

Christina Mutter | Aleksander Wiatr

The Virtual Research Environment of VerbaAlpina and its Lexicographic Function

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

18th Euralex International Congress, 17-21 July 2018, Ljubljana





Outline

1) Project description

- area under investigation
- data and methodology

2) Lexicographic function

- transcription (analogue/digital data)
- tokenization
- typification
- data access (interactive map/database)



3) Approaches to sustainability

- versioning
- citability
- long-term archiving



1) Project description

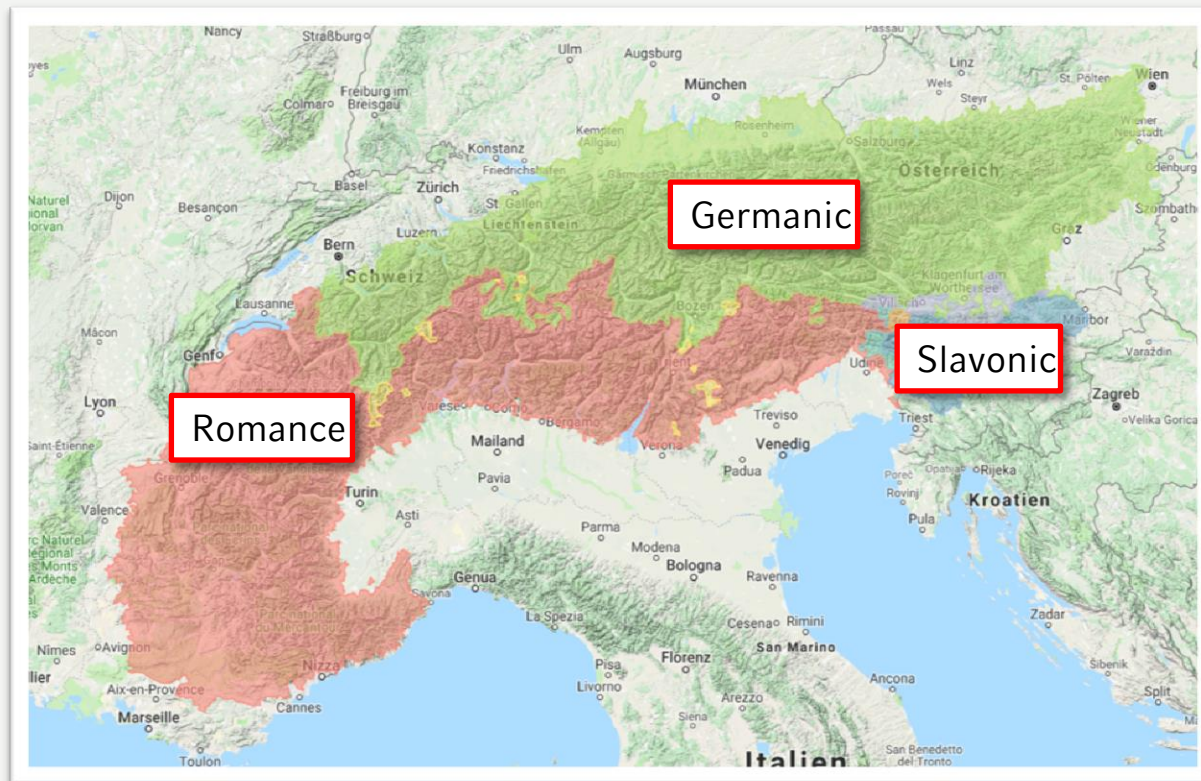
- *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit* (VerbaAlpina. The Alpine cultural region reflected through its multilingualism)
- Funded by the German Research Foundation (DFG)
- 1st term: 10/2014-10/2017, 2nd term: 11/2017-11/2020 (perspective until 2025)
- Investigation of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



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 km², 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 (Germanic, Romance and Slavonic)



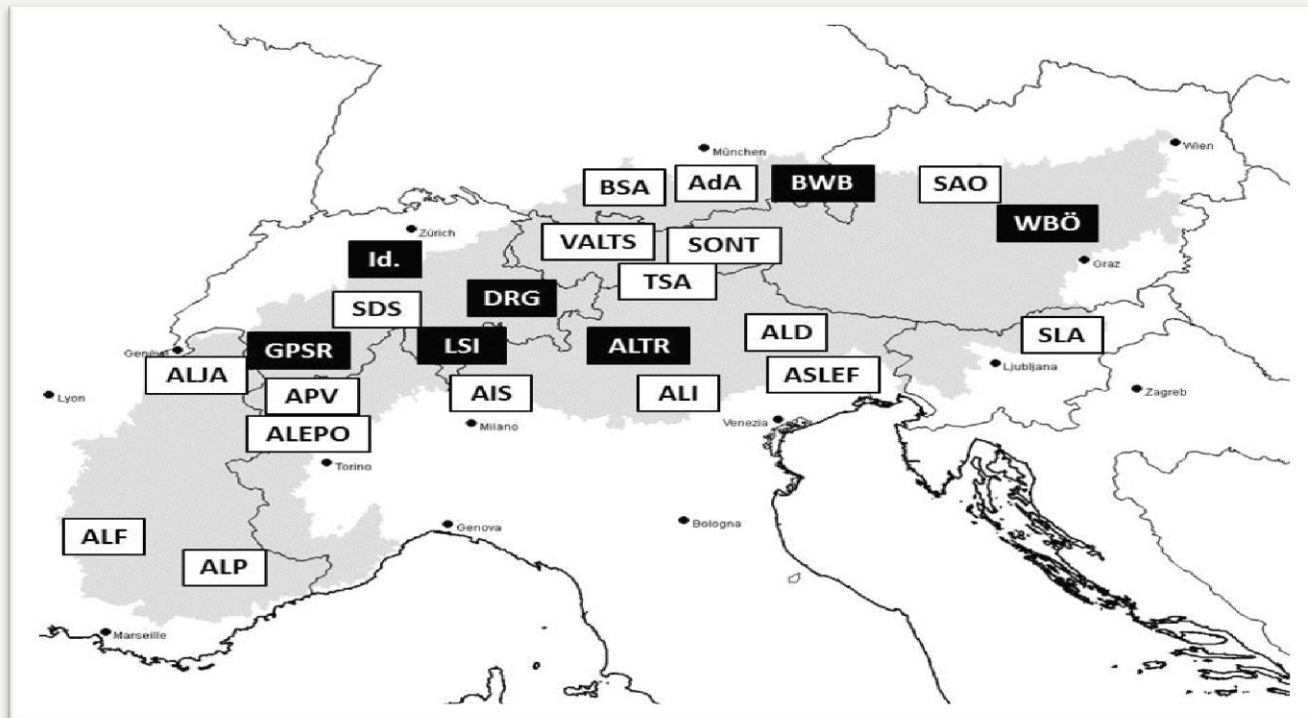


Three conceptual domains

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 <ul style="list-style-type: none"> • alpine pasture farming • milk processing 			nature <ul style="list-style-type: none"> • landscape formations • weather • fauna • flora 			modern life <ul style="list-style-type: none"> • ecology • tourism 			

Data and methodology

- Collection and analysis of data from linguistic atlases and from geo-referenced dictionaries from the past one hundred years





Data and methodology

– **Online-Crowdsourcing:** to even out, complete and correct inhomogenous data stock

→ Combination of three different approaches of digital geolinguistics:

- digitally published atlases (data gathered through traditional methods, e.g. ALD)
- atlases which document diverse languages and language families (e.g. WALIS)
- web-based atlases (e.g. AdA)



Data and methodology

- challenge:

lack of uniformity of data from individual data sources

→ unification of the different transcription systems

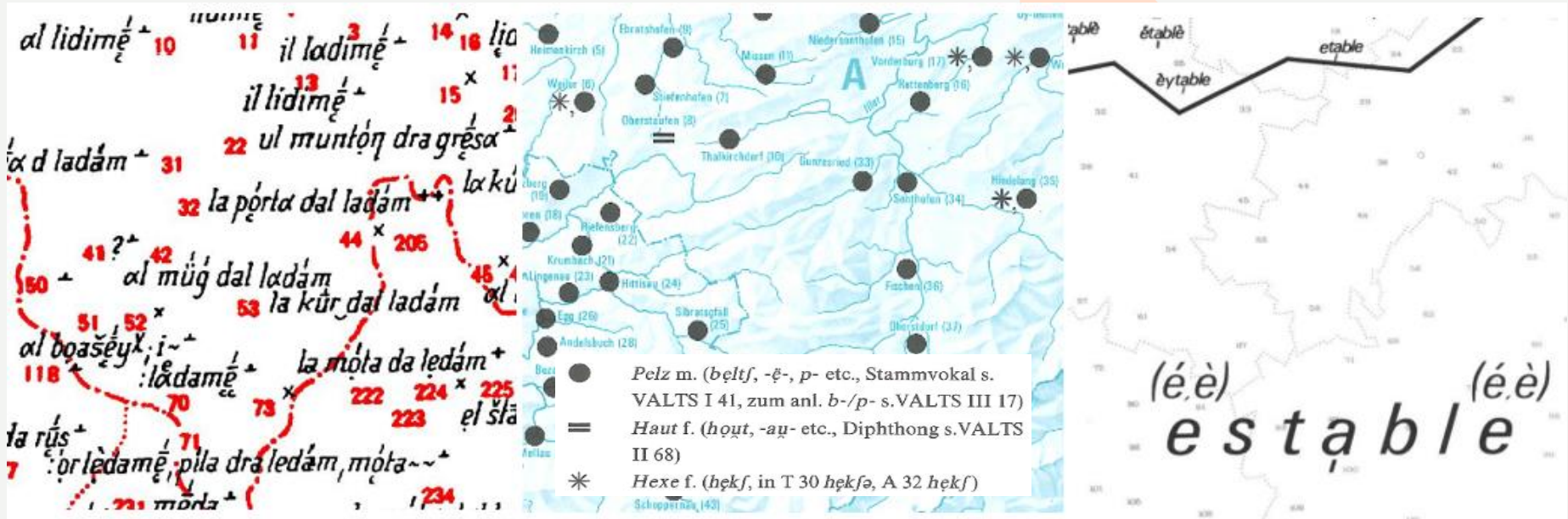
→ process of systematic data processing:

- **Transcription**

- **Tokenization**

- **Typification**

Lexicographic Process



Lexicographic Process: Transcription Tool

Meine Websites VerbaAlpina 14 + Neu Willkommen, AWiatr

1 von 1 20% 1 von 7 Automatischer Zoom

VI. BAND

IL BURRO

BUTTER — BEURRE

Ga. Nöb. 255 - ALF 130 - Bloch I -

BYun. 141

28,5 - 19,13 - 7,4

Legende:

Die Butter wird in den Poenavien Flehens, wo Di produziert oder wo das Öl beim Nöchen verwendet wird, selten oder gar nicht gebraucht. In den südlichen Mittelalpen, seltener Südalpen, ist die Butter als neuerdings eingeführtes Produkt bereits bekannt, aber dessen Verwendung in der Küche bleibt bei der kostenmäßigen Berücksichtigung wenig üblich. Einen Einblick in die tatsächlichen Verhältnisse gewährt uns auch die X. 1205 (Fabes), 1206 (abwachten), 1208 (Butterfressen).

Äquivalenz mit Butterinhalt

(duro rasciutto in una forma di pera di cacio)

Die Butter ist in einem bierenförmigen Mehl eingeschlossen. Vgl. Sitze u. X. II, 1206 (log)

207 a mandiflo 223 a mandiflo

216 lu butiro, la 225 a mandiflo

mandiflo (gelm) 226 lu mandiflo, u ba-

228 lu mandiflo ifira, * - fra (122)

226 lu mandiflo sevarinai mit But-

226 lu mandiflo kerinai, 2 Butter-

229 lu mandiflo inur inhalt)

229 lu mandiflo für den Gebrauch 227 lu mandiflo

SPRACH- UND SACHLICHE TRÄGER UND DER SÜDSCHWEIZ. KARTE 1207.

Konvention für die Transkription aller für VA relevanten Sprachtanten

Version vom 05.12.2016

Wir unterscheiden zwischen Basiszeichen und Diakritika.

Basiszeichen befinden sich auf der Grundlinie. Alle Zeichen, die sich nicht auf der Grundlinie befinden, werden als Diakritika betrachtet.

Basiszeichen, die in der ASCII-Tabelle vorhanden sind, werden beibehalten (= alle lateinischen Buchstaben, "nicht" deutsche Umlaute)

Folgende Basiszeichen, die nicht in der ASCII-Tabelle enthalten sind, werden wie folgt transkribiert (in der hinteren Spalte steht der Sprachcode, in dem das jeweilige Basiszeichen bzw. Diakritikon zum ersten Mal „entdeckt“ wurde)

α	Griechisches Alpha	a1	AIS
α	spiegelverkehrtes a	a2	TSA
β	Griechisches Beta	b1	AIS
β		b2	SLA
γ	Griechisches Gamma	g1	AIS
γ		g2	ALF
g		g3	SLA
?		g4	SLA
δ	Griechisches Delta	d1	AIS
δ		d2	SLA
i		i1	SLA
i		i2	SLA
i		i3	SLA
θ	Griechisches Theta	t1	AIS
θ		t2	SLA

AIS#1207 3 (il pane di burro)

AIS 1207 - il pane di burro - Informant_Nr 1 (Brigels-Breil)

Transkription Eintragen <vacat> Problem **Beleg**

Zugewiesene Konzepte **BUTTERBALLEN**



Lexicographic Process: Beta Code

Diacritics

Base sign

Diacritics

	erhebung	beta	ipa	Hex_ipa	Art
	AIS			 	Trennzeichen
	AIS	/	'	ˈ	Akzent
Beta Co	AIS	//	'	ˈˈ	Akzent
	AIS	\	,	ˌ	Akzent
	AIS	a	a	a	Vokal
	AIS	a!	a	a	Zeichen
	AIS	a%	añ	ah̆	Zeichen
	AIS	a%0	añ	ah̆	Zeichen
	AIS	a&	e	ɐ	Zeichen
	AIS	a(o	ɑ	Zeichen
	AIS	a*	o	ɑ	Zeichen
	AIS	a(-	o:	ɑː	Zeichen
- No l	AIS	a(-{i}	oi	ɑːĭ	Zeichen
	AIS	a(-	e:	ɐː	Zeichen
- Inst	AIS	a(:	æ	æ	Zeichen
	AIS	a({i)-	oi	ɑːĭ	Zeichen
- The	AIS	a(-	e:	ɐː	Zeichen
tran	AIS	a)	o	ɑ	Zeichen
	AIS	a)-	o:	ɑː	Zeichen
	AIS	a*	a	a	Zeichen
- For	AIS	a*(o	ɑ	Zeichen
	AIS	a*-	a:	aː	Zeichen
	AIS	a*~?	a:	aː	Zeichen
allo	AIS	a*:	æ	æ	Zeichen
	AIS	a*?~	a:	aː	Zeichen
	AIS	a*	e	ɐ	Zeichen
	AIS	a-	a:	aː	Zeichen
	AIS	a(-	o:	ɑː	Zeichen



? \ \

ibed

e





Lexicographic Process: Tokenization

Attestation in Beta Code	Attestation in IPA	Concept
una1 mu:g/a1 da1 va/c)/	una mydʒa da v ¹ atɕ	HERD OF COWS
TOKENIZATION		
una1	una	ARTICLE
mu:g/a1	mydʒa	HERD
da1	da	PREPOSITION
va/c)/	v ¹ atɕ	COW



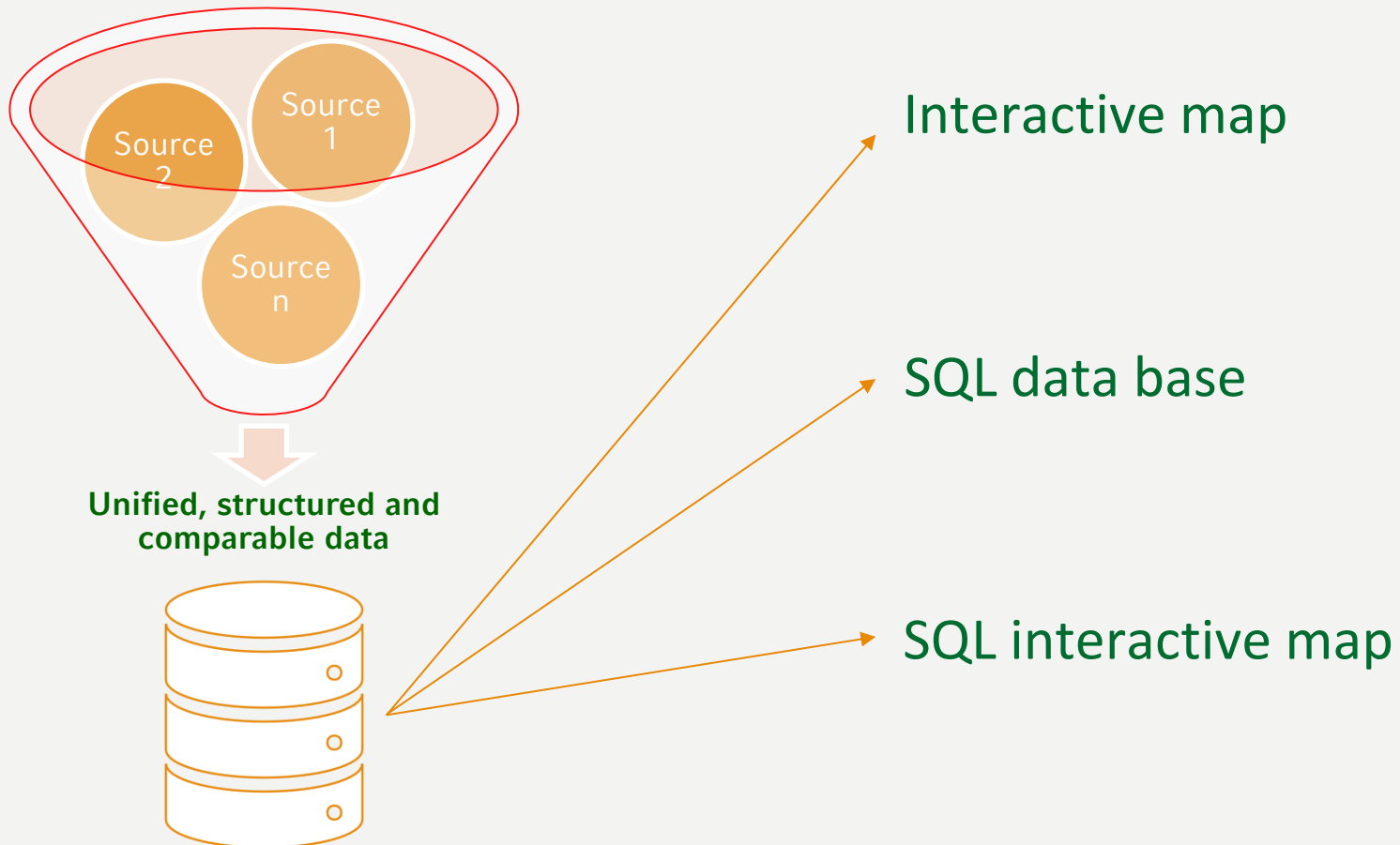
Lexicographic Process: Typification

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

	<i>barga</i>	<i>barg</i>	<i>margun</i>	<i>bargun</i>
language family	roa	roa	roa	roa
PoS	noun	noun	noun	noun
affix	-	-	+	+
gender	f	m	m	M
morpho-lexical type	1	2	3	3

*barga/*barca

Lexicographic Process



Data access: interactive map

Bitte Suchbegriff eingeben...

Kartographische Darstellung **Physisch** Hexagonal

Sprachliche Kerndaten

- Basistypen
- Morpho-lexikalische Typen
- Phonetische Typen
- Konzepte

Sprachbezogene Peripherie

- Informanten
- Ergänzende Daten
- Flächen und Regionen

Legende

Synoptische Karten

Menü schließen

Two view modes:

- Physical & hexagonal

Diverse filters:

- Onomasiologic
- Semasiologic
- Peripheral data

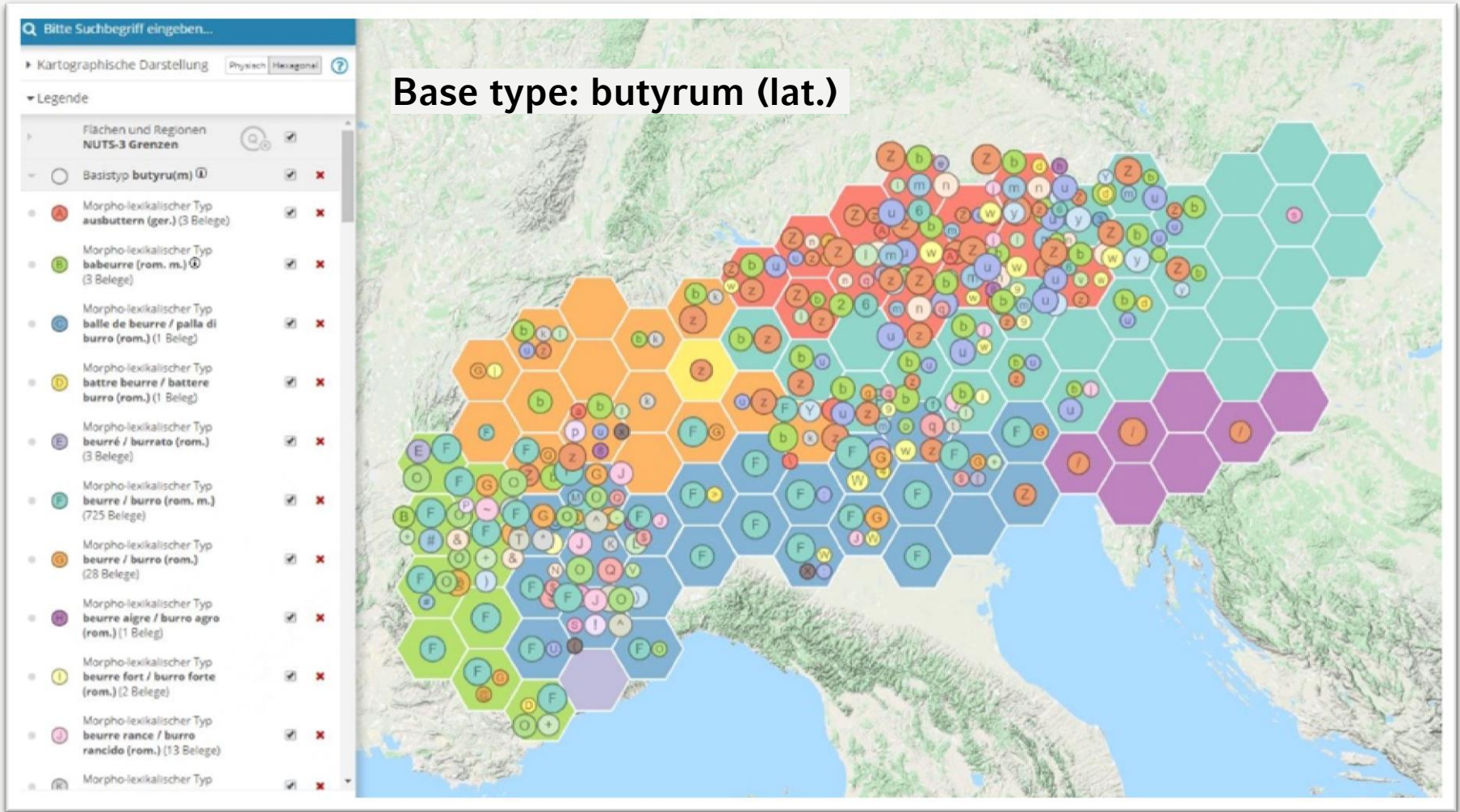


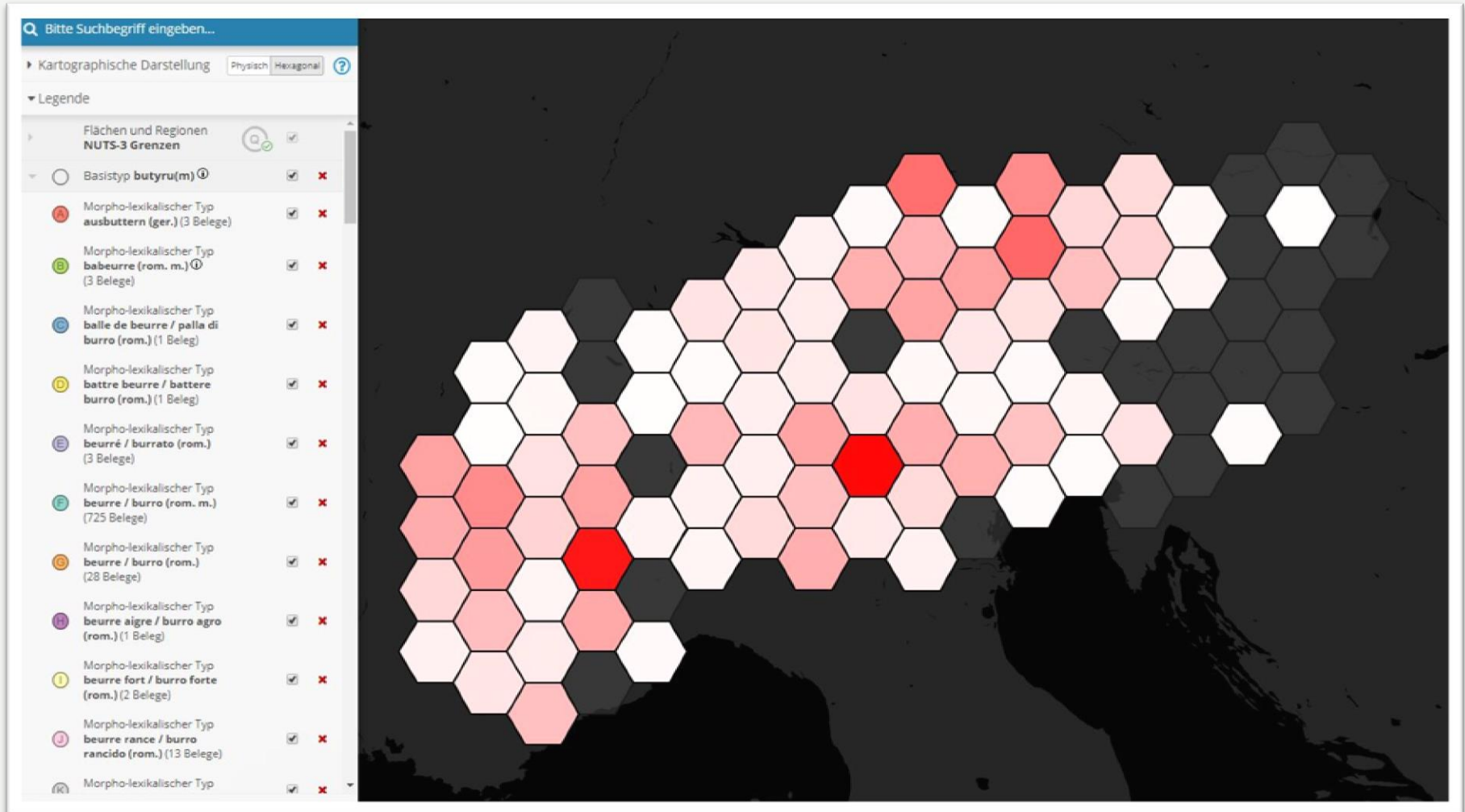
https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=133

Tools:

- Qu...
- Sy...

Base type: butyrum (lat.)





Data access: SQL-Queries

Id_Beleg	Beleg	Quelle_Beleg	Name_Konzept	Beschreibung_Konzept	Breitengrad	Laengengrad	Gemeinde	Publikationsjahr	Erhebungsjahr	Sprache_Infor
281	mĕjřǎ	ALF#SUPP_37#14#889#Barcelonette (B...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,650463	44,3863	Barcelonette	1897-1900	(NULL)	rom
281	mĕjřǎ	ALF#SUPP_37#14#889#Barcelonette (B...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,650463	44,3863	Barcelonette	1897-1900	(NULL)	rom
282	moeōs	ALF#SUPP_37#14#956#Sot (Haute-Sav...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,838646	46,045357	Sot-Fer-à-Cheval	1897-1900	(NULL)	rom
283	moeōs	ALF#SUPP_37#14#967#Chamonix (Haut...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,869433	45,923697	Chamonix-Mont-Blanc	1897-1900	(NULL)	rom
284	pāvĭo:	ALF#SUPP_37#14#967#Chamonix (Haut...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,869433	45,923697	Chamonix-Mont-Blanc	1897-1900	(NULL)	rom
284	pāvĭo:	ALF#SUPP_37#14#967#Chamonix (Haut...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,869433	45,923697	Chamonix-Mont-Blanc	1897-1900	(NULL)	rom
285	sǎĕ	ALF#SUPP_37#14#967#Chamonix (Haut...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	6,869433	45,923697	Chamonix-Mont-Blanc	1897-1900	(NULL)	rom
286	mĕjřo	ALF#SUPP_37#14#982#Maisette = Mais...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	7,1266143	44,9268317	Perrero	1897-1900	(NULL)	rom
286	mĕjřo	ALF#SUPP_37#14#982#Maisette = Mais...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	7,1266143	44,9268317	Perrero	1897-1900	(NULL)	rom
287	mĭǎĭdo	ALF#SUPP_37#14#982#Maisette = Mais...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	7,1266143	44,9268317	Perrero	1897-1900	(NULL)	rom
287	mĭǎĭdo	ALF#SUPP_37#14#982#Maisette = Mais...	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	7,1266143	44,9268317	Perrero	1897-1900	(NULL)	rom
289	tĕĉo	AIS#1192#1#1#Brgels-Brel	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,049995	46,766663	Brel/Brgels	1928-1940	(NULL)	rom
289	tĕĉo	AIS#1192#1#1#Brgels-Brel	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,049995	46,766663	Brel/Brgels	1928-1940	(NULL)	rom
291	tĕĉo	AIS#1192#1#3#Pitasch	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,216672	46,716663	Duvin	1928-1940	(NULL)	rom
291	tĕĉo	AIS#1192#1#3#Pitasch	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,216672	46,716663	Duvin	1928-1940	(NULL)	rom
293	tĕĉo	AIS#1192#1#5#Ems-Domat	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,449999	46,833329	Domat/Ems	1928-1940	(NULL)	rom
293	tĕĉo	AIS#1192#1#5#Ems-Domat	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,449999	46,833329	Domat/Ems	1928-1940	(NULL)	rom
295	tĕo	AIS#1192#1#7#Ardez	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	10,200002	46,766666	Ardez	1928-1940	(NULL)	rom
295	tĕo	AIS#1192#1#7#Ardez	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	10,200002	46,766666	Ardez	1928-1940	(NULL)	rom
297	tĕĉo	AIS#1192#1#11#Surrhein (Somvix)	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	8,933332	46,71666	Sumvitg	1928-1940	(NULL)	rom
297	tĕĉo	AIS#1192#1#11#Surrhein (Somvix)	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	8,933332	46,71666	Sumvitg	1928-1940	(NULL)	rom
299	tĕĉo	AIS#1192#1#13#Vrin	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,083334	46,649994	Vrin	1928-1940	(NULL)	rom
299	tĕĉo	AIS#1192#1#13#Vrin	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,083334	46,649994	Vrin	1928-1940	(NULL)	rom
301	tĕo	AIS#1192#1#14#Dahn (Práz)	SENNHÜTTE	GEBÄUDE, EINFACH, AUF DER ALM ZUR...	9,400005	46,733328	Sam	1928-1940	(NULL)	rom

- Direct access to the data base: structured & comparable data
- Textual mode
- Filtered and prepared data sets can be downloaded for further evaluation

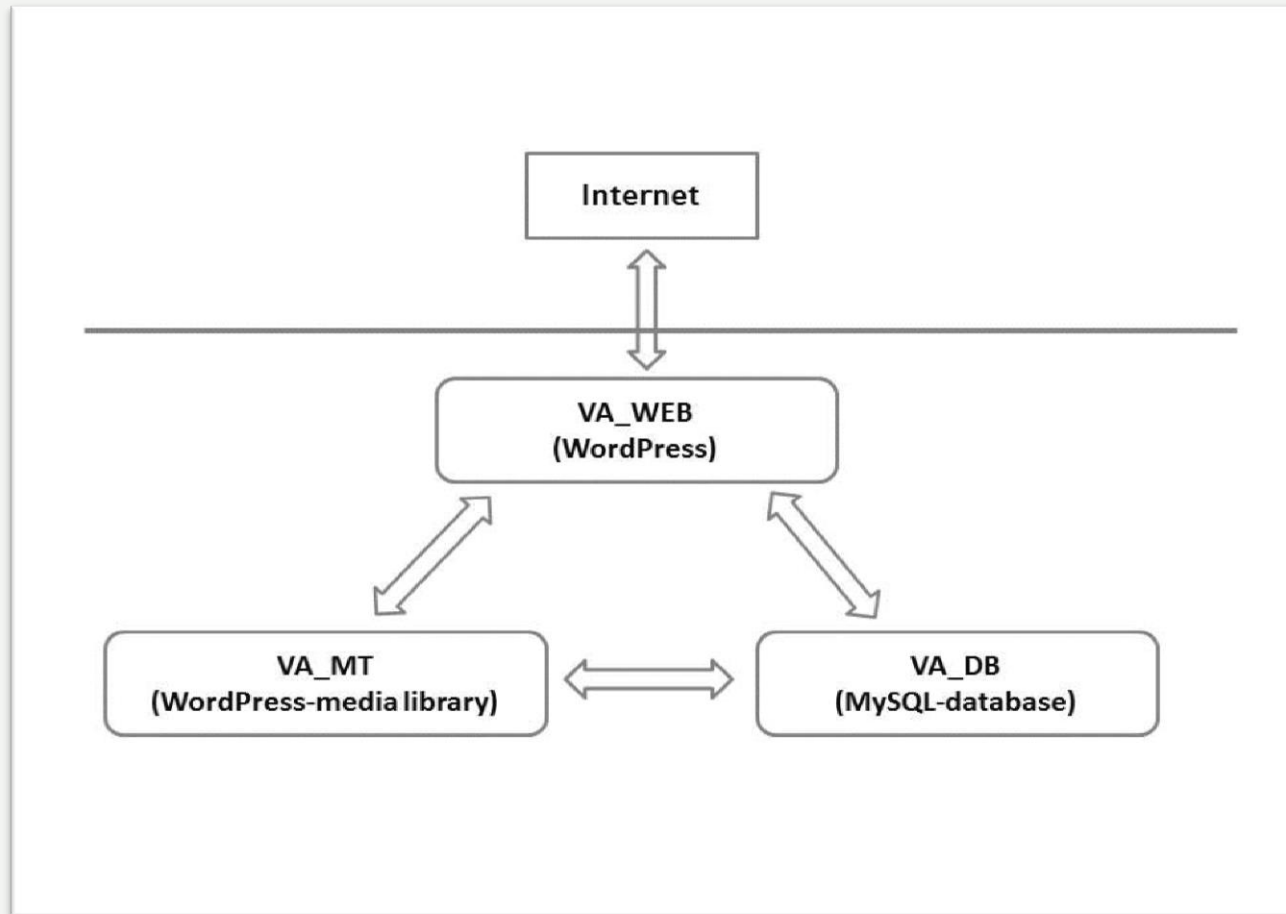


Approaches to sustainability

- 1) **Versioning**
- 2) **Citability**
- 3) **long-term archiving**



1) Versioning: The modules of VerbaAlpina





1) Versioning: The modules of VerbaAlpina

- VA_DB and VA_WEB are “frozen” every six months (15/06 and 15/12)
- versioning numbers for frozen copies (scheme: [year]/[sequence number], e.g. 18/1)
- every productive VA version is named XXX
- Possibility to switch between “productive” and “frozen” versions



2) Citability

- made possible by the versioning process
- date of last access is not necessary anymore
(cited versions are stable)
- cite in the following way:

VerbaAlpina (VA), <http://www.verba-alpina.gwi.uni-muenchen.de>,[version]

eg.: VerbaAlpina (VA), <http://www.verba-alpina.gwi.uni-muenchen.de>, 15/1

- graphic contents may also be cited: individual URLs for pages and pop-up windows



3) Long-term archiving

- Multiple copies of the project data archived by several institutions
 - currently: IT-Gruppe Geisteswissenschaften of Munich University (LMU Center for Digital Humanities) and archive.org
 - In the long term: University Library of Munich
- Documentation: data structuring, logical relationships between data and data categories, character encoding
- conversion from Google Maps to Leaflet is planned at the latest until the end of 2019

Thanks for your attention!

