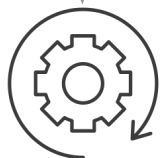
DOWNLOAD THIS CONCEPT: [RDF/XML](#) [TURTLE](#) [JSON-LD](#)

EXACTLY MATCHING CONCEPTS <http://vocab.getty.edu/page/aat/300410278> [vocab.getty.edu](#)

The workflow



- Photographs
- Initial metadata



- Orthorectified images
- Enriched metadata

indigo

UrbanChameleon web platform



- Long-term storage
- Basic metadata (ARCHE Schema)

OpenAtlas



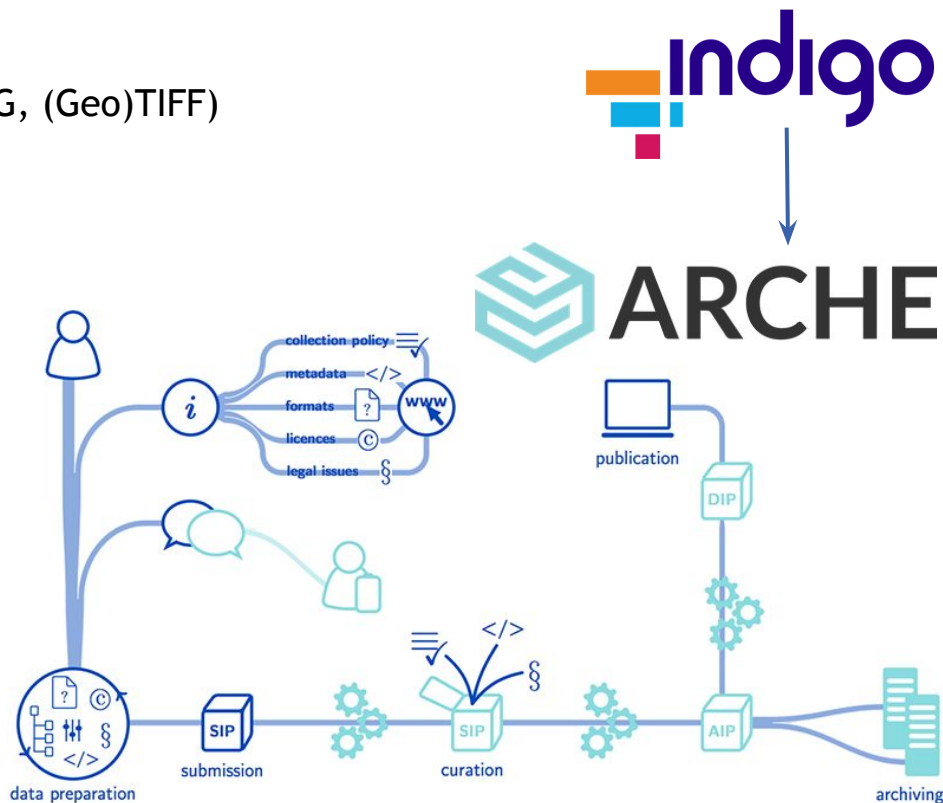
- Acts as data repository
- Enrich metadata
- CIDOC CRM



- Dedicated vocabulary for metadata properties & types

INDIGO to ARCHE

- Test batch from INDIGO
 - raw and processed images (NEF, JPEG, (Geo)TIFF)
 - sidecar files with metadata (XMP)
 - 2D polygons (GeoJSON)
 - 3D mesh and textures (PLY, JPEG)
- Curation strategy
 - Folder structure, file names
 - Required format conversions
 - Map metadata to ARCHE schema ([ExifTool](#)-based script)
 - Licensing



ARCHE ready for OpenAtlas

- ARCHE main API with core functionality ([CRUD](#))
- Files and Metadata exposed via [dissemination services](#)
 - EXIF Endpoint: Serve JSON from embedded metadata
 - Thumbnail service: Serve thumbnails
 - IIIF Endpoint: Serve images with standardised methods
- Extensive [documentation](#)

[INDIGO/4_Orthophotos/INDIGO_2023-03-27_G112/](#)

Overview

[Copy Resource Link](#)

[Switch to Expert-View](#)

● [INDIGO_2023-03-27_G112_Ortho_10mm.png](#)

Type(s): [acdh:Resource](#)



Available Date: 13 Jun 2023

Binary Size: 0.45 MB

Category: [Image](#)

Licensor: [Geert J. Verhoeven](#)

License: [In Copyright](#)

Owner: [Geert J. Verhoeven](#)

Access Restriction: [public](#)

Identifier(s): <https://arche-curation.acdh-dev.oeaw.ac.at/api/1491109>,
https://id.acdh.oeaw.ac.at/indigo/4_Orthophotos/INDIGO_2023-03-27_G112,
https://id.acdh.oeaw.ac.at/indigo/4_Orthophotos/INDIGO_2023-03-27_G112_Ortho_10mm.png

Part of: [INDIGO_2023-03-27_G112](#)

Dissemination Services

[RDF access](#) [BibLaTeXservice](#) [EXIF Endpoint](#) [View image](#) [Thumbnail service](#) [Download](#)
[IIIF Endpoint](#) [GUI access](#) [Add to CLARIN Virtual Collection](#)

ARCHE to OpenAtlas (1)

- Set up OpenAtlas instance for INDIGO
- Establish connection to
 - Vocab for types
 - ARCHE for files & metadata



Vocabs to OpenAtlas

- Vocabulary and concepts are exposed via API
- OpenAtlas fetches them in JSON-LD



OpenAtlas



```
@language: en
▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/graffitiConcepts"
▼ narrower:
  ▼ 0:
    ▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/burnersGraffiti"
      prefLabel: "burners (graffiti)"
    ▼ 1:
      ▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/handstylesGraffiti"
        prefLabel: "handstyles (graffiti)"
    ▼ 2:
      ▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/dedicationsGraffiti"
        prefLabel: "dedications (graffiti)"
    ▼ 3:
      ▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/worksGraffiti"
        prefLabel: "works (graffiti)"
    ▼ 4:
      ▼ uri: "https://vocabs.acdh.oeaw.ac.at/indigo/worksGraffiti"
        prefLabel: "works (graffiti)"
```

Associated Concepts <facet>

Type to search + Type Edit Delete

- Associated Concepts 0
 - attributes and properties by specific type 0
 - form attributes 0
 - concepts in the arts and humanities 0
 - graffiti concepts 0
 - backgrounds (graffiti) 0
 - shadows (graffiti) 0
 - burners (graffiti) 0
 - commissioned works (graffiti) 0
 - concept walls (graffiti) 0
 - dedications (graffiti) 0
 - fadings (graffiti) 0
 - fill ins (graffiti) 0
 - greetings (graffiti) 0
 - handstyles (graffiti) 0
 - memorials (graffiti) 0
 - messages (graffiti) 0
 - mirrored (graffiti) 0
 - productions (graffiti) 0
 - social science concepts 0
 - sociological concepts 0
 - fame 0
 - technology and related concepts 0
 - technology (general associated concept) 0

ARCHE to OpenAtlas (2)

The screenshot shows the OpenAtlas admin interface. At the top, there is a navigation bar with the 'indigo' logo and menu items: Source, Event, Actor, Place, Artifact, Reference, Types. A search bar is present with the text 'Search term'. Below the navigation bar, there is an 'Admin' section with a dropdown menu containing: Profile, Admin, Legal Notice, and Logout. A secondary navigation bar includes: Files, User 9, General, Email, Modules, Map, Content, and Data (highlighted in blue). The main content area is divided into three sections: 'Data integrity checks' with buttons for Orphans, Dates, Similar names, Links, and Link duplicates; 'Data transfer' with buttons for Import, CSV export, JSON export, XML export, SQL export, Execute SQL, and ARCHE; and 'Image processing' with buttons for Create resized images and Delete orphaned resized images. At the bottom, there is an 'API' section with an 'Edit' button and the text 'Public on'.



The screenshot shows the OpenAtlas interface for the 'ARCHE > Fetch' section. The top navigation bar is identical to the previous screenshot. Below the navigation bar, there is a search bar with 'Search term' and a dropdown menu with 'CA', 'DE', 'EN', 'ES', and a gear icon. The main content area shows 'Thanados 7.10.0' and 'Fetched entities 8'. There is an 'Import arche data' button and a search input field. Below this is a table with 10 columns: ID, Name, Image link, Image thumbnail link, Creator, Latitude, Longitude, Description, and License. The table contains 8 rows of data. At the bottom, there is a pagination bar showing 'Showing 1 to 8 of 8 entries' and 'Previous 1 Next'.

ID	Name	Image link	Image thumbnail link	Creator	Latitude	Longitude	Description	License
1390162	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21931166...	16.369365		In Copyrigh
1390163	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21936333...	16.36889833...		In Copyrigh
1390164	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21938	16.36876		In Copyrigh
1390166	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.22297833...	16.36774666...	I think the gu...	In Copyrigh
1390169	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21302666...	16.389155		In Copyrigh
1390171	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21306166...	16.38932		In Copyrigh
1390174	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21316	16.38937666...		In Copyrigh
1390176	INDIGO_202...	https://arche-...	https://arche-thumbnaill...	Geert Verho...	48.21321	16.38927		In Copyrigh

How features for INDIGO enriched OpenAtlas

- Record production of artefacts, e.g. graffiti ([#1500](#))
- Improved time tracking with hours, minutes and seconds ([#1574](#))
- Improved database model to store 3D geometries ([#1631](#))
- Importing vocabulary from Vocab ([#1663](#), [#1991](#))
- Connecting to ARCHE to get image files and metadata ([#1848](#))
- More detailed structure, e.g. track individual components of graffiti ([#1587](#), [#1647](#))



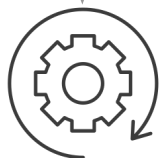
<https://redmine.openatlas.eu/>



The workflow



- Photographs
- Initial metadata



- Orthorectified images
- Enriched metadata

indigo

UrbanChameleon web platform



- Long-term storage
- Basic metadata (ARCHE Schema)

OpenAtlas



- Acts as data repository
- Enrich metadata
- CIDOC CRM



- Dedicated vocabulary for metadata properties & types

Thank you!

massimiliano.carloni@oeaw.at.at
martina.trognitz@oeaw.ac.at

bernhard.koschicek-krombholz@oeaw.ac.at
alexander.watzinger@oeaw.ac.at



Logos originate from the respective project pages.
Source and, if available, licence of external images are indicated.
The remaining content is licenced under [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).