

# „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)

Funded by



Deutsche  
Forschungsgemeinschaft  
German Research Foundation

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

# Services

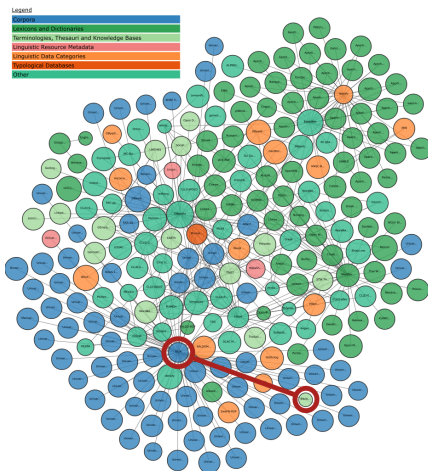
## Services (Examples)

- 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 ;
```

```
  skos:prefLabel "Kanauri"@de, "Kanauri"@en .
```

# 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

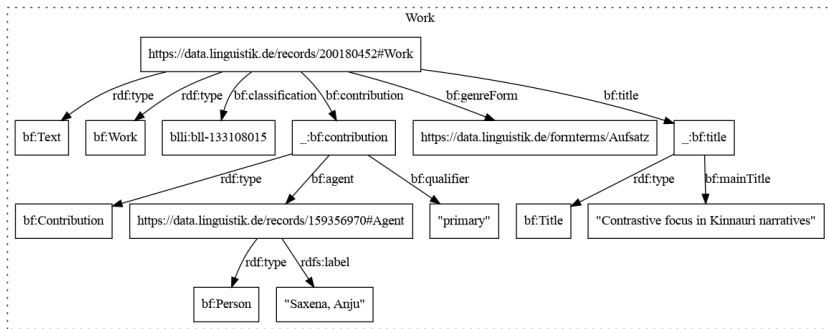
# Approach

- 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

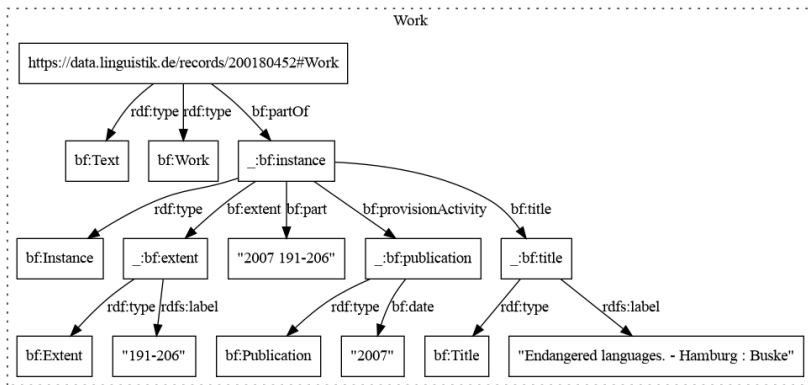
Using an existing PICA+ to Marc21 mapping, the official documentation and other implementations (e.g. SwePub as reference points).

## Example: Work level (1)

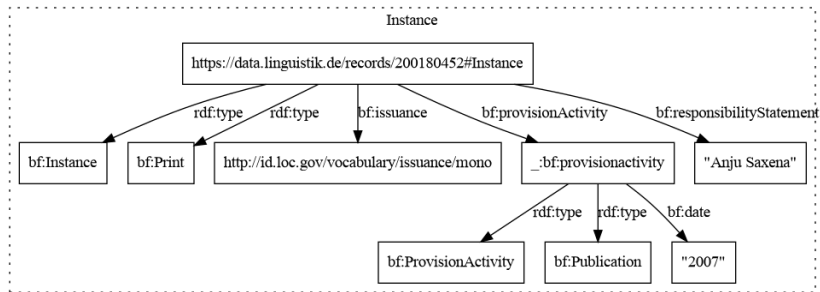




## Example: Work level (2)



## Example: Instance level



## 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://data.linguistik.de> (BLL Lod Edition)
- <https://www.linguistik.de/en/about/publications/>  
(Links to publications about BLL-LOD and other topics)

# Addendum: BIBFRAME classes in mapping

---

|                 |                   |                   |                      |
|-----------------|-------------------|-------------------|----------------------|
| AdminMetadata   | GenerationProcess | Multimedia        | Source               |
| Agent           | GenreForm         | Nbn               | SupplementaryContent |
| Audio           | Hdl               | Note              | Text                 |
| Cartography     | Instance          | Organization      | Title                |
| Classification  | Isbn              | ParallelTitle     | Topic                |
| CollectiveTitle | Issn              | Person            | Urn                  |
| Contribution    | Item              | Place             | VariantTitle         |
| Dissertation    | Lccn              | Print             | Work                 |
| Doi             | Local             | ProvisionActivity |                      |
| Electronic      | Manuscript        | Publication       |                      |
| Extent          | MixedMaterial     | Role              |                      |

---

**Tabelle:** BIBFRAME classes in mapping

# Addendum: BIBFRAME properties in mapping

---

|                       |                     |                   |                         |
|-----------------------|---------------------|-------------------|-------------------------|
| adminMetadata         | electronicLocator   | issuance          | qualifier               |
| agent                 | extent              | itemOf            | references              |
| assigner              | generationProcess   | mainTitle         | responsibilityStatement |
| classification        | genreForm           | note              | role                    |
| classificationPortion | grantingInstitution | noteType          | seriesEnumeration       |
| contribution          | hasInstance         | part              | seriesStatement         |
| count                 | hasItem             | partName          | source                  |
| creationDate          | hasSeries           | partNumber        | subject                 |
| date                  | heldBy              | partOf            | subtitle                |
| degree                | identifiedBy        | place             | supplementaryContent    |
| dissertation          | instanceOf          | provisionActivity | title                   |

---

**Tabelle:** BIBFRAME classes in mapping

## Addendum: Example use case: Language fact sheet

## Kanauri / Kinnauri

[https://data.linguistik.de/bll/bll-language-link#indiv\\_133108015](https://data.linguistik.de/bll/bll-language-link#indiv_133108015)

lexvo <http://lexvo.org/id/iso639-3/kfk>

**Glottolog** <http://glottolog.org/resource/language/id/kinn1249>

**WALS** [https://wals.info/languageid/lect/wals\\_code\\_knn](https://wals.info/languageid/lect/wals_code_knn)

**Phoible** <http://phoible.org/languages/kinn1249>

**Wikidata** http://www.wikidata.org/entity/O2383208

**dbpedia** [http://dbpedia.org/resource/Kinnauri\\_language](http://dbpedia.org/resource/Kinnauri_language)

**dbpedia** [http://dbpedia.org/resource/ISO\\_639:kfk](http://dbpedia.org/resource/ISO_639:kfk)

**Yago** [http://yago-knowledge.org/resource/Kinnauri\\_language](http://yago-knowledge.org/resource/Kinnauri_language)

**freebase** <http://rdf.freebase.com/ns/m.05nb9>

EU <http://publications.europa.eu/resource/authority/language/KFK>

**Wikimedia**: [http://commons.wikimedia.org/wiki/File:Rui-Porto-\(A870\)-rui-porto-B\(20\)-B\(20\)F\(970\).png](http://commons.wikimedia.org/wiki/File:Rui-Porto-(A870)-rui-porto-B(20)-B(20)F(970).png)

**Commons** <http://commons.wikimedia.org/wiki/Special:FilePath/A%208.jpg>

**Link** <https://data.linguistik.de>

**Link** <http://g.co/kg/m/05nb9>

**Comment:** Kinnauri, also known as Kanauri, Kanor, Koonawar, Kanawari or Kunawar (takri: कानौरी/कानीरि/कनौर/कुनावर),  
མི་རྒྱུད།, Tibetan:གཞིན་རྟེན།), is a Sino-Tibetan dialect cluster centered on the Kinnaur district of the Indian state of Himachal Pradesh.

Pradesh. Kaike, once thought to be Kinnauri, is closer to Tamangic. Bhoti Kinnauri and Tukpa, locally called Chhoyuli, are Bodish (Lahauli-Spiti).

**Comment:** Kinnauri, also known as Kanauri, Kanor, Koonawur, or Kunawar is a Tibeto-Burman language (Sino-Tibetan stock) spoken in the Kinnaur district of the Indian state of Himachal Pradesh.

**Altlabels:** Kanauri; Kinnairu; Kugugue; Kalpa; Kalpa (Ka); Kanai in Sangla; Kanaur; Kanauru Anuskad; Kanawari; Kanawari; Kanawari based on Bailey; Kanorenuu Skadd; Kanorenuu Skadd; Kanorin Skadd; Kanorug Skadd; Kanawari; Kanawari; Kanorug Skadd; Kinnari; Kinnaura Anuskad; Kinner; Kinner; Koonaware; Kunnaware-Milchng; Kunawari; Kunawar; Lower Kanauri; Lower Kinnauri; Lower Kinnari; Malhest; Milching; Milchan (Kanauri); Milchana; Milchaning; Milchan; Minchaning; Nichar (Ni); Pang; Pongl-Rarang-Trinda; Ropa (Ro); Sangla; Sangla (Sa); Sangla Kinnari; Takpa Boota; Theburskad (Thebor) in Ropa Gad; Theburskad; Tibas Skad; Tibas Skad; Tukpa; Umoor Kanauri