„BIBFRAME as a data model for aggregating heterogeneous data in a search portal“
SWIB 2021

Thorsten Fritze
t.fritze@ub.uni-frankfurt.de

UB Frankfurt / Fachinformationsdienst Linguistik

December 3rd, 2021
Project overview

Lin|gu|is|tik
FACHINFORMATIONSDIENST

UB Frankfurt
- Heike Renner-Westermann
- Vanya Dimitrova
- Thorsten Fritze

Uni Frankfurt (ACoLi)
- Prof. Dr. Christian Chiarcos
- Frank Abromeit
- Christian Fäth

All logos on this page are a copyright of their respective owners and may not be re-used or altered without permission.
Project goal & history

Goal

Nationwide information supply for study and research in the field of linguistics.

History

- 1950–2015: Sondersammelgebiet 7.11
- 2012–2017: Virtuelle Fachbibliothek Linguistik (ViFa)
- Since 2017: Fachinformationsdienst
- LOD since 2015 (2nd phase of ViFa)
Services

Services (Examples)
- Licenses for corpora, journals and databases
- Open Access Publishing (OJS, repository)
- Search Portal for relevant resources

Resource types in our search portal
- Traditional library resources (~62% articles)
- Curated websites
- Electronic language resources from the LLOD cloud (corpora, electronic dictionaries, software)
Linguistic Linked Open Data Cloud

The Linguistic Linked Open Data Cloud from lod-cloud.net
BLL & the BLL Thesaurus

**BLL**

- BLL: Bibliography of Linguistic Literature
- Maintained by the UB Frankfurt since the early 1970ties
- 500,000+ bibliographic references
- Dedicated thesaurus with 9,000+ subject terms
Core idea

Linking the subject terms of the thesaurus to other vocabularies that are already available as LOD.

Target vocabularies:
- OLiA for morpho-syntactic terms
- lexvo and Glottolog for language terms
- Phoible for phonology
BLL Thesaurus

Automatically converted from library authority records to SKOS.

:bll-133108015 a owl:Class, skos:Concept ;
  rdfs:label "Kanauri"@en, "Kanauri"@de ;
  rdfs:subClassOf :BLLConcept ;
  skos:broader :bll-133122719 ;
  skos:notation "02.13.22" ;
  skos:prefLabel "Kanauri"@en, "Kanauri"@de .
BLL Ontology

Intellectually re-specified terms of the BLL Thesaurus.

```
bllt:bll-133108015
   a owl:Class ;
   skos:notation "02.13.22" ;
   rdfs:label "Kanauri"@de, "Kanauri"@en ;
   rdfs:subClassOf bllt:BLLConcept, bllt:WestHimalayish ;
```
The old search portal

- Custom software for the most part (Apache Solr & Perl/HTML::Mason)
- Proprietary adhoc data scheme
- Separated data pool for some parts (Postgresql + Lucene)
- Clickable links for some relation types (e.g. Article ↔ Journal)
- No concept of authority records (just plain strings)
What is possible ...

While the BLL Ontology can be used to retrieve electronic resources, this technology doesn’t really allow to put it to full use.
What *should be* possible ...

- Ontological search
- Use the thesaurus as a unified „terminological backbone“ through the whole portal
- Employing authority data wherever possible
- Unified data pool
- Standardized openly documented data model
Some things we need from a data model

- Articles
- Electronic resources
- Authority records for everything
- All kinds of relations

Some things we don’t need from a data model

- Cataloging
- Library holdings (for the most part)
- Special purpose properties for music etc.
Data sources

Formats
1. PICA, Marc21, MAB, MODS
2. Endnote, BibTeX
3. DC Variants, dcat
4. Custom (e.g. MS Access)
5. HTML (!)

Source types
1. Catalogue data
2. Bibliographic data
3. Online resources
4. Open access repositories
5. Electronic resources
Approach

1. Direct mapping from PICA+ to BIBFRAME
2. Direct mapping from the old proprietary data scheme to BIBFRAME
3. Re-use of the LoC Marc to BIBFRAME mapping for all Marc data
4. Porting of existing mappings to BIBFRAME over time

Method

Using an existing PICA+ to Marc21 mapping, the official documentation and other implementations (e.g. SwePub as reference points.)
Example: Work level (1)

BIBFRAME for aggregating data
Example: Work level (2)
Example: Instance level

Instance

https://data.linguistik.de/records/200180452#Instance


rdf:type  rdf:type  bf:issuance  bf:provisionActivity  bf:responsibilityStatement

bf:ProvisionActivity  bf:Publication  "2007"
Some obstacles & tricky cases

- Using BLL inside of BIBFRAME
  - Solution: bf:ClassificationBLL
- Stating issue and volume info about journal article
  - Solution: sdo:issueNumber and sdo:volumeNumber (inspired by SwePub)
- Using BLL inside of BIBFRAME
  - Solution: bf:ClassificationBLL
- Modelling articles and specific resource types
  - Solution: Custom genreForms
Some notes on technology

- Python + RDFLib
- Django + Wagtail for the website
- Apache Solr for searching
- Triple store (several; not yet decided)
Thank you for your time!
For additional information, please refer to:

- https://linguistik.de (our Portal)
- http://www.acoli.informatik.uni-frankfurt.de/resources/olia/ (OLiA)
- https://www.linguistik.de/en/about/publications/ (Links to publications about BLL-LOD and other topics)
# Addendum: BIBFRAME classes in mapping

<table>
<thead>
<tr>
<th>AdminMetadata</th>
<th>GenerationProcess</th>
<th>Multimedia</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>GenreForm</td>
<td>Nbn</td>
<td>SupplementaryContent</td>
</tr>
<tr>
<td>Audio</td>
<td>Hdl</td>
<td>Note</td>
<td>Text</td>
</tr>
<tr>
<td>Cartography</td>
<td>Instance</td>
<td>Organization</td>
<td>Title</td>
</tr>
<tr>
<td>Classification</td>
<td>Isbn</td>
<td>ParallelTitle</td>
<td>Topic</td>
</tr>
<tr>
<td>CollectiveTitle</td>
<td>Issn</td>
<td>Person</td>
<td>Urn</td>
</tr>
<tr>
<td>Contribution</td>
<td>Item</td>
<td>Place</td>
<td>VariantTitle</td>
</tr>
<tr>
<td>Dissertation</td>
<td>Lccn</td>
<td>Print</td>
<td>Work</td>
</tr>
<tr>
<td>Doi</td>
<td>Local</td>
<td>ProvisionActivity</td>
<td></td>
</tr>
<tr>
<td>Electronic</td>
<td>Manuscript</td>
<td>Publication</td>
<td></td>
</tr>
<tr>
<td>Extent</td>
<td>MixedMaterial</td>
<td>Role</td>
<td></td>
</tr>
</tbody>
</table>

**Tabelle: BIBFRAME classes in mapping**
Addendum: BIBFRAME properties in mapping

<table>
<thead>
<tr>
<th>adminMetadata</th>
<th>electronicLocator</th>
<th>issuance</th>
<th>qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>agent</td>
<td>extent</td>
<td>itemOf</td>
<td>references</td>
</tr>
<tr>
<td>assigner</td>
<td>genreForm</td>
<td>mainTitle</td>
<td>responsibilityStatement</td>
</tr>
<tr>
<td>classification</td>
<td>grantingInstitution</td>
<td>note</td>
<td>role</td>
</tr>
<tr>
<td>classificationPortion</td>
<td>hasInstance</td>
<td>noteType</td>
<td>seriesEnumeration</td>
</tr>
<tr>
<td>contribution</td>
<td>hasItem</td>
<td>part</td>
<td>seriesStatement</td>
</tr>
<tr>
<td>count</td>
<td>hasSeries</td>
<td>partName</td>
<td>source</td>
</tr>
<tr>
<td>creationDate</td>
<td>heldBy</td>
<td>partNumber</td>
<td>subject</td>
</tr>
<tr>
<td>date</td>
<td>identifiedBy</td>
<td>place</td>
<td>subtitle</td>
</tr>
<tr>
<td>degree</td>
<td>instanceOf</td>
<td>provisionActivity</td>
<td>supplementaryContent</td>
</tr>
<tr>
<td>dissertation</td>
<td></td>
<td></td>
<td>title</td>
</tr>
</tbody>
</table>

**Tabelle:** BIBFRAME classes in mapping
Addendum: Example use case: Language fact sheet

Kanauri / Kinnauri

https://data.linguistik.de/bll/bll-language-link#indiv_133108015

lexvo  http://lexvo.org/id/iso639-3/kfk
Glottolog  http://glottolog.org/resource/languoid/id/kinn1249
WALS  https://wals.info/languoid/lect/wals_code_kmn
Phoible  http://phoible.org/languages/kinn1249
Wikidata  http://www.wikidata.org/entity/Q2383208
dbpedia  http://dbpedia.org/resource/Kinnauri_language
Yago  http://yago.knowitall.com/resource/Kinnauri_language
freebase  http://rdf.freebase.com/ns/m.9s5i69
Wikimedia Commons  http://commons.wikimedia.org/wiki/Special:FilePath/A%20signboard%20in%20Kinnauri%20language.jpg
Link  https://data.linguistik.de/bll/bll-language-link#indiv_133108015

Comment: Kanauri, also known as Kanaur, Koonarw, Kanawari or Kunawar (takri: कूनार, Koonari, Koonaw or Kunawari), is a Sino-Tibetan dialect cluster centered on the Kinnaur district of the Indian state of Himachal Pradesh. Kailke, once thought to be Kinnauri, is closer to Tamangic. Bhoti Kinnauri and Tukpa, locally called Chuyuli are Bodish (Lahaul-Spiti).

Comment: Kanauri, also known as Kanaur, Koonarw, or Kunawar is a Sino-Tibetan language (Sino-Tibetan stock) spoken in the Kinnaur district of the Indian state of Himachal Pradesh.

Altlabels: Kanauri; Kinnauri language; Kalpa; Kalpa (Ka); Kanar in Sangla; Kanaur; Kanauiry Aunskad; Kanawari; Kanauri; Kanawari based on Hailey; Kanoreo (Skad); Kanoreo (Skad); Kanorin Skad; Kanorin Skad; Kanawari; Kanowari; Kanawar Skad; Kanauri; Kanawari Yanaskd; Kimori; Koonawri; Kunawari; Kunawri; Kunawar; Kunawar Lower; Lower Kanauri; Lower Kana; Malhesti; Milcham; Milcham (Kanauri); Milchana; Milchana; Milcham; Minehlang; Minehlang; Nchar (Ni); Pangi; Pungi-Rangar-Randia; Ropa (Ro); Sangla; Sangla (Sa); Sangla Kinnauri; Tukpa Bootea; Theburskud (Thebor) in Ropa Gad; Theburskud; Tibas Skad; Tibas Skad; Tukpa; Upper Kanauri