

DARIAH-DE – a service and tool oriented federation architecture for distributed data and collections

Stefan Schmunk

State and University Library Göttingen, Germany

Denver, 13th September 2016

 @DARIAHde

- Introduction DARIAH-DE
- Research Data in the Arts and Humanities
- Research Data Federation Architecture
- Conclusion



Arts and Humanities
& Social Science

**Digital Humanities =
Applied digital enabled research
in the Arts and Humanities**



Computer Science



Libraries & Archives – Information Science

- one of 17 members of DARIAH-EU
- funded by the German Federal Ministry of Education and Research (BMBF)
- Preliminary phase: 2008 – 2011
- Development phase: 2011 – 2016
- Operating phase: 2016 – 2019
- 18 partners:
 - 5 universities, 5 academic institutions, 2 computer centres, 2 libraries, 2 academies of sciences, 1 commercial partner, 1 NGO
- Used by 80 projects and more than 3.000 scholars

Teaching (support, training)

- Workshops on methods, expert colloquia, Summer Schools
- Coordination of national and international curricular developments

Research Data (research data collections, research data management)

- Best practices for metadata, standardised exchange of data, ontologies
- Development of generic search, collection and schema registry

Research

- DH methods and practices
- Use Cases: Annotation, Big Data
- Tools and services
- Bibliography *Doing Digital Humanities*

Technical Infrastructure (development and provision of infrastructural services)

- collaborative research environments, virtual machines, monitoring, authentication and authorization infrastructure etc.

Definition Research Data

By research data in the Arts and Humanities is meant all those sources and outputs that are **based on a research question**, collected either by scholars, libraries, archives or any other institution, which may be described, evaluated and/or generated and stored in **machine-readable form** for **analyzing, archiving** purposes, **citability**, and for further **processing**.

- Metadata, bibliographical data, finding aids
- Digital and/or digitized data and/or digital representation of analog data
- Digital objects
- Full text, transcripts
- Enriched full text
- Images, movies, music & notes
- Authority files, controlled vocabularies, ontologies
- Etc.



Research Data Collection – epidat

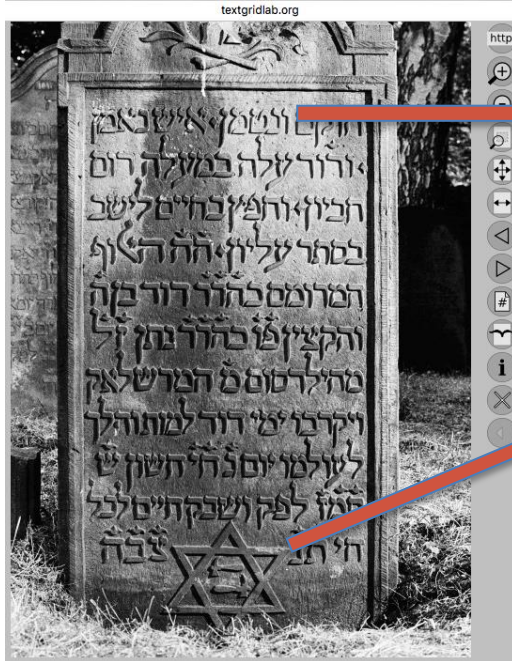
- **Data:** 30.000 inscriptions of more than 130 Jewish cemeteries in Germany and the Netherlands (1050-2000)
- **Contains:** Digitized Data, transcripts, maps, photographs, and archival sources
- **Methods:** Visual Analytics, Named-Entity-Recognition, Temporal-Spatial Visualization

*File and encoding format:

- data format
 - images: jpg png gif tiff eps
 - maps: svg pdf AutoCad [kml](#)
 - text
 - extern: x(html)
 - extern: TEI4 XML (using: [epidoc](#): Epigraphic Documents in TEI XML)
 - extern (wip): [TEI5 XML](#) (using: [example](#))
 - extern: plain/text (no format)
 - extern (on demand): PDF
 - intern: TUSTEP file with epidat structure
- metadata format
 - [TEI header](#)
 - DC (simple, [example](#))



Source: <http://www.steinheim-institut.de/cgi-bin/epidat?info=edv&art=art6>



Symbols

Photography

Burial site

Further information about Epidat:
Thomas Kollatz
kol@steinheim-institut.org

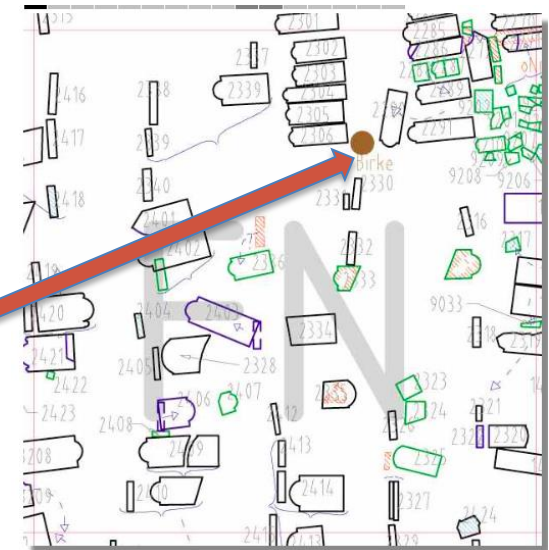
Name	Seitenanfang
David (genannt Hammerschlag) ben Natan [26.10.1686]	
Edition und Übersetzung	Text/Zeilensicht Textansicht Seitenanfang
ה[ס] Hier הוקם ונבנו : איש נאמן : ודוד עלה במעלה רום חביון : וחפץ בחיים לישב בסתר עליון : היה האלוף המרום כהר"ר דוד בן ה' והקצין פיו כהר"ר נתן דל' מהילדסום מ' המרשלאק ויקברו ימי דוד למות והלך לעולמו יום ג' ח"י חשוון ש' תמ"ז לפק ושבע חיים לכל חי הנצח"ה	errichtet (ein Mal) und ward geborgen ein getreuer Mann, >und David stieg hinauf: hoch ins Versteck, >der, der Leben begehrt, >wird sitzen im Schutze des Höchsten, es ist der Vornehme, der Erhabene, der geehrte Meister, Herr David, Sohn des Vornehmen und Einflußreichen, Vorstehers und Leiters, des geehrten Meisters, Herrn Natan, sein Andenken zum Segen, aus Hildesheim, genannt Hammerschlag. >Und es nahten Davids Tage dem Tod: und >er ging hin in seine Welt (am) Tag 3, 18. Cheschan des Jahres 447 der kleinen Zählung, >und er ließ das Leben wie alles Lebende. Seine Seele sei eingebunden in das Bündel des Lebens

Zl 3: 2Sam 15,30 Zl 4: Ps 34,13 Zl 4f: Ps 91,1 Zl 9: 1Kön 2,1 Zl 9f: Koh 12,5 Zl 11f: nach bBer 61b

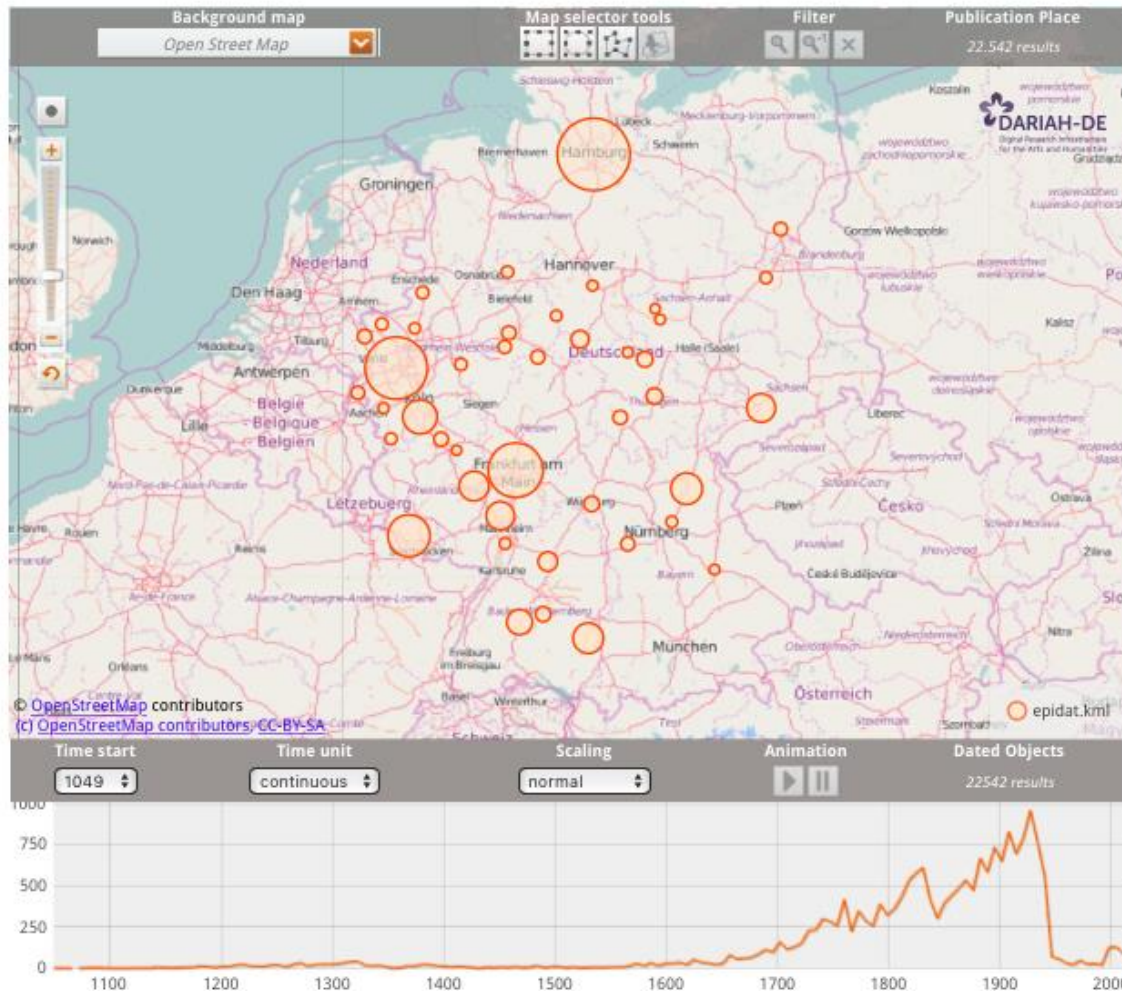
Scholarly Edition – synoptical view inscriptions

belegt Planquadrat FN

	P	O	N	M	L	K	I	H	G	F	E	D	C	B	A
A				NA	MA	LA									
B				NB	MB	LB	KB	IB							
C				NC	MC	LC	KC	IC	HC	GC					
D				OD	MD	LD	KD	ID	HD	GD	FD	ED	DD		
E				PE	OE	NE	LE	KE	IE	HE	GE	FE	EE	DE	
F				PF	OF	NF	MF	LF	KF	IF	HF	GF	FF	EF	DF
G				PG	OG	NG	MG	LG	KG	IG	HG	GG	FG	EG	DG
H				PH	OH	NH	MH	LH			HH	GH	FH	EH	DH
I											II	GI	FI	EI	DI
K											HK	GK	FK	EK	DK
L											HL	GL	FL	EL	DL
M											HM	GM	FM	EM	DM
N											GN	FN	EN	DN	CN
O											GO	FO	EO		
P											GP	FP	EP		
Q											QP	FP	EQ		
R											GR	FR			



Das DARIAH-DE spatio-temporal interface visualisiert die zeitlich-räumliche Belegung jüdischer Friedhöfe auf Grundlage der datierten epidat Inschriften.



Visualisation of the data collection (epidat)

DARIAH-DE Geo-Browser
geobrowser.de.dariah.eu

Temporal-Spatial-Contextualisation

Source: <http://www.steinheim-institut.de/cgi-bin/epidat?info=e4d&lang=de>

What is the situation for the Digital Humanities?

Research data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship”

C.L. Borgman (2015). Big Data, Little Data, No Data: Scholarship in the Networked World.

New challenges in the Humanities dealing with
Research data concerning:

- Discovering data
- Access to data
- Extended search strategies for data
- Using data tools (and generating data)
- Linking articles with data

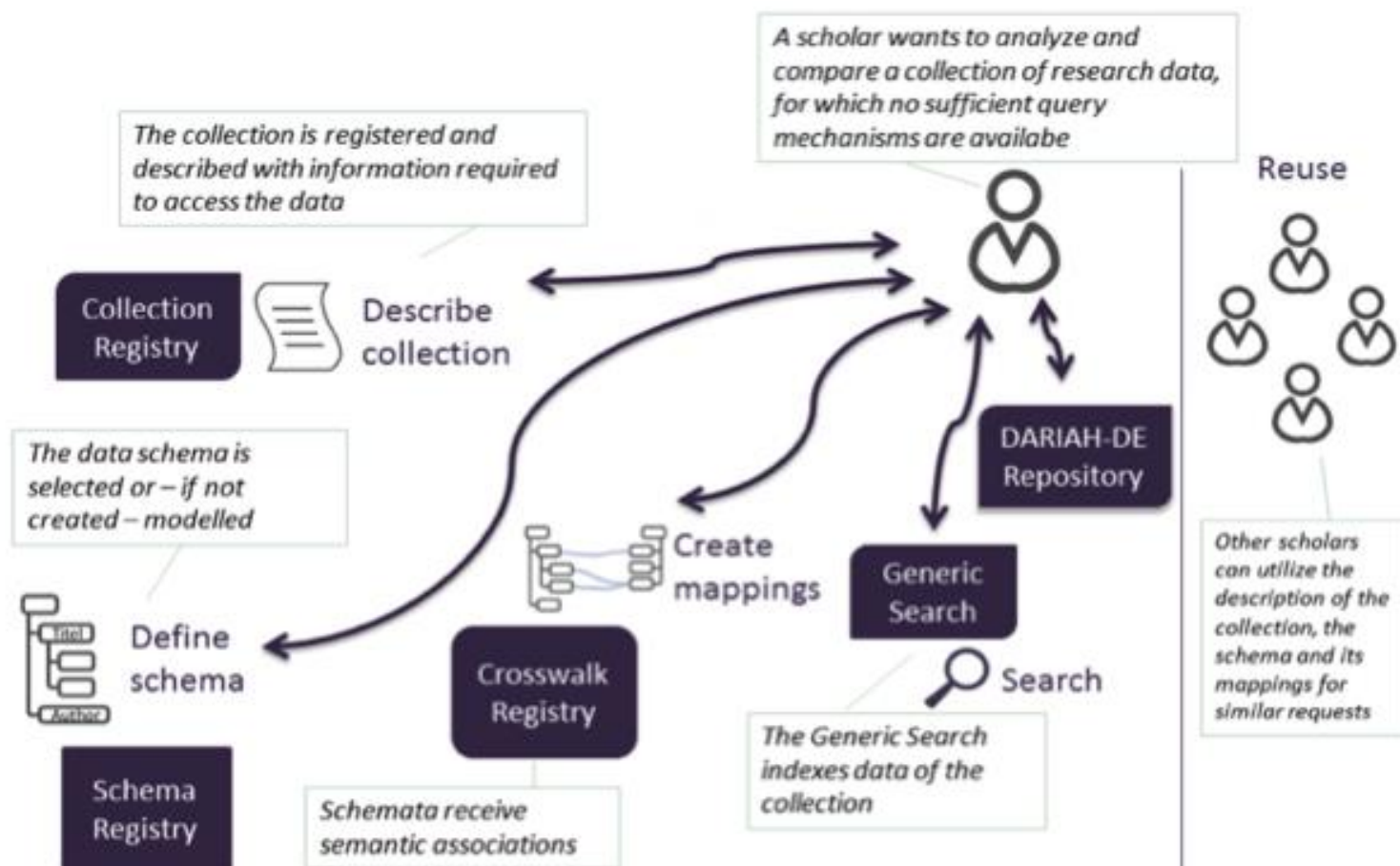


Tools needed for ...

- ... (Generic) search requests
- ... Defining metadata formats
- ... Generating crosswalks
- ... Describing research data and collections
- ... Storing data

Tool-oriented approach ...

Building a Data Federation Architecture



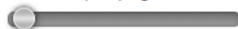
Source: <https://de.dariah.eu/data-federation-architecture>



Search options

☐ Show explanations

20 results per page



Source selection

All collections

Customize...

Select all

[Login to see your user collections](#)

138 collections selected:

Fontane Kasten C7

urglarglthenextcollection

publiziert

Weitere Fensterbank-Blumen aus dem

Entwicklerbüro: Ein Fischcurry

Fontane Notizbücher

[Search](#) / [Simple Search](#)

Simple Search

Extended Search

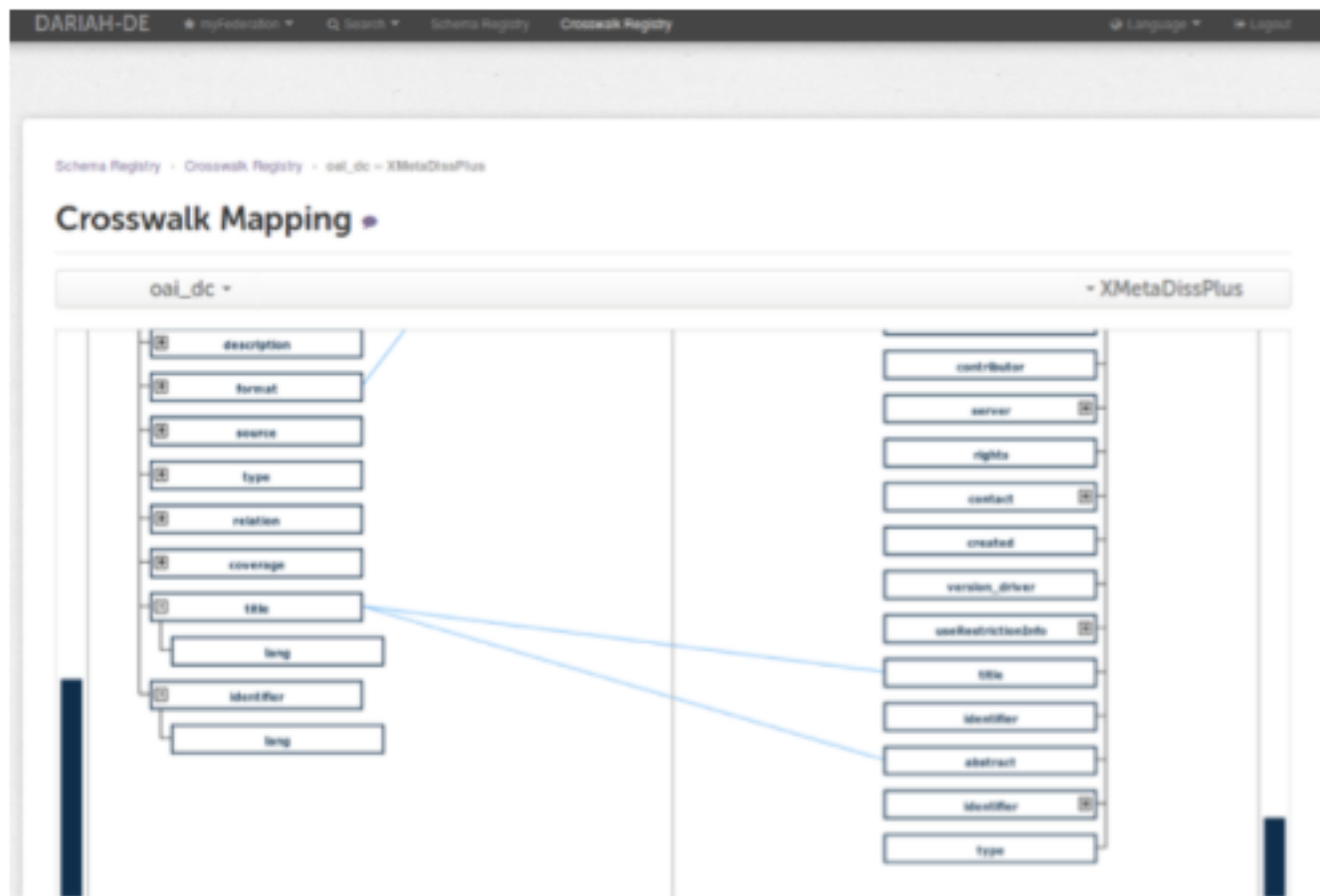
repository



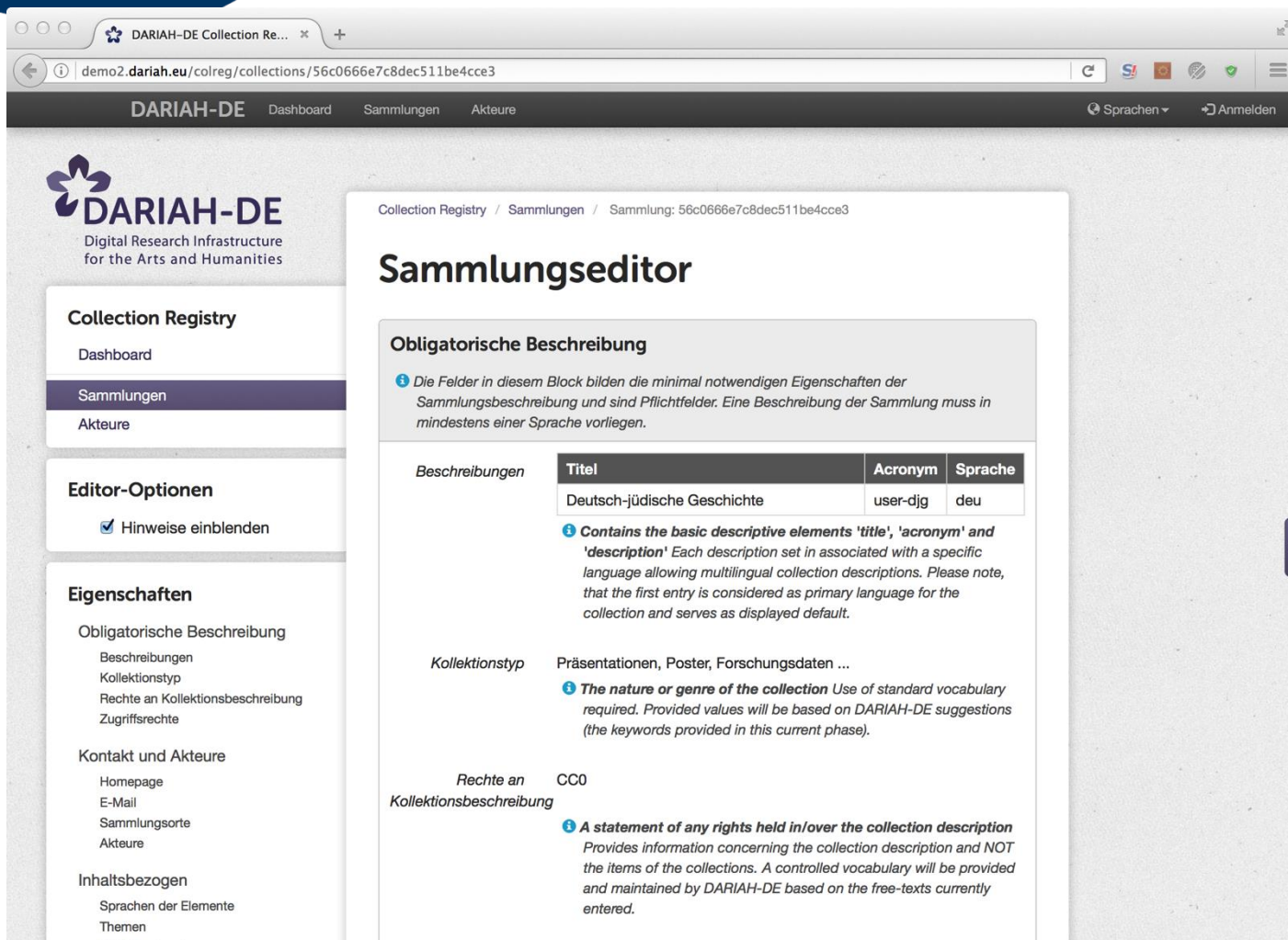
~ No results found for the provided query



Source: <http://dev3.dariah.eu/search/>



Source: <https://de.dariah.eu/schema-registry>



The screenshot shows the DARIAH-DE Collection Registry interface. The top navigation bar includes 'DARIAH-DE', 'Dashboard', 'Sammlungen', and 'Akteure'. The left sidebar contains a 'Collection Registry' menu with 'Dashboard', 'Sammlungen' (selected), and 'Akteure'. Below this is an 'Editor-Optionen' section with a checked option 'Hinweise einblenden'. The main content area is titled 'Sammlungseditor' and shows the 'Obligatorische Beschreibung' (Mandatory Description) section. This section includes a table for 'Beschreibungen' (Descriptions) with columns 'Titel', 'Acronym', and 'Sprache'. The table contains one entry: 'Deutsch-jüdische Geschichte', 'user-djg', and 'deu'. Below the table, there are three informational blocks: 'Kollektionstyp' (Collection Type) with a note about using standard vocabulary, 'Rechte an Kollektionsbeschreibung' (Rights in Collection Description) with a note about providing a statement of rights, and 'Eigenschaften' (Properties) with a note about the minimal required fields.

Collection Registry / Sammlungen / Sammlung: 56c0666e7c8dec511be4cce3

Sammlungseditor

Obligatorische Beschreibung

Die Felder in diesem Block bilden die minimal notwendigen Eigenschaften der Sammlungsbeschreibung und sind Pflichtfelder. Eine Beschreibung der Sammlung muss in mindestens einer Sprache vorliegen.

Beschreibungen	Titel	Acronym	Sprache
	Deutsch-jüdische Geschichte	user-djg	deu

Contains the basic descriptive elements 'title', 'acronym' and 'description' Each description set is associated with a specific language allowing multilingual collection descriptions. Please note, that the first entry is considered as primary language for the collection and serves as displayed default.

Kollektionstyp Präsentationen, Poster, Forschungsdaten ...

The nature or genre of the collection Use of standard vocabulary required. Provided values will be based on DARIAH-DE suggestions (the keywords provided in this current phase).

Rechte an Kollektionsbeschreibung CC0

A statement of any rights held in/over the collection description Provides information concerning the collection description and NOT the items of the collections. A controlled vocabulary will be provided and maintained by DARIAH-DE based on the free-texts currently entered.

Eigenschaften

Obligatorische Beschreibung

- Beschreibungen
- Kollektionstyp
- Rechte an Kollektionsbeschreibung
- Zugriffsrechte

Kontakt und Akteure

- Homepage
- E-Mail
- Sammlungsorte
- Akteure

Inhaltsbezogen

- Sprachen der Elemente
- Themen

Source: <https://de.dariah.eu/schema-registry>

~ DARIAH-DE
DARIAH-EU
DHd-Blog
DHd-Kanal
Twitter
TextGrid
Sign Out

DARIAH-DE Publikator

On this page you can edit your chosen collection. Please **add files** from your disc drive and add metadata to your collection and files. As metadata schema Dublin Core will be used, to provide a minimal standard and a certain flexibility for different research data formats. For that reason only few of the metadata fields are mandatory. Please note: the more metadata you provide with your collection and files, the bigger the chance that your data (and of course metadata) is found and re-used by other researchers and can be put into the research data life cycle again.

Edit Collection

← back to main view

– Manage Content and Edit Metadata

add file(s)
show optional metadata



view raw collection data

* Title ⓘ ⬆

Please provide mandatory metadata

* Creator ⓘ ⬆

Please provide mandatory metadata

* Rights ⓘ ⬆

Please provide mandatory metadata

Alle Inhalte dieser Seite unterliegen sofern nicht anders gekennzeichnet der CC-BY 4.0 Lizenz

Impressum Kontakt

Source: <https://devportal.de.dariah.eu/web/dariah-imp1/publish>

Aims

- Indexing research data / collections
- Mapping different metadata schemas
- Citation / long term preservation of research data
- Research(er) oriented view
- Comparison of collections and content
- Comprehensive view on heterogeneous collections

DARIAH-DE Repository Publish GUI

<https://de.dariah.eu/repository>

Collection Registry

<http://demo2.dariah.eu/colreg/>

Generic Search

<http://dev3.dariah.eu/search/>

TextGrid Repository

<https://textgridrep.org/en/>

Thank you very much
for your attention!

 @DARIAHde

Dr. Stefan Schmunk

schmunk@sub.uni-goettingen.de