

# Linked lexical data of different resources and countries in the Alpine region

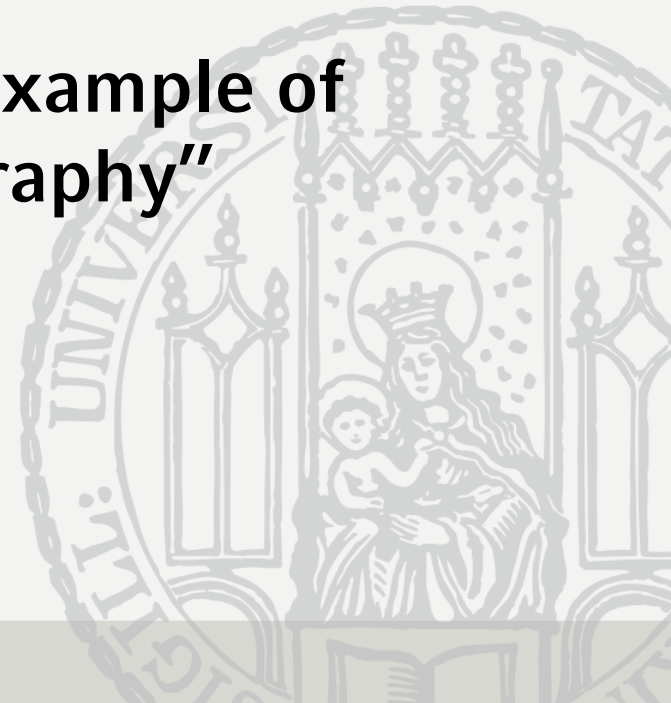
The project *VerbaAlpina* as an example of good practice in “smart lexicography”

Beatrice Colcuc | Christina Mutter

<https://www.verba-alpina.gwi.uni-muenchen.de/>

eLex Conference – Smart Lexicography

1 – 3 October 2019 – Sintra





## Outline

1. Challenges of modern lexicography
  - Changes in research work
2. Project overview
3. Data processing
  - Transcription
  - Tokenization
  - Typification
  - Data access
4. Visualization of data
5. Linking of data



# 1. CHALLENGES OF MODERN LEXICOGRAPHY

## Changes in research work

### PAST

- **Analogue World**
  - *Physical obstacles to sharing data*
  - *Publications mainly printed*
  - *Central role of libraries*
- **Individual Data Collection**
  - *Each research project by itself*

### PRESENT

- **Way from the analogue to the digital world**
  - *Digitalization → Change of mindset*
  - *Desire and need for sharing data*
- **Data collection with the aim of networking**
  - *Need for sustainability → sharing*
  - *Possibility to publish and to access all data → more transparency*



# 1. CHALLENGES OF MODERN LEXICOGRAPHY

## Changes in research work

### PAST

- **Semasiological vs. Onomasiological**
  - *Only one perspective possible*
- **Data from one source only**

### PRESENT

- **Semasiological and Onomasiological**
  - *Digitalization allows two perspectives at the same time*
- **Data from many sources**



# 1. CHALLENGES OF MODERN LEXICOGRAPHY

## Changes in research work

### FUTURE

- **Only digital data**
  - *Away from paper*
  - *Create more sharing possibilities*
- **Challenge of developing a uniform system for data processing**
- **Sustainability**
  - **F.***\_indable*
  - **A.***\_ccessible*
  - **I.***\_nteroperable*
  - **R.***\_eusable*



## 2. PROJECT OVERVIEW

- *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit* (VerbaAlpina. The Alpine cultural region reflected through its multilingualism)
- Funded by the German Research Foundation (DFG)
- 1<sup>st</sup> term: 10/2014-10/2017, 2<sup>nd</sup> term: 11/2017-11/2020 (perspective until 2025)
- Investigation of the multilingual Alpine region
- Combination of (geo-)linguistics and digital humanities



## Research Aims

- Selective and analytical investigation of the linguistically and dialectally highly fragmented alpine space in its historico-cultural and historical-linguistic unity
- Overcoming of the traditional limitation of geolinguistic investigation to nation-states
- recognition of connections regarding the etymology of the individual dialectal words
- Setting up a portal by using modern media technology: documentation, data collection, collaborative development
- cooperation with other projects is fundamental for VerbaAlpina

## Area under investigation: The Alpine region

- Area of investigation is limited to the territorial borders defined by the Alpine convention
- surface area of 190,600 km<sup>2</sup>, encompasses parts of six different countries (D, A, CH, I, F, SLO) and two entire countries (FL, MC)
- ethnographic and topographic homogeneity and strong linguistic heterogeneity → 3 language families







## Three conceptual domains

project years	1	2	3	4	5	6	7	8	9	
calendar year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
quarter	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
project phase	I			II			III			
focus	<b>culture</b> <ul style="list-style-type: none"> <li>• alpine pasture farming</li> <li>• milk processing</li> </ul>			<b>nature</b> <ul style="list-style-type: none"> <li>• landscape formations</li> <li>• weather</li> <li>• fauna</li> <li>• flora</li> </ul>			<b>modern life</b> <ul style="list-style-type: none"> <li>• ecology</li> <li>• tourism</li> </ul>			



## Data

### Multiple different sources

- printed atlases/dictionaries (georeferenced)
- Digital material from project partners
- Crowdsourcing





# Crowdsourcing-Tool

[www.lmu.de/verbaalpina](http://www.lmu.de/verbaalpina)

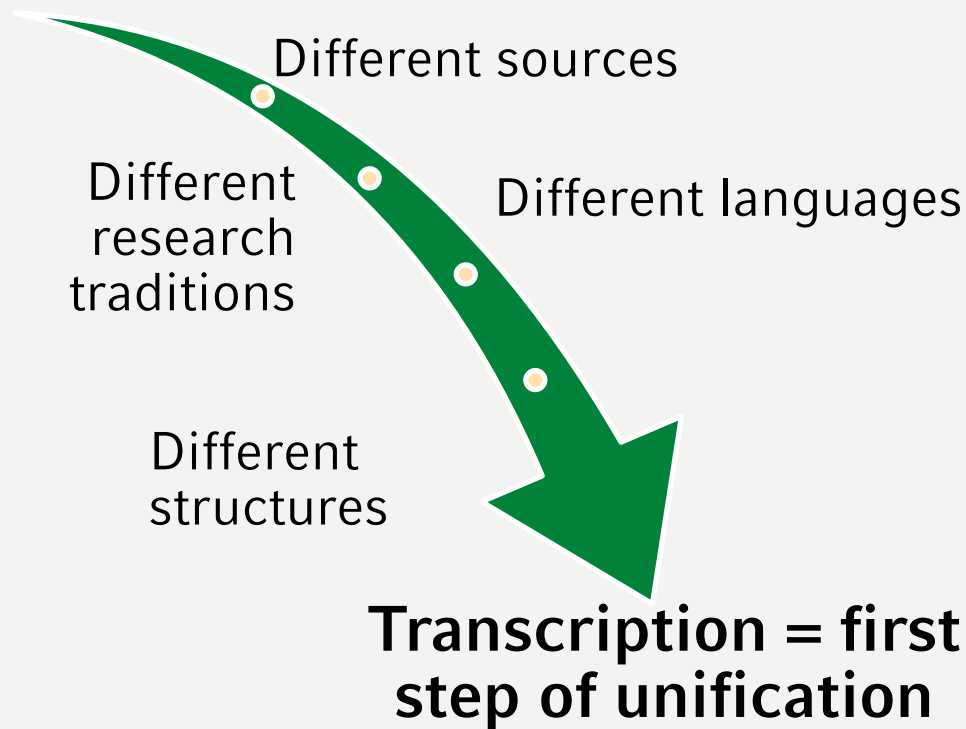
Wie sagt man zu *Begriff* in *Gemeinde* ? Ihre Antwort

Wählen Sie eine Gemeinde aus.



## 3. DATA PROCESSING

### Transcription



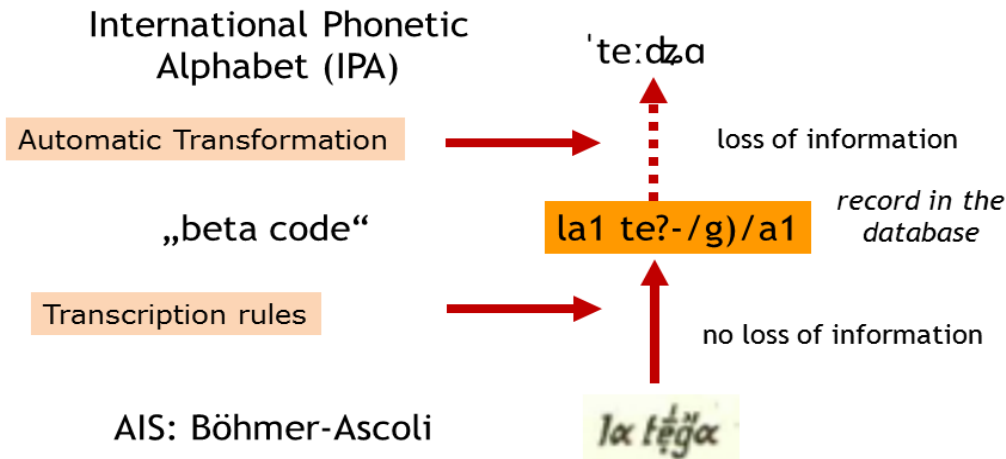
→ Relational  
Database: MySQL



### 3. DATA PROCESSING

## Transcription

Transcription system: the beta code



Principles of the transcription:

1. Reliability of the source
2. Easy comparability

### 3. DATA PROCESSING

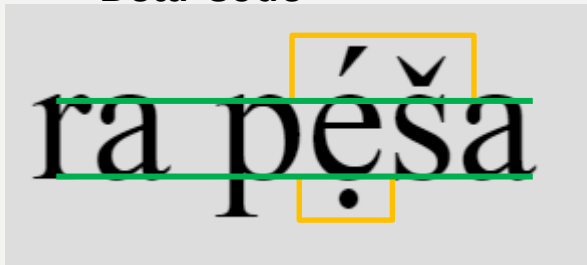
#### Betacode

##### Beta Code

Diacritics

Base sign

Diacritics



- The system uses only ASCII signs: the data can be transcribed with any computer and by any user
- For every source an unique codepage is created allowing the conversion into IPA

# 3. DATA PROCESSING

Transcription Tool

Original map

Transcription rules

Source

Stimulus

Informant number

Field for entering the transcription

Concept assignment

III. BAND SPRACH- UND SACHATLAS ITALIENS UND DER SÜDSCHWEIZ. KARTE 569.

**L'ABETE ROSSO (PICEA EXCELSA)**  
ROTTANNE, FICHTE-SAPIN ÉLEVÉ, ÉPICÉA  
Block L - Penzig 361 - Roll. XI, 106, 225  
71, 2 + 49, 3 = 0

**Legende:**  
Stems varietal = *Abies concolor*.  
Infl. long on Karte abete bianco.

**Einzelbemerkungen:**  
107. 254, 263, 267, 275, 285, 286, 289, 337, 345  
346, 372, 373, 376, 389, 404. Nur als *Abies*  
baum in Gärten.  
109. 357, 365, 377, 381, 409. Als *Abies*  
bezeichnet.  
119. 376 = Name für alle Nadelbäume.  
125, 143, 164, 247. Gärten.  
135. 137. Bildet angeblich keine Früchte.  
(was sonst von der *Abies* keine  
bezeugt sind).  
182. Erst vor 10 Jahren eingeführt.  
183. *Abies* verwendet.

Zeichen	Beschreibung	Beta-Code	Kommentar
α	Griechisches Alpha	a1	
ϐ	Spiegelverkehrtes a	a2	
æ	Ligatur ae	a3	
β	Griechisches Beta	b1	
ḃ	Durchgestrichenes	b2	

Transkription (z.B. hinzufügen)

1.)

Beleg Konzept(e) auswählen

Eintragen vacat Konzept nicht vorhanden Bezeichnung nicht bekannt Problem Neues Konzept anlegen





### 3. DATA PROCESSING

#### Tokenization

Attestation in BetaCode	Attestation in IPA	Concept
una1 mu:g/a1 da1 va/c)/	una mydʒa da v <sup>l</sup> atɕ	HERD OF COWS
TOKENIZATION		
una1	una	ARTICLE
mu:g/a1	mydʒa	HERD
da1	da	PREPOSITION
va/c)/	v <sup>l</sup> atɕ	COW

### 3. DATA PROCESSING

#### Typification

**Morpho-lexical type: Orthography, language family, POS, gender, affix + base type (lexical base)**

	barga	barg	margun	bargun
Language family	roa	roa	roa	roa
Part of speech	noun	noun	noun	noun
Affix	-	-	+	+
Gender	f	m	m	m
Morpho-lexical type	1	2	3	3

\*barga/\*barca



## 3. DATA PROCESSING

### Typification tool

Source and map

expressions

**Belege**

AIS 610\_1: la fragola

Nur Belege ohne Konzeptzuweisung anzeigen  
 Nur Belege aus der Alpenkonvention anzeigen

(Wenn "Strg" gedrückt ist, können mehrere Belege ausgewählt werden, wenn "Shift" gedrückt ist, werden alle orthographisch identischen Belege ausgewählt.)

Belege: amazû --- amaz u: (?) ✕ amõsä --- am õsa: (?) [1] ✕ amõsä --- am õsa: (?) [2] ✕ amõsä --- am õsa: (?) ✕

Beleg	Informanten	Bemerkungen	Konzept(e)	VA-Typ
amazû --- amaz u: (?)	249		ERDBEERE ✕	Korrigieren
amõsä --- am õsa: (?) [1]	150	Tokengruppe: amõ/sa-	ERDBEERE ✕	Korrigieren
amõsä --- am õsa: (?) [2]	150		ERDBEERE ✕	Korrigieren
amõsä --- am õsa: (?)	150		ERDBEERE ✕	Korrigieren

**VA-Typ zuweisen**

Typ auswählen Typ zuweisen Neuen Typ erstellen Typ bearbeiten

**Konzept zuweisen**

Konzept auswählen Konzept zuweisen Neues Konzept anlegen

Typisierung nicht notwendig

Belege sind ARTIKEL Bestätigen

**LA FRAGOLA ERDBEERE-FRAISE**

Legende:

11 amõ/sa-  
12 la fragola  
13 amõ/sa-  
14 la fragola  
15 amõ/sa-  
16 la fragola  
17 amõ/sa-  
18 la fragola  
19 amõ/sa-  
20 la fragola  
21 amõ/sa-  
22 la fragola  
23 amõ/sa-  
24 la fragola  
25 amõ/sa-  
26 la fragola  
27 amõ/sa-  
28 la fragola  
29 amõ/sa-  
30 la fragola  
31 amõ/sa-  
32 la fragola  
33 amõ/sa-  
34 la fragola  
35 amõ/sa-  
36 la fragola  
37 amõ/sa-  
38 la fragola  
39 amõ/sa-  
40 la fragola  
41 amõ/sa-  
42 la fragola  
43 amõ/sa-  
44 la fragola  
45 amõ/sa-  
46 la fragola  
47 amõ/sa-  
48 la fragola  
49 amõ/sa-  
50 la fragola  
51 amõ/sa-  
52 la fragola  
53 amõ/sa-  
54 la fragola  
55 amõ/sa-  
56 la fragola  
57 amõ/sa-  
58 la fragola  
59 amõ/sa-  
60 la fragola  
61 amõ/sa-  
62 la fragola  
63 amõ/sa-  
64 la fragola  
65 amõ/sa-  
66 la fragola  
67 amõ/sa-  
68 la fragola  
69 amõ/sa-  
70 la fragola  
71 amõ/sa-  
72 la fragola  
73 amõ/sa-  
74 la fragola  
75 amõ/sa-  
76 la fragola  
77 amõ/sa-  
78 la fragola  
79 amõ/sa-  
80 la fragola  
81 amõ/sa-  
82 la fragola  
83 amõ/sa-  
84 la fragola  
85 amõ/sa-  
86 la fragola  
87 amõ/sa-  
88 la fragola  
89 amõ/sa-  
90 la fragola  
91 amõ/sa-  
92 la fragola  
93 amõ/sa-  
94 la fragola  
95 amõ/sa-  
96 la fragola  
97 amõ/sa-  
98 la fragola  
99 amõ/sa-  
100 la fragola

map

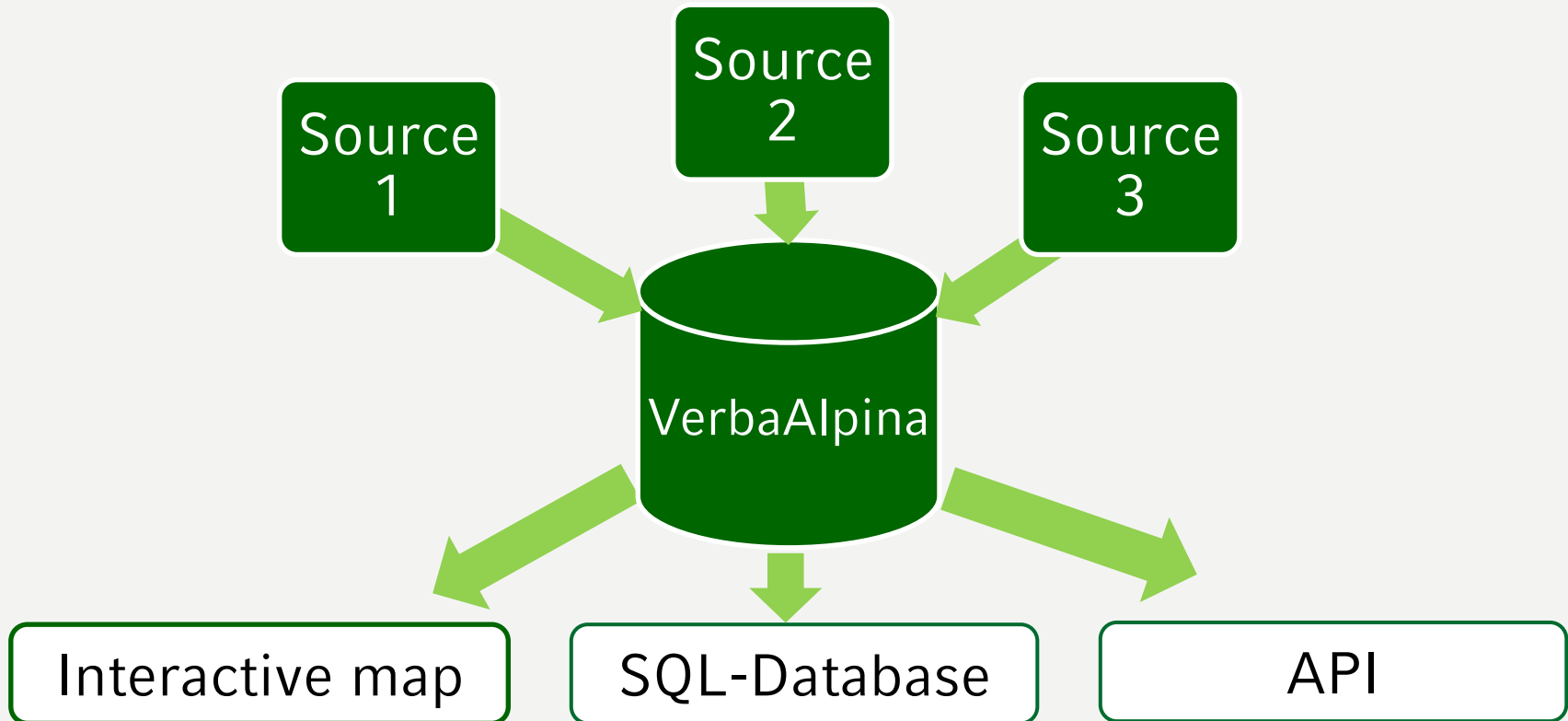
Types and concepts

Typification not necessary



### 3. DATA PROCESSING

#### Data access





## 4. VISUALIZATION OF DATA via the interactive map

- onomasiologically and semasiologically  
by using appropriate filters
- **4 different modes of visualization:**
  - geographical/physical
  - abstract
  - qualitative
  - quantitative



## 4. VISUALIZATION OF DATA via the interactive map

- onomasiologically and semasiologically  
by using appropriate filters
- **4 different modes of visualization:**
  - geographical/physical (NUTS 3 boundaries)
  - abstract
  - qualitative
  - quantitative



## 4. VISUALIZATION OF DATA via the interactive map

- onomasiologically and semasiologically  
by using appropriate filters
- **4 different modes of visualization:**
  - geographical/physical (NUTS 3 boundaries)
  - abstract (hexagons/honeycomb map)
  - qualitative
  - quantitative



# geographical qualitative (onomasiological perspective)

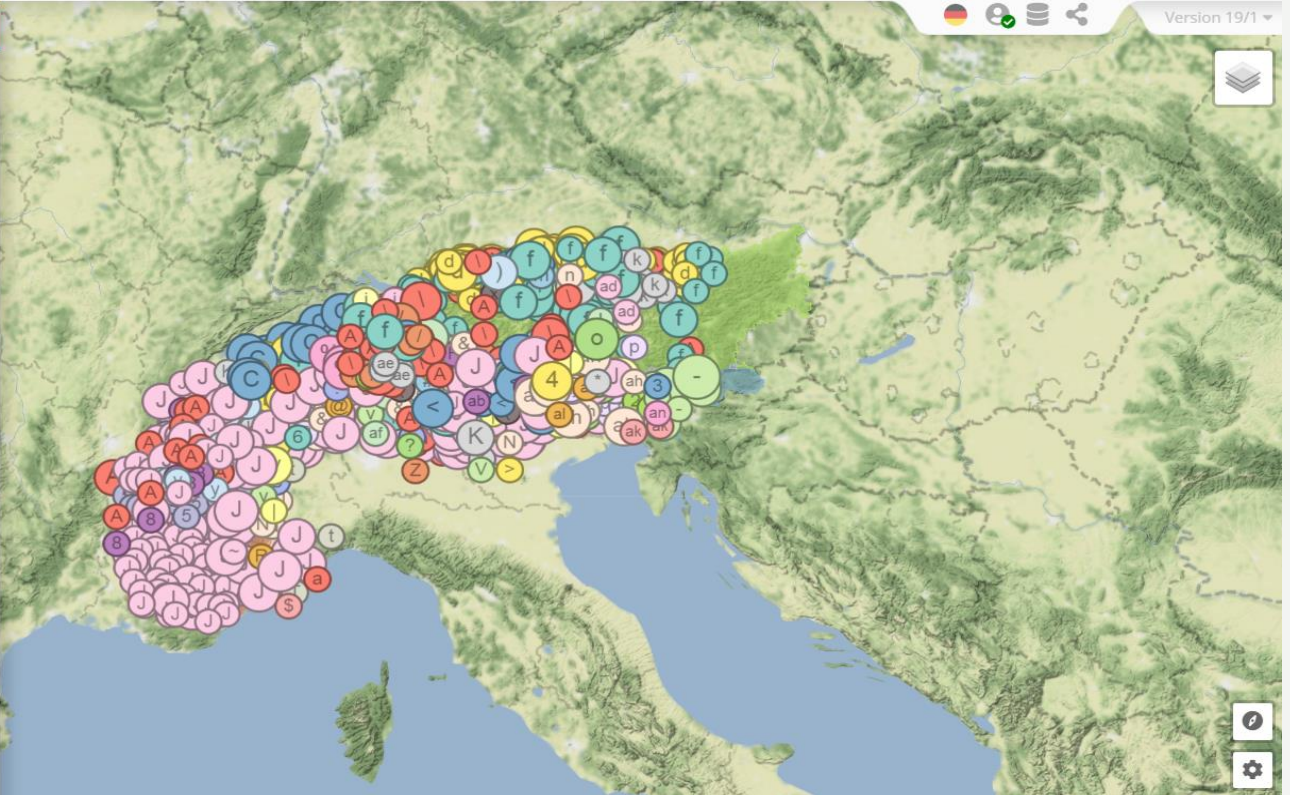
LMU HOME INTERAKTIVE KARTE LEXICON ALPINUM METHODOLOGIE BEITRÄGE MITMACHEN! PERSONEN DFG

Bitte Suchbegriff eingeben...

Kartographische Darstellung Physisch Hexagonal

Legende

- Flächen und Regionen Sprachgebiete
- Konzept BUTTER** (1 Beleg)
- (nicht typisiert) (82 Belege)
- Morpho-lexikalischer Typ **aigre / agro (roa)** (1 Beleg)
- Morpho-lexikalischer Typ **Anke (gem m.)** (351 Belege)
- Morpho-lexikalischer Typ **Anke (gem)** (91 Belege)
- Morpho-lexikalischer Typ **Ankeballlein (gem)** (1 Beleg)
- Morpho-lexikalischer Typ **Ankenmocklein (gem)** (1 Beleg)
- Morpho-lexikalischer Typ **balla di paintg (roa)** (1 Beleg)
- Morpho-lexikalischer Typ **balle / palla (roa f.)**



Version 19/1



# geographical qualitative (semasiological perspective)

LMU HOME INTERAKTIVE KARTE LEXICON ALPNUM METHODOLOGIE BEITRÄGE MITMACHEN! PERSONEN DFG

Bitte Suchbegriff eingeben...

Kartographische Darstellung Physisch Hexagonal

Legende

Flächen und Regionen Sprachgebiete

Basistyp **\*excōcta (lat) (\* = rekonstruiert)**

(kein Konzept) (4 Belege)

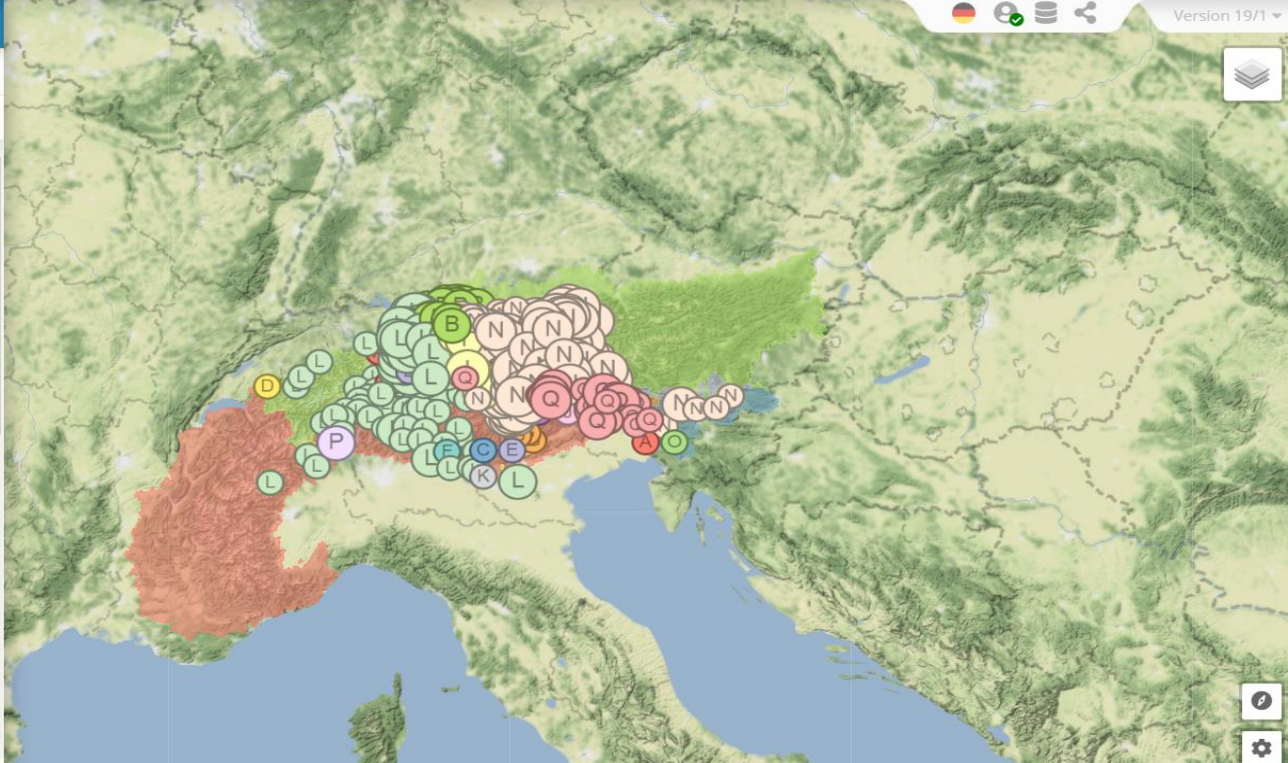
Konzept **EIWEISSTEILCHEN, DIE NACH DER ZWEITEN SCHEIDUNG BEI ERHITZEN DER MOLKE AUFSTEIGEN** (65 Belege)

Konzept **FLÜSSIGKEIT NACH ENTNAHME DER FESTEN MASSE, ZWEITE SCHEIDUNG** (1 Beleg)

Konzept **FLÜSSIGKEIT NACH ENTNAHME DER KÄSEMASSE, ERSTE SCHEIDUNG** (2 Belege)

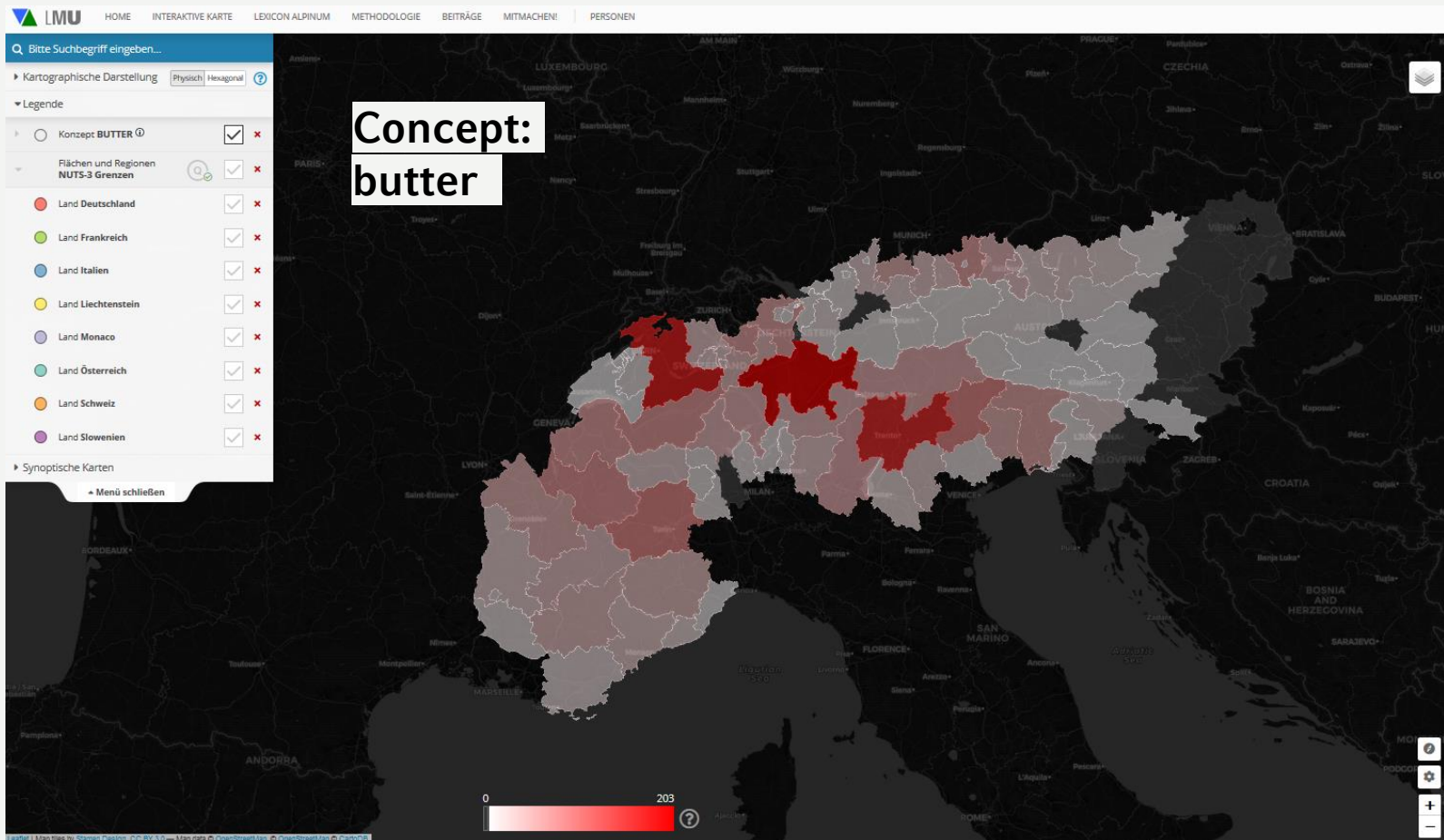
Konzept **FLÜSSIGKEIT NACH ENTNAHME DES QUARKS** (2 Belege)

Konzept **GEHILFE DES KÄSERS, PUTZT DIE GEFÄSSE UND DEN KESSEL**



Version 19/1

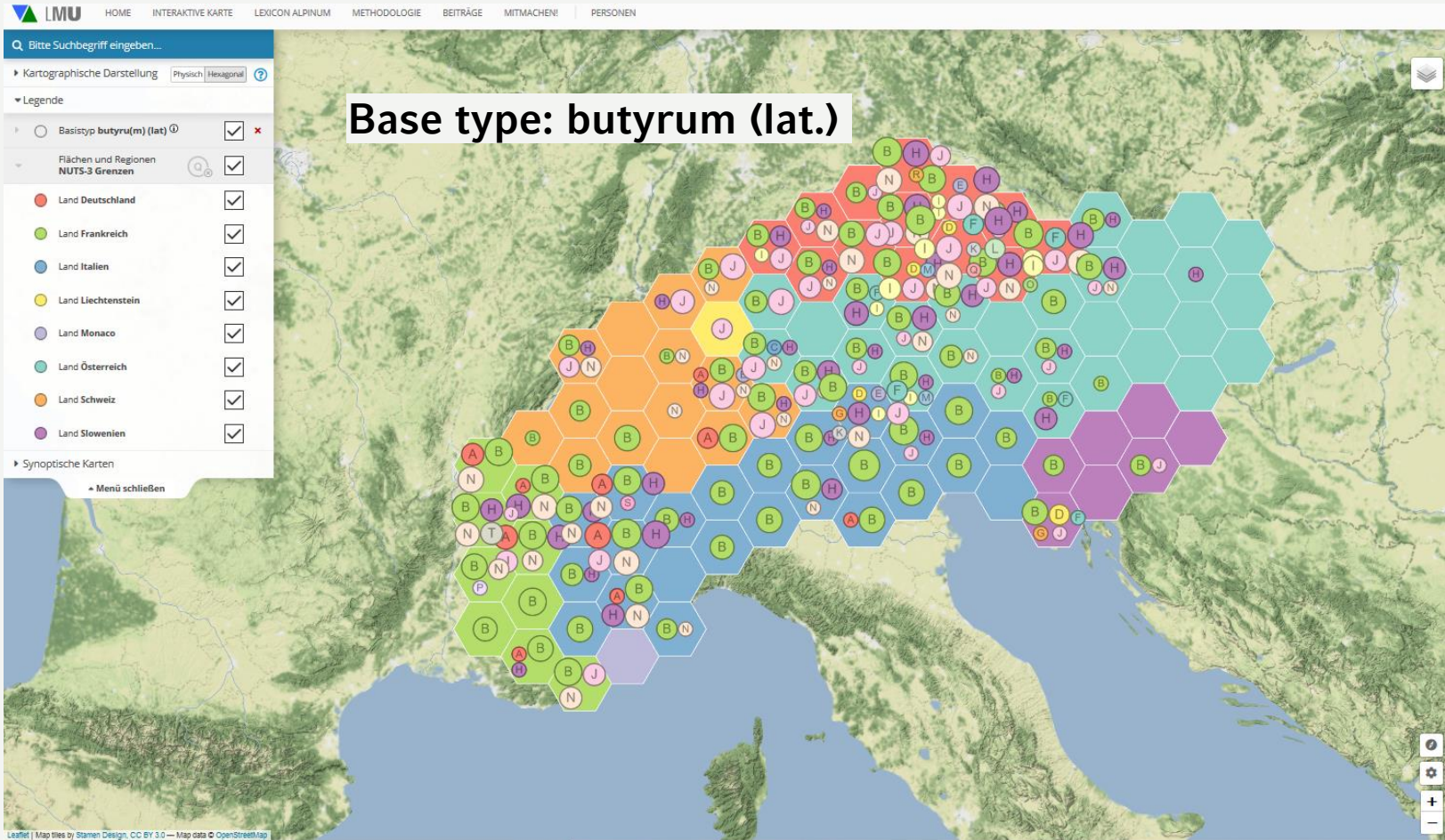
# geographical quantitative (onomasiological perspective)



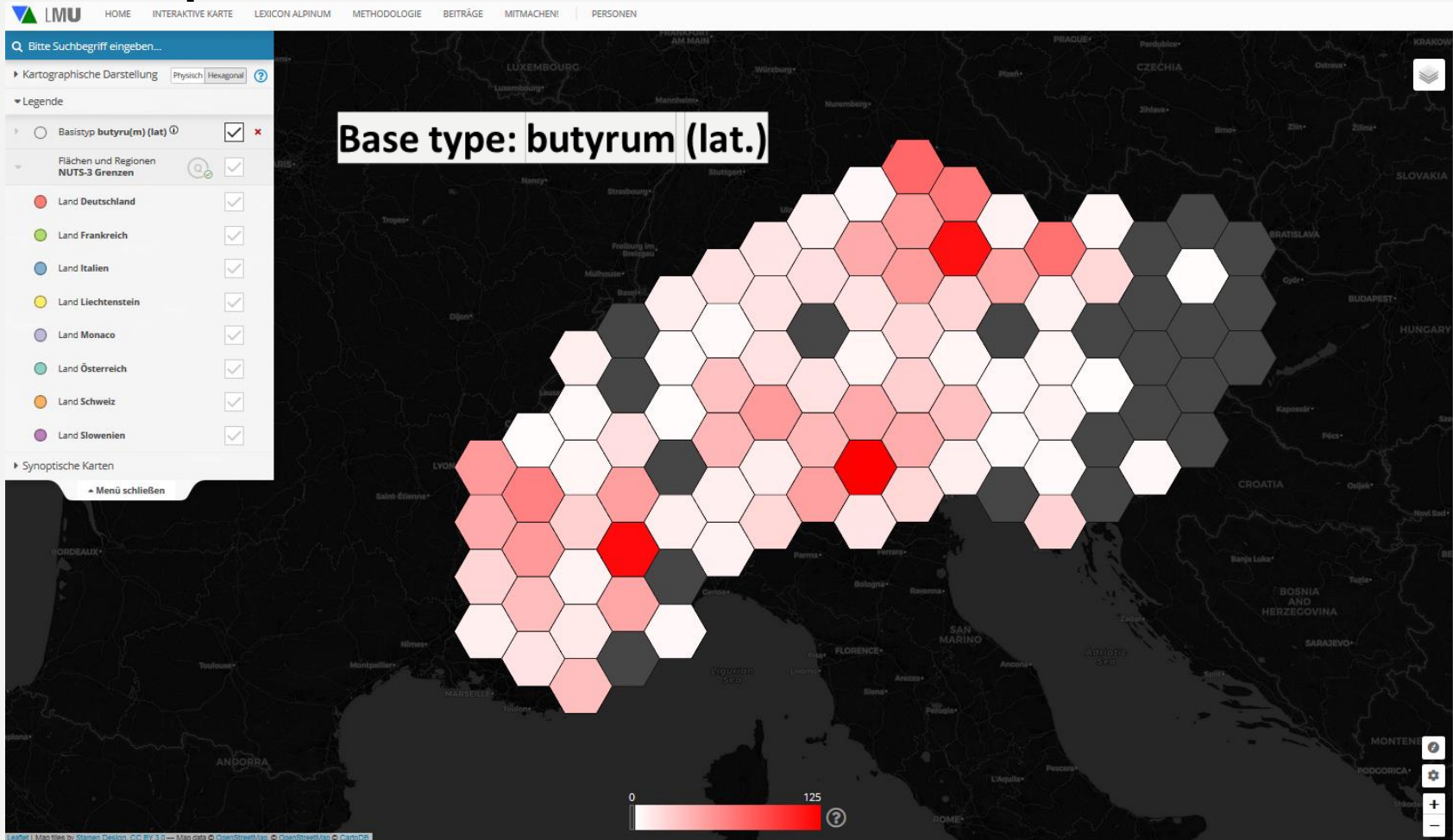
Concept:  
butter



# abstract qualitative (onomasiological perspective)



# abstract quantitative (onomasiological perspective)





## 5. LINKING OF DATA

### Guaranteed by:

- 1) Enrichment with **metadata**
- 2) Assignment of persistent **identifiers** of different knowledge data bases/norm data bases/reference dictionaries (GeoNames, Q-ID/L-ID, GND, DOI)



## 1) Enrichment with metadata

a) via the process of typification

→ assignment of basic types, morpho-lexical types, phonetic types

b) via the enrichment with sociolinguistic and sociodemographic data

→ language, municipality

(if available: age and sex of the informant)

## 2) Assignment of different identifiers

**OBJECTIVE:** Interconnection of VA data with data of other projects

- **GeoNames** of [www.geonames.org](http://www.geonames.org) (for municipalities)
- **Q-IDs** of Wikidata (for concepts)
- partly also: **L-IDs** of Wikidata (for morpho-lexical types) + **GNDs** of the German National Library (for concepts) (*Gemeinsame Normdatei*, „Integrated Authority File“)
- **ISO-Codes 639-3** (for languages)
- Identifiers of **reference dictionaries** (for morpho-lexical types + base types)
- **DOIs** (Digital Object Identifiers, assigned to every single data)

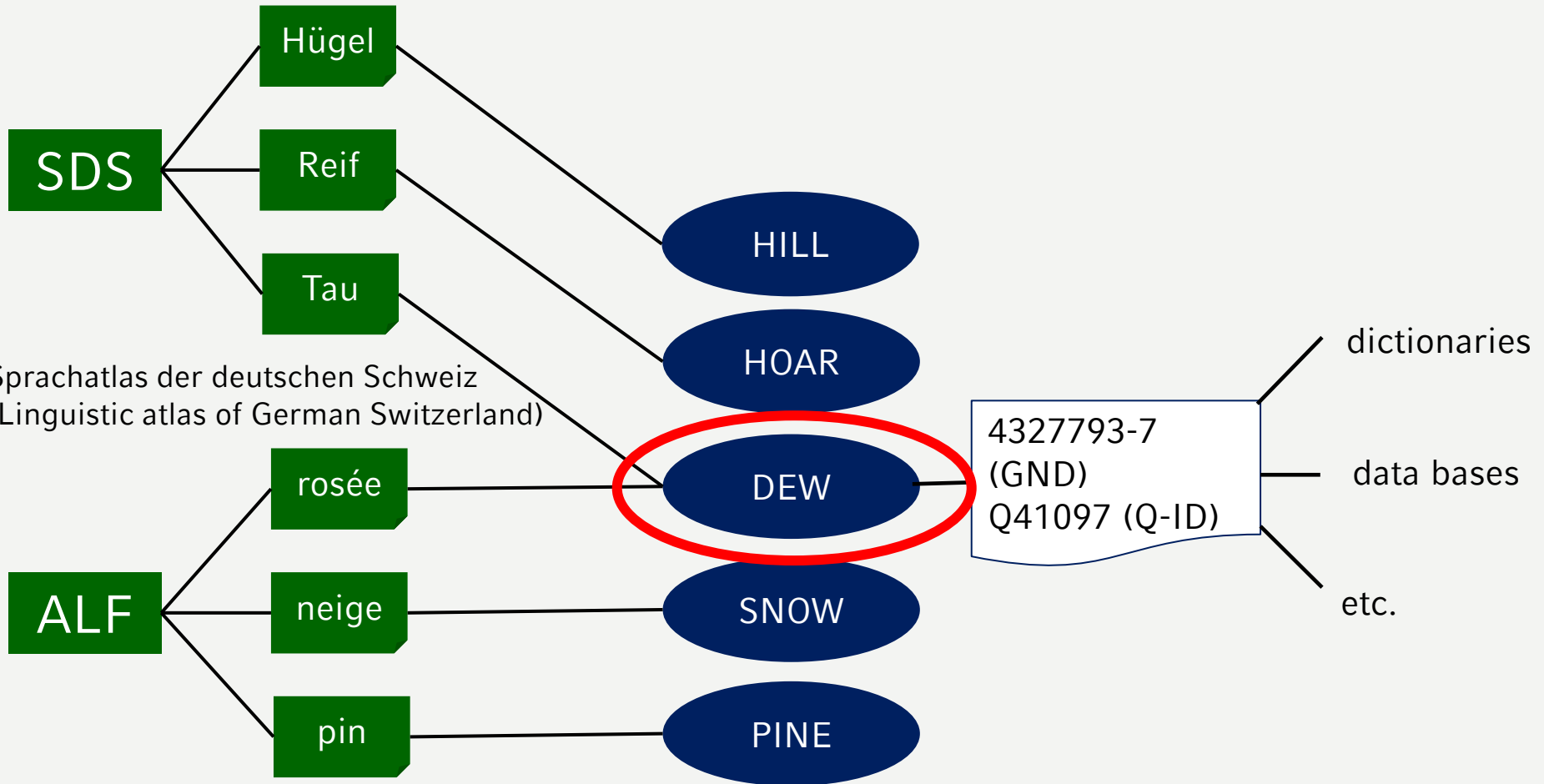
Sintra  
2262912

Q54050  
(Q-ID)

L73260  
(L-ID)

4135744-9 (GND)









- [Main page](#)
- [Community portal](#)
- [Project chat](#)
- [Create a new Item](#)
- [Create a new Lexeme](#)
- [Recent changes](#)
- [Random Item](#)
- [Query Service](#)
- [Nearby](#)
- [Help](#)
- [Donate](#)

- [Print/export](#)
- [Create a book](#)
- [Download as PDF](#)
- [Printable version](#)

- [Tools](#)
- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Permanent link](#)

Item [Discussion](#)



## butter (Q34172)

dairy product

edit

[In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	butter	dairy product	
German	Butter	aus Milch hergestelltes Streichfett	Salzbutter Tafelbutter Buttern Rohmilchbutter Sauerrahmbutter Kirnen Süßrahmbutter
French	beurre	aliment obtenu à partir de la matière grasse du lait	8029-34-3 Beurre d'escargot
Bavarian	Budan	No description defined	

[All entered languages](#)





# „Zirbelkiefer“ – Swiss stone pine



- [Main page](#)
- [Community portal](#)
- [Project chat](#)
- [Create a new item](#)
- [Create a new Lexeme](#)
- [Recent changes](#)
- [Random item](#)
- [Query Service](#)
- [Nearby](#)
- [Help](#)
- [Donate](#)

- [Print/export](#)
- [Download as PDF](#)

- [Tools](#)
- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Permanent link](#)
- [Page information](#)
- [Concept URI](#)
- [Cite this page](#)

Lexeme Discussion



(L73260) **Zirbelkiefer**  edit

de

Language [German](#)  
Lexical category [noun](#)

### Statements

grammatical gender	<>	feminine	edit
		▼ 0 references	
			+ add reference
			+ add value

hyphenation	<>	Zir-bel-kie-fer	edit
		▼ 0 references	
			+ add reference
			+ add value

+ add statement



Kontakt | A-Z | Träger / Förderer | Datenschutz | Impressum | Hilfe | Mein Konto | English

↓ Katalog

- Einfache Suche
- Erweiterte Suche
- Browsen (DDC)
- Suchverlauf
- Meine Auswahl
- Hilfe
- Datenschop
- Mein Konto
- Ablieferung von Netzpublikationen
- Informationsvermittlung

Login →

→ Über die Deutsche Nationalbibliothek

## KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

Gesamter Bestand
Musikarchiv
Exilsammlungen
Buchmuseum

→ Suchformular zurücksetzen

sw all "Butter" Finden →  Expertensuche ?

eingeschränkt auf

- Materialarten: Elektronische Datenträger
- Normdaten: Sachbegriffe

### Ergebnis der Suche nach: sw all "Butter"

← Zurück zur Trefferliste

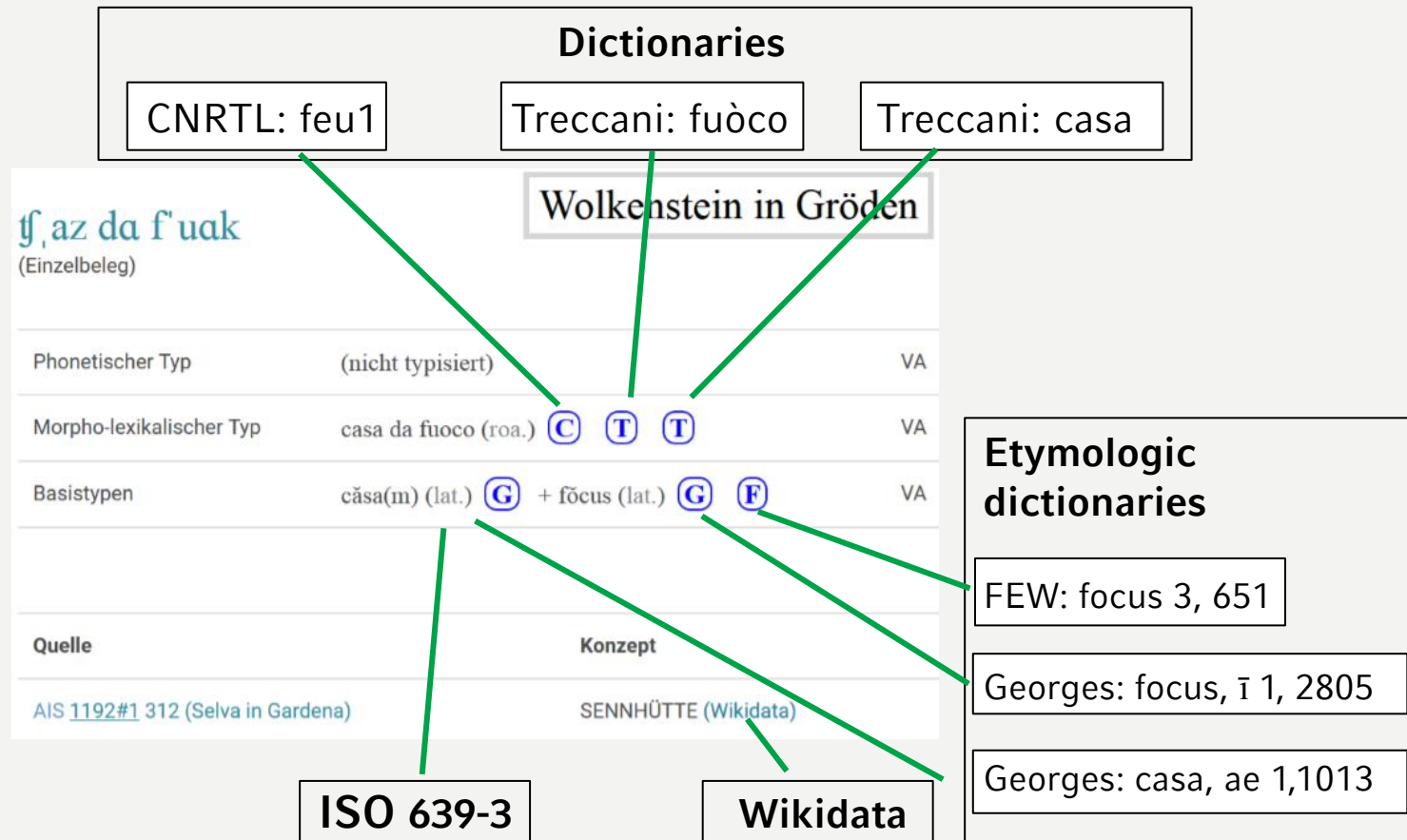
Treffer 1 von 11

<small>GND</small>	
<b>Link zu diesem Datensatz</b>	<a href="http://d-nb.info/gnd/4009236-7">http://d-nb.info/gnd/4009236-7</a>
<b>Sachbegriff</b>	Butter
<b>Quelle</b>	M
<b>Oberbegriffe</b>	Milchprodukt Streichfett
<b>DDC-Notation</b>	637.2 641.372
<b>Systematik</b>	32.7 Milchwirtschaft ; 31.11 Lebensmitteltechnologie
<b>Typ</b>	Allgemeinbegriff (saz)
<b>Andere Normdaten</b>	LCSH: Butter RAMEAU: Beurre LCSH: Cooking (Butter) RAMEAU: Cuisine (beurre)



## Identifiers of reference dictionaries

→ at present 43 dictionaries (e.g. Treccani, FEW online, DRG, Duden, etc.)





- **DOIs** (Digital Object Identifier)

→ assigned to every single data

(attestations, morpho-lexical types, concepts, municipalities)

→ accessibility also if URL (Uniform Resource Locator) changes

**Advantage:** sustainable citability of electronic resources



## Bibliography

Krefeld, T.: s.v. "Transkription", in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 16/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DT%2357](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DT%2357)

Krefeld, T. / Lücke, S.: s.v. "Betacode", in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 16/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DB%237](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DB%237)

Krefeld, T. / Lücke, S.: s.v. "Typisierung", in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 18/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DT%2358](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DT%2358)

Kümmet, S. / Lücke, S. / Schulz, J. / Zacherl, F.: s.v. "Forschungsdatenmanagement", in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DF%23112](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DF%23112)

Lücke, S.: s.v. "FAIR-Prinzipien", in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DF%23128](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DF%23128)

Lücke, S.: s.v. "Metadaten", in: VerbaAlpina-de 19/1, Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DM%23140](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DM%23140)

Lücke, S.: s.v. "Normdaten", in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DN%23114](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DN%23114)

Lücke, S. / Schulz, J.: s.v. "Digital Object Identifier (DOI)", in: VerbaAlpina-de 19/1 (Erstellt: 16/1), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DD%2373](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DD%2373)

Mutter, C.: s.v. "Wikidata", in: VerbaAlpina-de 19/1 (Erstellt: 18/1), Methodologie, [https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage\\_id%3D493%26db%3D191%26letter%3DW%23105](https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DW%23105)

# Thank you for your attention!

<https://www.verba-alpina.gwi.uni-muenchen.de/>

