



Anwendung der Föderationsarchitektur und generischen Suche

*Workshop 2: Daten sammeln, modellieren
und durchsuchen mit DARIAH-DE*

DHd 2017
Bern, 13.02.2017

1. Überblick und Hintergrund

2. Datenmodellierung

- *Hochladen von Schemata*
- *Explikation Hintergrundwissen*
- *Verfeinerung durch Regeln*

Live Session 1

3. Schema-Mappings

- *Wertkorrespondenzen*
- *Konzeptassoziationen*
- *Definition und Anwendung von Transformationsfunktionen*

Live Session 2

4. [Generische Suche]

<https://www.uni-bamberg.de/minf/forschung/dariah-de/>

Fakultät Wirtschaftsinformatik und Angewandte Informatik
Lehrstuhl für Medieninformatik



[Fakultäten](#) > [Wirtschaftsinformatik und Angewandte Informatik](#) > [Fächer und Lehrstühle](#) > [Angewandte Informatik](#) > [Lehrstuhl für Medieninformatik](#)
> [Forschung](#) > **DARIAH-DE**

- Studium
- Forschung
- Abgeschlossene Projekte
- DARIAH-DE**
- Methodenworkshop 2015
- P2P Information Retrieval
- Geographisches Information Retrieval
- Firmensuche
- Promotionen
- Transfer
- Veranstaltungen

DARIAH-DE: Digital Research Infrastructure for the Arts and Humanities

DARIAH ist eine Forschungsinitiative unter dem Dach des European Strategy Forum for Research Infrastructures (ESFRI) und wird in Deutschland als **nationales Teilprojekt "DARIAH-DE"** vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.



Für das nationale Verbundprojekt wurden wesentliche Zielsetzungen in Einklang mit der Mission von DARIAH-EU definiert: Neben der erweiterten **Anwendung digitaler Methoden** in den **Kultur- und Geisteswissenschaften** steht insbesondere auch der Aufbau einer nachhaltigen **Föderationsinfrastruktur** für den interdisziplinären Zugriff auf Forschungsdaten im Fokus der Initiative.

13.02.2017 - Workshop "Daten sammeln, modellieren und durchsuchen mit DARIAH-DE" im Rahmen der DHD 2017, Bern

Testaccounts (verwendete Accounts bitte entsprechend markieren)

https://docs.google.com/document/d/14Sql1_IG-ZM6chQwNfg7Xx8tS2N1ZFvQC-7OFTtdXM/edit?usp=sharing

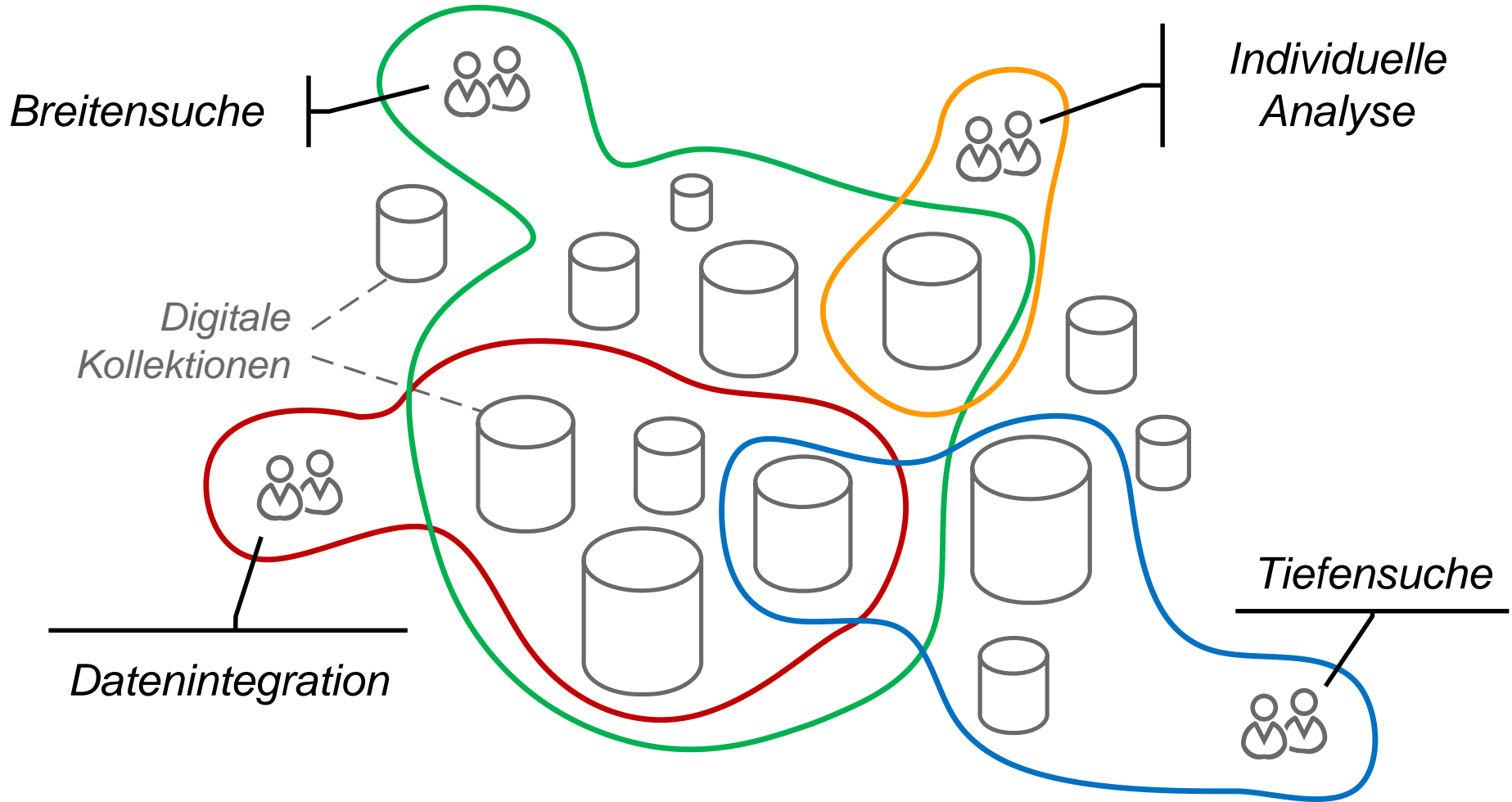


Anwendung der Föderationsarchitektur und generischen Suche

ÜBERBLICK UND HINTERGRUND

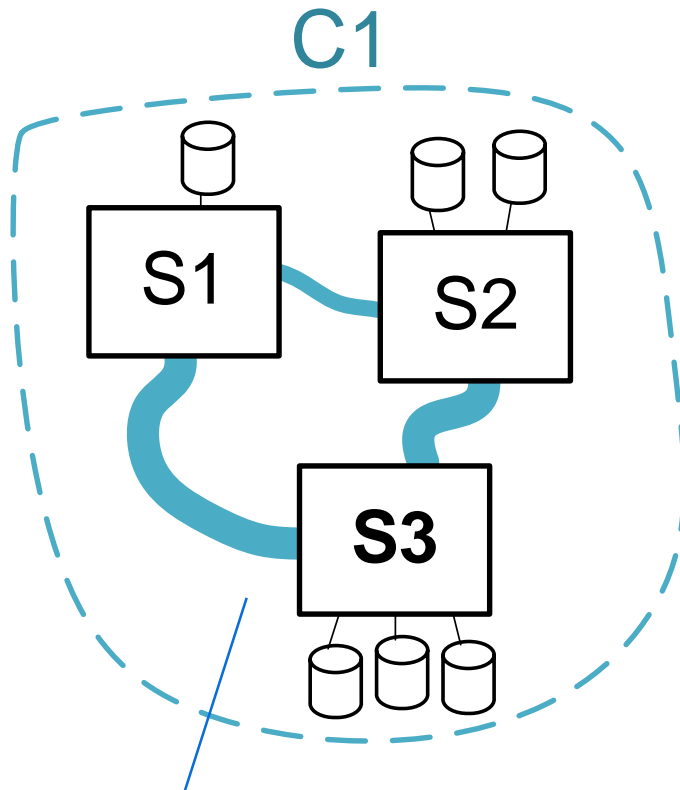
- *Datenintegration* versucht die Harmonisierung heterogener Darstellungsformen
 - *eine globale Sicht* für integrative Betrachtung
 - oft als einmaliger Aufwand realisiert z. B. ETL
 - Spezifität von Daten irrelevant, wenn nicht in Integrationssicht assoziierbar

Lenzerini, M (2002): Data Integration: A Theoretical Perspective. In: Abiteboul, S (Ed.): Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, p. 233. ACM, 2002. ISBN:9781581135077.



- *Datenföderation* beschreibt eine losere Zusammenführung von Daten
 - Daten bleiben im ursprünglichen Zustand erhalten
 - Aufwand je notwendiger Sicht erforderlich Standards
 - globale Sichten auch hier als einmaliger Aufwand
 - aber: mehrere „globale“ Sichten denkbar domänenspezifisch
 - Integrationsansätze nach spezifischen Bedürfnissen

Gradl, T.; Henrich, A. (2014): A novel approach for a reusable federation of research data within the arts and humanities, in: Digital Humanities 2014: Book of Abstracts. Lausanne: Ecole polytechnique fédérale de Lausanne; Université de Lausanne 2014, S. 382–384.

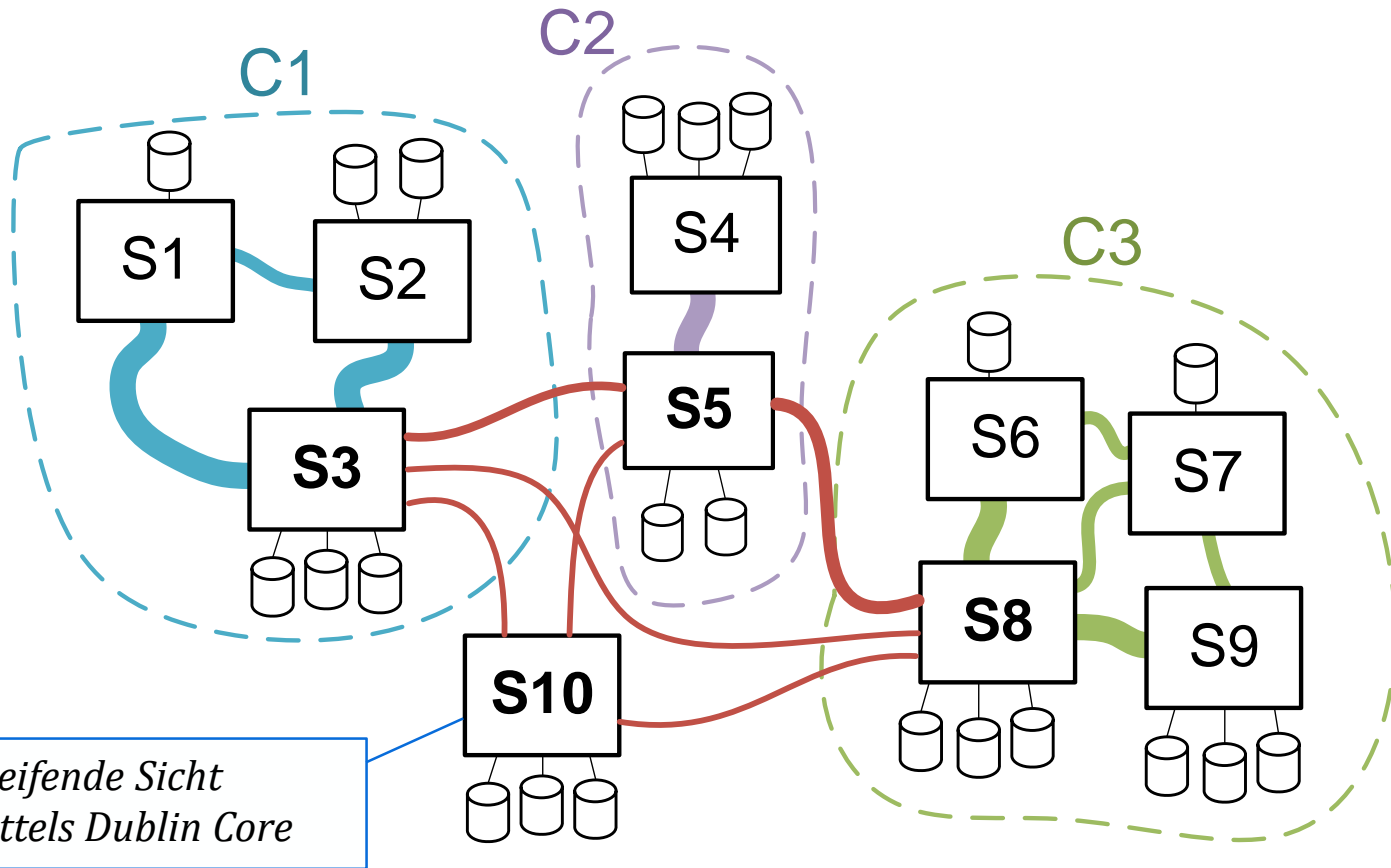


*Beispiel: drei
Repräsentationsformen
biographischer Daten*

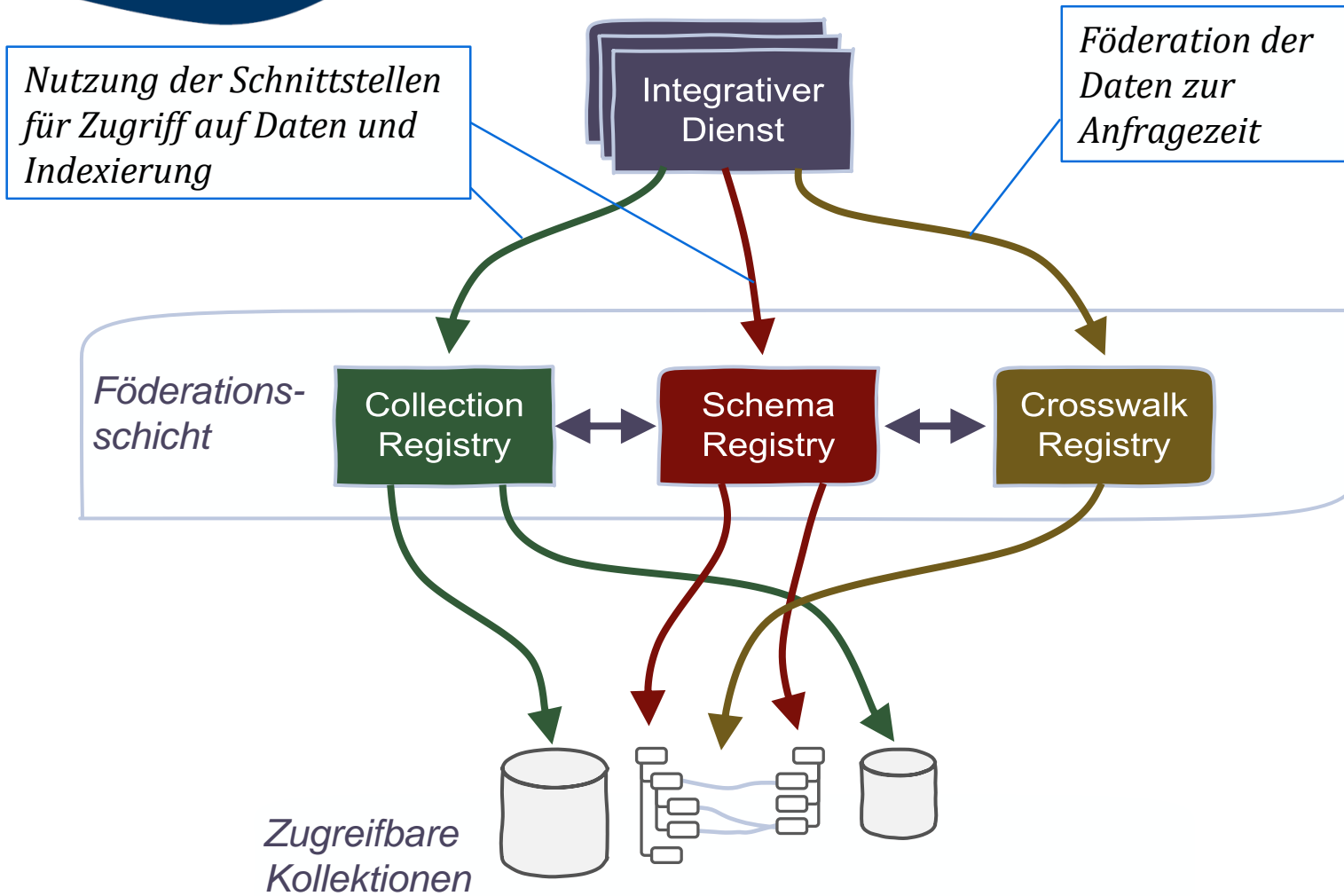
Fachwissenschaftler...

1. identifizieren relevante **Kollektionen**
2. wählen semantisch „passende“ **Export-Schemata** aus
3. **assoziiieren Konzepte** der gewählten Schemata

Ergebnis: Semantisch **eng**
korrelierte Datenbasis



Übergreifende Sicht
z. B. mittels Dublin Core

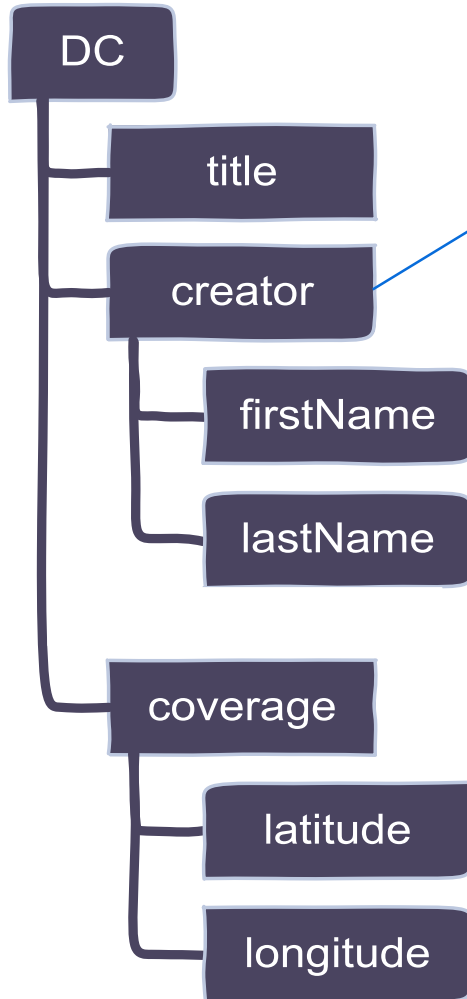


Anwendung der Föderationsarchitektur und generischen Suche

DATENMODELLIERUNG

- *Beschreibung der Daten* mit dem Ziel der Einarbeitung von Hintergrundwissen zu Daten in ihrem Erstellungskontext
 - durch Domänenexperten
 - Daten werden ohne Rücksicht auf spätere Verwendung erweitert
 - Kombination explizit vorliegender Strukturinformationen und Wissen um Instanzen

- Anwendung von Methoden aus dem Bereich *sprachbasierter Anwendungen*
 - Kernidee: Daten werden durch grammatikalische Regeln definiert und erklärt
 - Es werden *domänenspezifische Sprachen* (domain specific language, DSL) konstruiert und mit schematischen Informationen hinterlegt
 - Explikation von Wissen auf Instanzebene



Beispiele:

- *Substrukturen z. B. Listen*
- *Muster wie „Nachname, Vorname“*

Beispiel für Substruktur

```

...
<dc:coverage>
  LATITUDE: -46.069333 * LONGITUDE: 90.111167
  * MINIMUM AGE: 4.610 ka BP * MAXIMUM AGE:
  201.000 ka BP * MINIMUM DEPTH, sediment: 0.0
  m * MAXIMUM DEPTH, sediment: 11.7 m
</dc:coverage>
  
```

<http://doi.pangaea.de/10.1594/PANGAEA.51915>

EBNF

*Kontextfreie
Grammatik*

```
grammar PangaeaCoverage;
```

```
substruct      : subelem+;
```

```
subelem       : (longitude | latitude | start
                 | end | minDepth | maxDepth
                 | otherElem) SEPARATOR?;
```

*Modelliert durch
Domänenexperten*

```
longitude : 'LONGITUDE' ':' value;
```

```
latitude  : 'LATITUDE'  ':' value;
```

...

```
otherElem : key ':' value;
```

```
key       : ID;
```

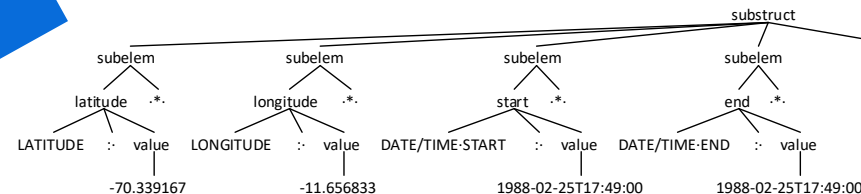
```
value    : DATE
```

```
          | ID;
```

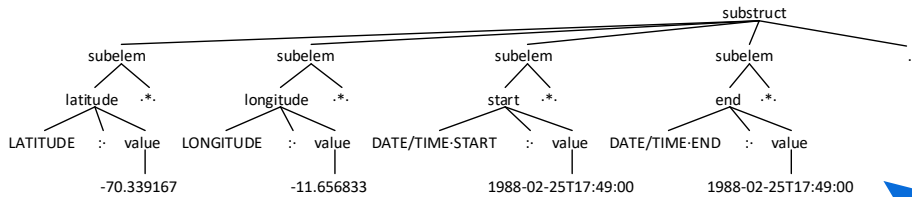
...

22 Zeilen insgesamt

Syntaxbaum



Syntaxbaum



Zuweisung

```
lon = @longitude.value;
```


Generierung untergeordneter Elemente

```
lat = {
  value = @latitude.value;
};
```

```
combined = CONCAT("[", @latitude.value, "-"
  , @longitude.value, "]);
```

Erweiterbarer Funktionsatz

DARIAH-DE Schema Registry Language ▾ Abmelden (tgradl)

 Schema Registry Schema-Editor

Schema Registry / Schemas und Mappings / Schema-Editor

Schema: **oai basis** Entwurf Bearbeiten Veröffentlichen Löschen

Beispieltransformation ⓘ

Sitzungen

Sitzung speichern Sitzung laden

Zurücksetzen

Beispieldaten

Input Ergebnisse Ausführen

Beispieldaten hier einfügen und die Schaltfläche 'Ausführen' betätigen, um die Transformation zu testen. Derzeit wird nur XML als Eingangsformat interpretiert.

Elementmodell

```

graph TD
    Dc[Dc] --- Title[Title]
    Dc --- Lang1[Lang]
    Dc --- Creator[Creator]
    Dc --- Lang2[Lang]
    Dc --- g["g: NameSplitter"]
    Dc --- f["f: NameSplitter"]
    g --- FirstName[FirstName]
    f --- FirstName
    
```

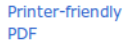
Live-Session 

© DARIAH-DE Datenschutz Impressum Kontakt

Anwendung der Föderationsarchitektur und generischen Suche

SCHEMA-MAPPINGS

- Einschränkung schemabasierter Crosswalks
 - Keine Abbildung kollektionsspezifischer Verwendung
 - Nur einfache Assoziationen zw. Elementen möglich

 [Printer-friendly PDF](#)

[Back to Electronic Publications](#)

Metadata Standards Crosswalk

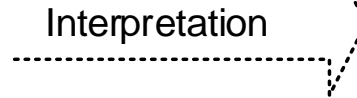
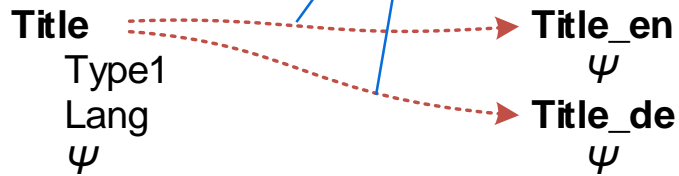
The crosswalk below includes only a partial list of the elements for each standard, focusing on the areas of overlap. This crosswalk is for planning purposes. For the full list of elements for any one of the standards below, click on the name of the standard at the top of the column.
(To print this crosswalk from your Web browser, use the tabloid 11 x 17-inch paper size setting, landscape print format, .25-inch margins. Alternatively, you can download the crosswalk as a PDF file.)

CDWA	CCO [1]	CDWA Lite [2]	VRA 4.0 XML	MARC/AACR	
1. OBJECT/ WORK (core)					
1.1. Catalog Level (core)		<cdwalite: recordType>	<vra: work> or <vra: collection>	655 Genre/Form 300a Physical Description - Extent	<g <e
1.2. Object/Work Type (core)	Work Type	<cdwalite: objectWorkType>	<vra: worktype> in <vra: work> or <vra: collection>	655 Genre - Form	<g
1.4. Components/Parts				300a Physical Description - Extent	<e
1.5. Remarks					
2. CLASSIFICATION (core)					
2.1. Classification Term (core)	Class	<cdwalite: classification>		050 084 "Other classification	<d

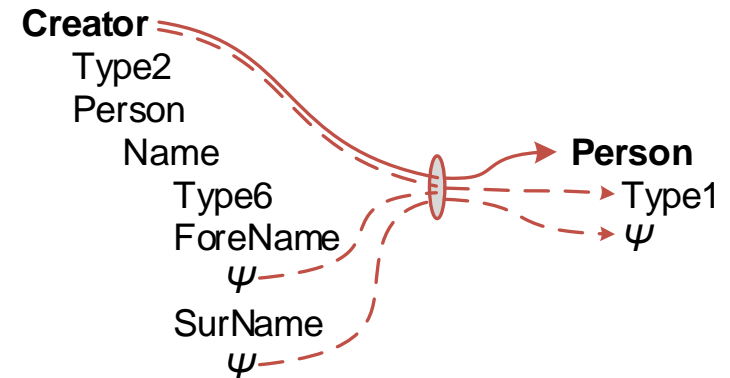
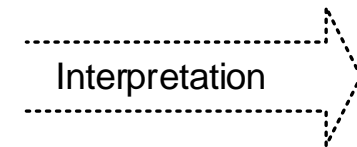
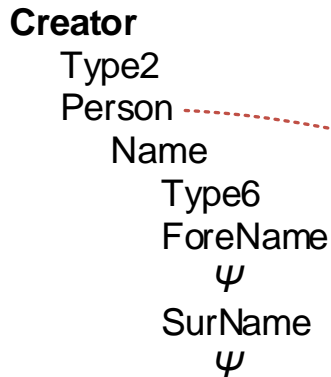
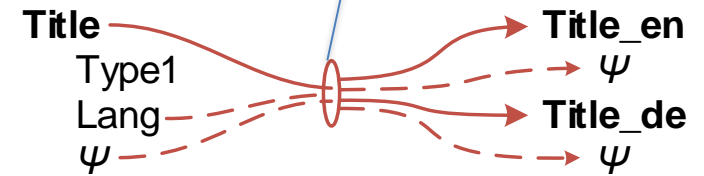
http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html

- wie bei Erweiterung von Schemata sind auch hier *Beschreibung und Transformation der Daten möglich*
- Hier aber: Fokus auf *Verwendungskontext*
- Modellierer: *Experte der Anwendungsdomäne*

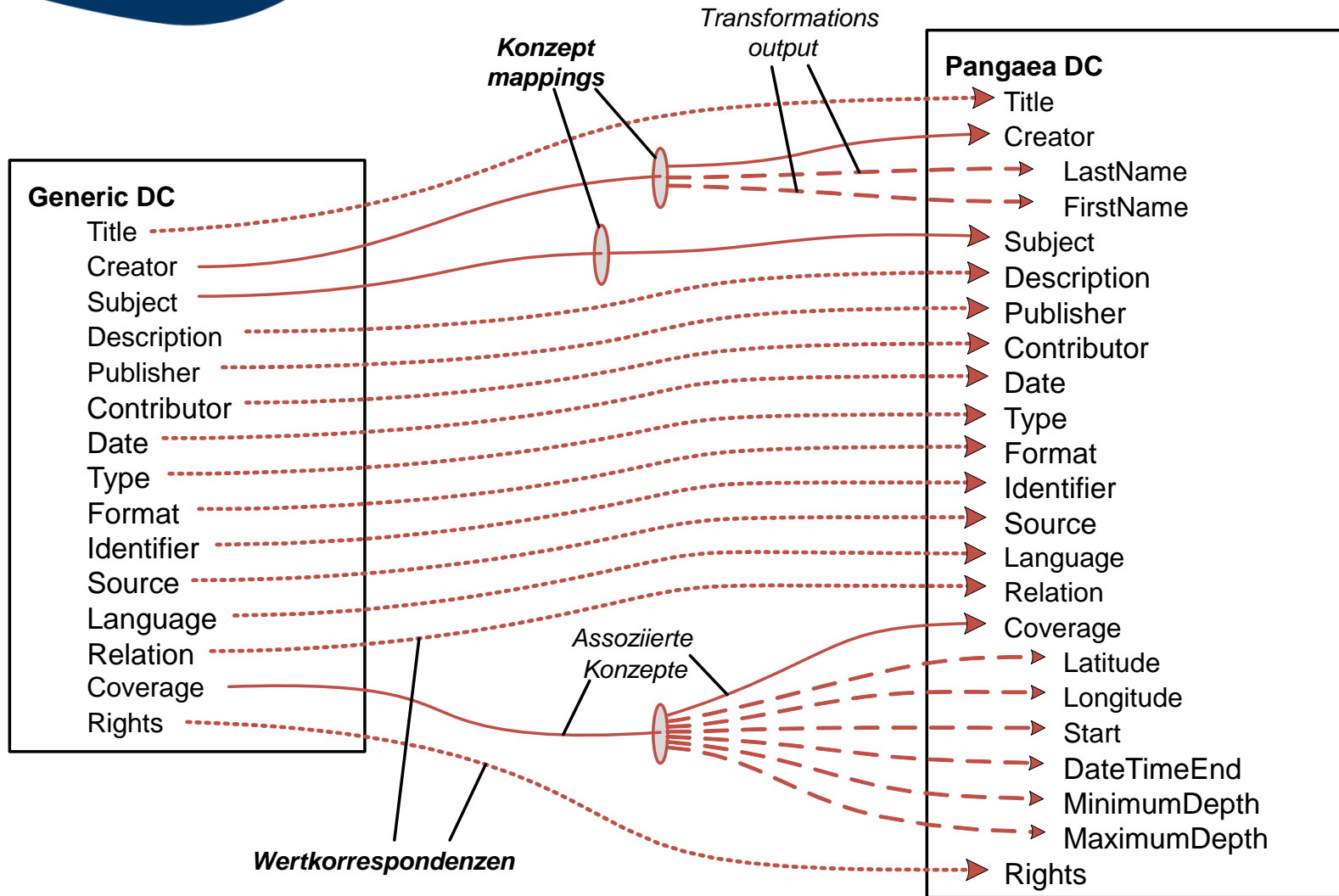
Wertkorrespondenzen




Assoziierte Konzepte



Schema-Mappings



DARIAH-DE Schema Registry Language ▾ Abmelden (tgradl)

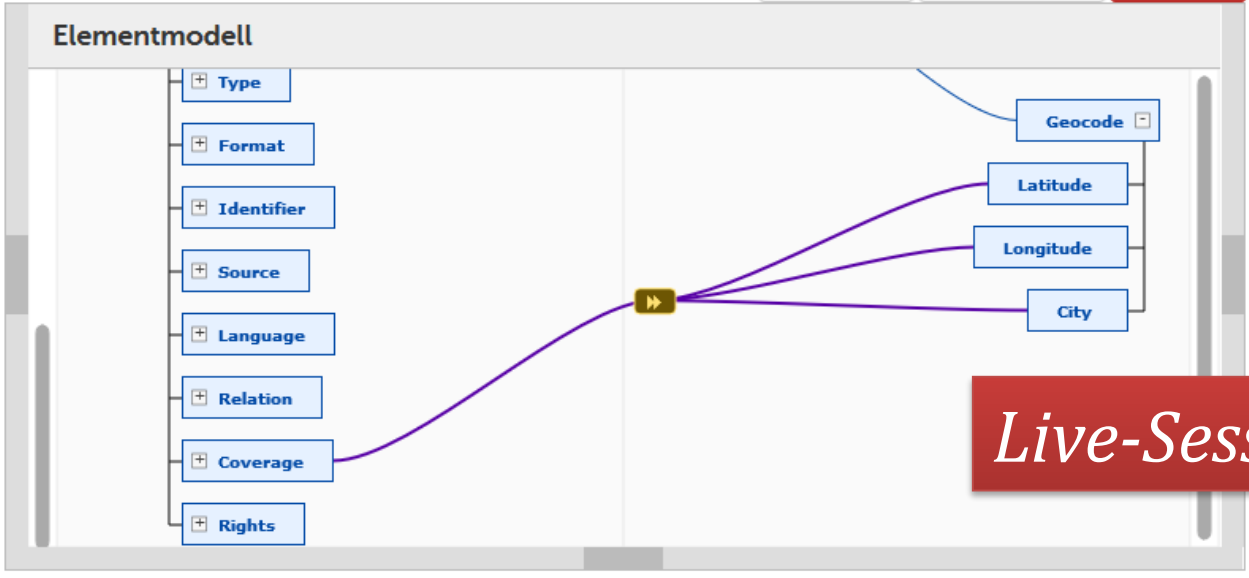
 Schema Registry **Mapping-Editor**

Schema Registry / Schemas und Mappings / Mapping-Editor

Quellschema: **oai basis** Zielschema: **geoschema** Entwurf

[Bearbeiten](#) [Veröffentlichen](#) [Löschen](#)

Elementmodell



The diagram shows a mapping editor interface. On the left, under 'Elementmodell', is a list of source schema elements: Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. On the right, under 'Zielschema', is a list of target schema elements: Geocode, Latitude, Longitude, and City. A yellow arrow points from the 'Coverage' element in the source schema to a central yellow box. From this central box, four purple lines connect to the 'Geocode', 'Latitude', 'Longitude', and 'City' elements in the target schema.

Live-Session

© DARIAH-DE Datenschutz Impressum Kontakt

Anwendung der Föderationsarchitektur und generischen Suche

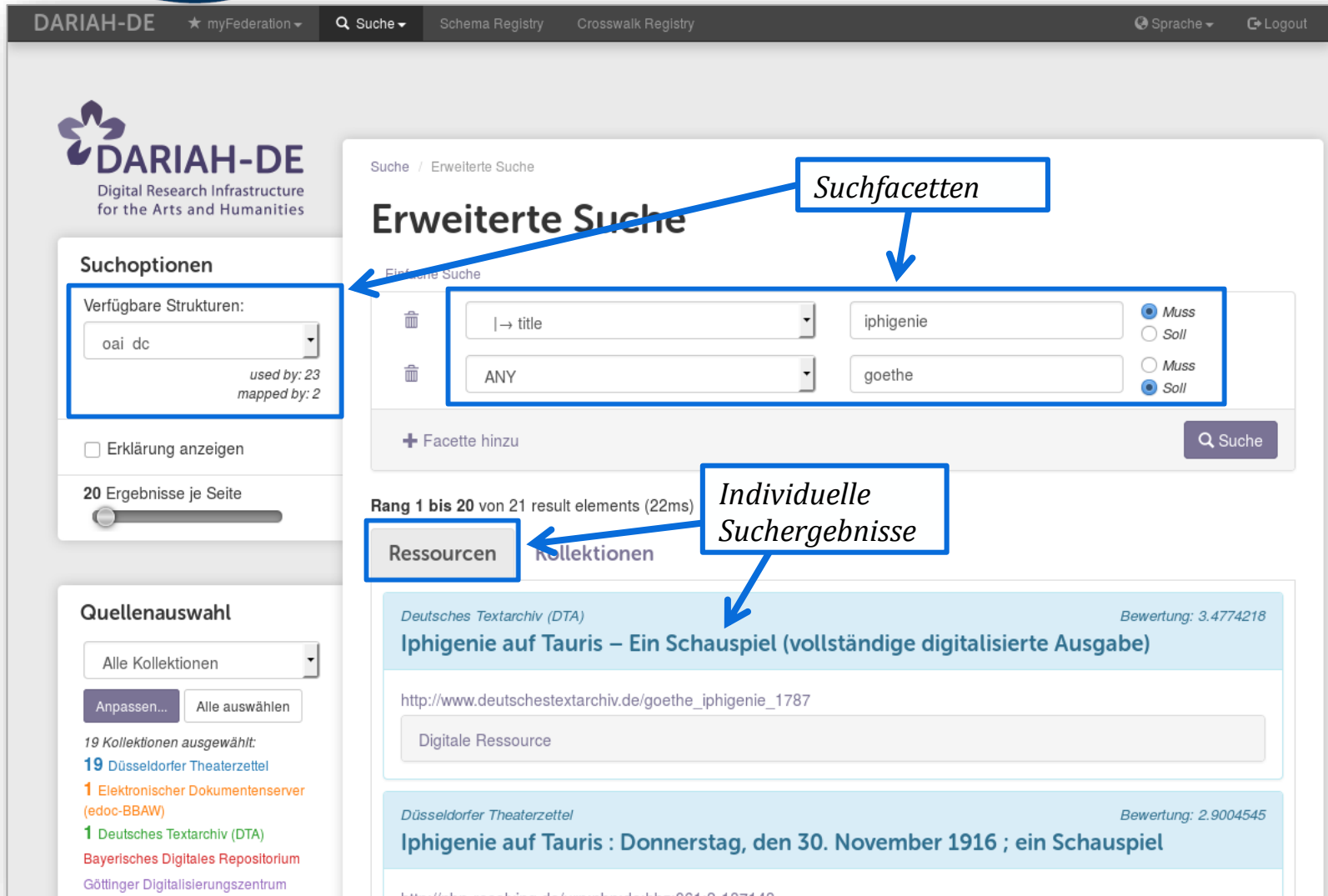
GENERISCHE SUCHE

The screenshot shows the DARIAH-DE search interface. At the top, there are navigation links for 'myFederation', 'Suche', 'Schema Registry', and 'Crosswalk Registry'. The main header reads 'Suche / Einfache Suche'. Below this, the search type is set to 'Einfache Suche'. A search input field contains the query 'jüdische'. The results are displayed as 'Rang 1 bis 20 von 102 result elements (12ms)'. On the left, there are sections for 'Suchoptionen' (with an unchecked 'Erklärung anzeigen' checkbox and a '20 Ergebnisse je Seite' slider) and 'Quellenauswahl' (with a dropdown menu set to 'Alle Kollektionen' and a list of 20 collections, including 'Bayerisches Digitales Repositorium'). A blue box highlights the 'Quellenauswahl' section with the label 'Auswahl von Kollektionen'. Another blue box highlights the 'jüdische' search input with the label 'Einfache Volltextsuche'. A third blue box highlights the 'Kollektionen' tab and the 'Bayerisches Digitales Repositorium' collection with the label 'Aggregation der Suchergebnisse je Kollektion'. The main area displays '98 Treffer' and a word cloud of search results. The most prominent words are 'jüdische', 'literatur', 'geschichte', 'sprache', 'interesse', 'politik', 'zeitung', 'religion', 'belletristik', and 'unpartheiisches'.

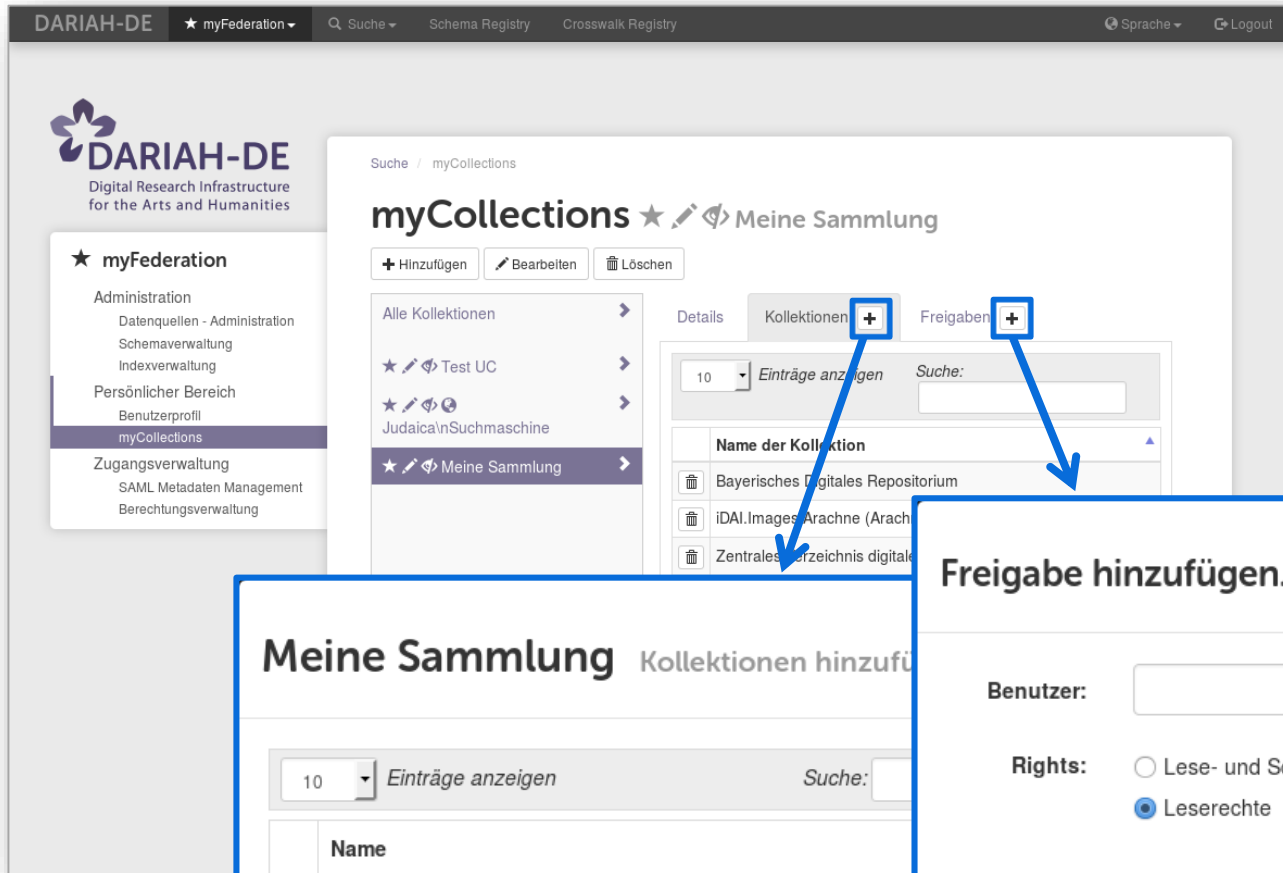
Auswahl von Kollektionen

Einfache Volltextsuche

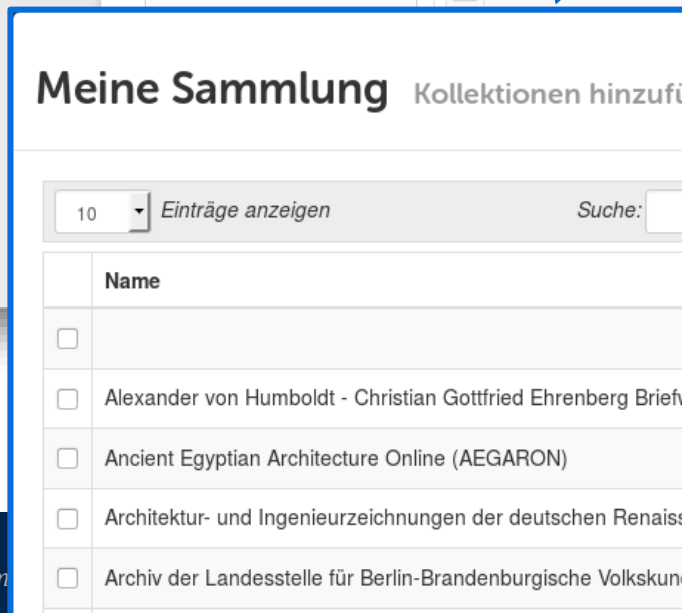
Aggregation der Suchergebnisse je Kollektion



The screenshot shows the DARIAH-DE search interface. At the top, there is a navigation bar with 'DARIAH-DE', 'myFederation', 'Suche', 'Schema Registry', 'Crosswalk Registry', 'Sprache', and 'Logout'. The main content area is titled 'Erweiterte Suche' (Advanced Search). On the left, there are three sidebars: 'Suchoptionen' (Search Options) with a dropdown for 'Verfügbare Strukturen' (Available Structures) set to 'oai dc', 'Erklärung anzeigen' (Show explanation) checkbox, and '20 Ergebnisse je Seite' (20 results per page); 'Quellenauswahl' (Source Selection) with a dropdown for 'Alle Kollektionen' (All Collections) and 'Anpassen...' (Adjust) and 'Alle auswählen' (Select all) buttons; and a list of 19 selected collections including 'Düsseldorfer Theaterzettel', 'Elektronischer Dokumentenserver (edoc-BBAW)', 'Deutsches Textarchiv (DTA)', 'Bayerisches Digitales Repitorium', and 'Göttinger Digitalisierungszentrum'. The main search area contains two search boxes: the first with 'title' selected and 'iphigenie' entered, and the second with 'ANY' selected and 'goethe' entered. To the right of these boxes are radio buttons for 'Muss' (Must) and 'Soll' (Should). Below the search boxes is a '+ Facette hinzu' (Add facet) button and a 'Suche' (Search) button. The results section shows 'Rang 1 bis 20 von 21 result elements (22ms)'. The first result is 'Iphigenie auf Tauris – Ein Schauspiel (vollständige digitalisierte Ausgabe)' from 'Deutsches Textarchiv (DTA)' with a rating of 3.4774218. The second result is 'Iphigenie auf Tauris : Donnerstag, den 30. November 1916 ; ein Schauspiel' from 'Düsseldorfer Theaterzettel' with a rating of 2.9004545. Blue arrows and boxes highlight key features: 'Suchfacetten' (Search facets) points to the search input area; 'Individuelle Suchergebnisse' (Individual search results) points to the first result; 'Ressourcen' (Resources) points to the 'Kollektionen' (Collections) link; and 'Erweiterte Suche' points to the search input area.

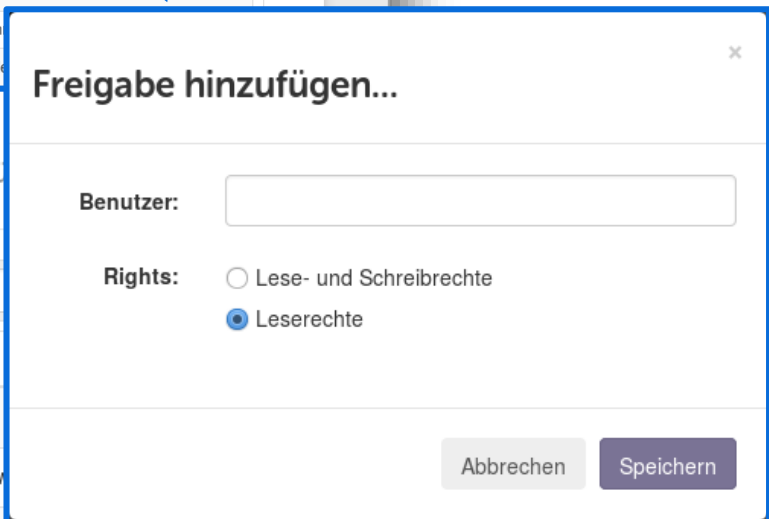


The screenshot shows the DARIAH-DE user interface. At the top, there's a navigation bar with 'DARIAH-DE', 'myFederation', and search options. The main content area is titled 'myCollections' and 'Meine Sammlung'. A sidebar on the left lists navigation options under 'myFederation'. The main area has buttons for '+ Hinzufügen', 'Bearbeiten', and 'Löschen'. Below these are links for 'Alle Kollektionen', 'Test UC', 'Judaica\InSuchmaschine', and 'Meine Sammlung'. A table of collections is visible, with a modal dialog 'Freigabe hinzufügen...' overlaid on top. The dialog has fields for 'Benutzer' and 'Rights' (with 'Leserechte' selected), and 'Abbrechen' and 'Speichern' buttons.



This inset shows a closer view of the 'Meine Sammlung' page. It features a search bar with '10 Einträge anzeigen' and a search input. Below is a table with the following entries:

Name
<input type="checkbox"/>
<input type="checkbox"/> Alexander von Humboldt - Christian Gottfried Ehrenberg Briefv...
<input type="checkbox"/> Ancient Egyptian Architecture Online (AEGARON)
<input type="checkbox"/> Architektur- und Ingenieurzeichnungen der deutschen Renaissance
<input type="checkbox"/> Archiv der Landesstelle für Berlin-Brandenburgische Volkskunde



This inset shows the 'Freigabe hinzufügen...' modal dialog. It contains the following elements:

- Benutzer:** A text input field.
- Rights:** Radio buttons for 'Lese- und Schreibrechte' and 'Leserechte' (which is selected).
- Buttons:** 'Abbrechen' and 'Speichern'.

myCollection bearbeiten

Name: Judaica\nSuchmaschine Ausgangspunkt: Benutzerkollection

Beschreibung:


„Branding“

Branded Search

Als 'branded search' veröffentlichen?

#ffc77f Kontrastfarbe: #000000


Organisationslogo



SALOMON LUDWIG
STEINHEIM INSTITUT
FÜR DEUTSCH-JÜDISCHE GESCHICHTE


Abbrechen Speichern


DARIAH-DE myFederation Suche Schema Registry Crosswalk Registry de.dariah.eu



**Judaica
Suchmaschine**

Index: 6 Kollektionen | 4226 Dokumente





powered by
DARIAH-DE
Generische Suche

Vielen Dank für die Aufmerksamkeit!

Collection Registry

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

Schema Registry

<http://schereg.de.dariah.eu>

Generische Suche

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

Repository

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