



Aggregating Performing Arts

Representing different Collections in one Database

Julia Beck | j.beck@ub.uni-frankfurt.de | @j4lib

Open Data - Open Access : New Frontiers for Archives and Digital Platforms dedicated to the Performing Arts





"The Memory Machine" Project documentation Rauminstallation, 2013 [Tanzfonds Erbe]

- Aggregates metadata from GLAM institutions from
 the performing arts domain (at the moment especially
 German-speaking institutions from Germany, Austria and Switzerland)
- Funded by the German Research Foundation
- What we are doing is best seen here:
- And here:

http://www.performing-arts.eu









- New version of the search portal coming up in the next days
- It consists of:



• Planned: venues/places









search find discover		Objekt / Werk Q Suchen Erweitert			
Suche / / Ähnliche Einträge					
		Datensatz exportieren			
	Schiller: Kabale und Liebe				
<u>8</u>	Materialtyp:	Fotografie			
	Umfang:	1 Exemplar 15,7 x 23 cm (6 3/16 x 9 1/16 in.)			
Protocial as a	Provenienz:	Theatermuseum der Landeshauptstadt Düsseldorf, Düsseldorf			
	Beitragende:	Kyser, Karl (SchauspielerIn), Gebühr, Cornelie (SchauspielerIn), Everth, Franz (SchauspielerIn), Atelier Hehmke-Winterer (Fotografie)			
"See also".	Sprache:	Deutsch			
Information about	Beschreibung:	In dieser Szene sind Cornelie Gebühr als Frau Miller, Franz Everth als Miller und Karl Kyser als Wurm zu sehen.			
		Szenenfoto			
data provider and		Schwarz-Weiß Fotografie auf Karton aufgezogen			
other related					
resources	Siene auch				

Datengeber: Theatermuseum der Landeshauptstadt Düsseldorf C Mehr von diesem Datengeber im FID Portal Link zum Katalog: Link zu diesem Titel beim Datengeber Theatermuseum der Landeshauptstadt Düsseldorf 🖸 Zugehörig: Kabale und Liebe (Inszenierungsbeschreibung)



Suche: (GND 2108840-8)



Theatermuseum (Düsseldorf)

Theatermuseum, Theaterarchiv gegründet 1981 Sitz in Düsseldorf Wirkungsgebiet: Nordrhein-Westfalen Schlagwort: Archiv, Museum, Theater, Tanz Vorgänger: Dumont-Lindemann-Archiv (Düsseldorf)

Namensvarianten

Theatermuseum Düsseldorf; Theatermuseum der Landeshauptstadt Düsseldorf

Weiterführende Links

- 🔀 International Standard Name Identifier (ISNI) 🖸
- 💵 Wikidata 🗹
- Gemeinsame Normdatei (GND) im Katalog der Deutschen Nationalbibliothek
- NACO Authority File
- 🎇 Virtual International Authority File (VIAF) 🖸
- BNF Bibliothèque nationale de France I
- DE-611 Kalliope Verbundkatalog C
- 🌱 Deutsche Digitale Bibliothek 🗹

Quelle:

Information about institution

	Suche einschränken		Treffe	r 1 - 20 von 3.622 für Suche '(GND 2108840-8)', Suchdauer: 0,03s	Sortieren	Relevanz	•
	Nur in Archivbeständen suchen		1	Form im Raum : Walter von Wecus, Das szenische Werk ; Ausstellun August bis 17. November 1993, anläßlich seines 100. Geburtstages	ng vom 29.		
	Datengeber			In Dokumente zur Theatergeschichte Wegus Walter von (Beitragende) Theatermuseum (Beitragende)			
	 Universitätsbibliothek Frankfurt am Main Image: Constraint of the second se	(301.880)		Düsseldorf : 1993 156 S, III Druckschrift			
	Verbund Deutscher	(99.884)		(Universitätsbibliothek Frankfurt am Main) (Verfügbarkelt KVK 🗷)			
0	bjects relat	ted to	2	Poet des Tanzes : der Choreograph Erich Walter ; [Ausstellung des Theatermuseumsin Zusammenarbeit mit der Stadt Duisburg ; Theate Düsseldorf vom 28.11.1993 - 16.1.1994 : Theater der Stadt Dui	ermuseum		
in	stitution			In Dokumente zur Theatergeschichte Kügler, Ilka (Beitragende), Theatermuseum (Beitragende)			



Suche: (GND 116713631)



RITA SACCHETTO AS SHE APPEARS IN PRIVATE LIFE

Sacchetto, Rita

Tänzerin, Sängerin, Schauspielerin geboren 15. Januar 1880 in München gestorben 18. Januar 1959 in Genua-Nervi

Biografische Angaben Gräfin Zamoyski; Geburtsjahr nach anderen Quellen: 1879 und 1886

Namensvarianten Sacchetto, Margherita; Zamoyski, Rita

Weiterführende Links

- 🔀 International Standard Name Identifier (ISNI) 🖸
- 📠 Filmportal 🗹
- 🥻 Virtual International Authority File (VIAF) 🖸
- 💵 Wikidata 🗹

Gemeinsame Normdatei (GND) im Katalog der Deutschen Nationalbibliothek

🗾 Bayerisches Musiker-Lexikon Online 🖸

BNF Bibliothèque nationale de France 🗹

DE-611 Kalliope Verbundkatalog

w Wikipedia (Deutsch)

w Wikipedia (English)

Quelle:

http://d-nb.info/gnd/116713631/about C (Lizenz: CC0 1.0 C) via lobid-gnd C







The Challenge



THE LANGEST AND REAL .

"The Laughing Audience and A Chorus of Singers" Copperplate by William Hogarth, 1733 [Theatre Museum of the State Capital of Düsseldorf]

Challenges in Aggregation





Typical differences and challenges in the metadata

- Different data formats and standards
- Different scope and detail of description
- Unstructured data / free text / "hidden information"
- No separation of entities (1 table row for all information)
- No identifiers for persons, places, events; no relations between entities
- No documentation

Challenges in Aggregation



Why are those differences a problem?

- Different data formats and standards; different "perspectives"
 - > Many conversions and mappings needed \rightarrow potential loss of information
- Different scope and detail of description
 - Confusion of the user
- Unstructured data / free text / "hidden information" / no relations / no entities
 - Harder to find the desired information
- No documentation
 - Loss of knowledge about the data
- Different spellings; singular vs plural; detail vs. abstraction
 - Normalization needed (for facetting)



Structured data is needed!

- There are some requirements for a functional search:
 - Obviously: the desired information was described in the data
 - the desired information is structured
 - > For more intelligent search: the information can be uniquely identified
 - Nice to have: ranking and facetting of results

Problematic: Expectations vs. actual existing data

If you have to uniquely identify entities from literals later on with the help of authority data and ontologies, these means a lot of manual work because automatic matching is prone to error!



What is possible with structured data:





Practical hints for data providers



Checklist

- Provide unique identifiers
- Use authority data / controlled vocabularies
- Rethink syntax and semantics of your data
- Choose property names wisely
- Keep internal information separate
- Separate entities and their description...
- o ... but keep relationships between your entities
- Good documentation is key



→ "Behind the scenes"



"The Taming of the Shrew [IV]" Set design draft by Traugott Müller, 1942 [Freie Universität Berlin, Institut für Theaterwissenschaft, Theaterhistorische Sammlungen]

Data Model



Europeana Data Model (EDM)

- flexible, extensible, universal, interoperable
- able to map different kinds of cultural heritage objects
- able to map data from all kinds of institutions (archives, libraries, museums)
- data can be published in EDM as Linked Open Data
- gives data providers possibility to also share their data in DDB (German Digital Library) and Europeana as they use the same data model
- we use a variant of EDM: Digital Manuscripts to Europeana (DM2E)
- extended with eclap namespace for performing arts



Workflow

www.performing-arts.eu



Technologies used: VuFind (discovery system) with Apache Solr (search engine),

BaseX (XML database) and OpenRefine

Workflow



From original data to linked data

- Preprocessing of metadata to normalize it and try to get rid of mistakes
 - preferably give the results back to the data provider's original system
- Parsing, mapping and transforming metadata while preserving the given information and making the most of it
 - requires thorough analysis and understanding of the data given
- Matching of metadata literals to authority data in order to enrich it
 - Half-automatic approach



Visit the search portal and please, give us your feedback!

Contact: Julia Beck | j.beck@ub.uni-frankfurt.de Project leader: Franziska Voß | f.voss@ub.uni-frankfurt.de