Prosopography and Networks

ACDH-CH DHRI-Heads Meeting, 2023-03-23

Space for running through slides, ESC for an overview

current projects

- VieCPro & ViecoElites -> projects around the viennese courts
- SiCProD -> prosopography of the court of Sigismund (Innsbruck)
- Jelinek -> bibliographic catalogue of Elfriede Jelineks work
- Frischmuth -> similar but different author ;)
- Tibschol -> scholarly exchange of tibetan scholars
- Nomansland -> similar but different area ;)
- InTaVia -> not APIS, but uses API to retrieve/convert data
- MINE/PAAS -> Prosopographie der Geschichte der Akademie

current developments

- major refactoring of APIS to work along the triple pattern
- major changes in the development workflow:
  - git flow
  - pre-commit hooks and GitHub actions for linting and automatic code formatting
- move to cluster2
- automatic deployment of review apps
- automatic tests in CI
- improved documentation

challenges

- refactoring cost more resources than anticipated
- documentation still not at an acceptable level
- missing tests
- hard to keep an overview
- long discussions in the jfx

vision

- prosopographical data & semantic web group rather than the 'APIS' group:
  - use tools developed in InTaVia to finally get the Omnipot up and running
  - establish prefect (or another solution) as ETL in the institute
  - experiment with available triplestores and conversion tools to use the most powerful (while affordable) solution

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

prefect

- ETL (Extract, Transform and Load) solution
- open-source components
- ...but focused on hosted cloud solution
- @ACDH-CH authorization via mTLS (mutual TLS)
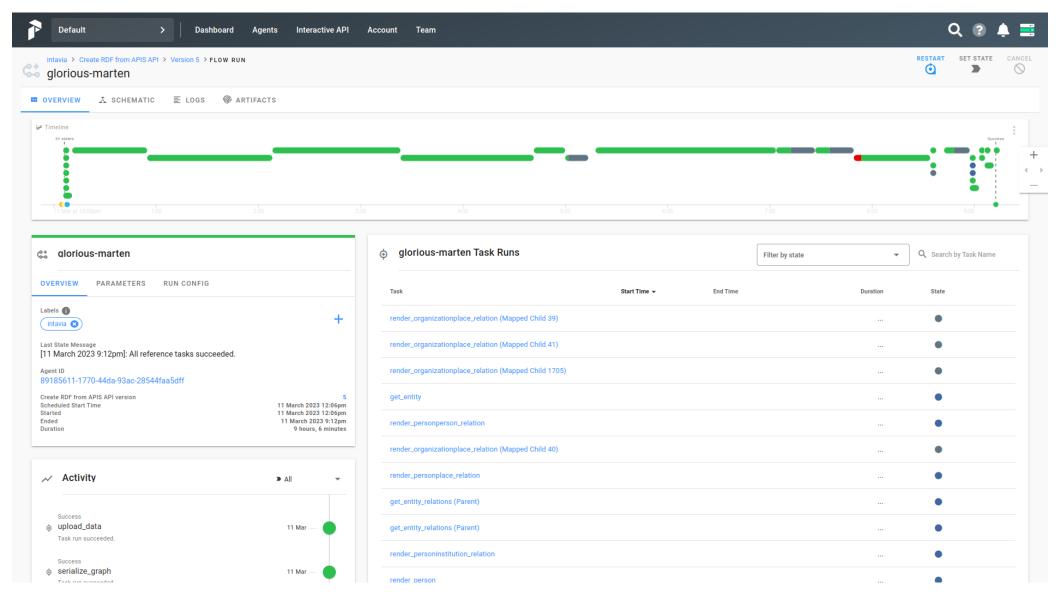- graphql endpoint + web UI based

DAG (Directed acyclic graph)

- one DAG clause (flow)
- many tasks (nodes in the graph)
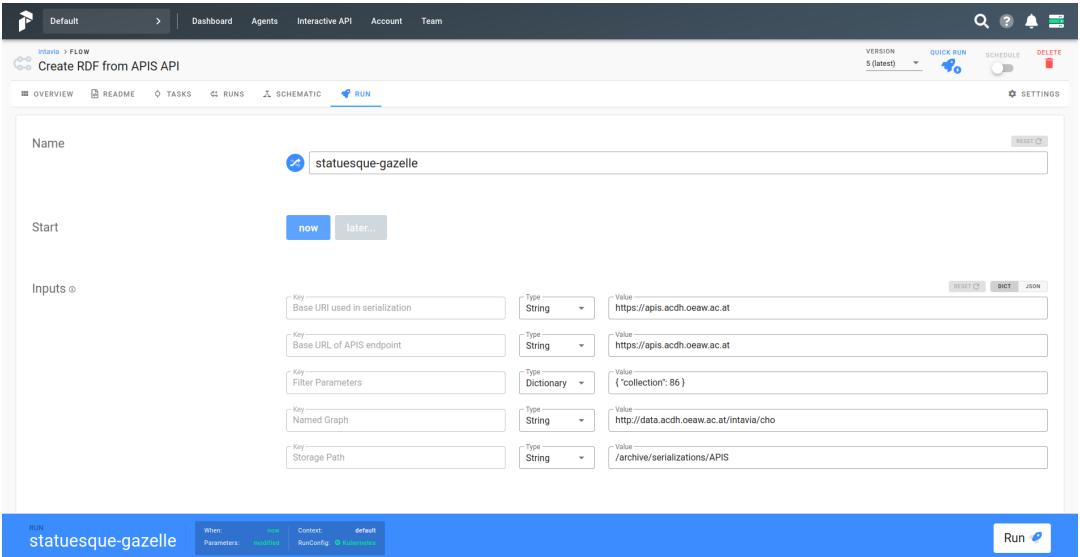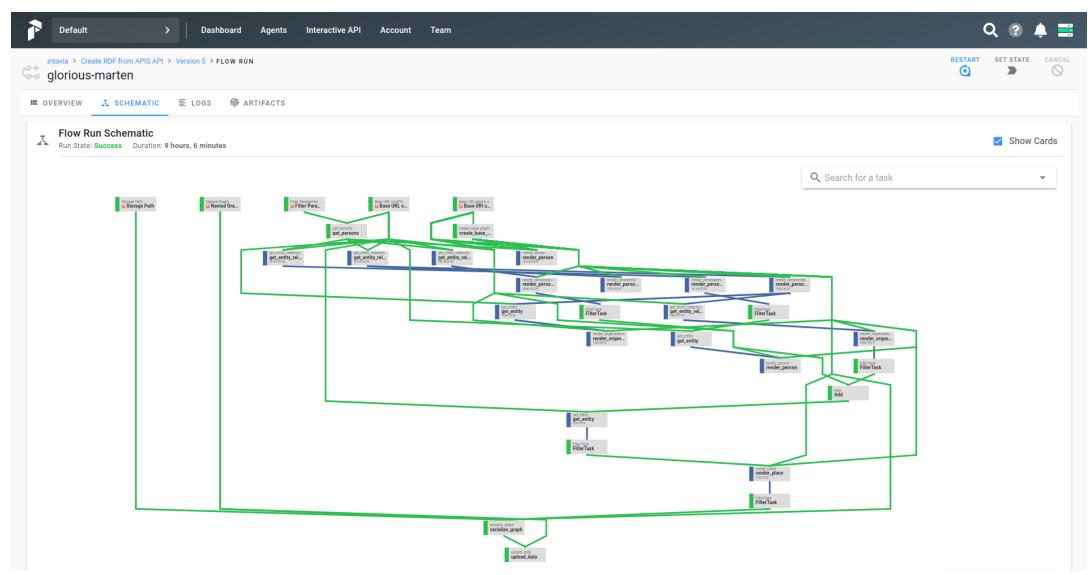- ready made tasks (eg for creating a PR in GitHub)

prefect UI

ACDH-CH DHRI-Heads Meeting, 2023-03-23
Matthias Schlögl, ACDH-CH (Austrian Academy of Sciences)

prefect config run

prefect schema

prefect example

```python
1  from datetime import timedelta
2  from string import Template
3  from prefect import task, Flow, Parameter, context
4  from prefect.storage import GitHub
5  from prefect.run_configs import KubernetesRun
6  from SPARQLWrapper import SPARQLWrapper, JSON
7  import os
8  import requests
9
10
11 TEMP_FOLDER = '/tmp/'
12
13 @task(log_stdout=True)
14 def download_source_data(sources):
15     logger = context.get("logger")
```

prefect live

InTaVia

- H2020 project
- 4 countries: Slovenia, Austria, The Netherlands and Finland
- started 2020-11-01
- less than 1 year to go

Goals

- integrate 4 national biographical dictionaries
- further enrich them (LOD & NLP)
- provide a online platform for Visual Analytics on top of the data

Research questions

- is it possible to integrate different prosopographical resources and create a almost homogenous resource?
- can we use NLP and/or the LOD to enrich the less rich data points?
- does VA provide useful means to analyse such a dataset?

Techstack

- FastAPI (including pydantic for type checking)
- Blazegraph as DB backend
- Jinja2 for creating SPARQL templates
- Redis for caching

models

```python
1  import base64
2  import datetime
3  from enum import Enum
4  import os
5  import re
6  import typing
7  from geojson_pydantic import Point, Polygon
8  from pydantic import BaseModel, Field, HttpUrl, NonNegativeInt, PositiveInt
9  from rdf_fastapi_utils.models import FieldConfigurationRDF, RDFUtilsModelBaseClass
10
11 BASE_URL = os.getenv("BASE_URL", "http://intavia-backend.acdh-dev.oeaw.ac.at")
12
13 source_mapping = {
14     "intavia.eu/apis/personproxy": "Austrian Biographical Dictionary",
15     "data.biographynet.nl": "BiographyNET",
16
```

WWW.OEAW.AC.AT

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

SPARQL templates I

```
1  PREFIX bds:  <http://www.bigdata.com/rdf/search#>
2  PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
3  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX idm: <https://www.intavia.eu/idm/>
5  PREFIX idmcore: <http://www.intavia.eu/idm-core/>
6  PREFIX bioc: <http://ldf.fi/schema/bioc/>
7  PREFIX owl: <http://www.w3.org/2002/07/owl#>
8
9  SELECT ?entity ?entityType ?entityTypeLabel ?entityLabel ?gender ?genderLabel ?nationalityLa
10 ?event ?linkedIds ?count ?geometry ?role_type (?entity as ?source) ?mediaObject ?biographyOb
11
12 {% include 'add_datasets_v2_1.sparql' %}
13
14 WITH {
15 SELECT DISTINCT ?entity ?entityTypeLabel
```

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

SPARQL templates II

```
 1 BIND("no label provided" AS ?defaultEntityLabel)
 2 OPTIONAL {?entity crm:P1_is_identified_by ?appellation .
 3 {?appellation a crm:E33_E41_Linguistic_Appellation .} UNION {?appellation a crm:E35_Title}
 4 ?appellation rdfs:label ?entityLabelPre .}
 5 OPTIONAL {?entity bioc:has_occupation ?occupation . ?occupation rdfs:label ?occupationLabel
 6 OPTIONAL {?entity owl:sameAs ?linkedIds}
 7 OPTIONAL {?entity bioc:has_gender ?gender
 8    OPTIONAL {?gender rdfs:label ?genderLabel }}
 9 OPTIONAL {?entity bioc:has_nationality ?nationality . ?nationality rdfs:label ?nationalityLa
10 OPTIONAL {?entity bioc:bearer_of ?role .
11       ?role ^bioc:had_participant_in_role|^bioc:occured_in_the_presence_of_in_role ?event
12       ?role a ?role_type
13 }
14 OPTIONAL {?entity crm:P168_place_is_defined_by/crm:P168_place_is_defined_by ?geometry}
15 OPTIONAL {?entity ^crm:P70_documents ?mediaObject }
```

InTaVia IDM-Json Backend [2] OAS3

/v2/openapi.json

Development version of the InTaVia backend.

Servers

/v2 ▾

Events endpoints ⌃

| GET | **/api/events/search** Query Events | ⌄ |

| POST | **/api/events/retrieve** Bulk Retrieve Events | ⌄ |

| GET | **/api/events/{event_id}** Retrieve Event V2 | ⌄ |

Entities endpoints ⌃

| GET | **/api/entities/search** Query Entities | ⌄ |

| POST | **/api/entities/retrieve** Bulk Retrieve Entities | ⌄ |

| GET | **/api/entities/{entity_id}** Retrieve Entity V2 | ⌄ |

| POST | **/api/media/retrieve** Bulk Retrieve Media Objects | ⌄ |

ACDH-CH DHRI-Heads Meeting, 2023-03-23
Matthias Schlögl, ACDH-CH (Austrian Academy of Sciences)

WWW.OEAW.AC.AT

ACDH-CH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

intavia

Search & Curation   Visualization & Analysis   Storytelling & Presentation   Import data

Visual Analysis, Curation & Communication
for In/Tangible European Heritage

| Search term | Search |

🔍 find, create & curate

The InTaVia knowledge graph contains data on Europe's cultural history, including data on individual artists, cultural objects, and groups or organizations. Search for these entities in our knowledge base (with a focus on Slovenia, Austria, the Netherlands and Finland) or with a global reach via data from Wikipedia. You can also upload your own data, and curate (i.e., edit, assemble, or enrich) all kinds of data for further operations of visual analysis and narration.

📊 visualize, analyze & explore

A visualization studio allows you to explore and analyze selected data on cultural actors, objects or organizations - and to further curate your data. This components offers a geo-analytical map-based view, a timeline view for chronological analysis, a set-based view for prosopographical data aspects and a network-analytical view. You can further annotate the resulting visualizations and store them for future use, or export them to a presentation and storytelling component.

💬 communicate, show & tell

The story creator module allows you to communicate your selected culture and history data by narrative means. Connect and combine rich media elements (images, text, or A/V data) with data visualizations (maps, timelines, sets, or network diagrams, created in the visualization studio) to compelling stories. These stories or presentations can be further enriched with gamification elements, and exported to a story viewer component for the reception on different target devices.

Search & Curation ›   Visualization & Analysis ›   Storytelling & Presentation ›

WWW.OEAW.AC.AT

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE
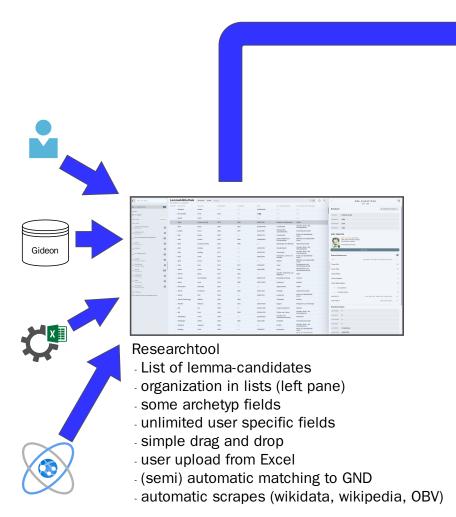
ACDH-CH D
Matthias Sch

IRS

Integriertes Redaktionssystem

- meant to replace end-of-life system Gideon
- should cover whole process of Bio creation
- built on well proven open-source technologies
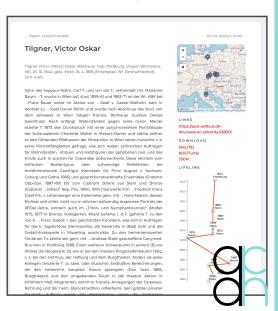- ...and obviously published under MIT license

IRS components

Workflowtool
- direct import from Researchtool
- organization in deliveries ("Lieferungen"),
- edition steps (columns) and orga (batches)
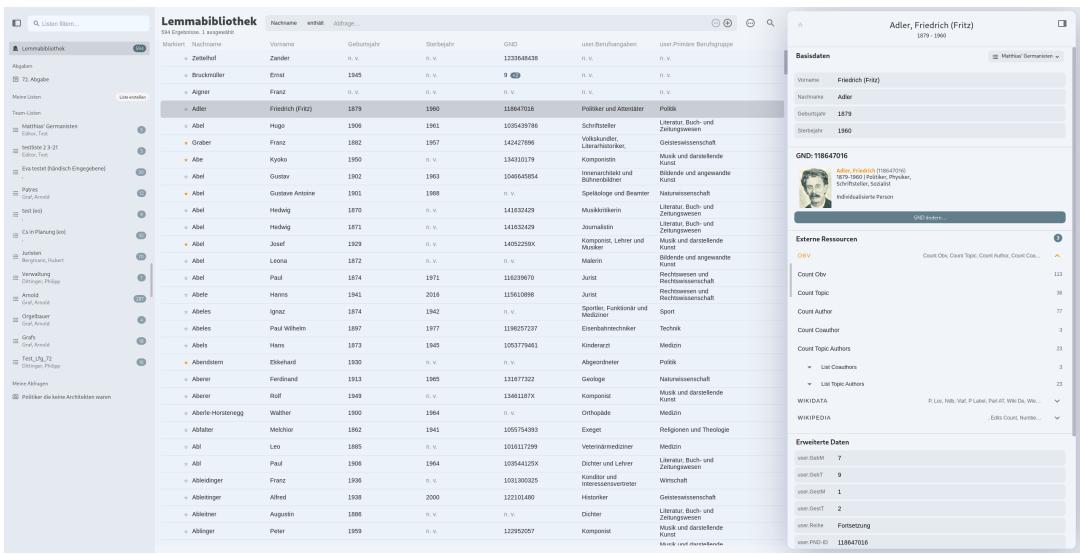- simple drag and drop
- customizeable search

Researchtool
- List of lemma-candidates
- organization in lists (left pane)
- some archetyp fields
- unlimited user specific fields
- simple drag and drop
- user upload from Excel
- (semi) automatic matching to GND
- automatic scrapes (wikidata, wikipedia, OBV)

Editor
- built on prosemirror
- comments
- annotations
- versions

Current state

WWW.OEAW.AC.AT

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

IRS Researchtool

WWW.OEAW.AC.AT

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

IRS Workflowtool

IRS Editor

ÖBL Webpage new

WWW.OEAW.AC.AT

IRS Researchtool live

ACDH-CH DHRI-Heads Meeting, 2023-03-23
Matthias Schlögl, ACDH-CH (Austrian Academy of Sciences)
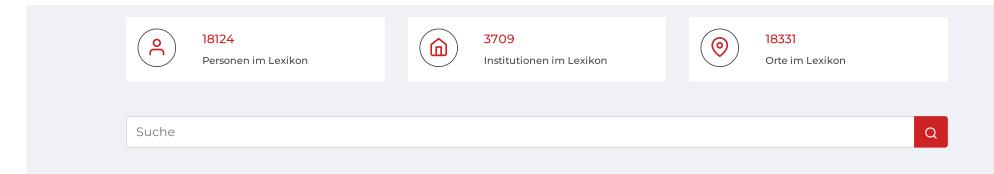
ÖBL webpage new

O^e BL

SUCHE   ÜBER DAS ÖBL   KONTAKT

Volltextsuche

Sie befinden sich auf einer Testseite des ÖBL. Die Seite befindet sich gerade im Aufbau und enthält noch nicht alle Biographien (Stand Herbst 2017). Teilweise können Informationen wie die Referenzliteratur, oder der Autor vom Original abweichen. Für Referenzzwecke wechseln sie bitte zu https://biographien.ac.at.

Österreichisches Biographisches Lexikon

Das „Österreichische Biographische Lexikon 1815–1950" erfasst bedeutende Persönlichkeiten, die im jeweiligen österreichischen Staatsverband bzw. in der österreichisch-ungarischen Monarchie geboren wurden, gelebt oder gewirkt haben und zwischen 1815 und 1950 verstorben sind.

18124
Personen im Lexikon

3709
Institutionen im Lexikon

18331
Orte im Lexikon

Suche

FEATURED

Andreas Streit
Bildende und angewandte Kunst Architekt

ÖAW AUSTRIAN ACADEMY OF SCIENCES

Thanks

ACDH-CH - AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE