DDB-EDM to FaBiO:
The Case of the German Digital Library

Mary Ann Tan[1,2], Tabea Tietz[1,2], Oleksandra Bruns[1,2], Jonas
Oppenlaender[1,2], Danilo Dessì[1,2], and Harald Sack[1,2]

[1] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany
{ann.tan|firstname.lastname}@fiz-karlsurhe.de
[2] Karlsruhe Institute of Technology, Institute AIFB, Germany

Abstract. Cultural heritage portals have the goal of providing users
with seamless access to all their resources. This paper introduces initial
efforts for a user-oriented restructuring of the German Digital Library
(DDB). At present, cultural heritage objects (CHOs) in the DDB are
modeled using an extended version of the Europeana Data Model (DDB-
EDM), which negatively impacts usability and exploration. These chal-
lenges can be addressed by leveraging ontologies, and building a knowl-
edge graph from the DDB's voluminous collection. Towards this goal, an
alignment of bibliographic metadata from DDB-EDM to FRBR-Aligned
Bibliographic Ontology (FaBiO) is presented.

Keywords: Cultural Heritage · Digital Library · Ontology.

1 Introduction

The German Digital Library[3] or *Deutsche Digitale Bibliothek* (DDB) was of-
ficially launched in 2014. It is Germany's contribution to the European Union
initiative entitled "i2010," whose aim was to set up a multi-sector cultural her-
itage (CH) portal called Europeana[4]. Other than being the national aggregator
for the Europeana, the DDB is also tasked with making its valuable collection
available to the general public. DDB currently contains around 38M CHOs from
hundreds of providers belonging to 7 sectors, namely: GLAM institutions (Gal-
leries, Libraries, Archives, Museums), multimedia libraries, cultural sites, and
research institutions.

Common challenges facing CH portals are heterogeneity and representation
complexity of the CHOs. These challenges are partly addressed using Semantic
Web (SW) technologies and Linked Open Data (LOD). In particular, knowledge
graphs (KG) show significant advantages in enhancing user experience [5]. In the
current state of the DDB, users have to carefully scrutinize the search results
before they can become useful to the general public. Therefore, building a KG
out of the DDB's collection is the goal for a user-oriented restructuring of the
current implementation.

---

[3] Deutsche Digitale Bibliothek, https://www.deutsche-digitale-bibliothek.de
[4] Europeana, https://www.europeana.eu

As a Europeana aggregator, the DDB uses the European Data Model (EDM) to represent CHOs. EDM defines non-specialized classes and properties to model the metadata of CHOs. Therefore, diverse types of CHOs ranging from the conventional (manuscripts) to the highly specialized (wooden skewers), tangible (bird cage) to the intangible (bird songs), movable (porcelain) to the immovable (monuments) are modeled without distinction. This limits the portal's usability. Another drawback is that EDM allows data providers to use their preferred metadata element sets, unnormalized literals, varied vocabularies and linked datasets, further adding to the difficulty and confusion of the users. By favoring simplicity and flexibility, EDM has failed to model the inherent complexity of the CHOs and negatively impacted usability [3]. Furthermore, with the sheer amount of metadata, in the same order of magnitude as other knowledge bases such as Wikidata, the DDB is also affected by big data issues, such as scalability and veracity. To allow for integration with linked data, the DDB extends EDM (DDB-EDM) to include Dublin Core (DC) Metadata Terms (*dcterms*). But, this extension has proven to be insufficient for retrieval [4]. To model a variety of CHOs from the aforementioned sectors in a KG, existing ontologies can be used. However, to move the CHOs from DDB-EDM to a KG, alignments between DDB-EDM and existing ontologies are required. As an initial step in constructing the DDB-KG, this paper focuses on the library sector with the alignment of bibliographic metadata from DDB-EDM to FRBR-Aligned Bibliographic Ontology (FaBiO) [2]. The resources resulting from this effort are publicly available[5].

2   From DDB-EDM to FaBiO

DDB-EDM. As in EDM, all CHOs are instances of the *edm:ProvidedCHO* class. For each object, its digitized versions are represented by the class *edm:WebResource*. In the DDB, digitized items from the libraries are highly granular. A book may be composed of several objects or instances of *edm:ProvidedCHO*, such as, cover, title page, chapter, section, etc. The distinction of these objects is reflected in the properties *dc:type* and *edm:hasType*, from hereon referred to as *object type terms*. **FaBiO.** FaBiO is intended "for recording and publishing bibliographic records of scholarly endeavors," and is based on the Functional Requirements for Bibliographic Records (FRBR). In FRBR, a book is not just a single concept but abstracted in several levels using *entities* connected by a particular *relationship*. For example, a book containing Friedrich Schiller's first drama, "*Die Räuber*" (The Robbers), starts out as a "distinct intellectual or artistic" creation, from hereon referred to as **Work** (*fabio:Work*). The work is linked by *fabio:hasRealization* to a *fabio:Expression*. An **Expression** is a distinct combination of forms to convey the content, in this case, words. When the words are *embodied* in a physical object, indicated by *fabio:hasEmbodiment*, the book becomes a **Manifestation** (*fabio:Manifestation*), identified by its ISBN. The copy of the book one finds on the shelf of the library, identified by its call number, is the **Item** (*fabio:Item*). A manifestation is linked to its items

---

[5] DDB Object Type Terms to FaBio Classes, https://git.io/JERRo
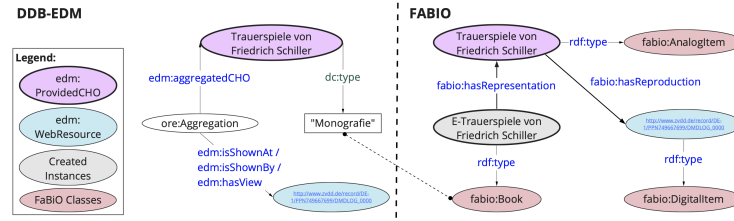
Fig. 1. Alignment from DDB-EDM to FaBiO

through *fabio:hasExemplar*. Work, Expression, Manifestation, and Item (WEMI) are collectively referred to as **Endeavors**. They are expected to improve and simplify retrieval by enabling users to search on specific levels, e.g. information at expression or manifestation level, rather than the highly granular book parts and copies. This prevents information overload while exploring. FaBiO also provides new relationships and their inverse: (1) *fabio:hasManifestation* between Work and Manifestation, (2) *fabio:hasPortrayal* between Work and Item, and (3) *fabio:hasRepresentation* between Expression and Item.

These relationships simplify representations by allowing omissions, e.g. a rare book may be described with only Item and Work instances since it has neither translations (Expression) nor editions (Manifestation). FaBiO defines extensive hierarchical sub-classes under each of the endeavors, which allows for specialization of heterogeneous objects. Moreover, FaBiO may be used in conjunction with Document Component Ontology (DoCO) to model object granularity.

Alignment. In line with the recommendation of the Europeana working group on library metadata alignment [1], all *edm:ProvidedCHO* instances are aligned to *fabio: AnalogItem*, and *edm:WebResource* as its digital reproduction. Figure 1 shows an example using "*Trauerspiele von Friedrich Schiller*"[6] (*TvFS*).

In order to model the heterogeneity of the objects, the *object type terms* are used as reference for mapping to FaBiO. All terms in the library dataset are assigned to FaBiO endeavor sub-classes manually when applicable. There is no 1-to-1 correspondence between a term and a FaBiO class; this is because a term has several conceptual equivalences: document type, document structure, production process, purpose, manifestation, and subject headings. Terms that refer to a document type are matched with existing FaBiO sub-classes on the Work and Expression levels. Those that are not assigned to a specific *Work* or *Expression* sub-class use the default superclass (See "E-Die Räuber" in Figure 2). Terms that refer to production process and manifestation determine the sub-classes under *fabio:Manifestation*. It is possible for terms to be assigned to multiple sub-classes, for instance, *Audiovisuelles Material* (Audiovisual Material), which is both a *fabio:AudioDocument* and *fabio:Movie* (sub-classes of Expression).

DDB CHOs in FaBiO. As an example, the partial mapping of "*Dritter Aufzug*"[7] is shown in Figure 2. The visual representation, including *Manifes-*

---

[6] Friedrich Schiller's Dramas, https://bit.ly/365OBHz

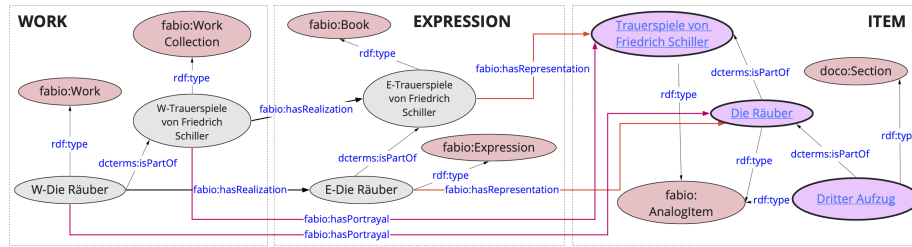[7] "*Dritter Aufzug*" (Third Act), https://bit.ly/3h5Uf2F

Fig. 2. "*Dritter Aufzug*" in FaBiO

tation instances, is hosted on GitHub[8]. "*Dritter Aufzug*" is the 3rd act of "*Die Räuber*", which is part of "*TvFS*." The *dc:type* value of "*Dritter Aufzug*" is *Abschnitt→doco:Section*, while "*TvFS*" is *Monografie→fabio:Book*. The *dc:type* for "*Die Räuber*" is "*Beigefügtes oder enthaltenes Werk*" (Contained Work). A *fabio:Work* instance for "*Die Räuber*" is created since the title is considered a Work (*Werk*) in the German Integrated Authority File aka *Gemeinsame Normdatei* (GND)[9]. On the *Work* level, "*TvFS*" is designated as a *fabio:WorkCollection* since it includes a *Work* ("*Die Räuber*").

3   Conclusion and Future Work

The task of building a KG from the rich and complex collection of the DDB is fraught with challenges that can been addressed by adapting sector-specific ontologies. This paper introduces the alignment of bibliographic metadata from DDB-EDM to FaBiO to make retrieval simpler and more efficient through the Endeavors, their sub-classes, and additional relationships. This alignment may serve as a reference for other CH portals that are currently using EDM. Rigorous evaluation will be conducted to assess scalability and efficiency of the DDB-KG implementation. Finally, further investigations will be conducted to formulate a top-level ontology enabling interoperability across objects from different sectors, making access and exploration of German cultural heritage effortless for both experts and the general public.

References

1. Angjeli, A., Bayerische, M., et al.: D5.1 Report on the alignment of library metadata with the European Data Model (EDM) Version 2.0. Tech. rep., Europeana (2012)
2. Peroni, S., Shotton, D.: FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Journal of Web Semantics **17**, 33–43 (12 2012)
3. Peroni, S., Tomasi, F., Vitali, F.: Reflecting on the Europeana Data Model. In: IRCDL 2012. pp. 228–240 (2012)
4. Schulze, F.: Objektbeschreibende Metadaten in der Deutschen Digitalen Bibliothek. Master's thesis, Humboldt-Universität, Berlin (2020)
5. Tietz, T., Waitelonis, J., Zhou, K., Felgentreff, P., Meyer, N., Weber, A., Sack, H.: Linked stage graph. In: SEMANTICS 2019 Posters & Demos (2019)

---

[8] TvFS in FaBiO, https://git.io/JEGpa

[9] GND is maintained by the German National Library, https://bit.ly/2SHmATp