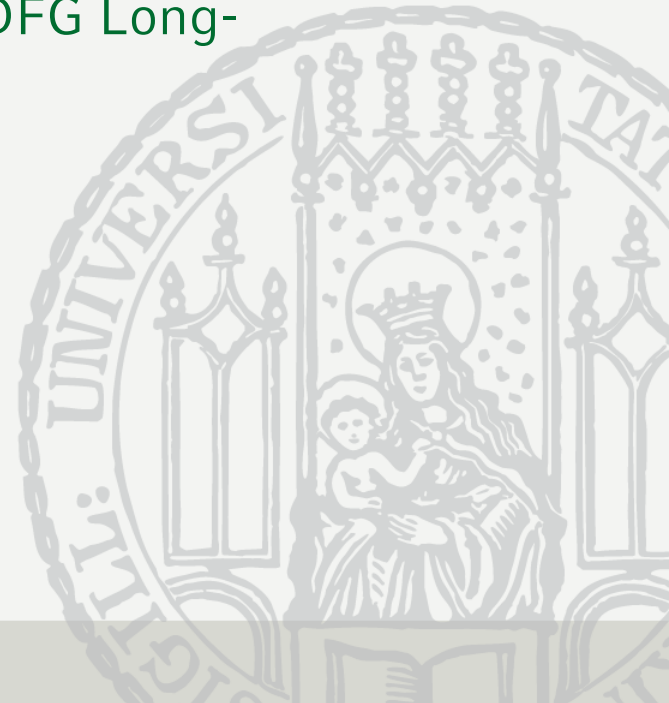


Stephan Lücke

FAIRness in Digital Corpus Linguistics

Best Practices Using the Example of the DFG Long-Term Project VerbaAlpina

Workshop at the
RTG Dimensions of Constructional Space
Friedrich-Alexander-Universität Erlangen-Nürnberg
6.7.2023 (via Zoom)



Overview

1. Introduction to VerbaAlpina
2. Our data
3. Data management
4. Online presentation

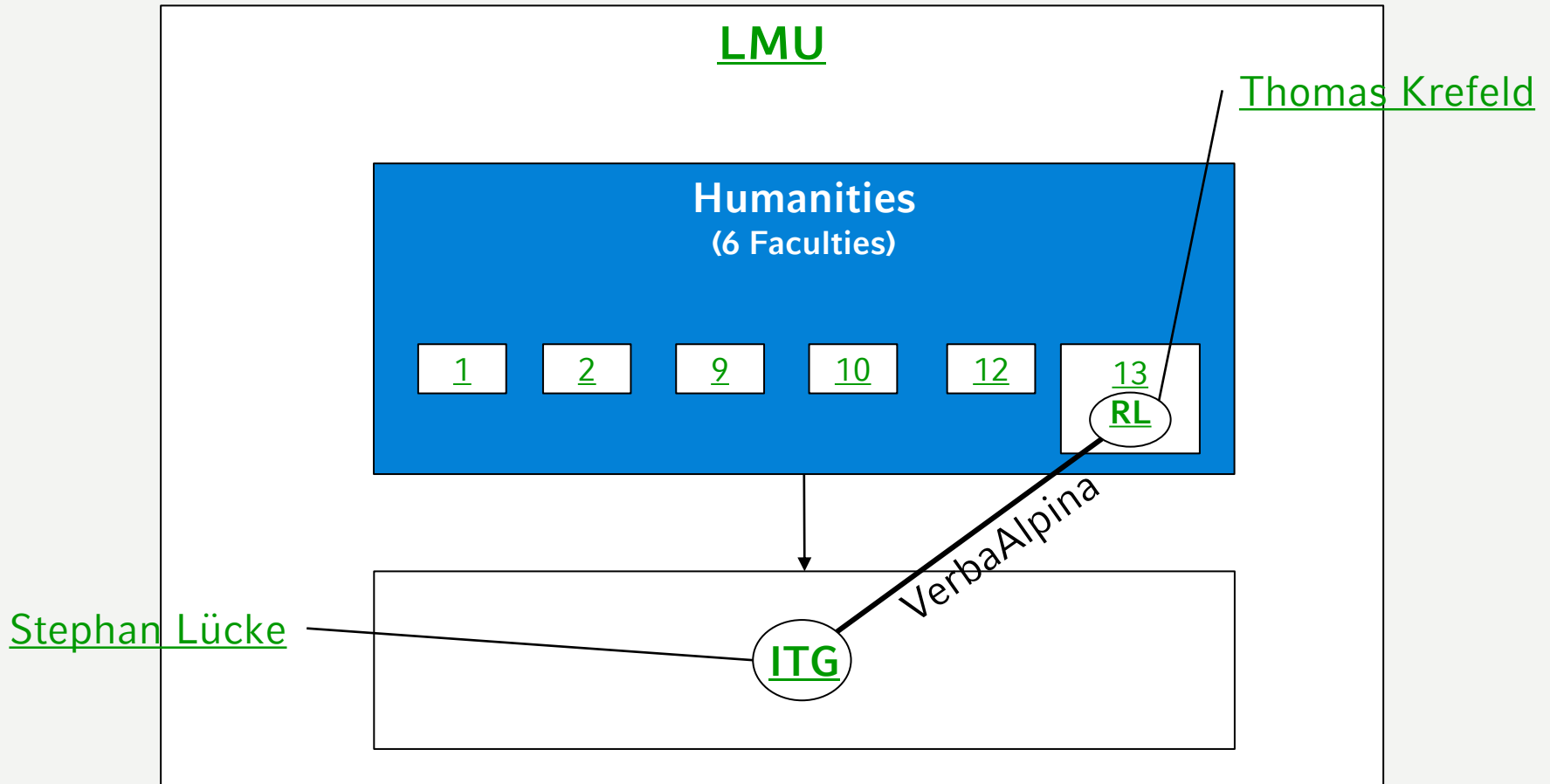
INTRODUCTION TO VERBAALPINA

- **Longterm DH-project** funded by the German Research foundation from 2014 to 2023
- Project located in the field of **geolinguistics**
- <https://www.verba-alpina.gwi.uni-muenchen.de/>

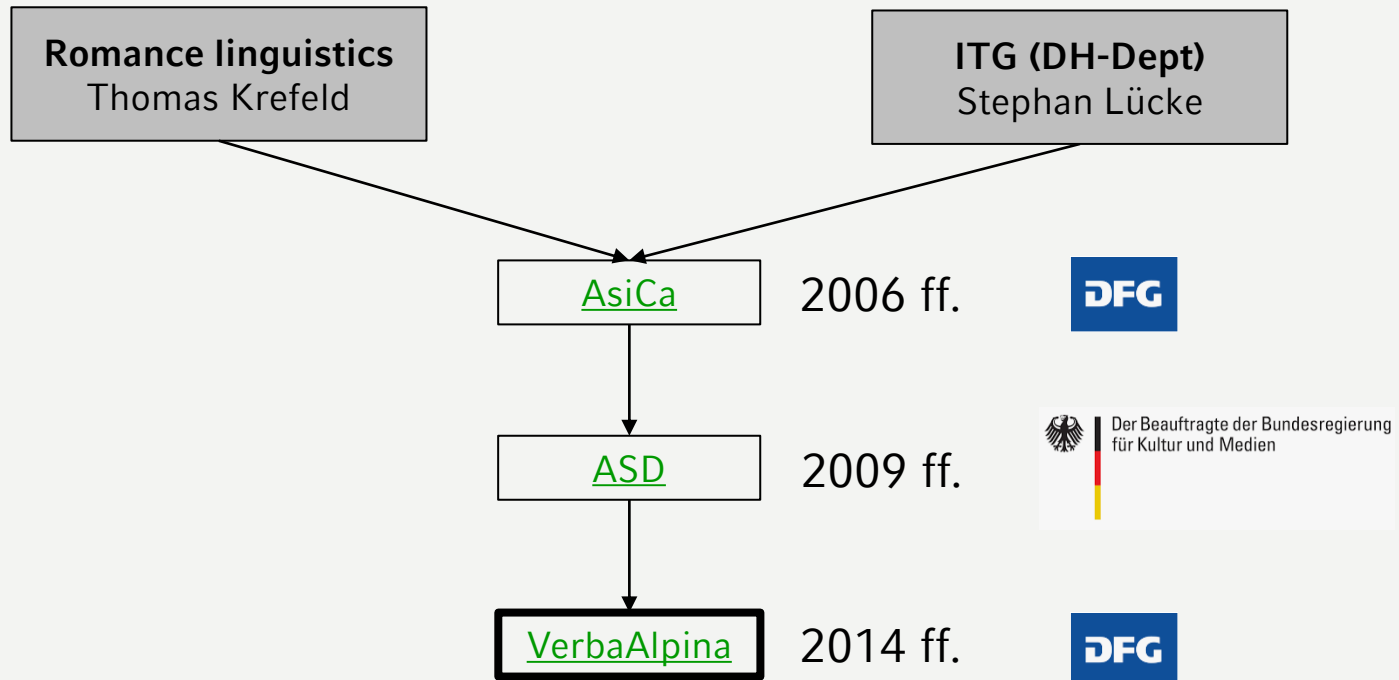
Funded by

DFGDeutsche
Forschungsgemeinschaft
German Research Foundation

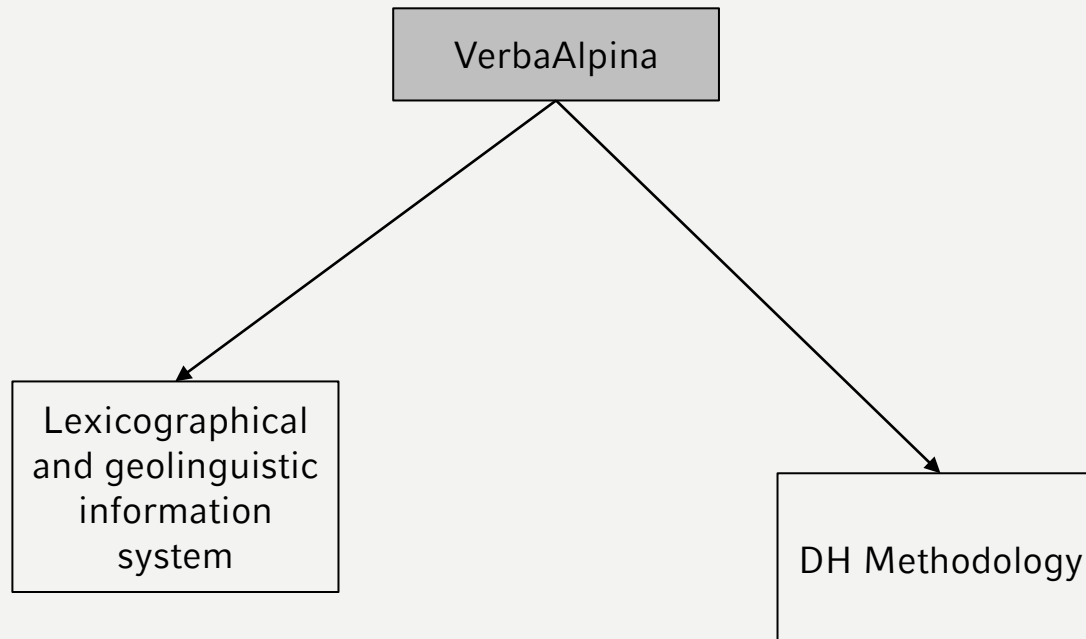
Institutional embedding



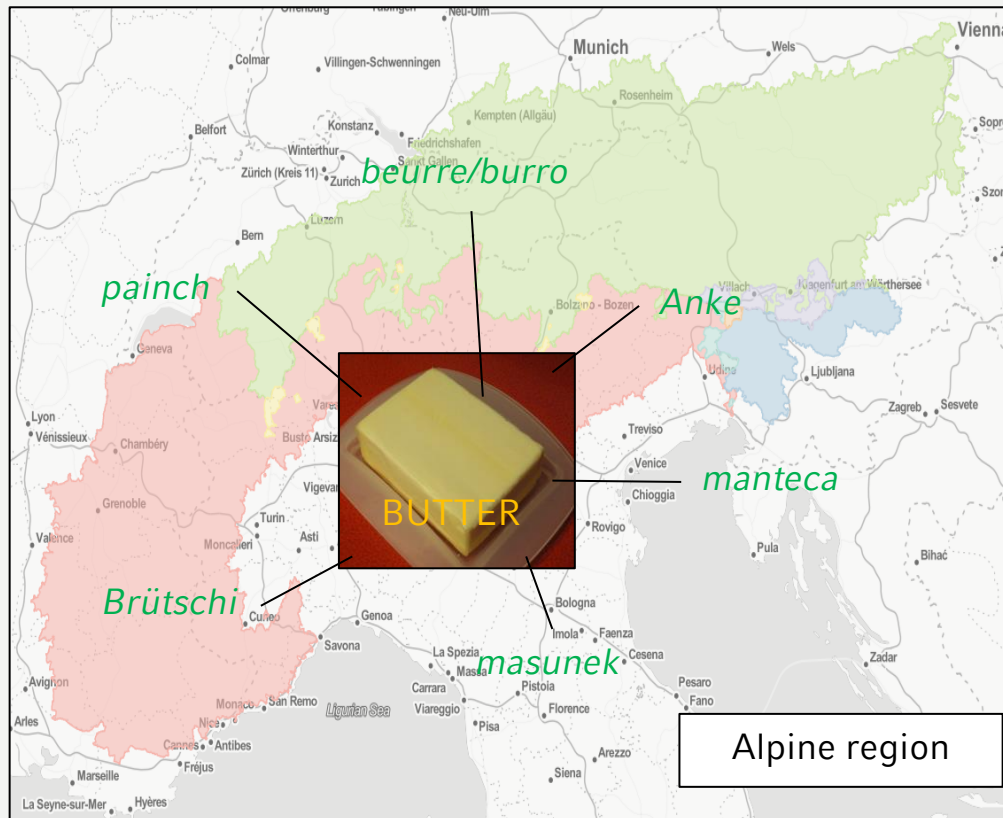
(My) major DH-Projects



Goals



Goal 1: Collection of dialect terms designating specific concepts within the Alpine region



- What dialect terms do exist?
- Important: Distinction between *terms* and **CONCEPTS**
→ *butter* != **BUTTER**
- Alpine region defined by border of so called „alpine convention“

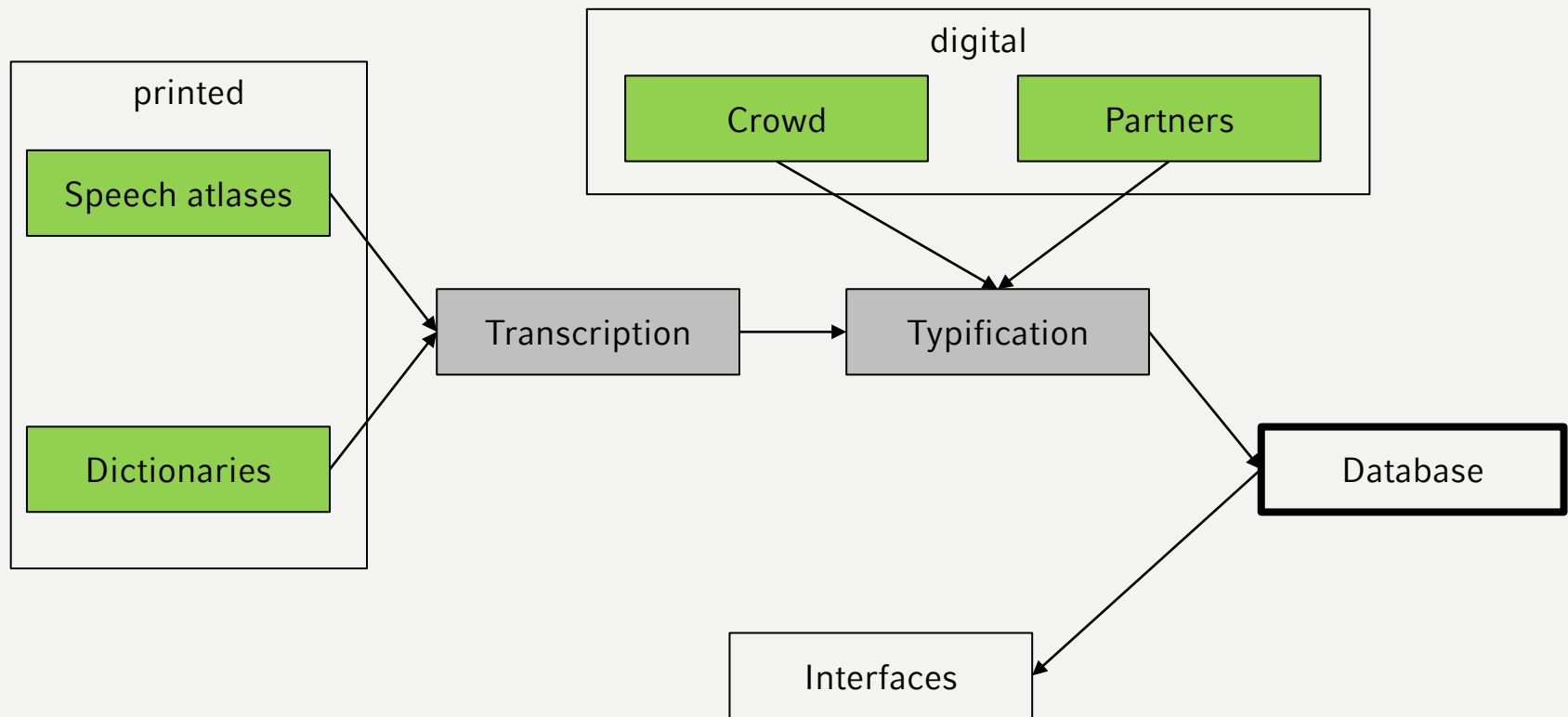
Timeline and selected concept domains

	1	2	3	4	5	6	7	8	9
2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv	i, ii, iii, iv
I Traditional environment - Alpine agriculture - Dairy processing				II Nature - Landscape formation - Weather - Fauna - Flora			III Modern life - Environment - Tourism		

NOW

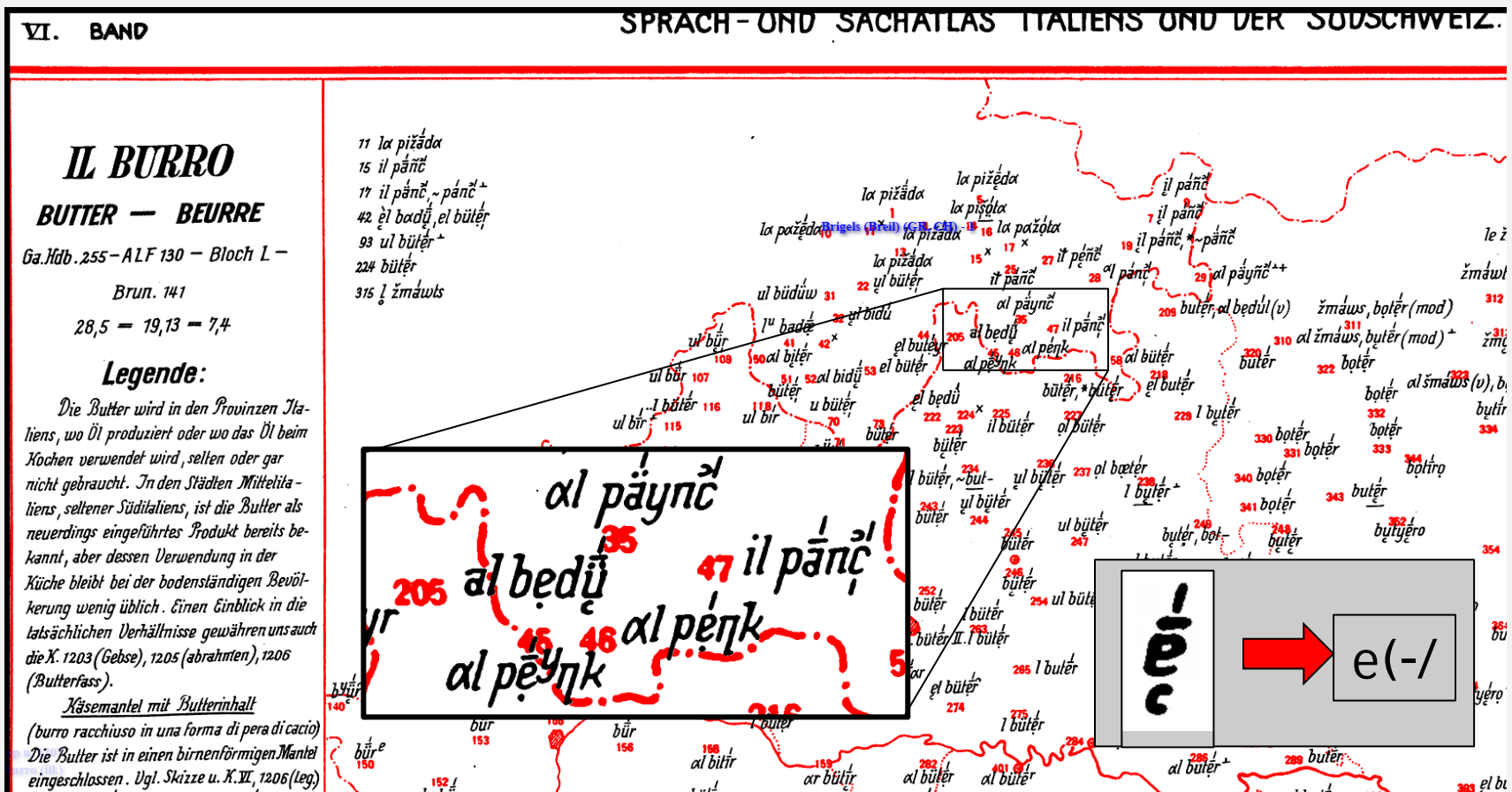
OUR DATA

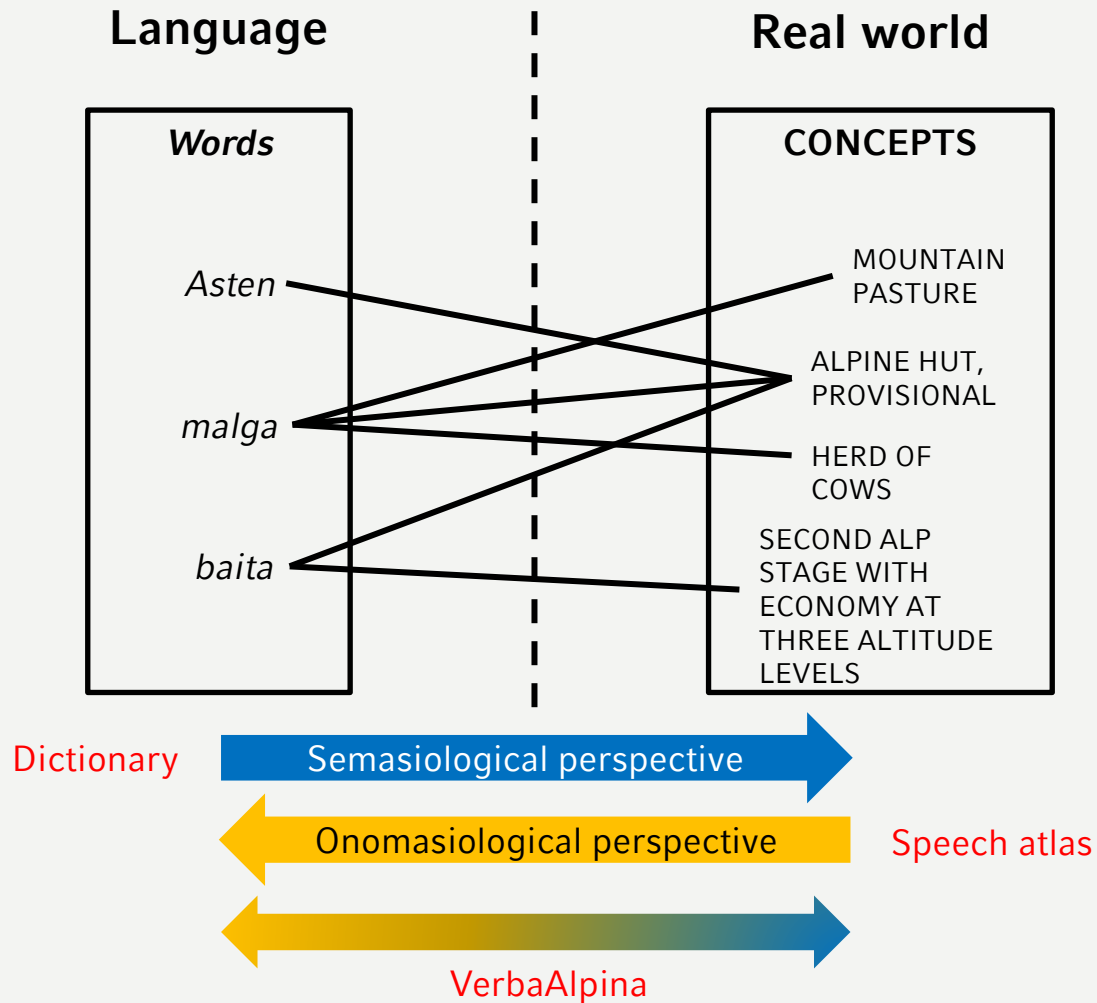
Data sources and workflow



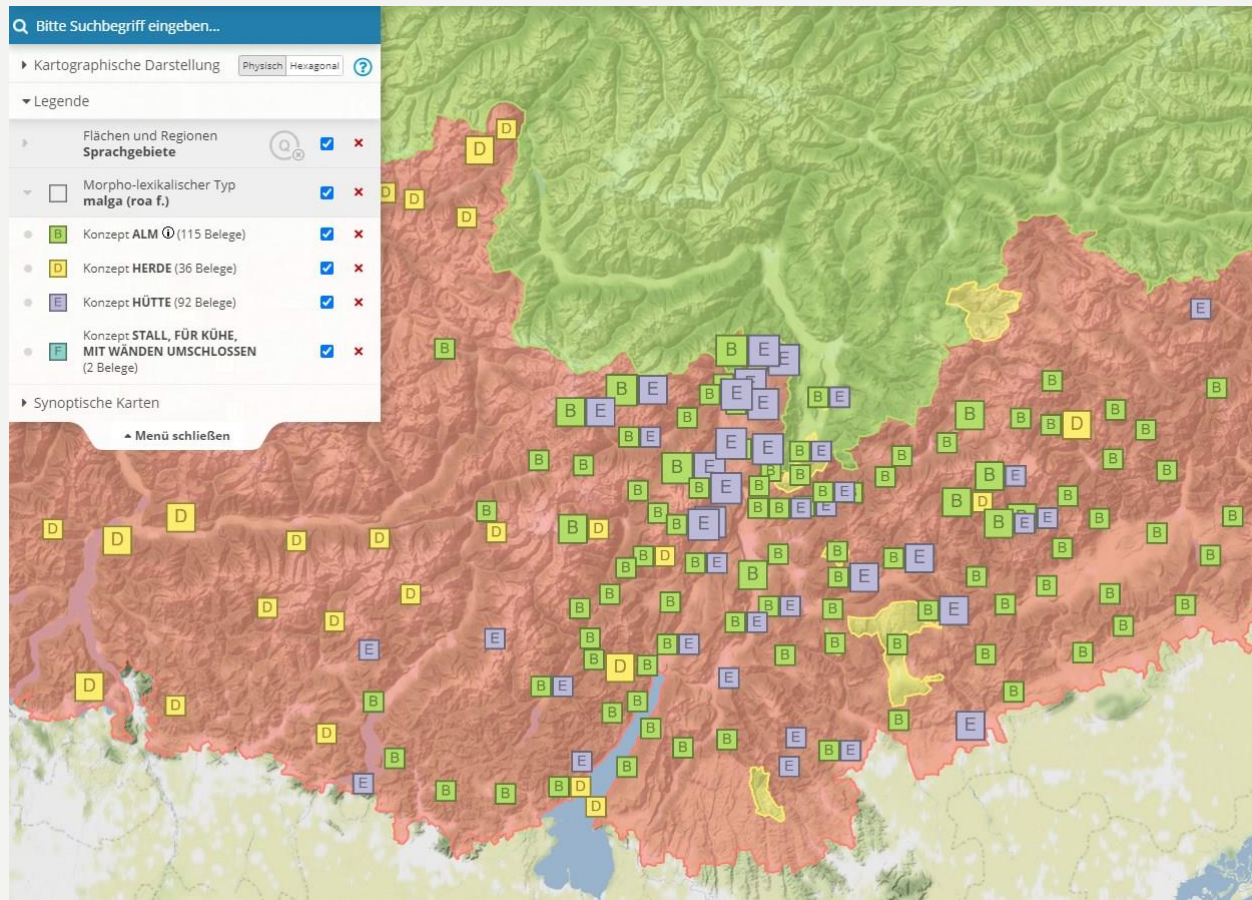
Example of a speech atlas:

The „Sprach- und Sachatlas Italiens und der Südschweiz“ („AIS“, Jaberg/Jud 1928-1940); Map 1207, „IL BURRO“



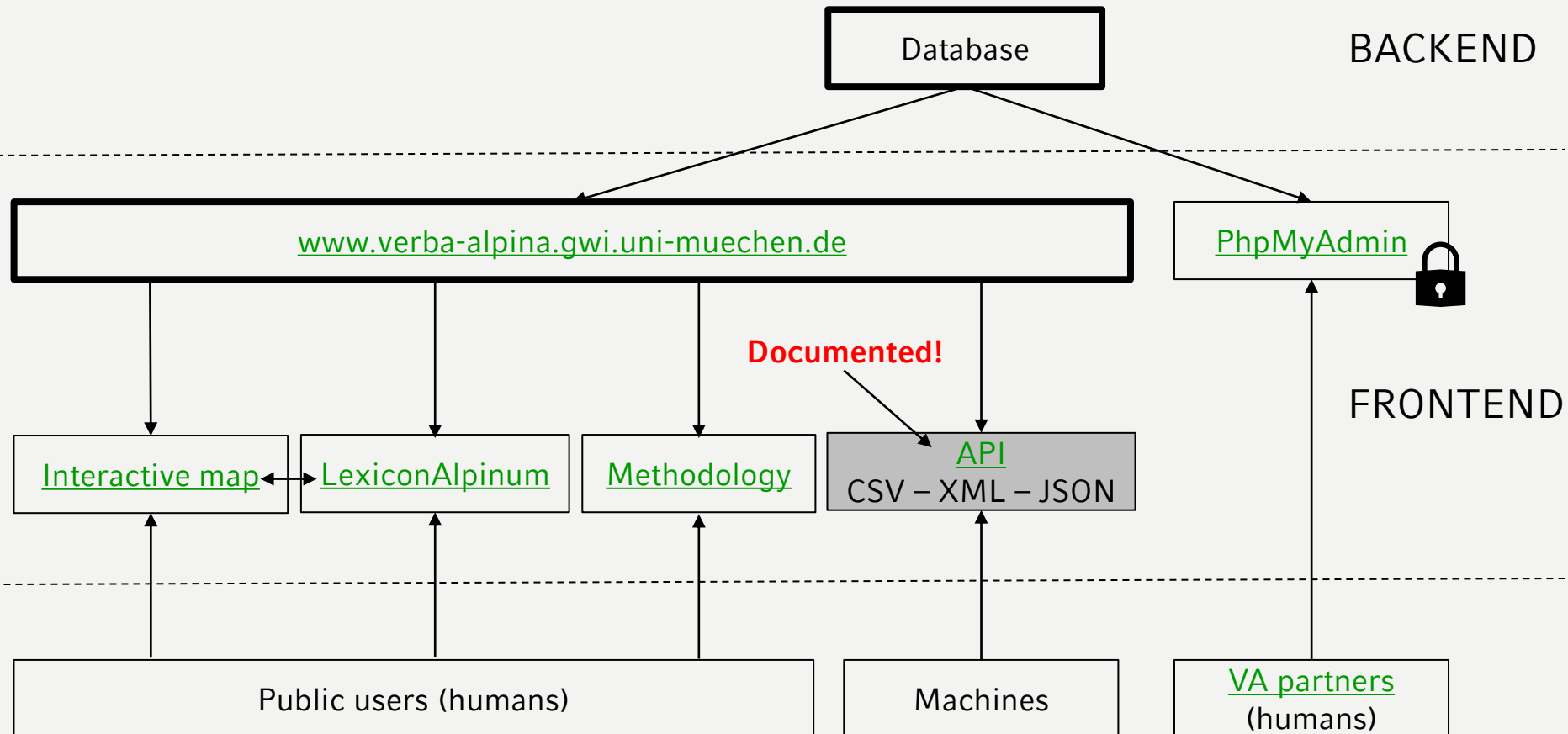


Diatopic variation of the meaning of romance term „la malga“



Geovariation von malga (roa. f.) (https://www.verba-alpina.gwi.uni-muenchen.de?page_id=133&db=xxx&tk=3466)

Access to our data: VerbaAlpina interfaces



Used technologies

BACKEND

MySQL

Database

FRONTEND

WordPress

PHP

JavaScript

CSS

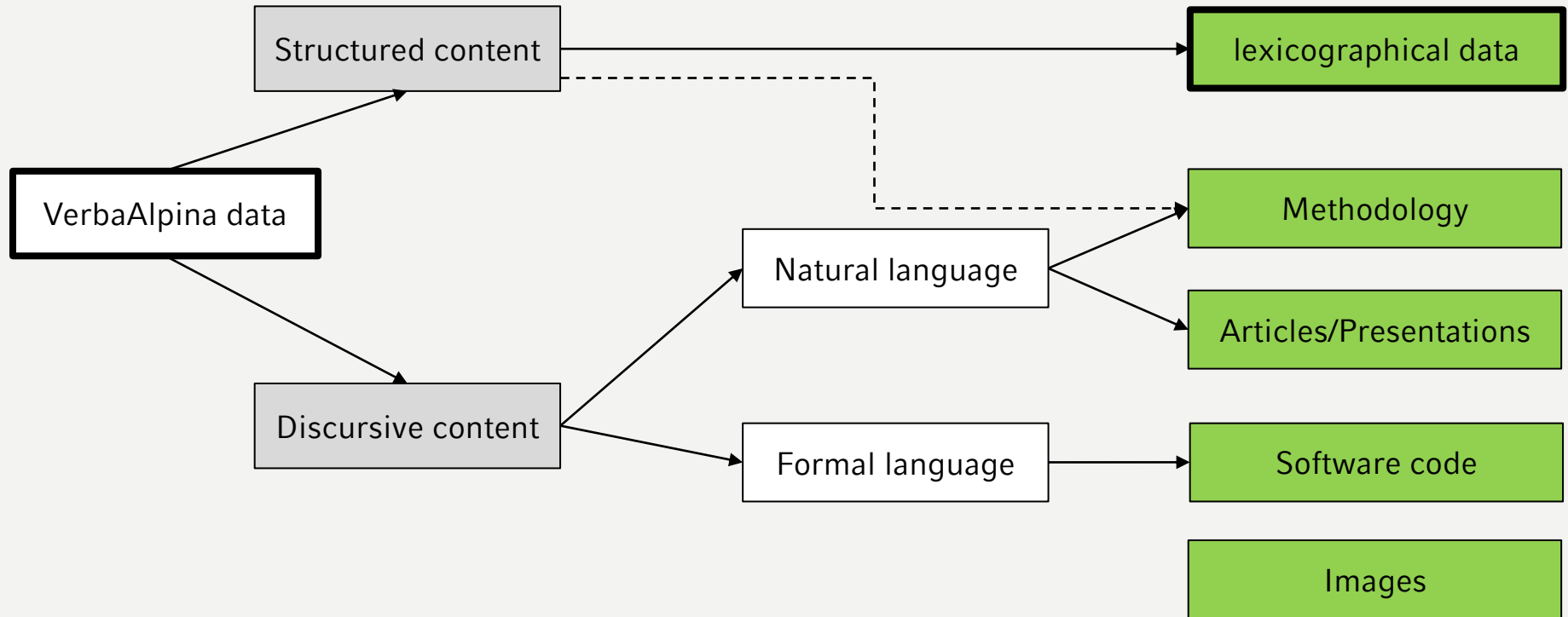
WebGL

Webinterface

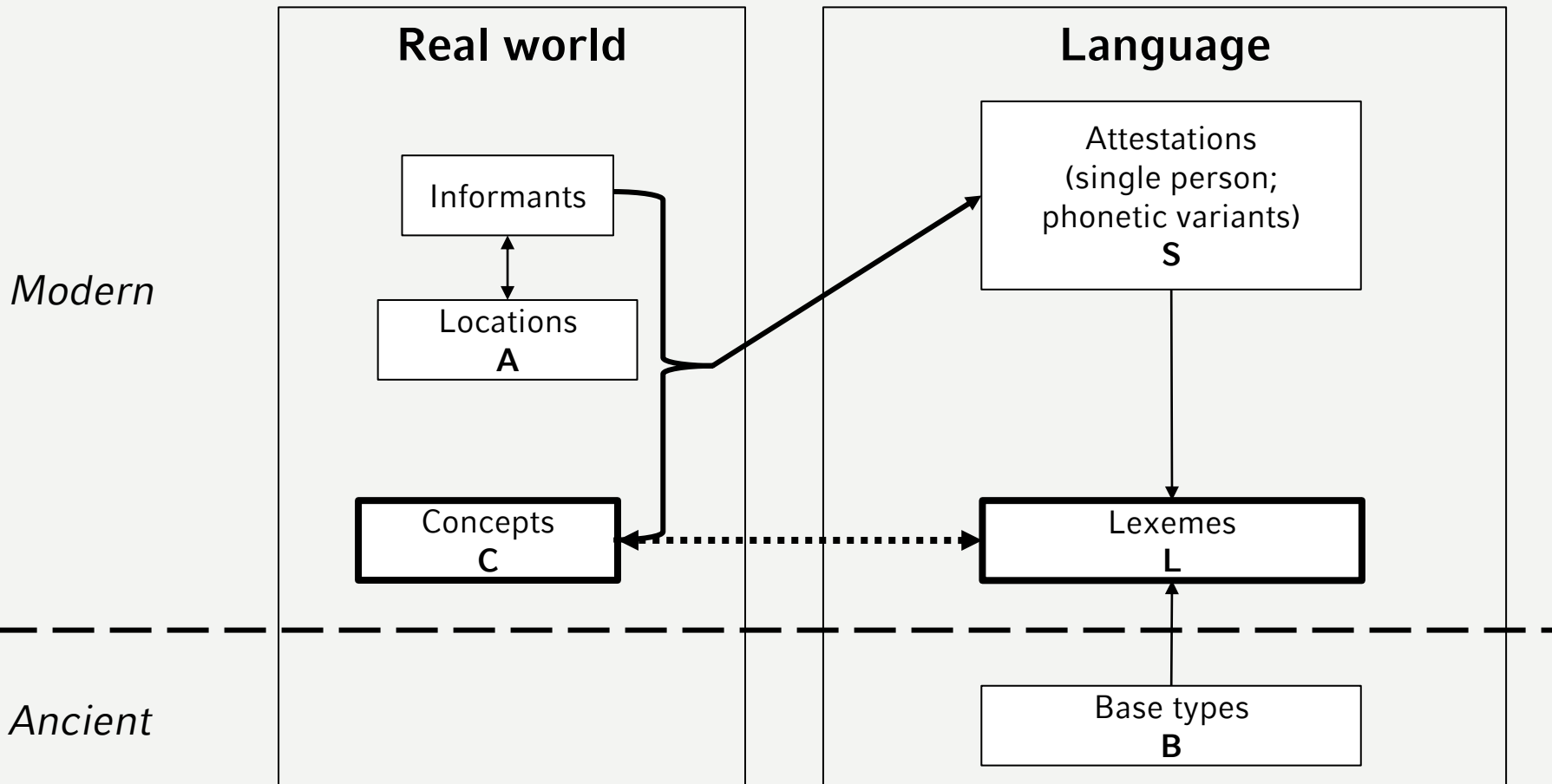
map

VerbaAlpina data

Research data



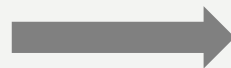
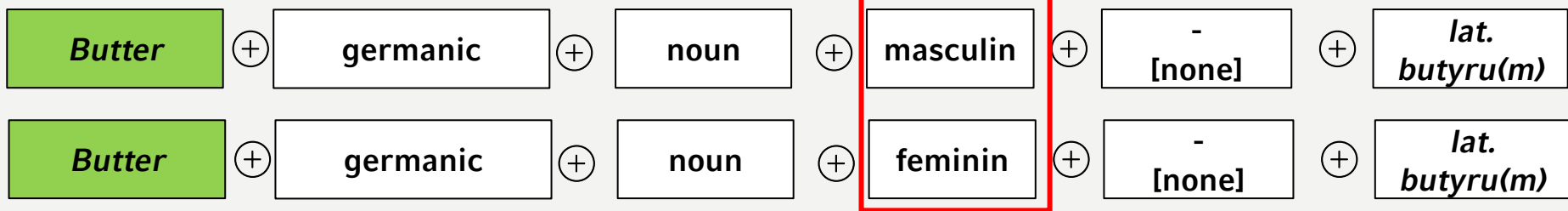
Structured data: lexicon



Morpholexical type („lexemes“): Definition

Orthography ⊕ Language family ⊕ Part of speech ⊕ Gender ⊕ Affigation ⊕ Base type

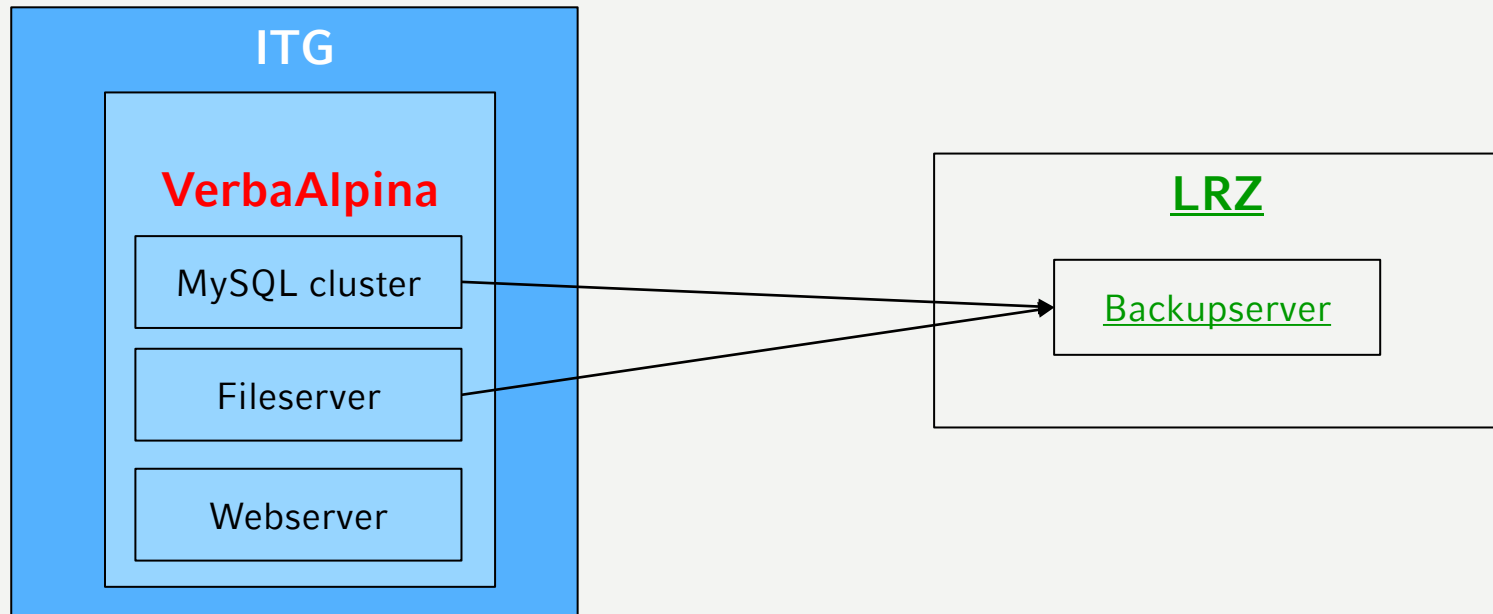
Example



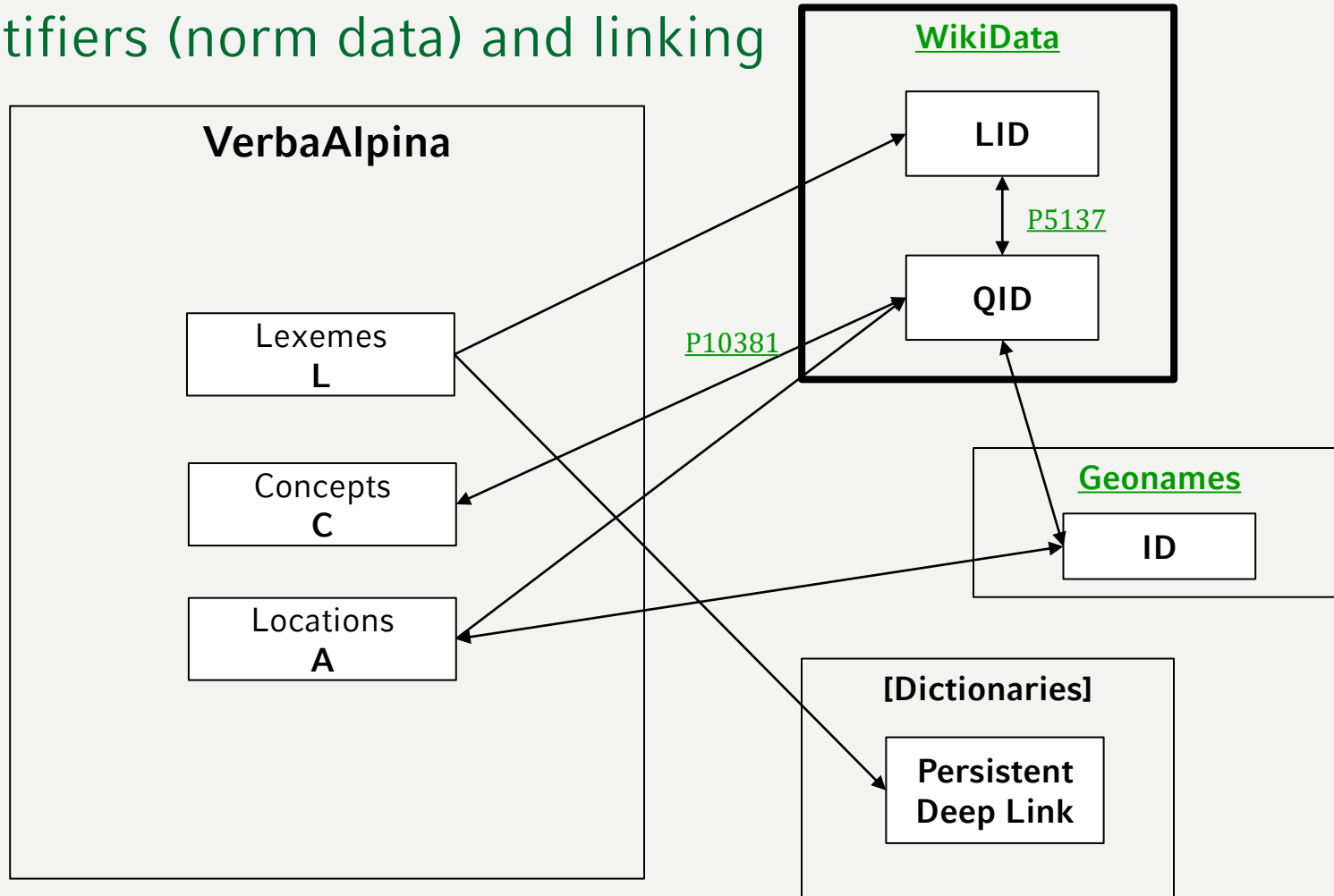
Two different lexemes

DATA MANAGEMENT

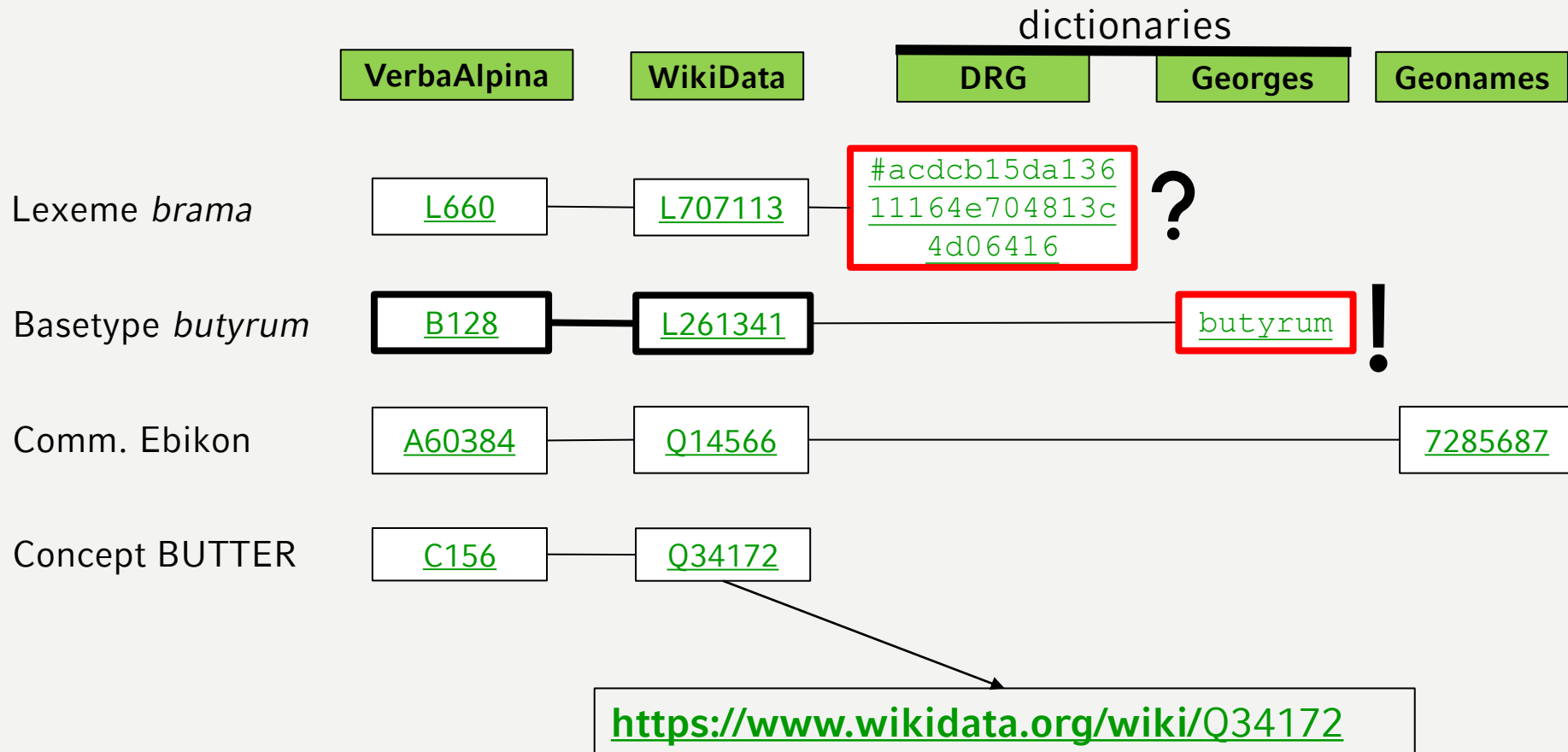
(technical) Infrastructure



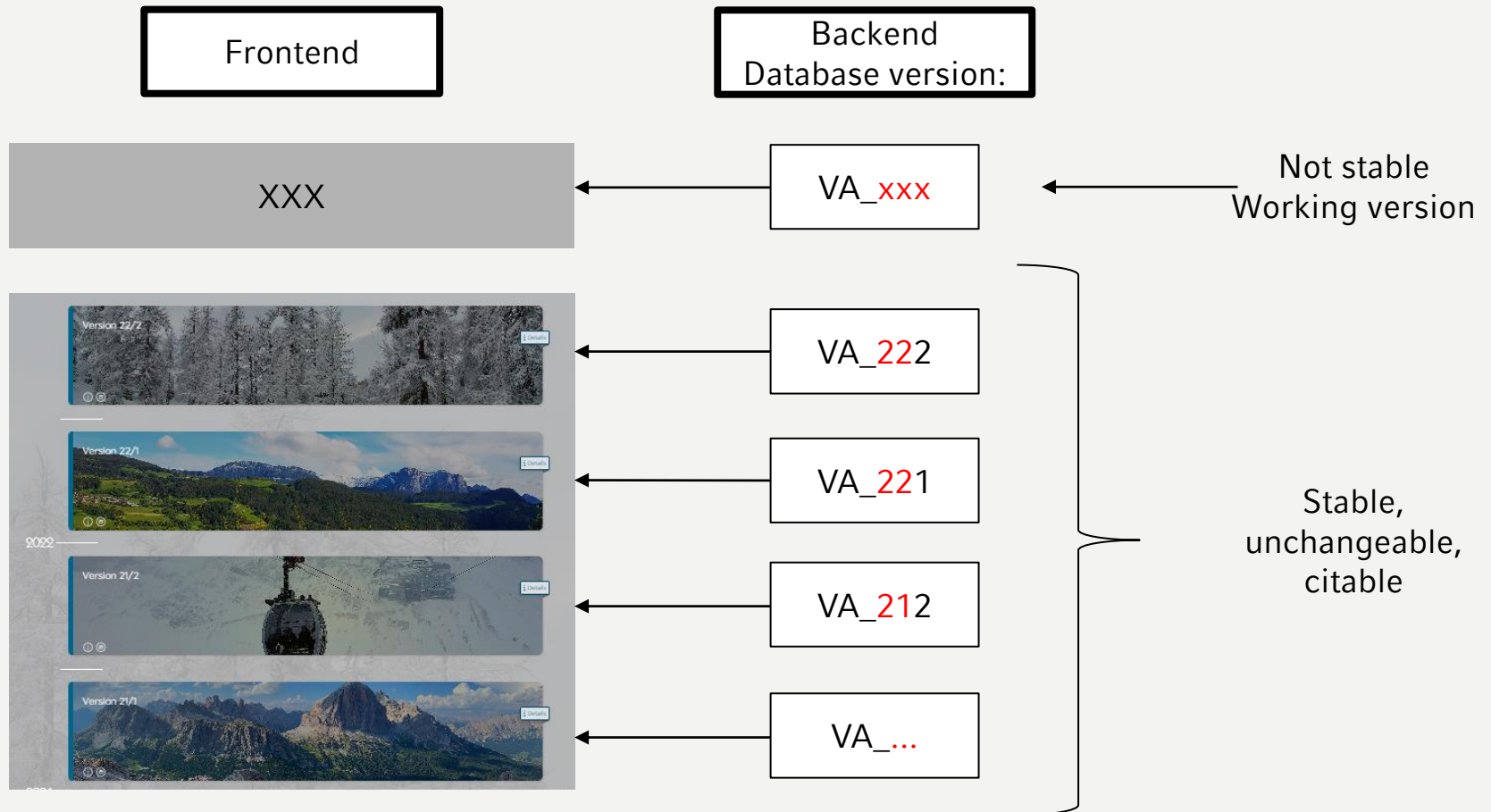
Identifiers (norm data) and linking



Identifiers (norm data): Examples



VerbaAlpina Versioning



Referencing VerbaAlpina

Citation suggestion (example: article in Methodology)

Krefeld, T. / Lücke, S. (2022): s.v. "Betacode", in: VerbaAlpina-de 22/2 (Erstellt: 16/1, letzte Änderung: 22/1), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D222%26single%3DM7

Deep links (examples)

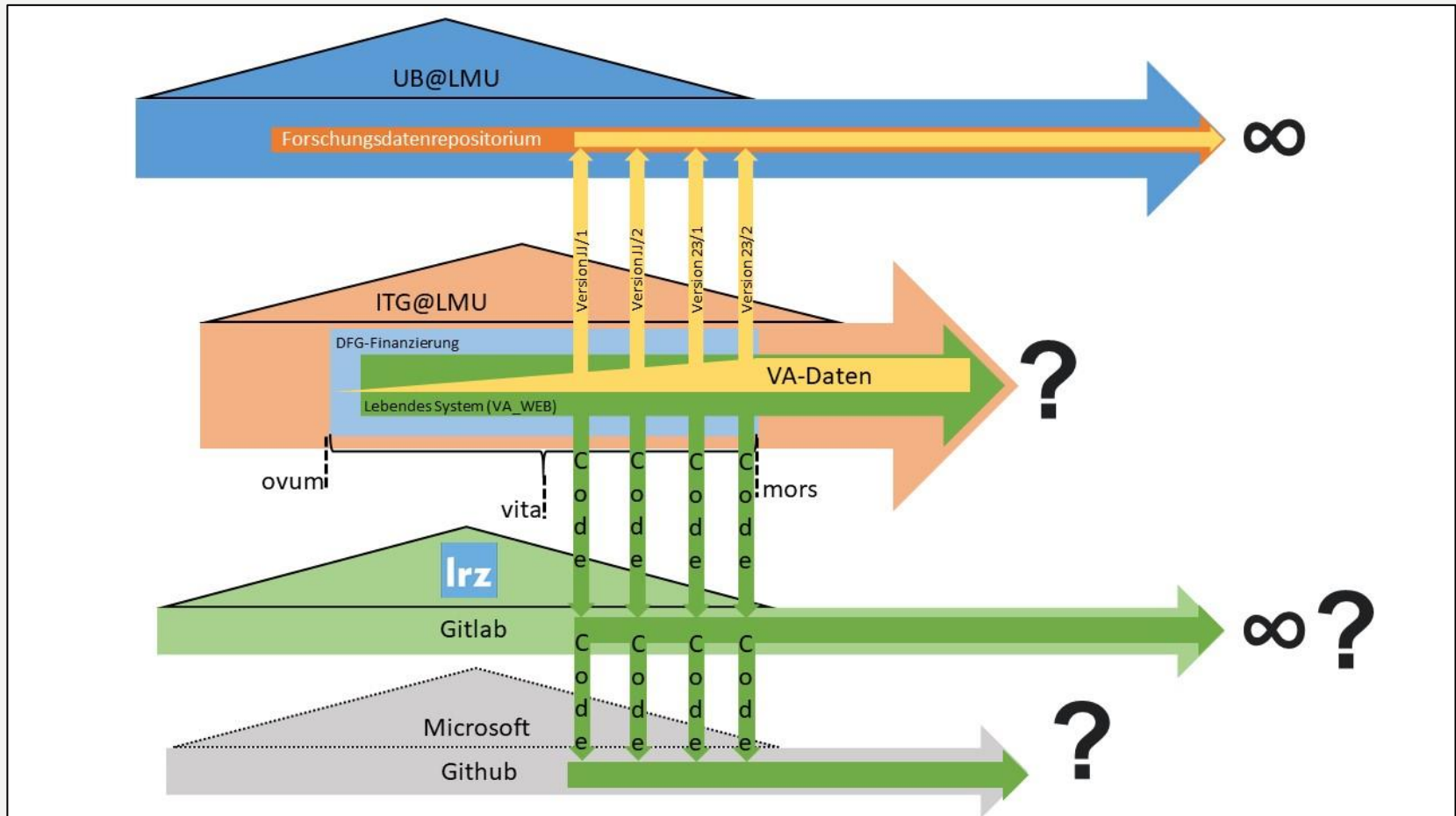
https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=222&single=M7

**Methodology
s.v. Betacode**

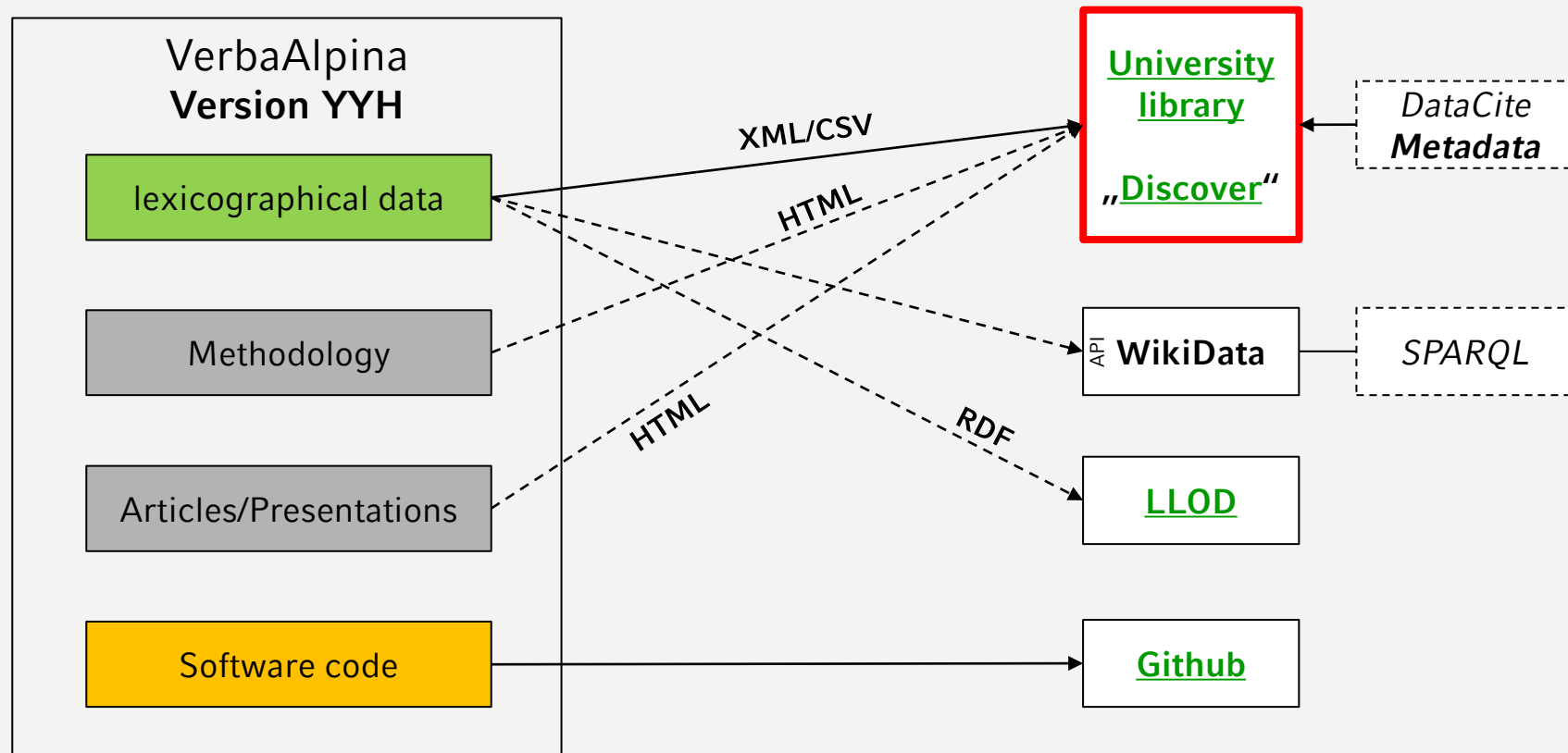
https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=12180&db=222&single=581

**LexiconAlpinum
s.v. Anke (L581)**

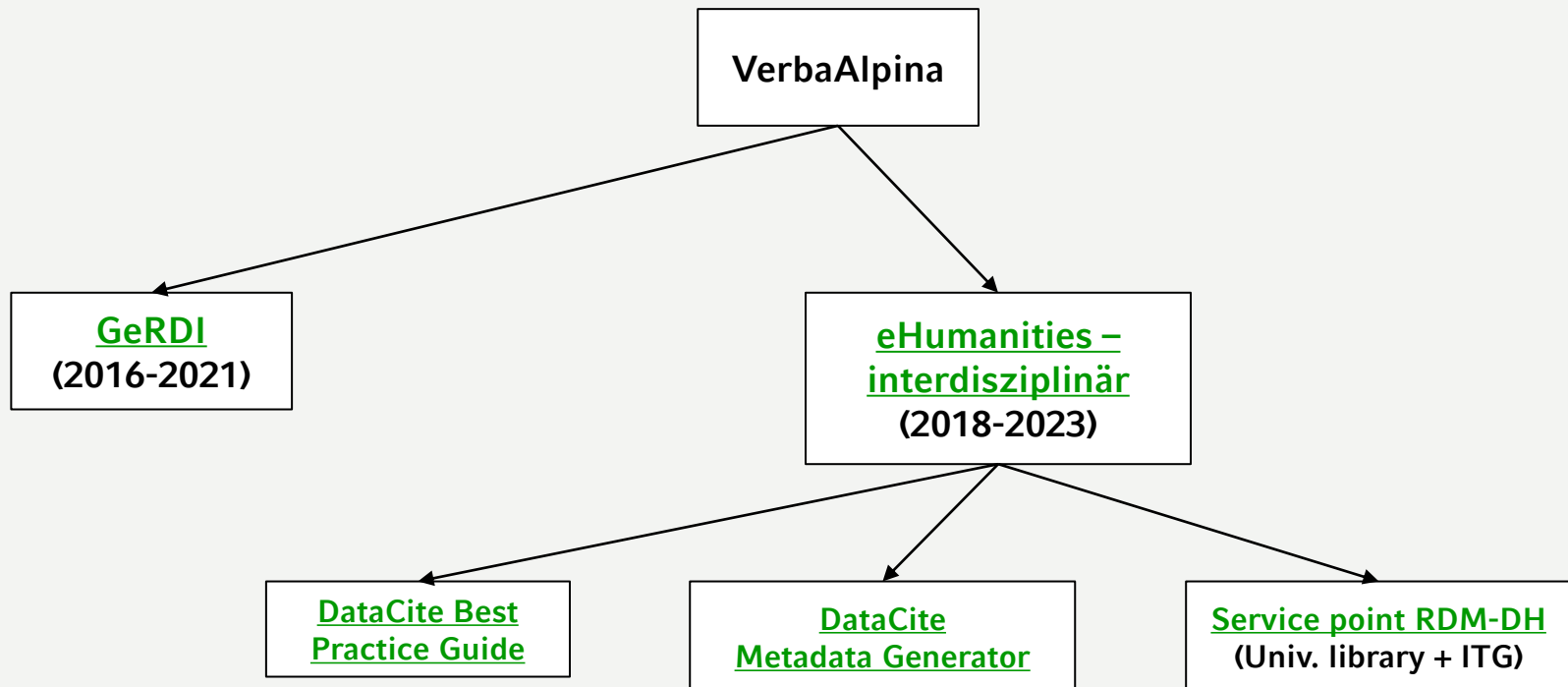
Sustainability: Long term perspective



VerbaAlpina data export



VerbaAlpina: Involvement in research data management projects



The FAIR principles

- Formulated in 2014 by „Force11
- FAIR means data should be:
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- Mandatory for research data management according to DFG guidelines and EU Horizon 2020

What does FAIR mean in detail? Vague!

Force11 →

3. Data Objects can be **Interoperable** only if:

3.1. (Meta) data is machine-actionable [8]

3.2. (Meta) data formats utilize shared vocabularies and/or ontologies [9]

3.3 (Meta) data within the Data Object should thus be both syntactically parseable and semantically machine-accessible [10]

Forschungs-
daten.info

→

Interoperabilität

I1. (Meta-) Daten verwenden eine formale, zugängliche, gemeinsame und allgemein anwendbare Sprache für die Wissensrepräsentation.

I2. (Meta-) Daten verwenden Vokabulare, die den FAIR-Prinzipien folgen.

I3. (Metadaten) enthalten qualifizierte Verweise auf andere (Meta-) Daten.

WikiFair →

- **Interoperable**: Known data model and interfaces

DFG →

schungsdaten und zentralen Materialien – den **FAIR-Prinzipien** („Findable, Accessible, Interoperable, Re-Usable“) folgend – zugänglich in anerkannten Archiven und Repositorien. Einschränkungen können sich im Kontext

FAIRness of VerbaAlpina

	Findable	Accessible	Interoperable	Reusable
Metadata	x		x	
Persistent Identifiers			x	
Interfaces		x	x	
Open license				x
Professional infrastructure		x	x	
External norm data	x		x	
Data export				x
Institutional embedding	X (?)	x		x
Versioning			x	x
Detailed Documentation			x	x
Long term availability				x

Attribution in detail subject for debate!

ONLINE PRESENTATION

VerbaAlpina: Live demonstration

VerbaAlpina interfaces

[Interactive Map](#)

[LexiconAlpinum](#)

[Crowdsourcing Tool](#)

[Database](#)

[Methodology](#)

VerbaAlpina @UB

[UB Discover](#)

VerbaAlpina @Wikidata

VerbaAlpina
[property](#)

[SPARQL](#)
[query](#)

Thank you!