

# There is no “I” in “Infrastructure”

Creating a shared data-centric DH Infrastructure  
for Cultural Heritage Research in Saxony/Germany

DH2023, 14.07.2023, Graz

Dirk Goldhahn (goldhahn@saw-leipzig.de)  
Peter Mühleder (muehleder@saw-leipzig.de)  
Franziska Naether (naether@saw-leipzig.de)

Saxon Academy of Sciences and Humanities, Leipzig, Germany

# Institutional Context

## Saxon research center and network for digital humanities and cultural heritage

- Network of non-university research institutions
  - Smaller humanities research institutions
  - Diverse research areas (history, social- and political sciences, linguistics)



- 'DH-lab' at the Saxon Academy of Sciences and Humanities
  - Three permanent full time positions:
    - Project development, coordination, research
    - Research data management & digital research methods
    - Research software development ***\*that's me\****

# DIKUSA – Digital Cultural Data in Saxony

State funded, 2022-2025; 6 smaller scale DH projects (1 researcher per institution) and the DH-lab developing a shared data infrastructure

- **Possibilities and Limits of Jewish Participation at Saxon Universities: Graduates from Chemnitz, Dresden, Freiberg, and Mittweida (1850-1950)**  
Leibniz Institute for Jewish History and Culture – Simon Dubnow
- **Migration of artists to and from Saxony in the 17th century**  
Leibniz Institute for the History and Culture of Eastern Europe
- **Women migrated from, to and within Germany - building an experience-historical knowledge base**  
Hannah Arendt Institute for Totalitarian Studies
- **Expansion of the historical gazetteer (HOV) to the central gateway for standardized location data in Saxony**  
Institute for Saxon History and Cultural Anthropology
- **Landscape as Cultural Heritage. Transformation of a Mining Landscape in Saxony in the 20th Century**  
Saxon Academy of Sciences and Humanities
- **Sorbian cultural monuments in Saxony. Development of a virtual working environment for their digital recording and presentation**  
Sorbian Institute

2023; one adjacent project:

- **Ecclesiastical practice in the GDR - Implementation of a digital research environment**  
Saxon Academy of Sciences and Humanities

# DIKUSA – Digital Cultural Data in Saxony

People and their biographies  
Artworks and their provenance  
Cultural monuments  
Places / landscapes

Social networks  
Mapping  
Timelines  
Transformation processes

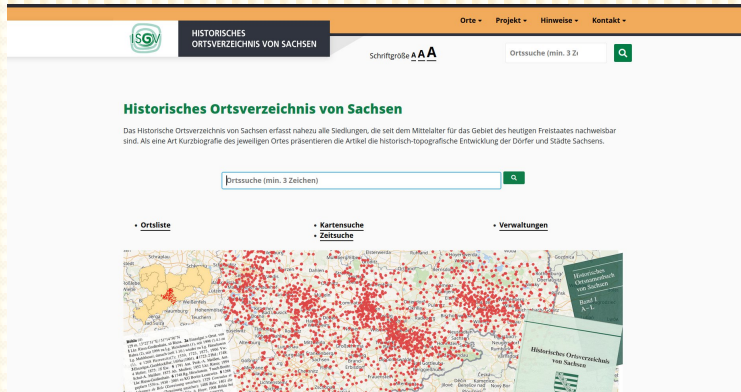
# DIKUSA – Data-centric infrastructure

**Main goal:** support the generation of robust, interoperable, reusable historical research data sets

- Focus on structured data (biographic records, bibliographic information, historic networks, history/provenance of cultural artifacts, geographic information)
- **NOT:** Text, digital editions, linguistic corpora; media, or fancy ML stuff

# DIKUSA – Data-centric infrastructure

- **Secondary goal:** improve the quality and (re)usability of existing local data resources



HOV - Historical gazetteer of Saxony  
*ISGV*



Bellum et Artes. War, Art and Diplomacy in Central Europe During the Thirty Years' War. (*Exhibition and database*)  
*GWZO*



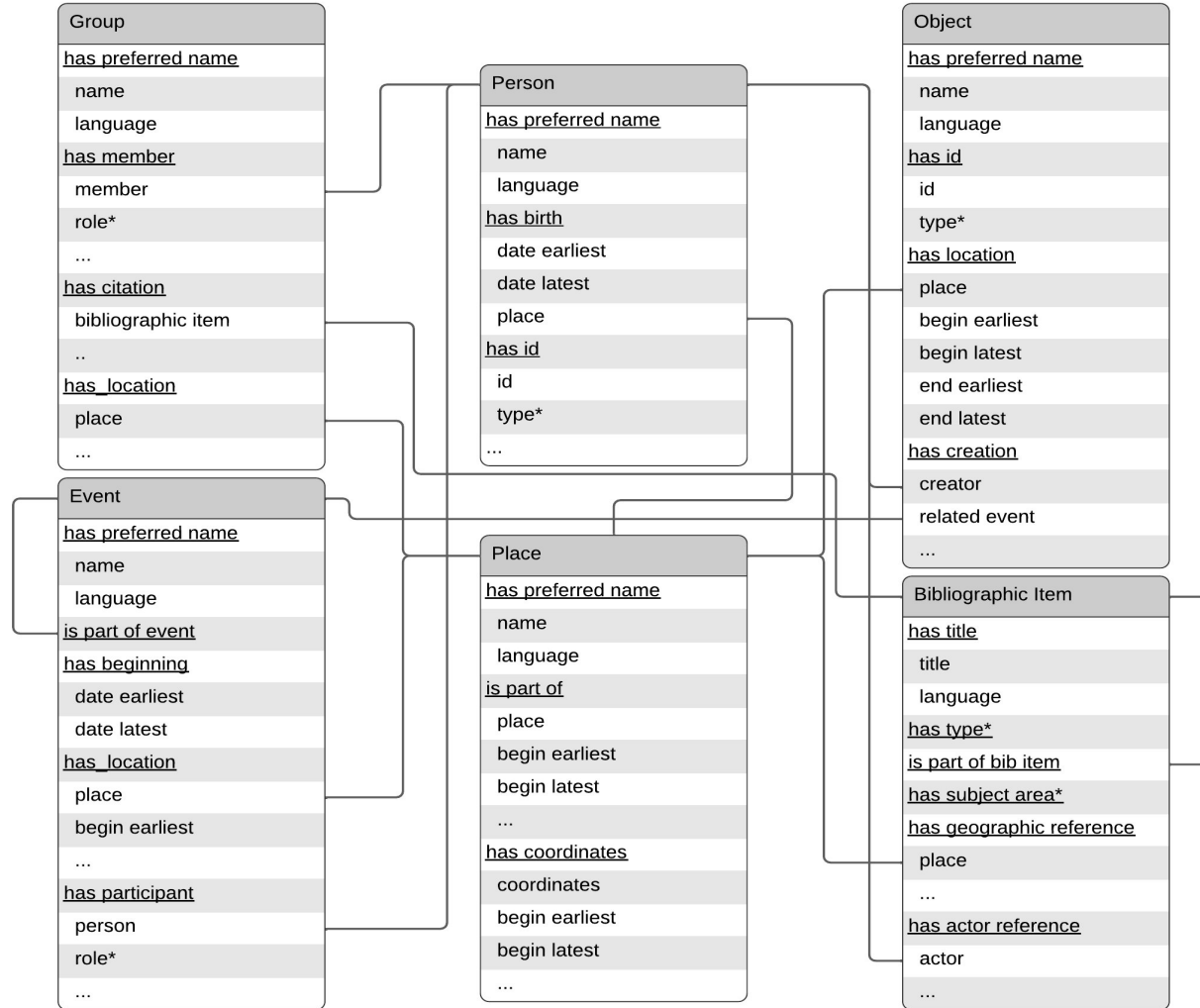
VKF - Collection of historical, geo-referenced maps  
*SLUB Dresden*

# DIKUSA – Data-centric infrastructure

- Building the fundamentals
  - Creating a shared data model for general entities such as persons, institutions, artifacts, locations etc.
  - Developing a service for data reconciliation
  - Providing a flexible tool for data curation



# Shared data model



## Interoperability through

- Linking to authority data
- mapping to CIDOC-CRM
- Use of RDF(\*) as exchange format to include provenance information

+ Creation of common SKOS vocabularies



# Data Reconciliation

## Dikúdex – A modular reconciliation REST api

- Provides standardized access/mappings to authority data for persons, places and organizations (e.g. GND, Geonames, Wikidata)
- Standardized properties
- Results are clustered based on identifiers
- Plugins for local data resources:
  - HOV – Historical gazetteer of Saxony
  - (Bellum & Artes database )

```
▼ defaultTypes:
  0: "person"
  1: "place"
  2: "organization"
▼ extendProperties:
  ▼ organization:
    0: "description"
    1: "alternative_names"
    2: "location"
    3: "coordinates"
    4: "founding_date"
    5: "termination_date"
    6: "predecessor"
    7: "successor"
    8: "thumbnail"
  ▼ person:
    0: "description"
    1: "alternative_names"
    2: "birth_date"
    3: "birth_place"
    4: "death_date"
    5: "death_place"
    6: "occupation"
    7: "place_of_activity"
    8: "thumbnail"
  ▼ place:
    0: "description"
    1: "alternative_names"
    2: "coordinates"
    3: "area_code"
    4: "country"
    5: "state"
    6: "population"
    7: "inception"
    8: "abolished"
    9: "thumbnail"
```

search query

zetkin

+ Create new item without reconciliation

Dikudex reconciliation candidates

Zetkan en

Geonames: 3186556

Zettin en

Geonames: 11862777

Zetkin, Konstantin de Kostja Zetkin en

GND: 125000243 Wikidata: Q1785273

Zetkin, Ossip de Ossip Zetkin en

GND: 128348143 Wikidata: Q2034705

Zetkin, Maxim de Maxim Zetkin en

GND: 125000235 Wikidata: Q1913855

Zetkin, Clara de Clara Zetkin en

GND: 118636618 Wikidata: Q75823

Zayatkon en

Geonames: 1284124

Zetkin-Milovidova, Emilia de

GND: 1067168672

☒ Zetkin, Clara (de)

☒ Clara Zetkin (en)

Description GND

Dt. Politikerin und Feministin, Mitglied des Reichstages Juni 1920-30.03.1933 -

Description GND

Gründete 1891 die sozialdemokrat. Frauenzeitschr. "Die Gleichheit", 1920-1933 KPD-Reichstagsmitglied, 1932 Alterspräs. des Dt. Reichstags -

Description Wikidata

Communist activist, and advocate for women's rights -

☐ GND: 118636618

☐ Wikidata: Q75823

alternative\_names ☺

birth\_date ☺

☐ GND: 1857-07-05

☐ Wikidata: 1857-07-05

☐ Wikidata: 1857-06-05

+ Create item and import data

candidate details

clustered reconciliation candidates

# Data set curation

## **‘Weedata’** - A flexible knowledge base for smaller DH projects

- Light-weight web application (Sveltekit frontend + Django backend)
- Easy to deploy, configure and maintain
- Focus on data-sharing/accessibility and compatibility with other research tools (Zotero, Visualisation tools like gephi)
- Flexible, wikidata-inspired general data model

Data item

Statement (subject - predicate - object)

Qualifier statements (statement about statements)

Data item, e.g. a person

weedataItemsVocabulariesMediaData modelAdministration

Q3

Dadler, Sebastian

Person

de

Dadler, Sebastian

★

de

Sebastian Dadler

★

de

Sebastian Dattler

☆

en

Sebastian Dadler

☆

Add image

Identifiers

wikidata id

Q1286570

+

Statements

Expand qualifier

Collapse qualifier

birth

P65

+

date

1586-3-6

location

Straßburg

Qualifier

+

⌵

⌵

death

P57

+

location

Hamburg

date

1657-7-6

Qualifier

+

⌵

⌵

profession

P59

+

Artist

Qualifier

+

⌵

⌵

Medailleur

Qualifier

+

⌵

⌵

short description

P47

+

commentary

P48

+

multilingual labels

linked identifiers

statement metadata (provenance, confidence, comments)

edit history with undo

item statements and qualifier

weedataItemsVocabulariesMediaData modelAdministration

searchHelp

Q1 Klimt, Gustav Person

de Klimt, Gustav

en Gustav Klimt

Add image

Identifiers

gnd id 118563289 wikidata id Q34661

Statements

bith date P13

1862-7-14

Qualifier

birth place P15

Statement history

#132 update

Statement bith date 1860-7-14 → 1862-7-14

2023-07-13 15:06 - admin undo

#131 update

Statement bith date 1862-07-14 → 1860-7-14

2023-07-13 15:06 - admin undo

#6 create

Statement bith date 1862-07-14

2023-07-06 06:20 - admin

1

statement level  
change history

Items Vocabularies Media Data model Administration search Help admin

Verzeichnis ⓘ

Packmohr, Horst Person ⇄

Packmohr, Horst ★

Add image

fiers

d 176895529 / +

ments + Expand qualifier Collapse qualifier

name P19 +

de Horst Qualifier +

name P18 +

de Packmohr Qualifier +

Reconcil

**Provenance statements**

Items +

External urls +

<https://wiki.de.dariah.eu/KPIDD/> 🗑

Zotero references +

Dietrich, Christian, and Uwe Schwabe, eds.  
Freunde Und Feinde. Friedensgebiete in Leipzig  
Zwischen 1981 Und Dem 9. Oktober 1989.  
Dokumentation. Leipzig, 1994. 🗑

Free Text +

Who is Who wiki import 🗑

provenance statements on item  
or statement level

- Weedata items
- external URLs
- Zotero libraries
- free text comments

# Simple data modelling

## Entity classes

Type	Id	Label
person	C15	Person
place	C16	Place
	C17	Collection
	C18	Object
organization	C21	Institution

## Properties

Type	Id	Label	Scope
item	P25	artist	Person
blank	P20	birth	
item	P32	collection holder	
text	P36	commentary	
item	P26	commissioner	
blank	P27	creation	
date	P37	date	
blank	P57	death	
date	P34	dissolution	
date	P33	first mention	
identifier	P61	gnd id	
text	P24	historic name	
identifier	P16	hov id	
item	P63	location	Place

(OWL) RDF export / import



# Manage controlled vocabularies

weedataItemsVocabulariesMediaData modelAdministration

V19

Profession

enProfession★

✓

Concepts ⊕

Architekt✓

Artist✓

Bernsteinschnitzer✓

Bildhauer✓

Bühnenmaler✓

Dirigent✓

Drechsler✓

Elfenbeinschnitzer✓

Gartenarchitekt✓

Glasmaler✓

Glasschneider✓

Goldschmied✓

Graveur✓

Holzbildhauer✓

Holzschnitzer✓

Kapellmeister✓

Komponist✓

Kunstdrechsler✓

Kunsthändler✓

Elfenbeinschnitzer Q39

deElfenbeinschnitzer

deElfenbeindrechsler

enIvory carver

Descriptions ⊕

SKOS (RDF) Import / Export

<https://www.example.me/weedata/items/3>

Browser

Application

*HTML representation / web-app*

*JSON object*

The screenshot shows a web application interface for a person's profile. The top navigation bar includes tabs: weedata, Items, Vocabularies, Media, Data model, and Administration. The main content area is titled "Q3 Dadler, Sebastian" with a "Person" tag. Below the title is a table of labels in German and English. Further down, there are sections for "Identifiers" (showing a Wikidata ID), "Statements" (with expand/collapse buttons), and specific data fields for birth, death, profession, short description, and commentary. Each statement field includes a qualifier and a value.

Label	Value
de Dadler, Sebastian	★
de Sebastian Dadler	★
de Sebastian Dattler	☆
en Sebastian Dadler	☆

Identifiers

wikidata id **Q1286570**

Statements

birth **P65**

date 1586-3-6 location Straßburg

death **P57**

location Hamburg date 1657-7-6

profession **P59**

Artist

Medailleur

short description **P47**

commentary **P48**

The screenshot shows a JSON object representing the data from the web application. The object is a dictionary with various keys and values, including labels, identifiers, statements, and metadata. The JSON is displayed in a dark-themed editor with syntax highlighting.

```
{
  "id": 3,
  "labels": {
    "0": {
      "id": 93,
      "label": "Dadler, Sebastian",
      "lang": "de",
      "is_pref_label": true
    },
    "1": {
      "id": 94,
      "label": "Sebastian Dadler",
      "lang": "de",
      "is_pref_label": true
    },
    "2": {
      "id": 95,
      "label": "Sebastian Dattler",
      "lang": "de",
      "is_pref_label": false
    },
    "3": {
      "id": 98,
      "label": "Sebastian Dadler",
      "lang": "en",
      "is_pref_label": false
    }
  },
  "entity_class": 15,
  "free_text": "",
  "statements": {
    "0": {
      "id": 46,
      "subject_item": 3,
      "subject_statement": null,
      "predicate": {
        "id": 57,
        "datatype": "blank",
        "created": "2023-07-03T08:20:23.176344Z",
        "labels": {
          "0": {
            "id": 88,
            "label": "death",
            "lang": "en",
            "is_pref_label": true
          }
        },
        "descriptions": [],
        "is_bidirectional": false,
        "is_system": false,
        "export": false,
        "identifier_url": null,
        "is_meta": false,
        "unit": null,
        "meta_property": [],
        "filter_classes": []
      },
      "default_qualifier": {
        "0": {
          "id": 17
        }
      }
    }
  }
}
```

# More stuff

- Display map tiles (e.g. historical maps) for editing geoinformation
- Set tasks for specific users
- Add and manage media files (images and documents)

# Challenges and Takeaways

- DH in a challenging environment - many (small) projects – many databases – minimal resources (hardware and people) and many ‘legacy projects’
- Our goal is to creating a data centric infrastructure that helps with data quality, curation processes and accessibility
- ‘Pick your battles’ -> core ontology, reconciliation API, data curation tool as a flexible foundation for current and future projects
- Project documentation is important (people leave after projects end – retain as much knowledge as possible)
- Only possible because of DH Lab with permanent positions – DH Lab as social/knowledge infrastructure, internal documentation is key!

# Thank you

[kompetenzwerkd@saw-leipzig.de](mailto:kompetenzwerkd@saw-leipzig.de)

<https://www.github.com/kompetenzwerkd>