Christina Mutter

FAIR linguistic data thanks to norm data – Wikidata as part of the research project VerbaAlpina

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

Celtic Knot Conference 2020
9-10 July 2020 - online

Outline

1. Project Overview

- research aims
- area under investigation
- conceptual domains
- data

2. FAIR principles and their role in research

3. What VerbaAlpina does to make its data FAIR

3.1. Assignment of norm data

- norm data created by VerbaAlpina
- persistent identifiers of external institutions

(Q-ID + L-ID of Wikidata, GND, GeoNames etc.)

1. 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)
- 1st term: 10/2014-10/2017, 2nd term: 11/2017-10/2020 (perspective until 2025)
- Investigation of the multilingual Alpine region
- Combination of (geo-)linguistics and Digital Humanities (DH)

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^2, 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

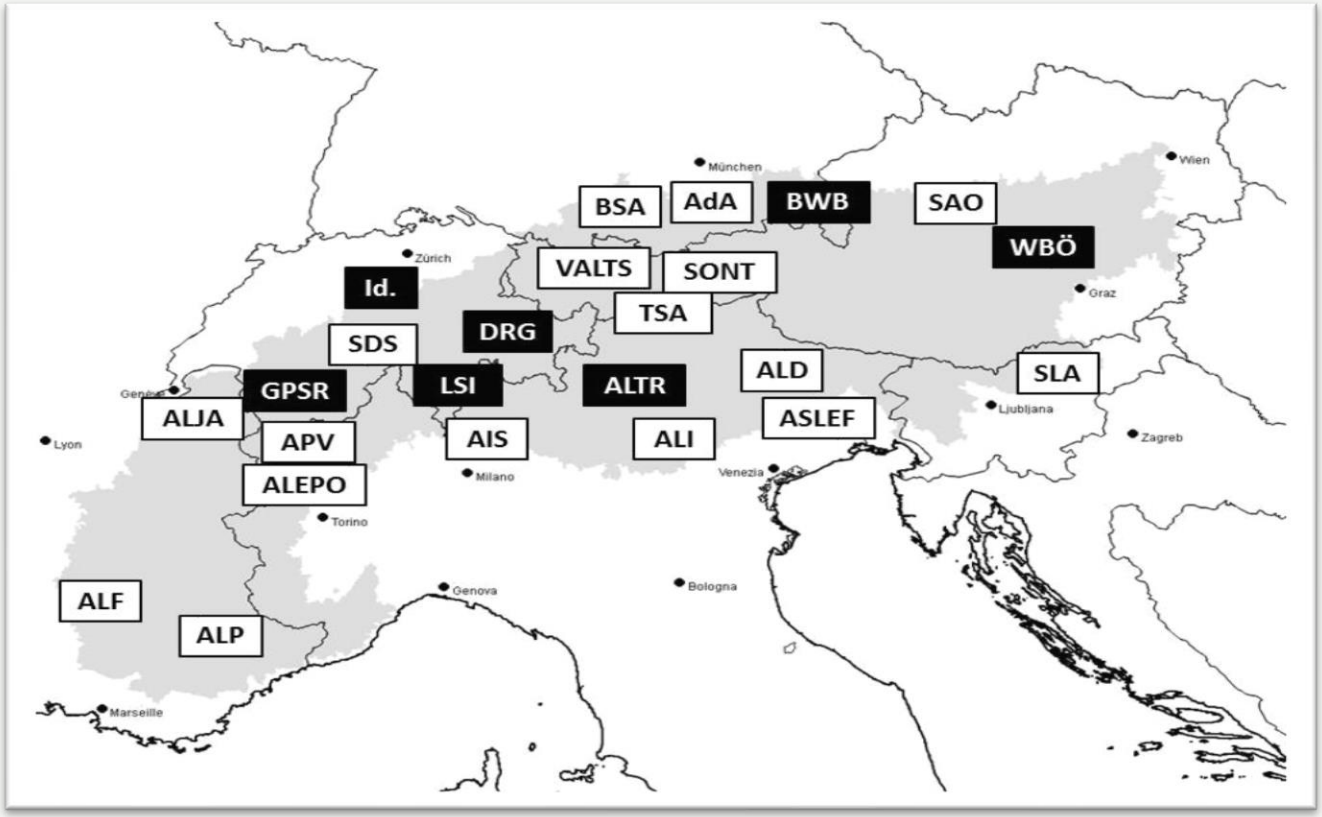
project phase	I	II	III
focus	**culture** • alpine pasture farming • milk processing	**nature** • landscape formations • weather • fauna • flora	**modern life** • ecology • tourism

Data

Multiple different sources

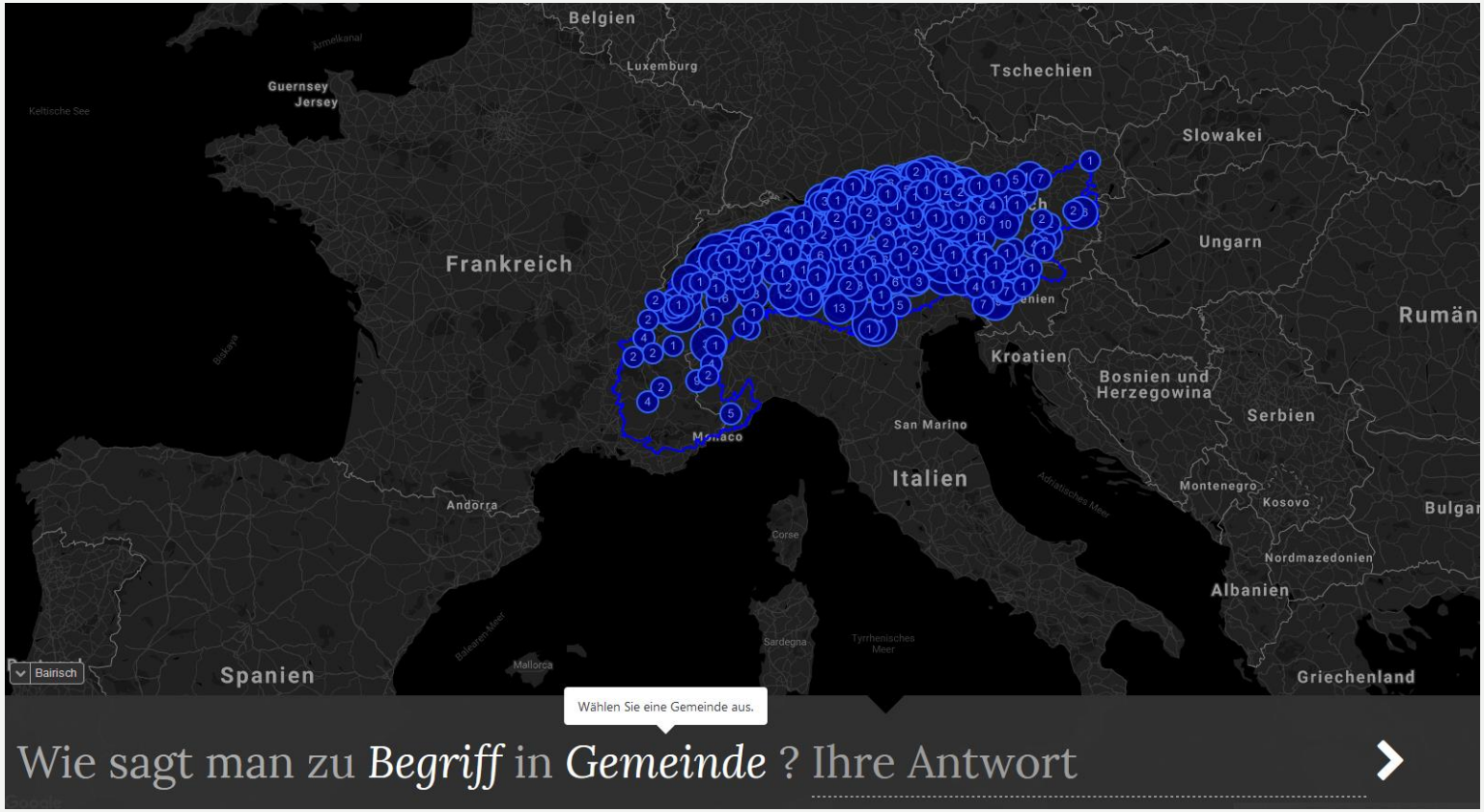
- printed atlases/dictionaries (georeferenced)
- digital material from project partners
- crowdsourcing

Atlases and Dictionaries in the Alpine region



Crowdsourcing-Tool

www.lmu.de/verbaalpina



2. FAIR principles and their role in research

Research data have to be FAIR:

F_indable
A_ccessible
I_nteroperable
R_eusable

→ principles postulated by Wilkinson et al. 2016 as the guiding principles for scientific data management

F_indable → via library catalogues and data aggregators

A_ccessible → via open access licences

I_nteroperable → via compatibility of databases and their interconnection

R_eusable → results from F, A, I

3. What VerbaAlpina does to make its data FAIR

F_indable

- Cooperation with the University Library of Munich University (as part of the project "e-humanities-interdisziplinär") and until 2019 also with the project GeRDI (Generic Research Data Infrastructure)

A_ccessible

- Creative Commons licence (compatible with open access) for all data managed by VerbaAlpina

  (up to version 18/1: CC BY SA 3.0, from 18/2: CC BY SA 4.0)

I_nteroperable

- through a fine granulation of the data stock via
  - structured data processing (transcription, tokenization, typification)
  - assignment of norm data (Q-ID, L-ID, GND, GeoNames etc.)
  - enrichment with metadata in DataCite and CIDOC CRM format
  - assignment of persistent identifiers (e.g. DOIs, Digital Object Identifiers)

- access to primary data and metadata (via interactive map, Lexicon Alpinum, API)

R_eusable

- results from F, A, I

I_nteroperable

- through a fine granulation of the data stock
  - structured data processing (transcription, tokenization, typification)
  - assignment of norm data (Q-ID, L-ID, GND, GeoNames etc.)
  - enrichment with metadata in DataCite and CIDOC CRM format
  - assignment of persistent identifiers (e.g. DOIs, Digital Object Identifiers)
- access to primary data and metadata (via interactive map, Lexicon Alpinum, API)

R_eusable

- results from F, A, I

3.1. Assignment of norm data

- **Norm data created by VerbaAlpina**

  For the 3 core entities

  - morpho-lexical types → L
  - concepts → C
  - municipalities → A

  e.g.: L1435, „babeurre (m.) (roa.)“
  C612, „ALMHÜTTE“ (*chalet*)
  A60171, „Sils in Engadin/Segl“

- **Persistent identifiers of external institutions**
  (knowledge data bases/norm data bases/reference dictionaries)

Persistent identifiers of external institutions

- **Q-IDs** of Wikidata (for concepts), partly also **L-IDs** of Wikidata (for morpho-lexical types)

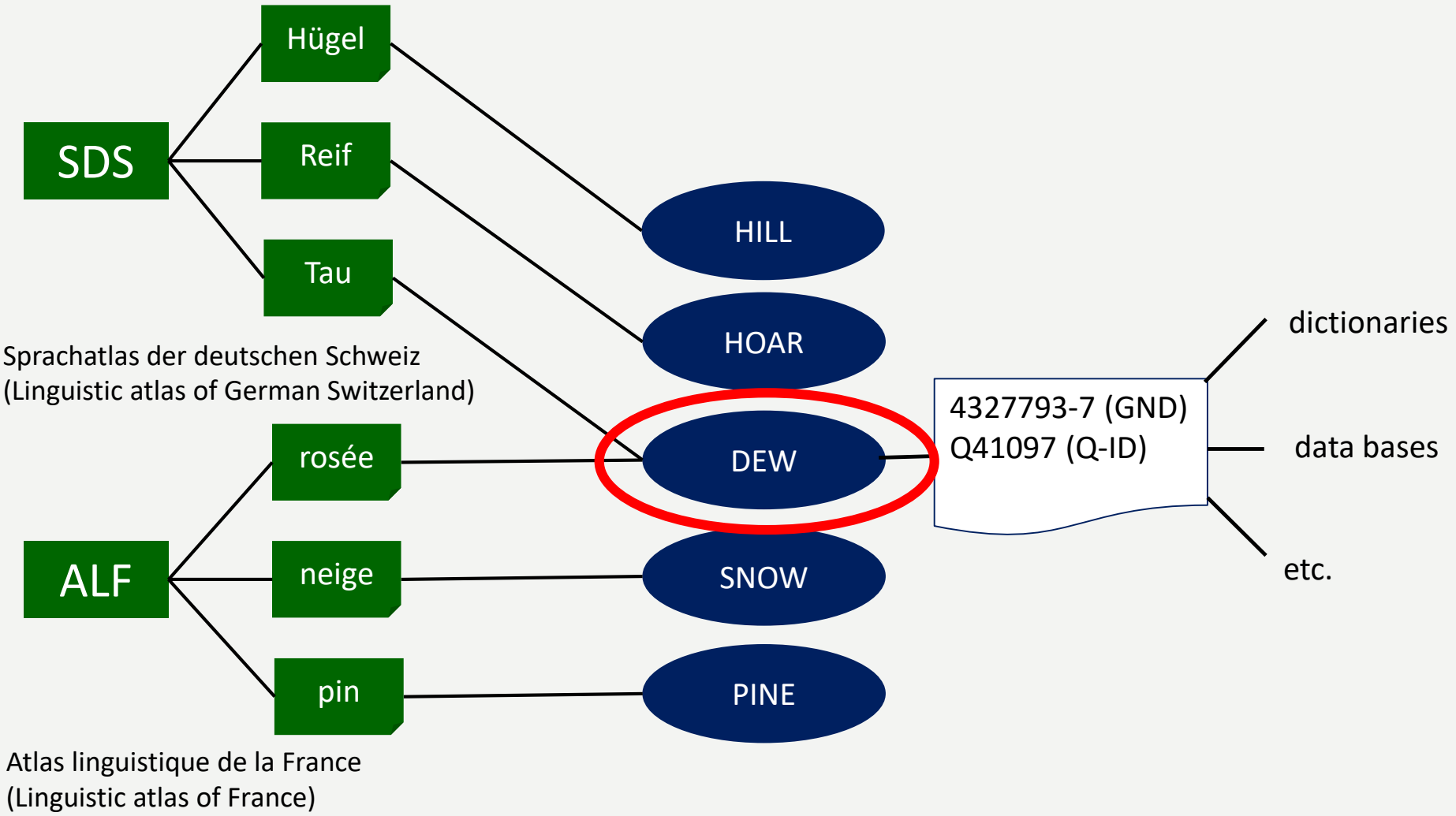
- partly **GNDs** of the German National Library (for concepts) (*Gemeinsame Normdatei*, „Integrated Authority File“)
- **GeoNames** of www.geonames.org (for municipalities)
- **ISO-Codes 639-3** (for languages)
- Identifiers of **reference dictionaries** (for morpho-lexical types + base types)
- **DOIs**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

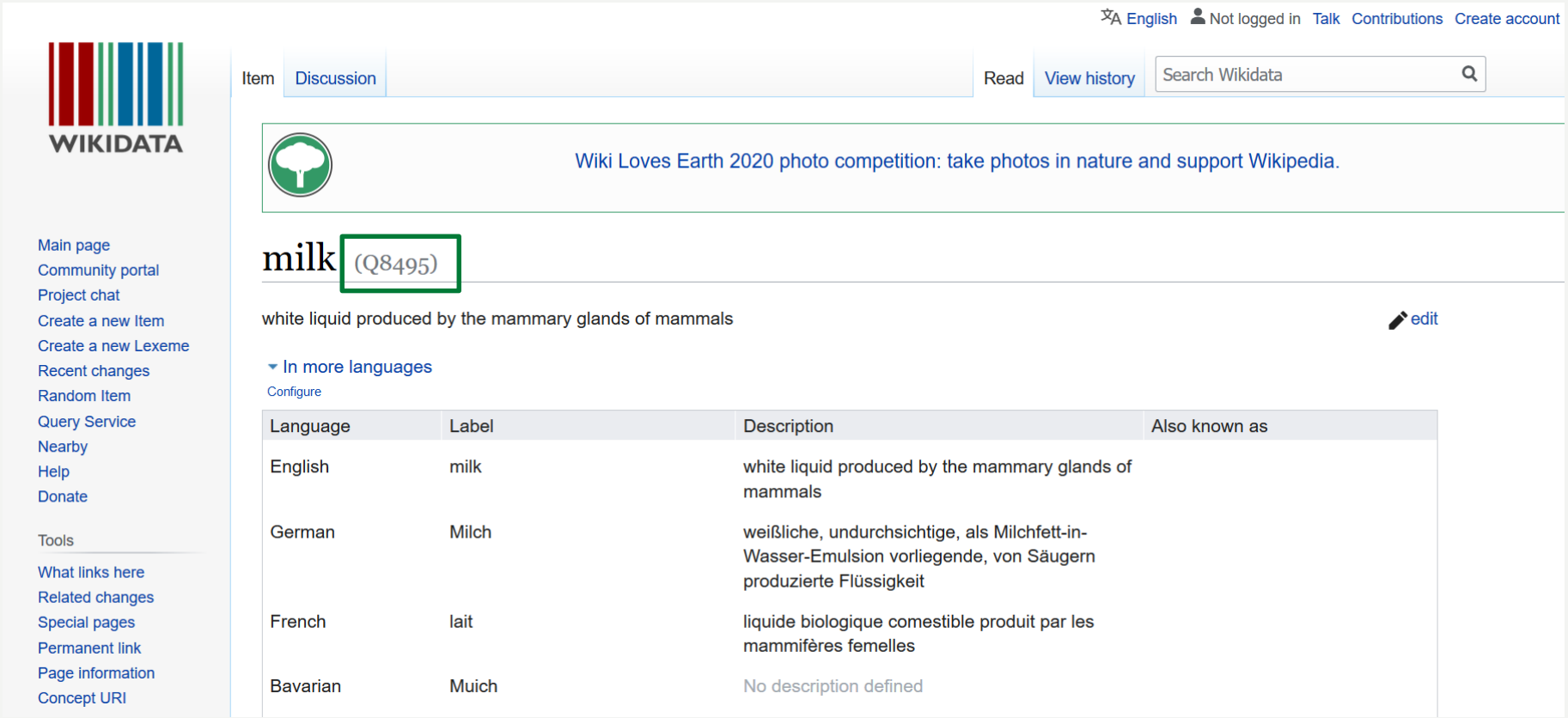
Limerick
2962943

Q54050
(Q-ID)

L73260
(L-ID)

4135744-9 (GND)


Hügel
SDS
Reif
Tau
Sprachatlas der deutschen Schweiz
(Linguistic atlas of German Switzerland)
HILL
HOAR
DEW
4327793-7 (GND)
Q41097 (Q-ID)
dictionaries
data bases
etc.
rosée
ALF
neige
pin
SNOW
PINE
Atlas linguistique de la France
(Linguistic atlas of France)



English Not logged in Talk Contributions Create account

WIKIDATA

Item Discussion Read View history Search Wikidata

Wiki Loves Earth 2020 photo competition: take photos in nature and support Wikipedia.

Main page
Community portal
Project chat
Create a new Item
Create a new Lexeme
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Tools

What links here
Related changes
Special pages
Permanent link
Page information
Concept URI

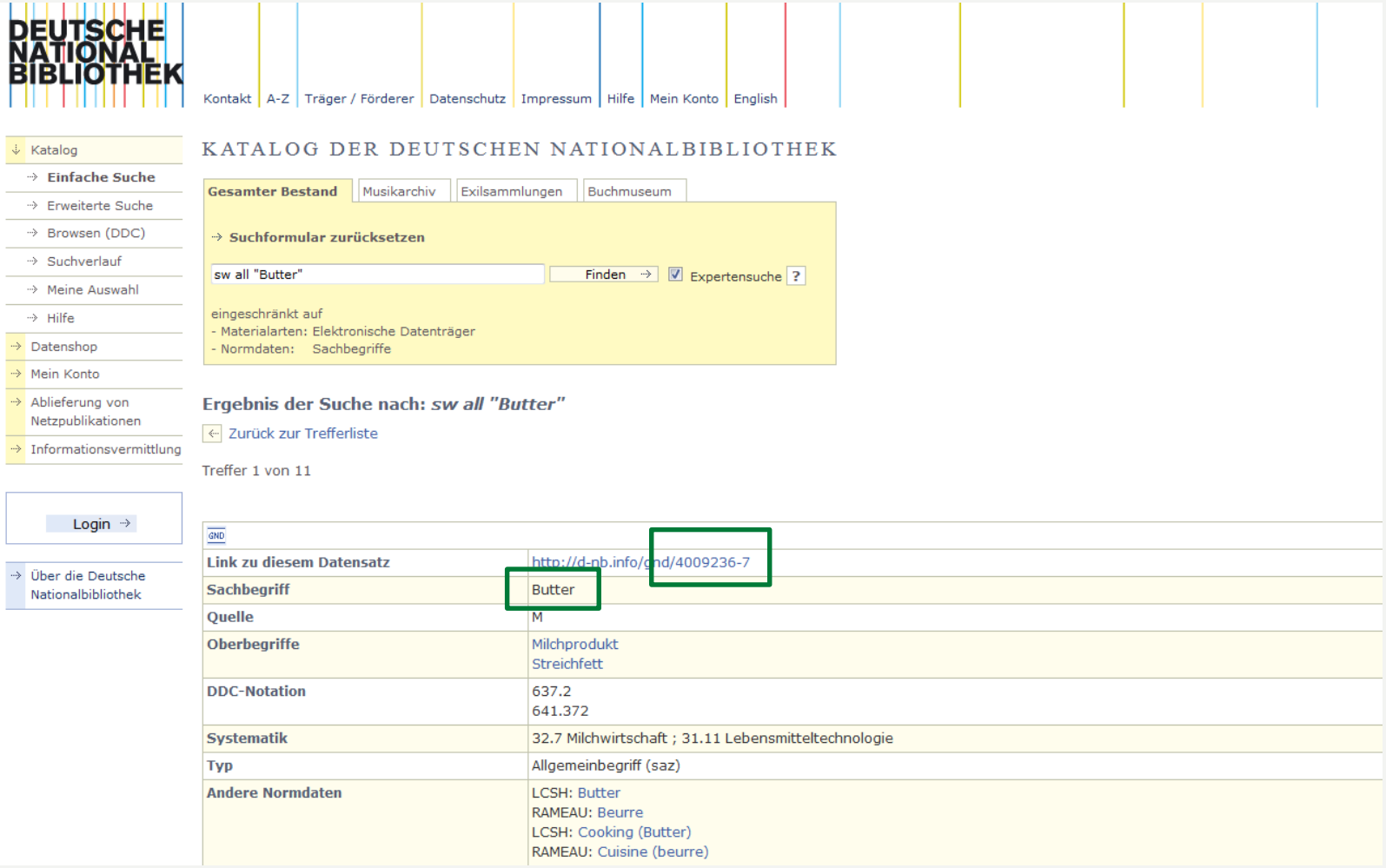
milk (Q8495)

white liquid produced by the mammary glands of mammals edit

In more languages

Configure

Language	Label	Description	Also known as
English	milk	white liquid produced by the mammary glands of mammals	
German	Milch	weißliche, undurchsichtige, als Milchfett-in-Wasser-Emulsion vorliegende, von Säugern produzierte Flüssigkeit	
French	lait	liquide biologique comestible produit par les mammifères femelles	
Bavarian	Muich	No description defined	



DEUTSCHE NATIONAL BIBLIOTHEK

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

- Katalog
- Einfache Suche
- Erweiterte Suche
- Browsen (DDC)
- Suchverlauf
- Meine Auswahl
- Hilfe
- Datenshop
- 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"*

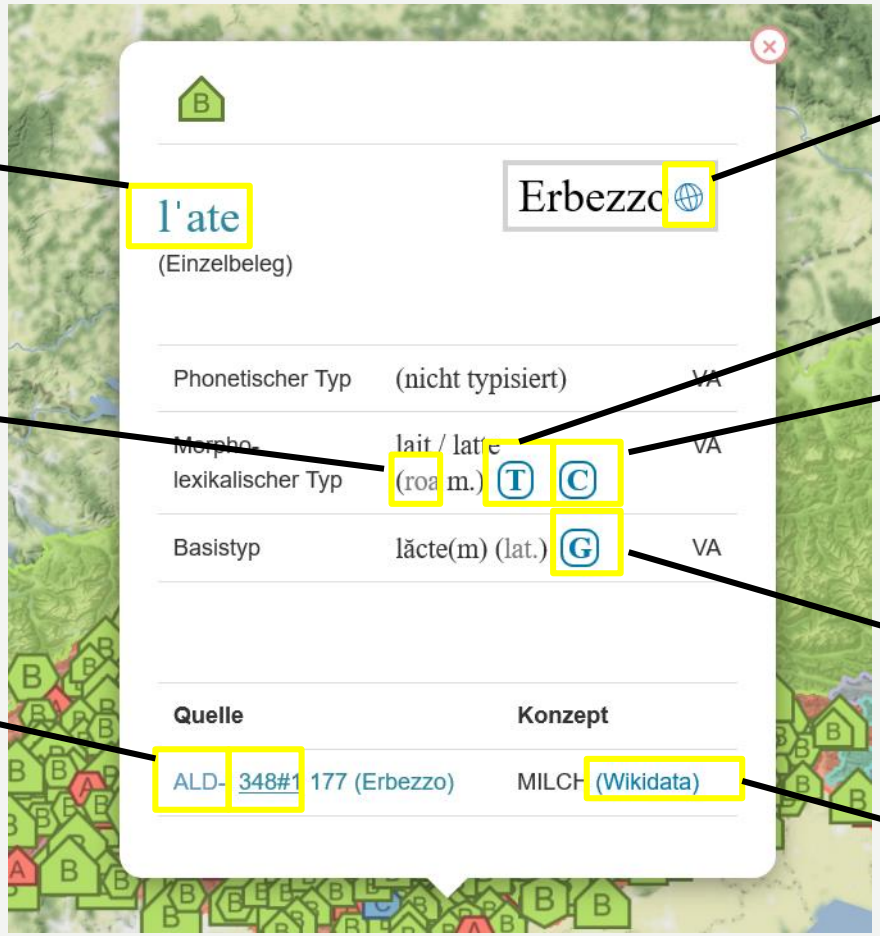
Zurück zur Trefferliste

Treffer 1 von 11

GND

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

Detail view of one specific data point



IPA

Darstellung: IPA VA

ISO Code

source + link

Goebl, Hans (Wiesbaden): Atlant linguistich dl ladin dolomitich y di dialec vejins I, vol. 1-7 (sprechend: http://ald.sbg.ac.at/ald/ald-i/index.php), 1998, vol. 1-7, Reichert

Link

GeoNames

Dictionaries

Treccani: latte

CNRTL: lait

Etymologic dictionary

Georges: lac 2, 525

Wikidata

References

- Alpine Convention. Contracting parties. https://www.alpconv.org/en/home/ [accessed 26 June 2020].
- Force11 (eds.), "The Fair Data Principles", https://www.force11.org/group/fairgroup/fairprinciples [accessed 26 June 2020]
- 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
- Thomas Krefeld & Stephan Lücke (2019): Kleinsprachen, Dialekte und die FAIR-Prinzipien (am Beispiel von VerbaAlpina), Version 1 (02.10.2019, 15:57). In: Korpus im Text, Serie A, 48389. url: http://www.kit.gwi.uni-muenchen.de/?p=48389&v=1
- 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
- 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
- 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
- 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
- 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
- M. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. O. Bonino da Silva Santos, P. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. Evelo, R. Finkers, and B. Mons. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 03 2016. doi: 10.1038/sdata.2016.18

Thank you for your attention!

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

You`re welcome to get in touch with me:

E-Mail: christina.mutter@lmu.de

Telegram: @ChristinaMutter