FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

DDB-KG:
The German Bibliographic Heritage in a Knowledge Graph

HistoInformatics@JCDL2021 - 30.September.2021

Mary Ann Tan, Tabea Tietz, Oleksandra Bruns
Jonas Oppenlaender, Danilo Dessì, Harald Sack

Search and Retrieval in Libraries

284 B.C.
Catalog Book
Zenotodus
Library of Alexandria

Printed
Book
Catalog

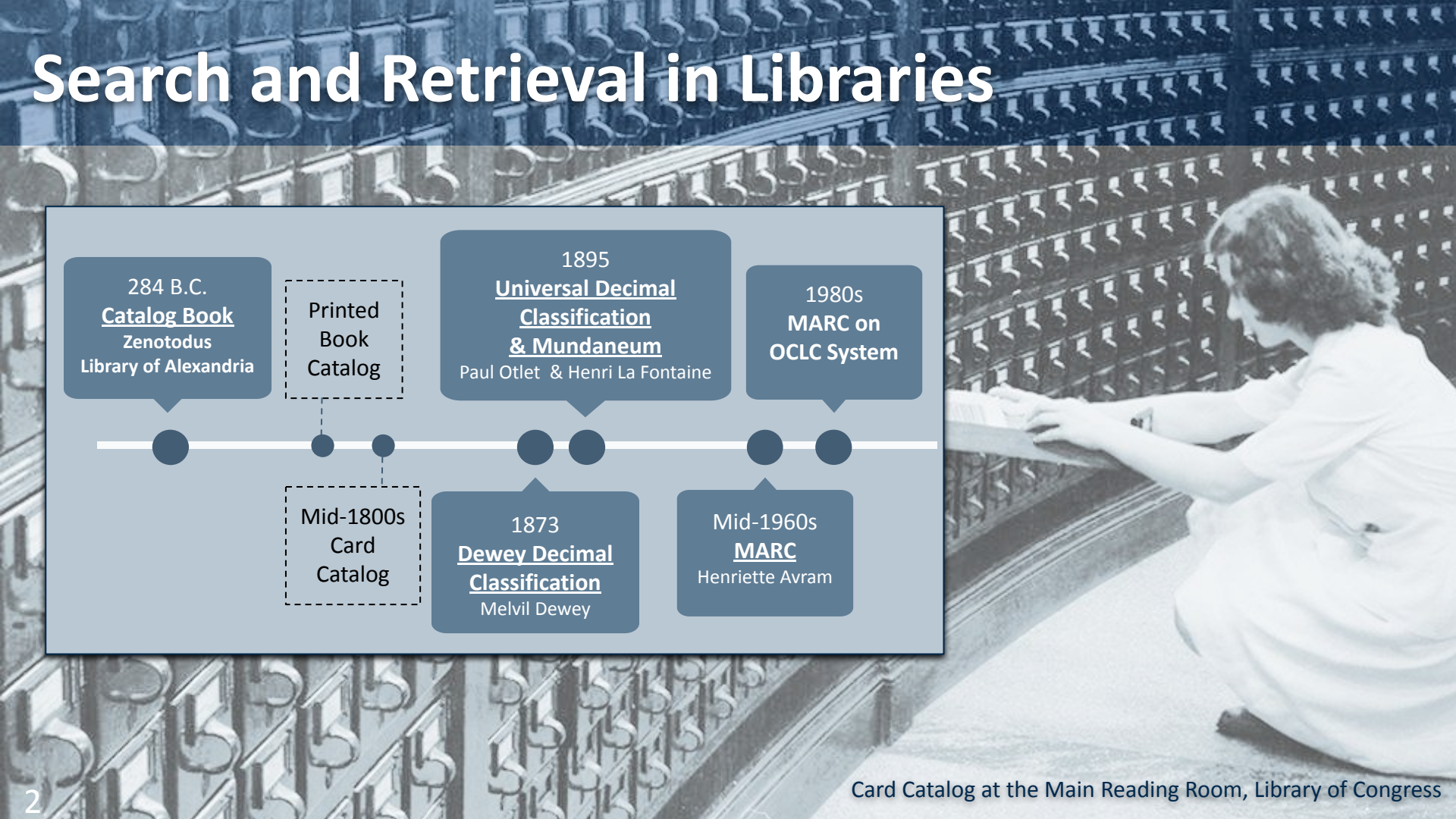
Mid-1800s
Card
Catalog

1873
Dewey Decimal
Classification
Melvil Dewey

1895
Universal Decimal
Classification
& Mundaneum
Paul Otlet & Henri La Fontaine

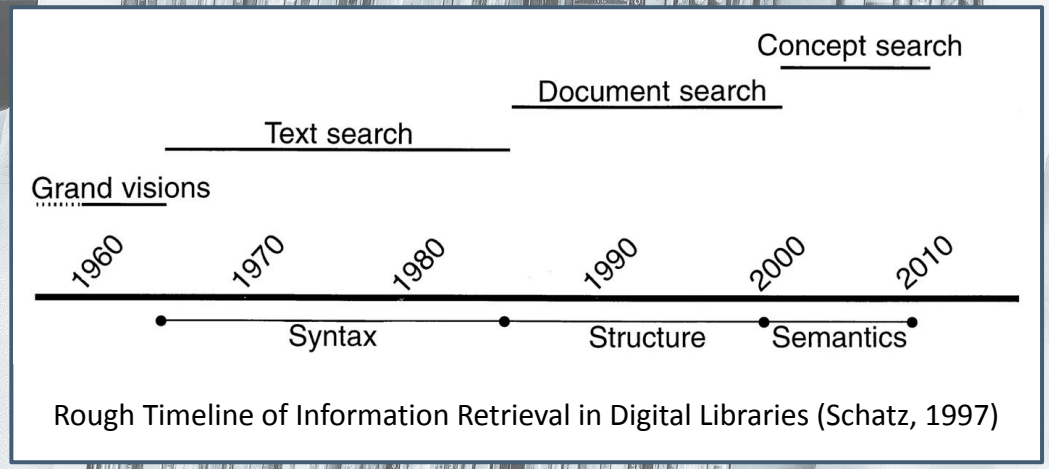
Mid-1960s
MARC
Henriette Avram

1980s
MARC on
OCLC System



Card Catalog at the Main Reading Room, Library of Congress

Digital Libraries



Rough Timeline of Information Retrieval in Digital Libraries (Schatz, 1997)

Digital Libraries by Gerd Altmann, Pixabay

German Digital Library (DDB) - A Cultural Heritage Portal

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

europeana
think culture

Cultural Heritage Objects (CHOs)

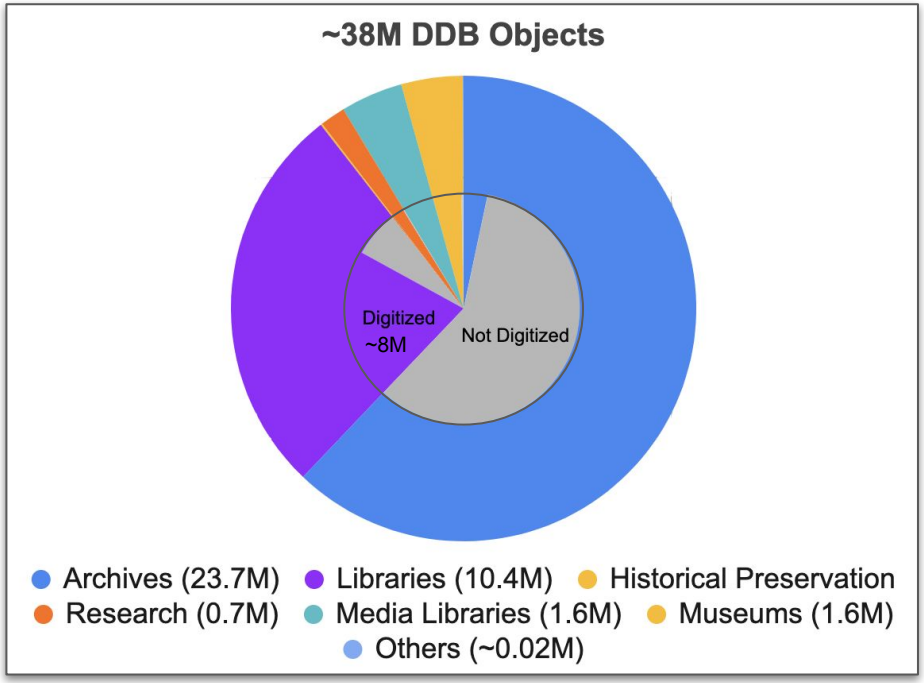
DDB and the Europeana

GOALS of the **DDB** or the **German Digital Library**:

✓ **Aggregate** metadata of Germany's CHOs for **Europeana**

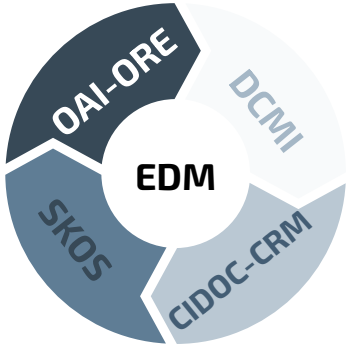
? Make digitized CHOs available to the general public.

- ❑ Easy to search
- ❑ Returns relevant results



DDB-EDM: What's under the hood?

Europeana Data Model (EDM)



- resource vs digital representation
- resource vs metadata
- allow contradicting metadata
- support granularity
- support resources for contextual descriptions

DDB-EDM

- Dublin Core Metadata Terms (dcterms)
- Properties to represent:
  1. Hierarchical elements
  2. Sequence ordering
  3. Document components

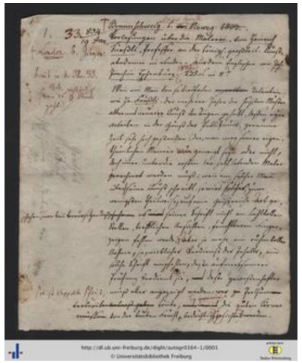
How is the DDB now?

Friedrich Schiller (1759 - 1805)
- playwright, poet, philosopher
- *Sturm und Drang* movement

Challenges - One class to rule them all!

All objects are instances of *edm:ProvidedCHO*!

1. Heterogeneous CHOs



- Goethe's Manuscript

- Audio: Der Vogel im Wald

- A photo of a model of Barberini Faun

- A photo of Jägerhof Hotel

2. Granularity and hierarchical organization of CHOs

Other Challenges

3. **Generic properties for important attributes**
   - Edition, page count, volume no. are all stored in *dc:description*.
4. **Related CHOs are not linked.**
   - e.g. no way to encode bibliographic relationships (Tillet, 2004) in DDB-EDM.
   - Exploration of the collection is limited.
5. **Keyword-based search and retrieval on unnormalized terms.**
   - Facets are based on keyword.
   - Keywords are a mixture of literals, controlled vocab, linked data.

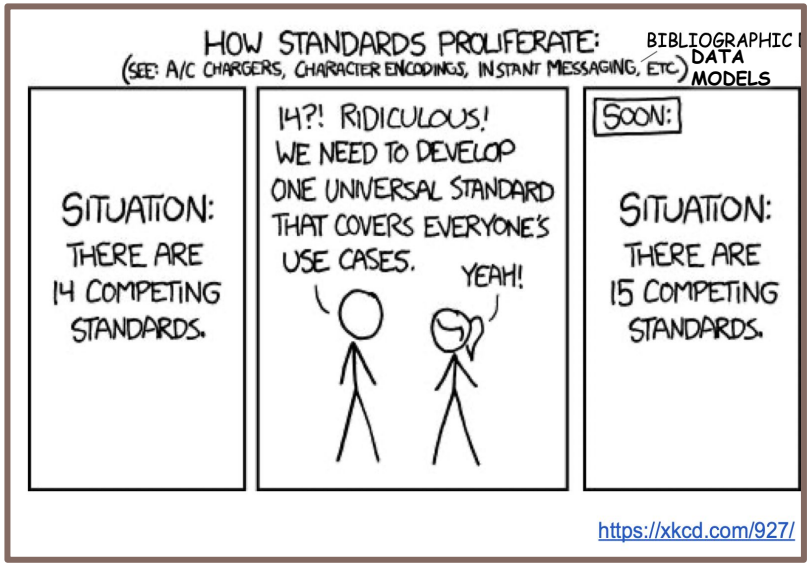
How to address the challenges?

Semantic Web technologies (Berners-Lee et al., 2001)

- **Ontologies** to encode domain-specific concepts
  - Sector-specific ontologies + upper-level ontology
    - *FinnOnto* - Hyvönen et al., 2008
    - *ArCO* - Carriero et al., 2019
- **Knowledge Graphs** for representation, storage and exploration of data
  - Semantic search
  - Exploration
    - integration with existing linked open data such as GND*, Wikidata, DBPedia, etc.) (*refer* - Tietz et al., 2016)

* Gemeinsame Normdatei (Authority File) is maintained by the German National Library (DNB)

Bibliographic Ontologies



Source: Suominen and Hyvönen, 2016

	BIBO	BF	FaBiO*
Used By	BNB DNB	LOC	ArCO
Use Case	Publishing	Cataloging	Publishing
Hierarchical?	✘	✓	✓
FRBR?	✘	W+E, M, I	✓

* FaBiO is part of SPAR Ontologies, which includes Document Component Ontology (DoCO)

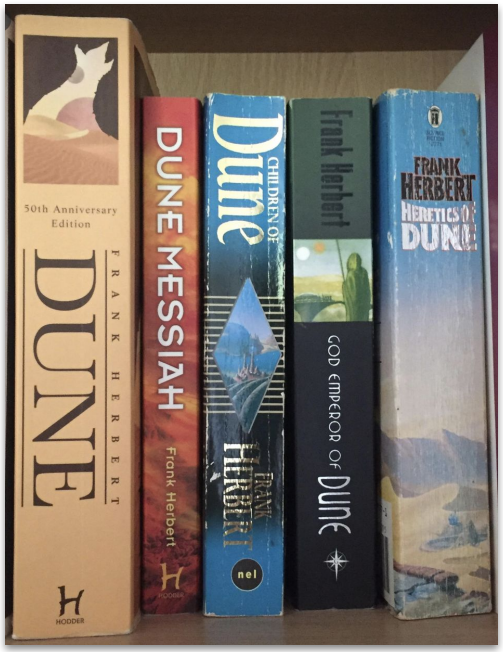
FRBR (/ˈfɜːrbər/) - How to model a book?

FRBR Functional **R**equirements for **B**ibliographic **R**ecords

A book is:

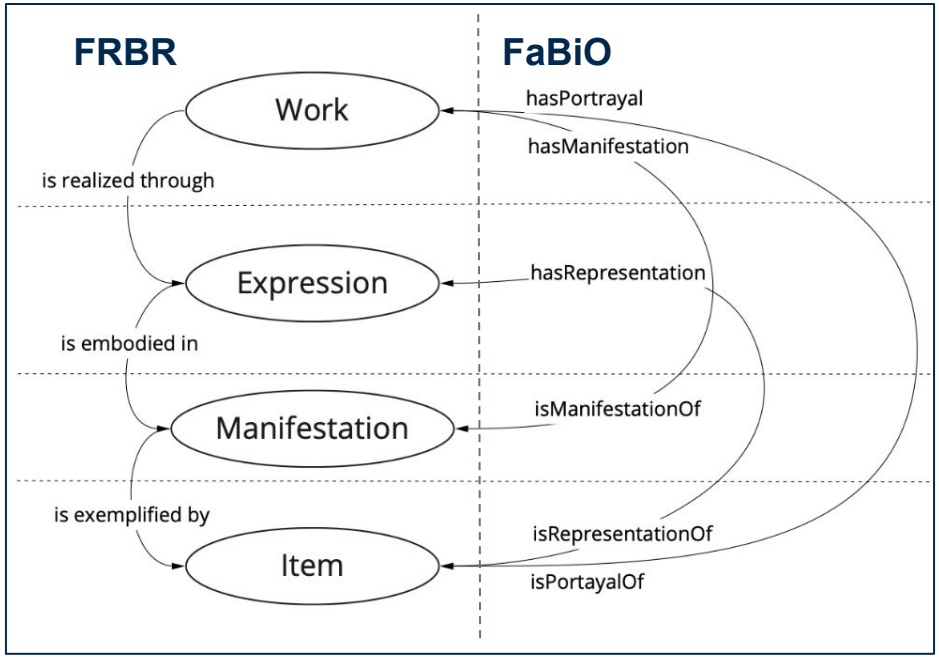
- a copy of “Dune” that you have on your book shelf (Item)
- can be identified by an ISBN referring to “Dune” by Frank Herbert (Manifestation)
- a specific translation of “Dune” to German (Expression)
- “Dune” authored by Frank Herbert (Work)

How to relate the book to an ebook or film adaptation?



FRBR and FaBiO (Peroni and Shotton, 2012)

4 Layers of Abstractions:



FRBR **extension** in FaBiO

- additional relationships
- allow “skipping” of layers
- e.g. rare manuscripts with neither expression nor manifestation.

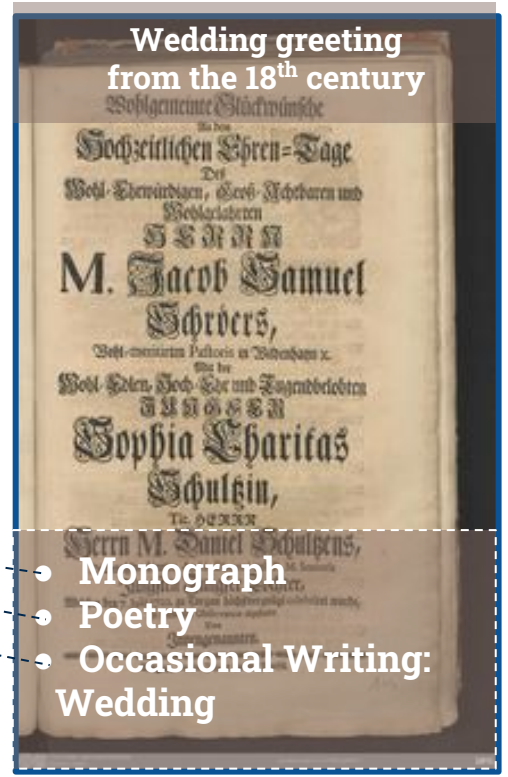
DDB-EDM to FaBiO Mapping

DDB objects types
770 Object Type Terms
dc:type, edm:hasType

FaBiO has 238 Classes

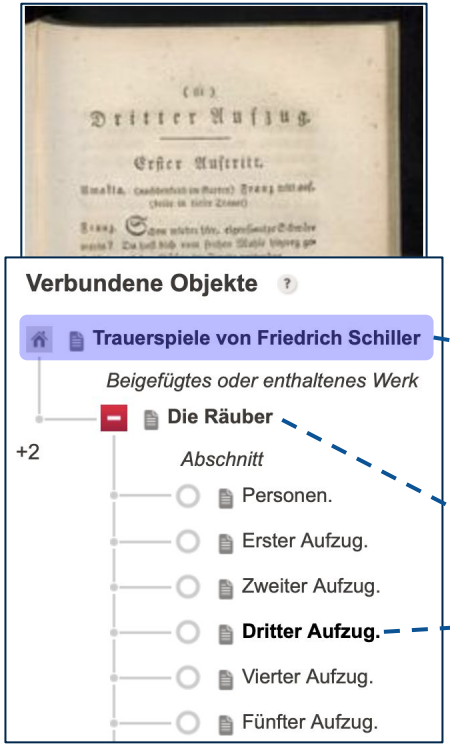
Challenges:

- 1 object, many “types” or terms
- terms are mixture of concepts:
    1. document types
    2. subject headings
    3. production process or manifestation
    4. document hierarchy



Painstaking manual mapping… NOT 1-to-1!

Granularity - An Example



Friedrich Schiller's Dramas

↳ The Robbers

↳ Cast

↳ ...

↳ **Third Act**

Primary

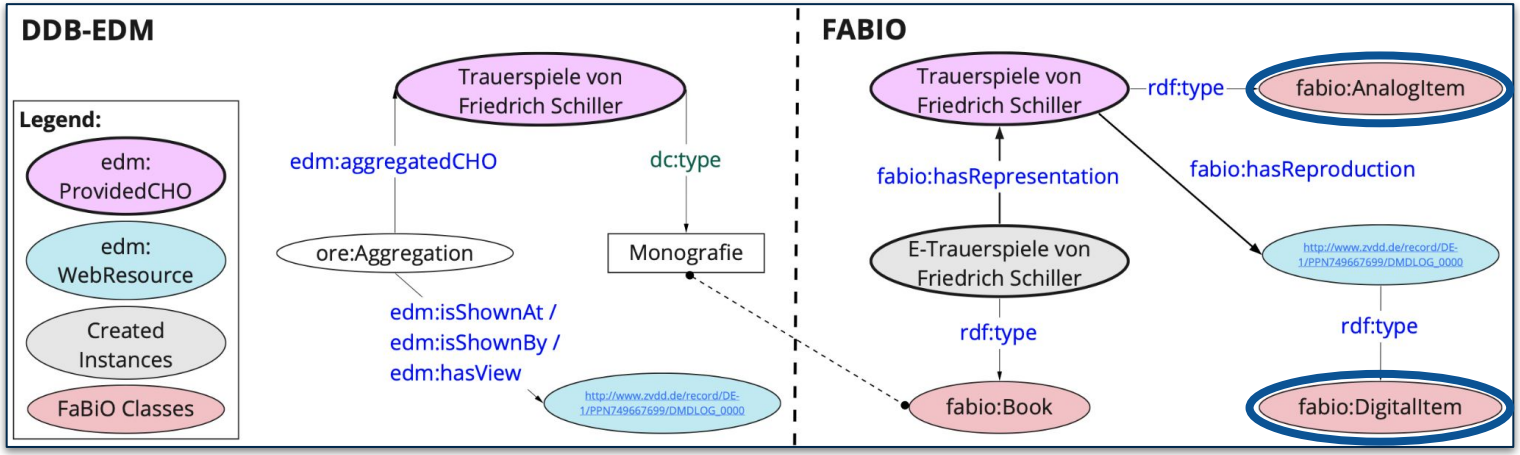
- top of the hierarchy.
- often considered as a Work or Work Collection.

Secondary or Partial

- part of other CHOs.
- could be a distinct Work.

From edm:ProvidedCHO to FaBiO Endeavors - 1 of 2

1. Item instances are automatically created.



2. Item instances of *secondary* objects are linked to their parent (*primary* object) item instances.

From edm:ProvidedCHO to FaBiO Endeavors - 2 of 2

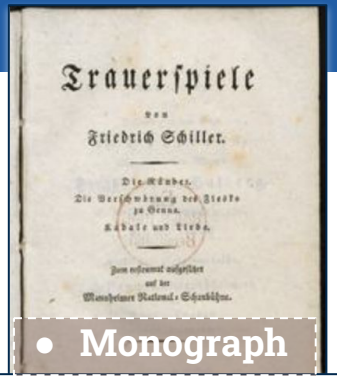
3. **Work** instance is created when the object title is in the German Authority File or *Gemeinsame Normdatei* (GND).

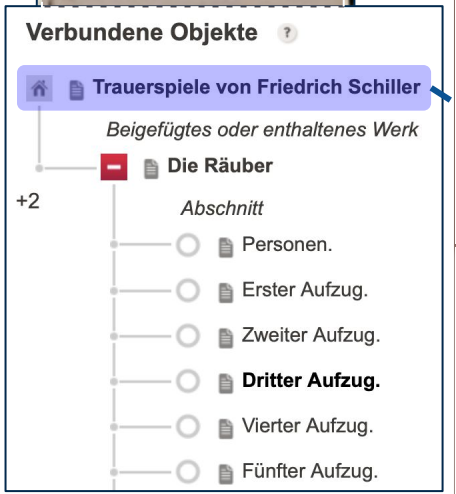
4. **Expression** and **Manifestation** instances are created for primary objects.
   - class assignment is based on the mapping.

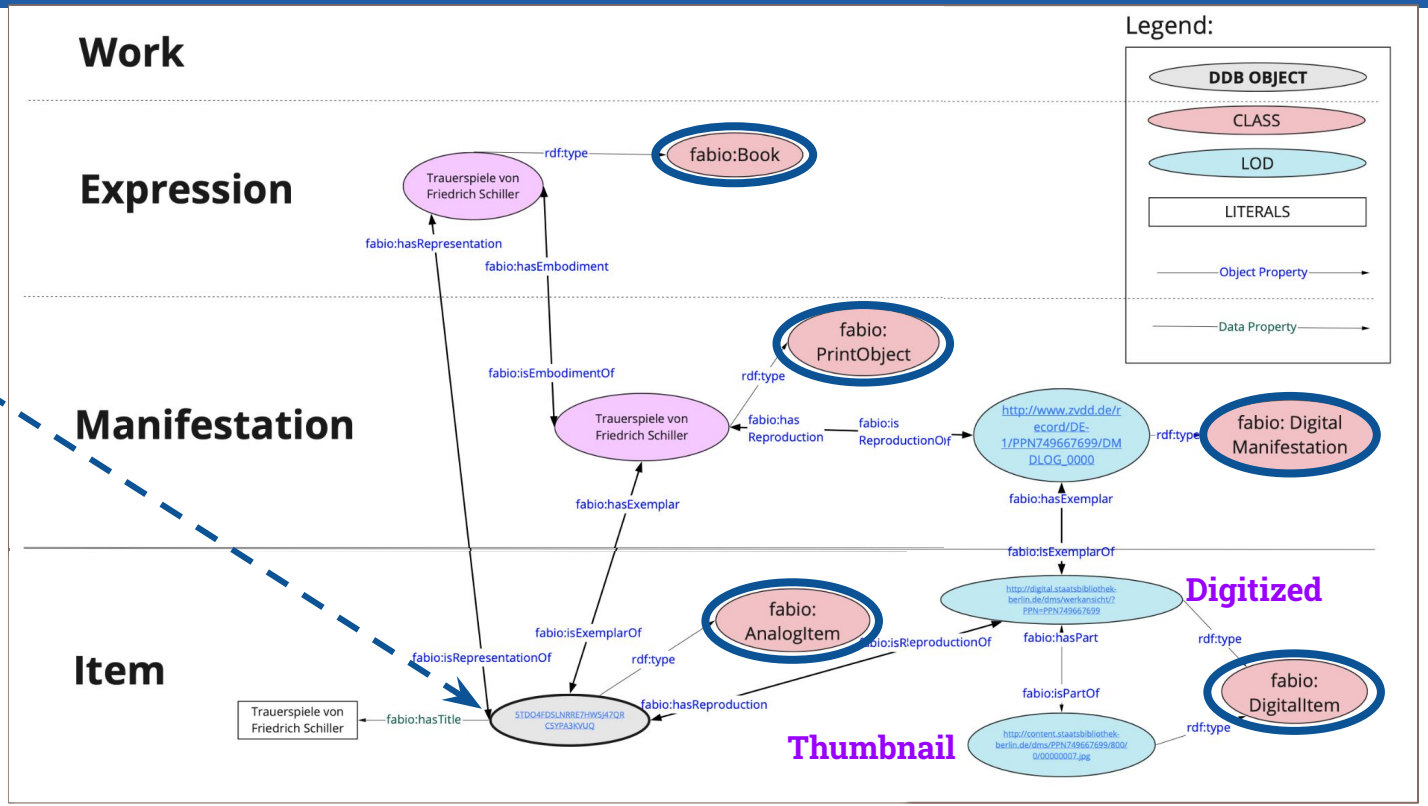
1	Terms (de)	Total Objects	en	Normalized	Concept	Work	Expression	Manifestation	Item
16	Akten	1	Records	Akte	T	fabio:ArchivalDocument		fabio:AnalogManifestation	fabio:AnalogItem
17	Akten Kriegsbriefe	282	Archival Documents / War Letters		T	fabio:ArchivalRecord	fabio:Letter	fabio:AnalogManifestation	fabio:AnalogItem

Term-FaBiO classes and attributes mappings (https://bit.ly/3qBxxCo).
- mappings are also available in JSON files on GitHub (https://git.io/JzVTS)
- when endeavor sub-class is not set, default is assigned (e.g. fabio:Expression)

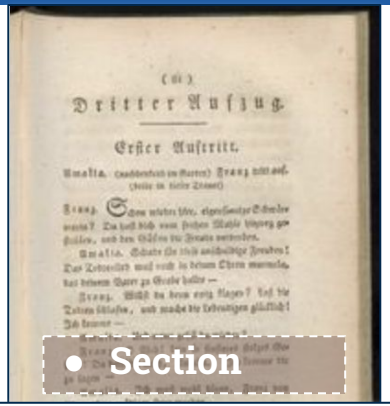
Example: Friedrich Schiller's Plays (Primary)



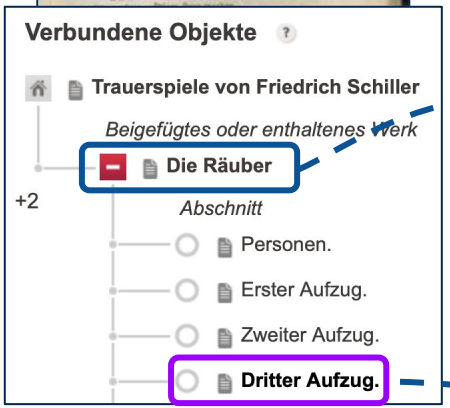


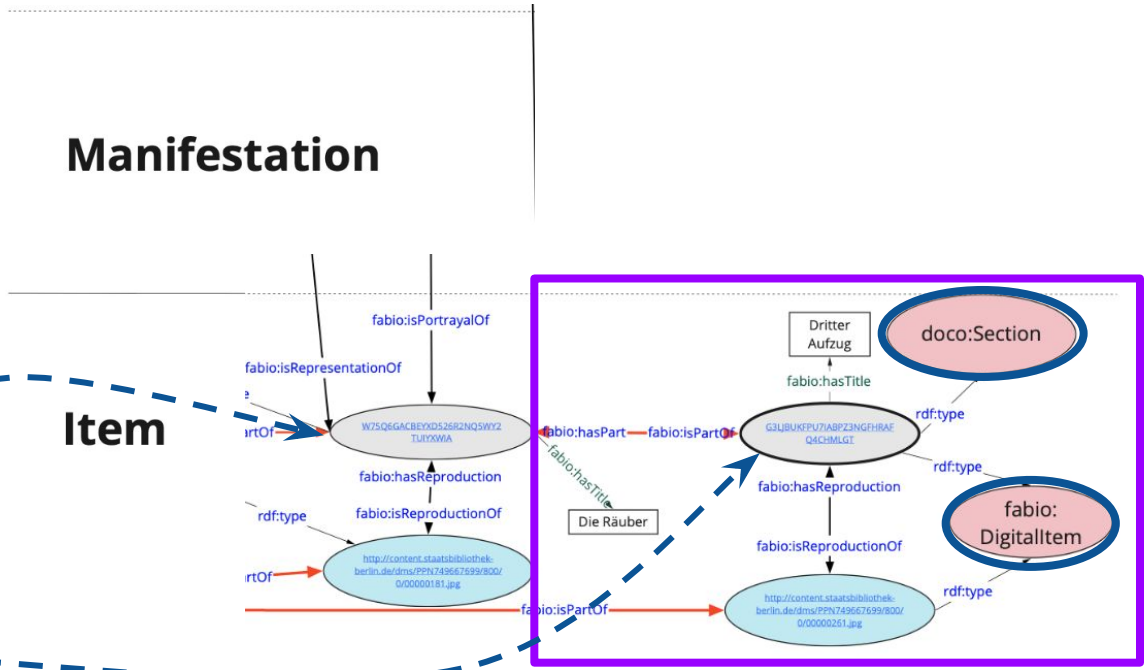


Example: Third Act of The Robbers (Secondary)



Complete modeling: https://bit.ly/3h5Uf2F





Proof of Concept

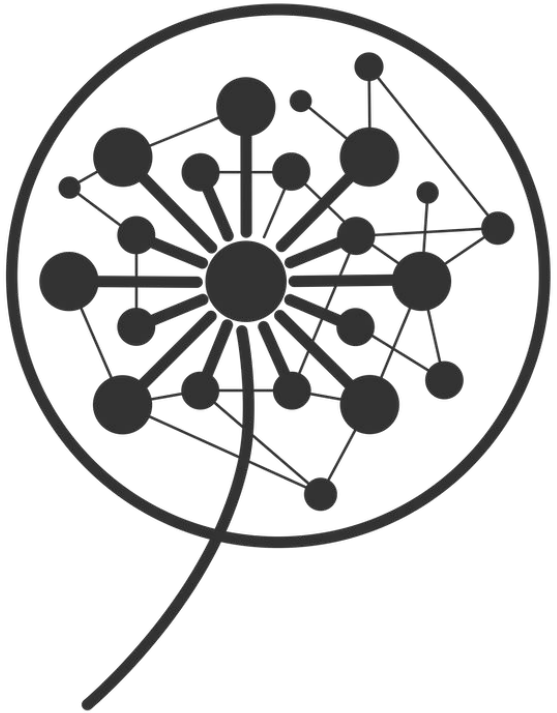
SPARQL Endpoint hosted with Virtuoso triple store instance:

- from 22k objects from the libraries, 2M triples

Search on specific abstraction levels, for example:

- Works of an author
- Translations of a book
- Digitized copies of a book

https://ise-fizkarlsruhe.github.io/ddbkg/docs/examples/



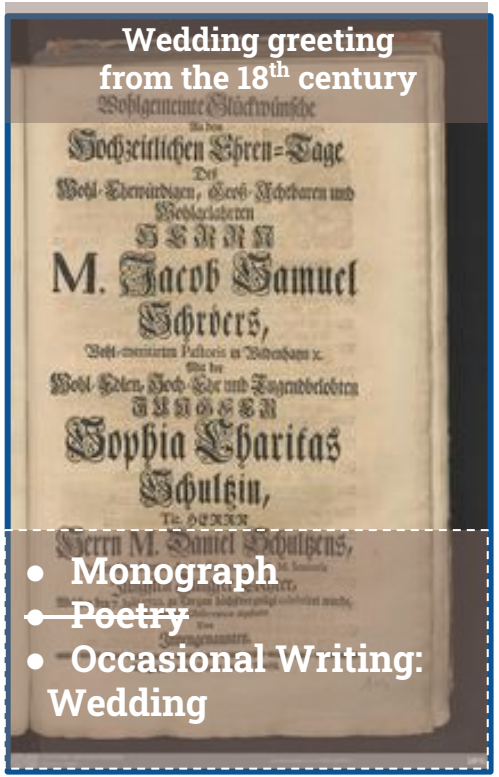
Conclusion

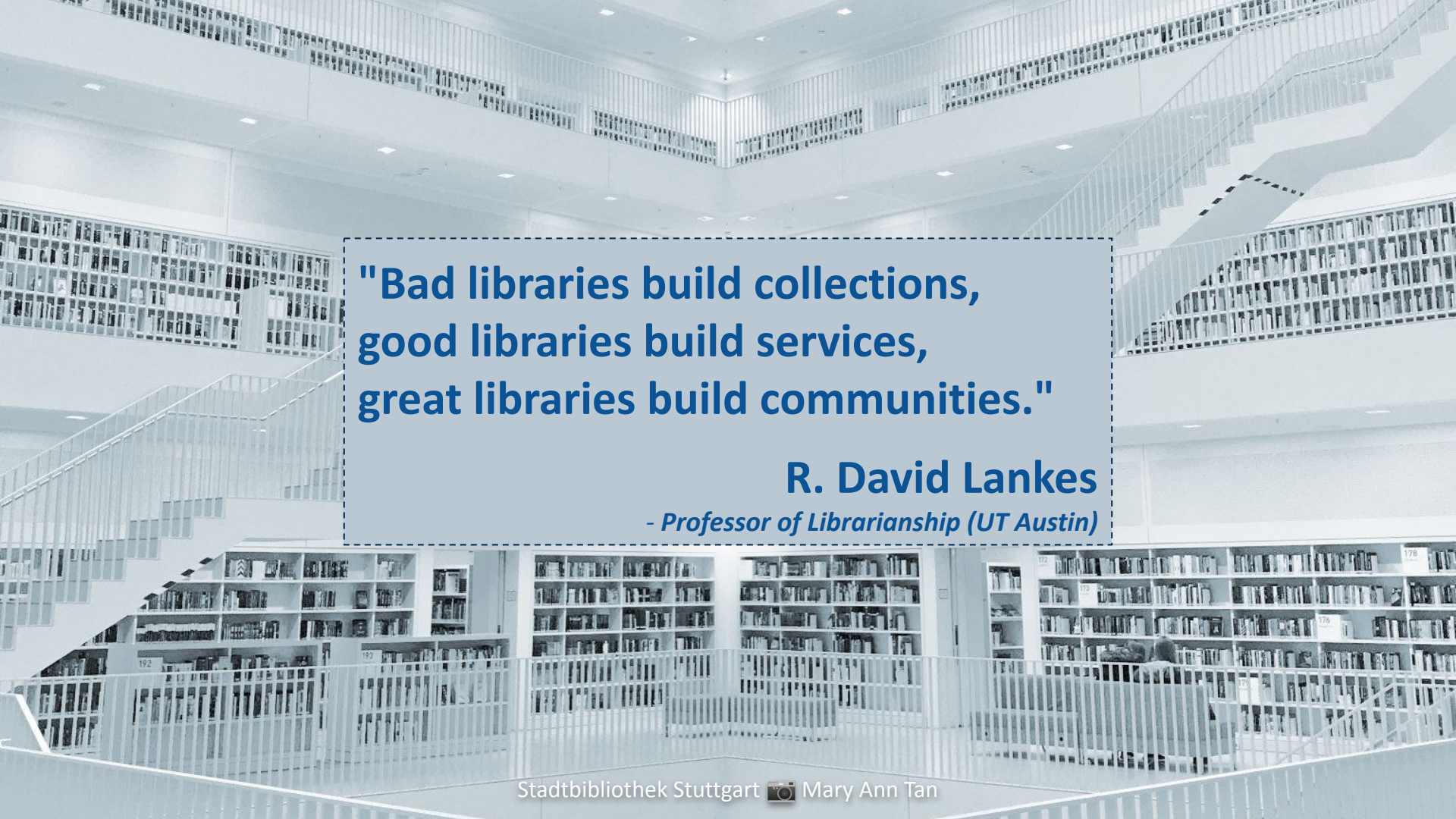
Challenges addressed by this study:

1. Heterogeneity
   - ☞ library-specific ontology (FaBiO) with extensive sub-classes for object types
2. Granularity
   - ☞ distinction between primary and secondary objects, using DoCO.
3. Lack of exploration
   - ☞ encoding of bibliographic relationships, linkage to external resources through KG.
4. Unnormalized keywords
   - ☞ integration with linked open data
5. Inefficient retrieval
   - ☞ search on specific abstraction levels

Coming Soon...

- ❑ Linkage to external resources (Wikidata, DBPedia, VIAF, Getty)
- ❑ Other sectors: Museums, Archives, Film Museum, Research.
- ❑ Upper ontology for interoperability across sectors.
- ❑ Metadata quality assessment.
- ❑ Metadata data enrichment.
- ❑ Integration of DDB-KG with the current portal.




"Bad libraries build collections,
good libraries build services,
great libraries build communities."
R. David Lankes
- Professor of Librarianship (UT Austin)
Stadtbibliothek Stuttgart 📷 Mary Ann Tan

References

T. Berners-Lee, J. Hendler, O. Lassila, The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities, ScientificAmerican.com (2001).

V. A. Carriero, A. Gangemi, M. L. Mancinelli, L. Marinucci, A. G. Nuzzolese, V. Presutti, C. Veninata, ArCo: The Italian Cultural Heritage Knowledge Graph, The Semantic Web – ISWC 2019 (2019) 36—52.

E.Hyvönen,Semantic Portals for Cultural Heritage,in:S.Staab,R.Studer(Eds.), Handbook on Ontologies, International Handbooks on Information Systems, Springer, 2009, pp. 757–778.

E. Hyvönen, K. Viljanen, J. Tuominen, K. Seppälä, Building a national semantic web ontology and ontology service infrastructure –the finn onto approach, in: S. Bechhofer, M. Hauswirth, J. Hoffmann, M. Koubarakis (Eds.), The Semantic Web: Research and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 95–109.

S. Peroni, D. Shotton, FaBiO and CiTO: Ontologies for describing bibliographic resources and citations, Journal of Web Semantics 17 (2012) 33–43.

B. Schatz, Information Retrieval in Digital Libraries: Bringing Search to the Net, Science, Vol 275, Issue 5298, 1997, pp. 327-334.

O. Suominen, N. Hyvönen, From MARC Silos to Linked Data Silos, in: Semantic Web in Libraries, Bonn, Germany, 2016.

T. Tietz, J. Jäger, J. Waitelonis, H. Sack, Semantic annotation and information visualization for blog posts with refer., in: VOILA@ ISWC, 2016, pp. 28–40.

B. Tillet, What is FRBR?: A Conceptual Model for the Bibliographic Universe, 2004.

Images

Pp. 2 : A woman using the card catalog at the main reading room of the Library of Congress, circa 1940. Photo by Jack Delano

Pp. 3 : Digital Libraries. Photo by Gerd Altmann, Pixabay

: Fig 1. Rough Timeline of Information Retrieval in Digital Libraries (Schatz, 1997)

Pp. 7 : Portrait of Friedrich Schiller from Ludovike Simanowiz (1794). Source: Wikipedia

Pp. 8 : (1) Manuscript of Johann Wolfgang von Goethe, 19.January 1804, Jenaische Allgemeine Literaturzeitung

(2) "Der Vogen im Walde", sang by Vera Schwarz, composed by Taubert Wilhelm

(3) A photo by Jörg P. Anders of a model done by Vicenzo Pacetto of Barberini Faun used for reconstruction.

(4) A photo of Jägerhof Hotel

Pp. 11 : From MARC Silos to Linked Data Silos (Suominen & Hyvönen, 2016), pp. 6.

Pp. 12 : Dune - 5 books, 5 different editions, posted by carrot926 @ Reddit

Pp. 14/22 : Congratulatory greeting on the occasion of wedding of M. Jacob Samuel Schröers and Sophia Charitas Schultzin on 7. July 1722

Pp. 15/19 : Dritter Aufzug (3rd Act of The Robbers by Friedrich Schiller)

Pp. 18 : Trauerspiele von Friedrich Schiller (Friedrich Schiller's Dramas)

Pp. 23 : Stadtbibliothek Stuttgart (Municipal Library of Stuttgart). Photo by Mary Ann Tan

Thank you for your attention!

DDB : https://www.deutsche-digitale-bibliothek.de
DDB-KG : https://ise-fizkarlsruhe.github.io/ddbkg/
SPARQL Endpoint : https://ddbkg.fiz-karlsruhe.de
GitHub : https://github.com/ISE-FIZKarlsruhe/ddbkg

Mary Ann Tan
ann.tan@fiz-karlsruhe.de

Oleksandra Bruns
oleksandra.bruns@fiz-karlsruhe.de
@sashavses

Danilo Dessì
danilo.dessi@fiz-karlsruhe.de
@danilodess

Tabea Tietz
tabea.tietz@fiz-karlsruhe.de
@tabea_t

Jonas Oppenlaender
@duesynapse

Harald Sack
harald.sack@fiz-karlsruhe.de
@lysander07

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Mitglied der Leibniz Gemeinschaft