

Mapping Pica+ to BIBFRAME: first observations

Background in a nutshell

In the German National Library, we keep the distinction between the internal format (Pica +), the cataloging format (Pica3) and several exchange formats (MARC 21, RDF, specific formats such as ONIX, etc.). The highest level of granularity and flexibility is given in the internal format, it serves as a hub ("turntable") - therefore mappings are created on the basis of Pica + (see the field directory for bibliographic data¹). Thus, a mapping from the DNB internal format to the BIBFRAME model has been started to explore the possibilities and limits of the new approach. We aim to represent the DNB data in BIBFRAME and - as a further step - make them available as a part of the DNB Linked Data Service.

In 2009 we started to work on providing the DNB data as Linked Data. Since 2010, the Integrated Authority File (GND²) is published as Linked Data. The Linked Data Service of the DNB has been extended to bibliographic data in January 2012. Since then, the DNB data are available under the license Creative Commons Zero in RDF.

The modelling of the core element set follows the recommendations of the DINI AG KIM Working Group on Bibliographic Data³. This working group consists of representatives from library networks and some single libraries in the German speaking countries. It has the goal of achieving harmonization between RDF representations of bibliographic data in the German speaking countries. In the DNB Linked Data Service, existing RDF vocabularies were re-used (see the website of the DNB-Linked Data Service⁴).

In the first draft of the Pica +/ BIBFRAME-mapping, we tried to map those elements to the BIBFRAME vocabulary that already exist in the DNB Linked Data representation. It was continuously documented in which cases a mapping is not a 1:1 match without loss of information, which further requirements we impose on the specification of the vocabulary, and where ambiguities exist.

The early observations fall into the following categories:

- Requirements of DNB data
- Suggestions and feature requests for the website bibframe.org
- Questions / comments regarding the vocabulary
- Questions / comments regarding the modeling
- General questions / comments

¹ https://wiki.dnb.de/download/attachments/43090988/titeldaten_dnb.pdf?version=2&modificationDate=1369848387000 (in German only)

² http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html

³ <https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe> (in German only)

⁴ http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html

Requirements of DNB data

Missing work level in the DNB data: With the exception of DMA bibliographic data, currently there are no records for the work level in DNB. However, the BIBFRAME model assumes the work level. Therefore for the first prototypical implementation a provisional solution was chosen to generate a temporary work level to each instance. The work-specific information is extracted from the bibliographic record in this case.

Example:

```
<rdf>
<bf:instance about="http://d-nb.info/1028889461">
  <bf:instanceOf resource="http://d-nb.info/_bf_temp/work_1028889461">
</bf:instance>

<bf:work about="http://d-nb.info/_bf_temp/work_1028889461">
  <bf:hasInstance resource="http://d-nb.info/1028889461">
</bf:work>
</rdf>
```

Suggestions and feature requests for the website bibframe.org

- **Search** inside the vocabulary (free text, by category, with wildcards, etc.)
- **Definitions:** Most BIBFRAME elements are listed without definitions in the vocabulary, leading to confusion regarding the use.
Example: <http://bibframe.org/vocab/local.html>
- **Deleted elements** should still be available, and should be given a notice (including period of validity) that the element is deprecated and should no longer be used.
Example: <http://bibframe.org/vocab/pubdate.html>

Questions / comments regarding the vocabulary

- **Title of the work:** Why do we have the element bf:uniformTitle? If it is identical to the title of the work , bf:title in the context of a work would be sufficeint.
- **Language of the work:** what is the difference between <http://bibframe.org/vocab/language.html> (used with work) and <http://bibframe.org/vocab/languageOfWork.html> ?
- **Expressed as and Expression of:** is this pair intended to accommodate the FRBR expression level? How would a designation that a bf:work corresponds to the FRBR work-level exactly look like?
- **bf:ProviderEntity:** Is the use of bf:ProviderEntity-subproperties allowed without the ProviderEntity-Bnode?
- **Data provenance:** the description of the bibliographic record should be clearly distinguished from the bibliographic description of the title itself.
- **Accompanying material:** We miss an element for Pica+ 034K - this would be analogous to MARC 300 \$e "Accompanying material".
Proposal (bfp = BIBFRAME proposed): bfp: accompanyingMaterial

- **ISBN:** The vocabulary provides a distinction between ISBN13 and ISBN10. Currently the URIs do not solve ("oops"). Is a distinction necessary or would it be sufficient to create a common element bf:isbn?
- **ISSN:** BIBFRAME should distinguish between an authorized ISSN, ISSN of the item in hand, etc.
- **Notation of the elements:** Most elements are in Camelback notation. Some are listed with hyphens instead (eg. publisher-name). Here we wish to have a unified naming convention.
- **Inconsistencies with current versions of Point Papers:**
The examples and elements of the vocabulary on bibframe.org do not always correspond to the drafts of the point papers. Especially with regard to the listed examples, this is confusing.

Questions / comments regarding the modeling

- **Types of work:**
Regarding <http://bibframe.org/vocab/Work> : The types of work are not disjoint, also the types of sub-types (i.e. in Language Materials: Book vs. dissertation.). Is it meant to assign them more than once?
- **sameAs:**
Will BIBFRAME provide general properties such as sameAs, or is it allowed to use elements as owl:sameAs or umbel:isLike in a BIBFRAME description? This is needed to refer cases like the culturegraph⁵-URI and 006Z (ZDB identifier).
- **URIs vs. Literals:**
Often literals are expected by the current BIBFRAME vocabulary where from our point of view a URI should be recommended. At least the expected value should be expanded to include the possibility of a link using a URI. In the case of carrier type and media type corresponding types are provided in BIBFRAME vocabulary, but it is not obvious how to use them.
Examples are:
<http://bibframe.org/vocab/identifierAssigner.html>
<http://bibframe.org/vocab/carrierType.html>
<http://bibframe.org/vocab/mediaType.html>
<http://bibframe.org/vocab/frequency.html>
<http://bibframe.org/vocab/language.html>
<http://bibframe.org/vocab/languageOfWork.html>
<http://bibframe.org/vocab/modeOfIssuance.html> (RDA-Terms could be reused)
- **Data Type:**
Should the data type for a date be given, for example for Year of publication?
- **Lightweight Abstraction Layer (LAL):**
General question to the formation and incorporation of the LAL in the bibliographic description: Is a LAL expected for all controlled vocabularies? I.e. for subject headings, personal names, corporate bodies, etc.
- **Internal links:**
It is not clear whether the links from one resource to another resource 039B, 039C, 039D, 039E, 039S (MARC 76X - 78X) should refer to bf:instance or to bf:work. Obviously the model is expecting a reciprocal relation of Work / Instance with bf:hasInstance and bf:instanceOf. Is this really the intention, including all possible

⁵ <http://www.culturegraph.org>

consequences (regarding reasoning), and despite unclear / incomplete possibilities to represent them?

- **Language Information:**

Which content does bf:languageOfWork expect? Does this refer to the language of the original work (010@ \$c, MARC 041 \$h)?

What is expressed with bf:language in the context of bf:work?

Therefore, clear definitions are also needed for bf:languageOfWork and for bf:language.

As said above, the language should be expressed with a URI

- **Frequency of publication:**

In addition to bf:frequency (018@ first \$a, MARC 008 CR Pos. 18) there should also be an element such as "formerFrequency" for the second and third frequency entry. Otherwise there is no way to represent the current publication frequency.

General questions / comments

- **Focusing of the vocabulary:**

The BIBFRAME vocabulary is based on the MARC 21 structure with fields and subfields. The issue of backward compatibility to the MARC format seems to play the central role in developing the BIBFRAME vocabulary. Other formats and a format-independent approach seem to be largely ignored. The question whether the MARC data can operate the BIBFRAME format seems to be more important than the question of whether the BIBFRAME format meets the requirements of current and future information structure needs. For a transitional period, this accent is certainly understandable and pragmatic, in the interest of timely results. However, it should be stronger supplemented by the requirements of future aspects.

- **MARC codes:**

Will the MARC codes and controlled vocabularies at id.loc.gov stay maintained in the future? May BIBFRAME data be linked to MARC codes and use URIs that are based on MARC codes?

- **Alignments:**

In order to achieve interoperability between the BIBFRAME vocabulary and existing and established vocabularies, alignments to existing ontologies have to be created. What's the current status and what alignments are planned?

- **Inclusion of library-external sectors:**

The BIBFRAME model has the ambition to serve both the library community and foreign sectors. The current members of the Early Experimenters Group all come from the library community. At what time and in what manner should representatives from museums, archives, publishing institutions, etc. become involved in the development of BIBFRAME?

- **Inclusion of other non-English speaking partners:**

With the exception of the German National Library, all current members of the Early Experimenters Group come from the Anglo-American world. To strengthen internationality and language independence of the BIBFRAME initiative, other partners should be invited from countries whose native language is not English, or not exclusively English.

Appendix: Example Data

Please note that these examples are produced according to a very first mapping, which does not claim neither to contain the whole BIBFRAME-approach nor to represent a complete Pica+-record.

Example 1:

Based on: <http://d-nb.info/101777644X>

current DNB-RDF-representation: <http://d-nb.info/101777644X/about/rdf>

Instance-record, generated from the Pica+-bibliographic record:

```
<bf:Instance rdf:about="http://d-nb.info/101777644X">
  <bf:modeOfIssuance>Einbändiges Werk</bf:modeOfIssuance>
  <bf:InstanceOf rdf:resource="http://d-nb.info/bf_temp/Work_101777644X"
/>
  <bf:isbn13>9783446427020</bf:isbn13>
  <bf:isbn10>3446427023</bf:isbn10>
  <bf:ean>9783446427020</bf:ean>
  <bf:nb>11,N51</bf:nb>
  <bf:nb>13,A16</bf:nb>
  <bf:local>3934024</bf:local>
  <bf:local>
    <bf:identifizierungScheme>VLB / Netzpublicationen</bf:identifizierungScheme>
    <bf:identifizierungValue>3934024</bf:identifizierungValue>
  </bf:local>
  <bf:publisher-number>Best.-Nr.: 556/42702</bf:publisher-number>
  <bf:responsibilityStatement>Gerhard Gietl/Werner
Lobinger</bf:responsibilityStatement>
  <bf:extent>XI, 288 S.</bf:extent>
  <bf:dimensions>24 cm</bf:dimensions>
  <bf:illustrativeContentNote>graph. Darst.</bf:illustrativeContentNote>
  <bf:note>Literaturangaben. - Zusätzliches Online-Angebot unter
www.hanser-fachbuch.de/ebookinside</bf:note>
  <bf:title>Leitfaden für Qualitätsauditoren</bf:title>
  <bf:providerDate>2012</bf:providerDate>
  <bf:frequency
rdf:resource="http://marc21rdf.info/terms/continuingfre%23/u" />
  <bf:editionStatement>4. Aufl.</bf:editionStatement>
  <bf:providerName>Hanser</bf:providerName>
  <bf:providerPlace>München</bf:providerPlace>
</bf:Instance>
```

Corresponding work-record, generated from the same Pica+-bibliographic record:

```
<bf:Work rdf:about="http://d-nb.info/bf_temp/Work_101777644X">
  <bf:hasInstance rdf:resource="http://d-nb.info/101777644X" />
  <bf:Person>
    <bf:hasGNDLink rdf:resource="http://d-nb.info/gnd/123458587" />
    <bf:label>Gietl, Gerhard</bf:label>
```

```

        <bf:resourceRole
rdf:resource="http://id.loc.gov/vocabulary/relators/aut" />
        </bf:Person>
    </bf:associatedAgent>
    <bf:associatedAgent>
        <bf:Person>
            <bf:hasGNLink rdf:resource="http://d-nb.info/gnd/132115530" />
            <bf:label>Lobinger, Werner</bf:label>
            <bf:resourceRole
rdf:resource="http://id.loc.gov/vocabulary/relators/aut" />
            </bf:Person>
        </bf:associatedAgent>
        <bf:subject rdf:resource="http://d-nb.info/gnd/4061963-1" />
        <bf:class-ddc>658.4013</bf:class-ddc>
        <bf:title>Leitfaden für Qualitätsauditoren</bf:title>
    </bf:Work>

```

Example 2:

Based on: <http://d-nb.info/1028050682>

current DNB-RDF-representation: <http://d-nb.info/1028050682/about/rdf>

Instance-record, generated from the Pica+- bibliographic record:

```

<bf:Instance rdf:about="http://d-nb.info/1028050682">
    <bf:modeOfIssuance>Einbändiges Werk</bf:modeOfIssuance>
    <bf:InstanceOf rdf:resource="http://d-nb.info/bf_temp/Work_1028050682"
/>
    <bf:isbn13>9783648036570</bf:isbn13>
    <bf:isbn10>3648036572</bf:isbn10>
    <bf:ean>9783648036570</bf:ean>
    <bf:nbn>12,N48</bf:nbn>
    <bf:nbn>13,A16</bf:nbn>
    <bf:local>4183123</bf:local>
    <bf:local>
        <bf:identifizierScheme>VLB / Netzpublicationen</bf:identifizierScheme>
        <bf:identifizierValue>4183123</bf:identifizierValue>
    </bf:local>
    <bf:responsibilityStatement>von Inge Wulf ; Stefan
Müller</bf:responsibilityStatement>
    <bf:extent>467 S.</bf:extent>
    <bf:dimensions>21 cm</bf:dimensions>
    <bf:illustrativeContentNote>graph. Darst.</bf:illustrativeContentNote>
    <bf:note>Literaturangaben. - Zusätzliches Online-Angebot unter
www.haufe.de/arbeitshilfen oder über Smartphone bzw. Tablet per QR-
Code</bf:note>
    <bf:title>Bilanztraining</bf:title>
    <bf:providerDate>2013</bf:providerDate>
    <bf:frequency
rdf:resource="http://marc21rdf.info/terms/continuingfre%23/u" />
    <bf:editionStatement>14., überarb. Aufl.</bf:editionStatement>
    <bf:providerName>Haufe-Gruppe</bf:providerName>
    <bf:providerPlace>Freiburg, Br., Berlin, München [i.e.]
Planegg</bf:providerPlace>
</bf:Instance>

```

Corresponding work-record, generated from the same Pica+-bibliographic record:

```
<bf:Work rdf:about="http://d-nb.info/bf_temp/Work_1028050682">
  <bf:hasInstance rdf:resource="http://d-nb.info/1028050682" />
  <bf:associatedAgent>
    <bf:Person>
      <bf:hasGNDLink rdf:resource="http://d-nb.info/gnd/122803345" />
      <bf:label>Wulf, Inge</bf:label>
      <bf:resourceRole
rdf:resource="http://id.loc.gov/vocabulary/relators/aut" />
      </bf:Person>
    </bf:associatedAgent>
  <bf:associatedAgent>
    <bf:Person>
      <bf:hasGNDLink rdf:resource="http://d-nb.info/gnd/131723863" />
      <bf:label>Müller, Stefan</bf:label>
      <bf:resourceRole
rdf:resource="http://id.loc.gov/vocabulary/relators/aut" />
      </bf:Person>
    </bf:associatedAgent>
  <bf:subject rdf:resource="http://d-nb.info/gnd/4011882-4" />
  <bf:subject rdf:resource="http://d-nb.info/gnd/4121026-8" />
  <bf:class-ddc>657.30943</bf:class-ddc>
  <bf:title>Bilanztraining</bf:title>
</bf:Work>
```