



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

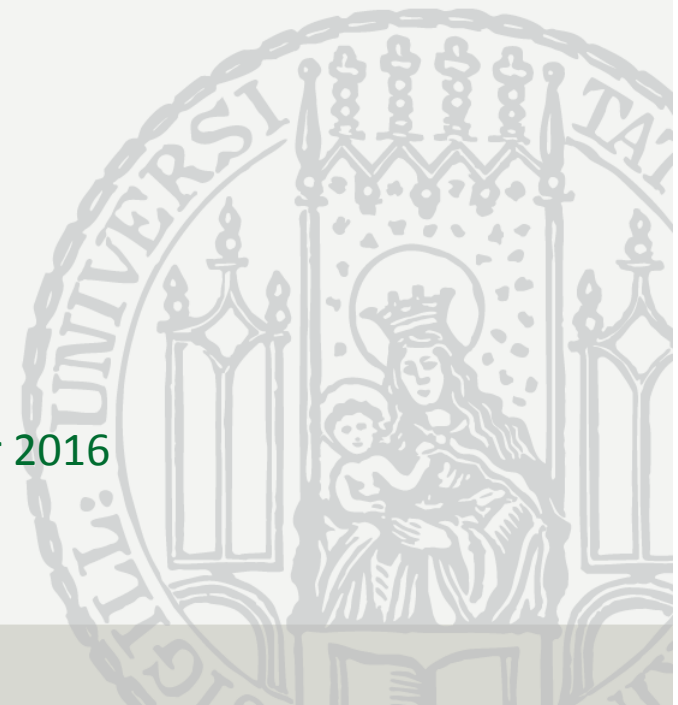
VERBA
ALPINA

Stephan Lücke

VerbaAlpina: Projektpräsentation

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

dhmuc-Workshop “Digitale Daten”, 28./29. Januar 2016





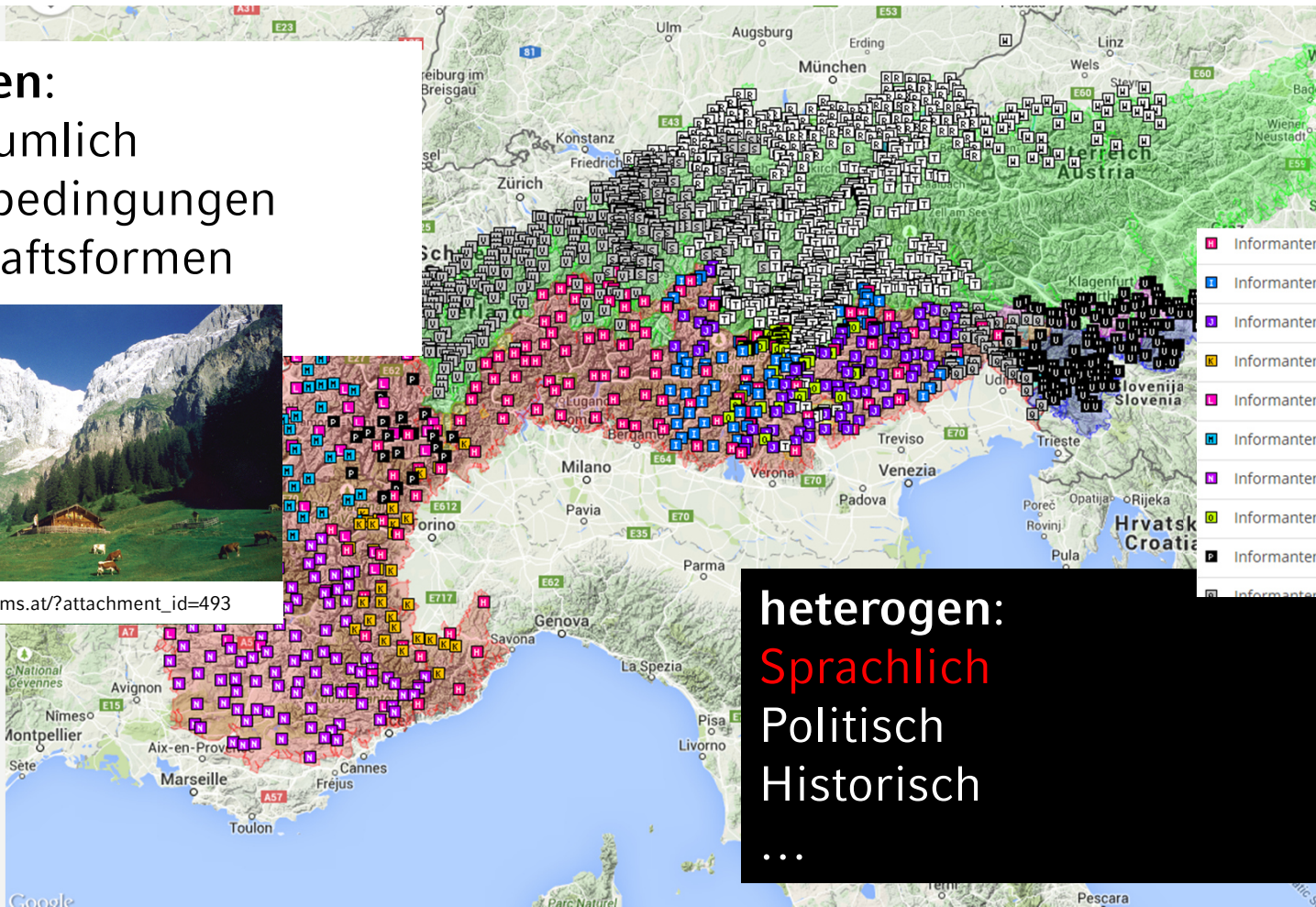
Übersicht

- 1) Grundidee
- 2) Eckdaten (administrativ)
- 3) Methode und Konzept
 - Arbeitsschritte
 - Typisierung
 - Crowd-Sourcing
- 4) Die Technik
- 5) Nachhaltigkeit



Das Untersuchungsgebiet: Homogen vs. heterogen

homogen:
 naturräumlich
 Lebensbedingungen
 Wirtschaftsformen



- Informanten AIS
- Informanten ALD-I
- Informanten ALD-II
- Informanten ALEPO
- Informanten ALF
- Informanten ALJA
- Informanten ALP
- Informanten ALTR
- Informanten APV
- Informanten ASLFE

heterogen:
 Sprachlich
 Politisch
 Historisch





Im Fokus

LEXIK

Bezeichnung ↔ KONZEPT → Raum & Zeit

(Phase 1: Almwirtschaft)

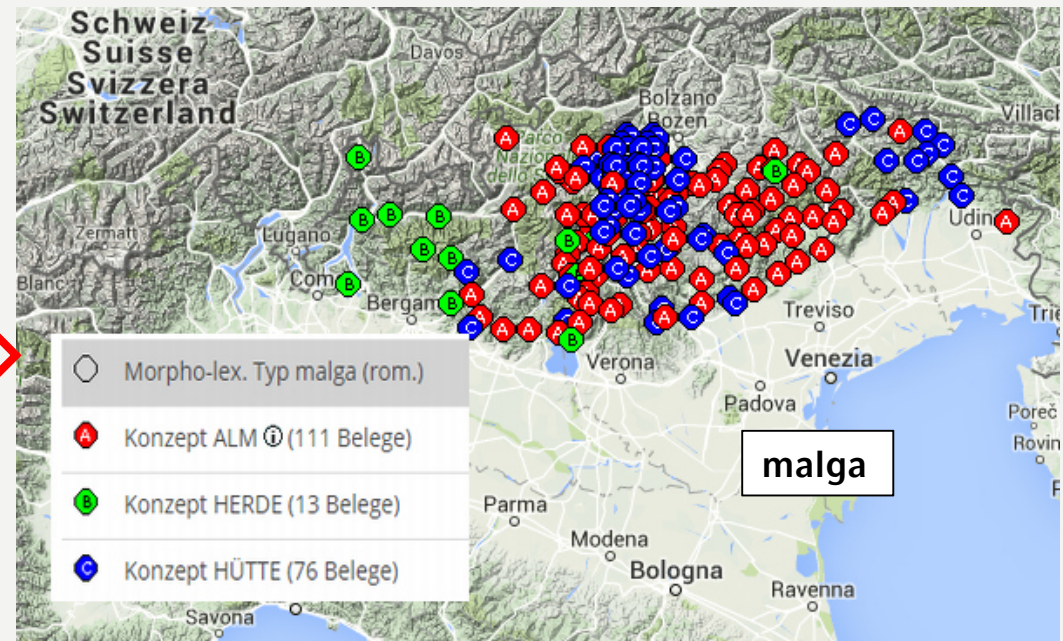


Semasiologische Perspektive, georeferenziert: Bezeichnung → KONZEPT

Bezeichnung	KONZEPT
malga	ALM
malga	ALMHÜTTE
malga	KUHHERDE

Vergleich:
Lexikon

Georeferenzierung
→ Verteilung im Raum



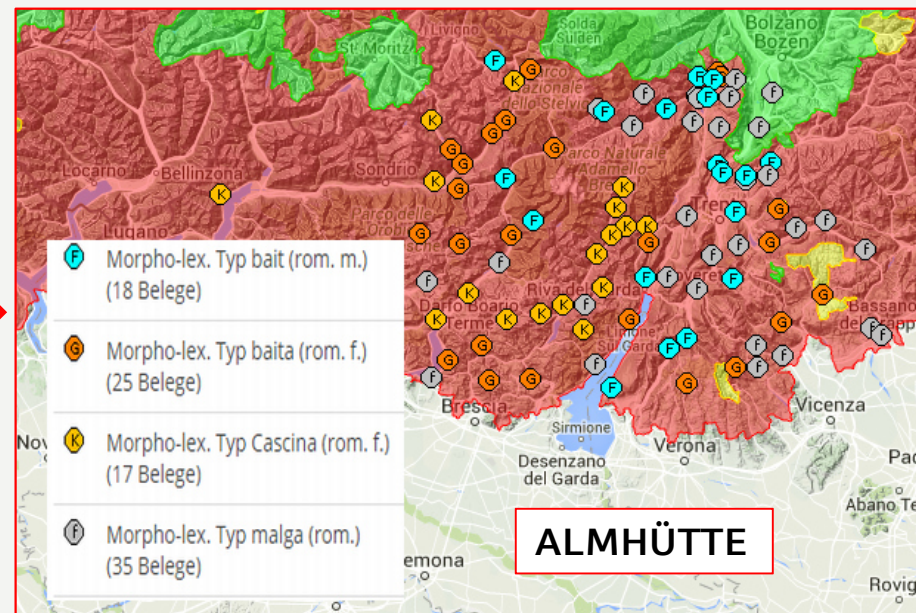


Onomasiologische Perspektive, georeferenziert: KONZEPT → Bezeichnung

KONZEPT	Bezeichnung
ALMHÜTTE	bait (romanisch, maskulin)
ALMHÜTTE	baita (romanisch, feminin)
ALMHÜTTE	cascina (romanisch, f)
ALMHÜTTE	malga (romanisch, f)

Vergleich:
Sprachatlas

Georeferenzierung
→ Verteilung im Raum

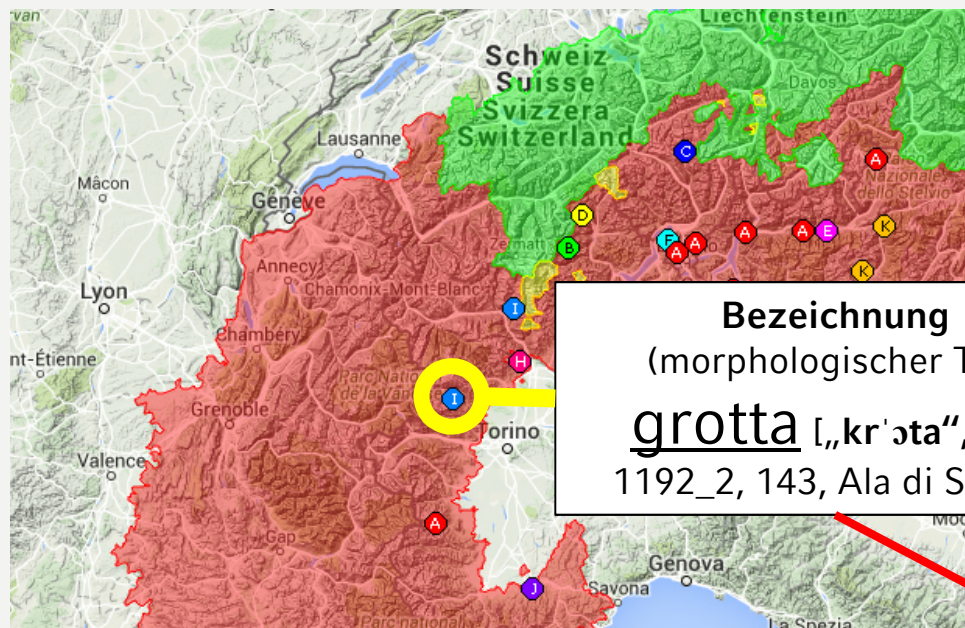


ALMHÜTTE





Die „dritte Dimension“: Die Sache / Der Referent / Das konkrete Objekt



Bezeichnung
(morphologischer Typ:
grotta [„kr'ɔta“, AIS
1192_2, 143, Ala di Stural])

KONZEPT
(immateriell: MILCH-
UND KÄSEKELLER,
SEPARATES
HÄUSCHEN)



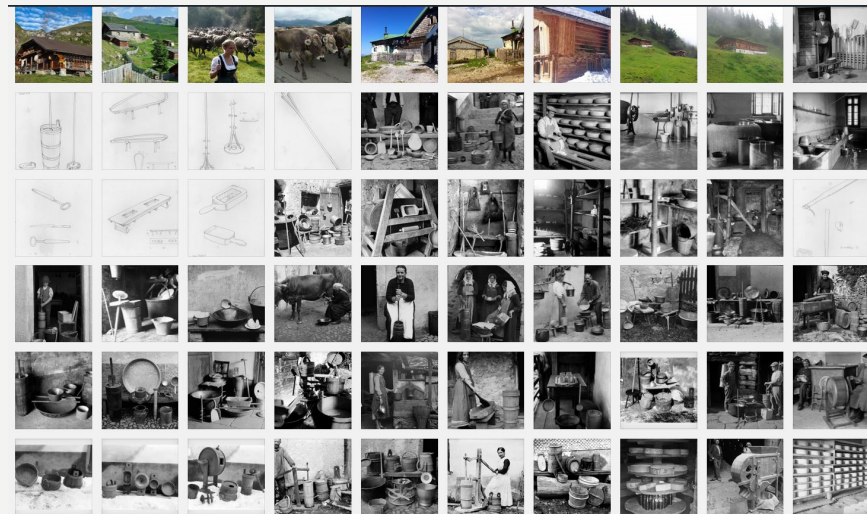
Foto: Claudine Remacle

Referent
(materiell; konkrete,
individuelle Instanz)



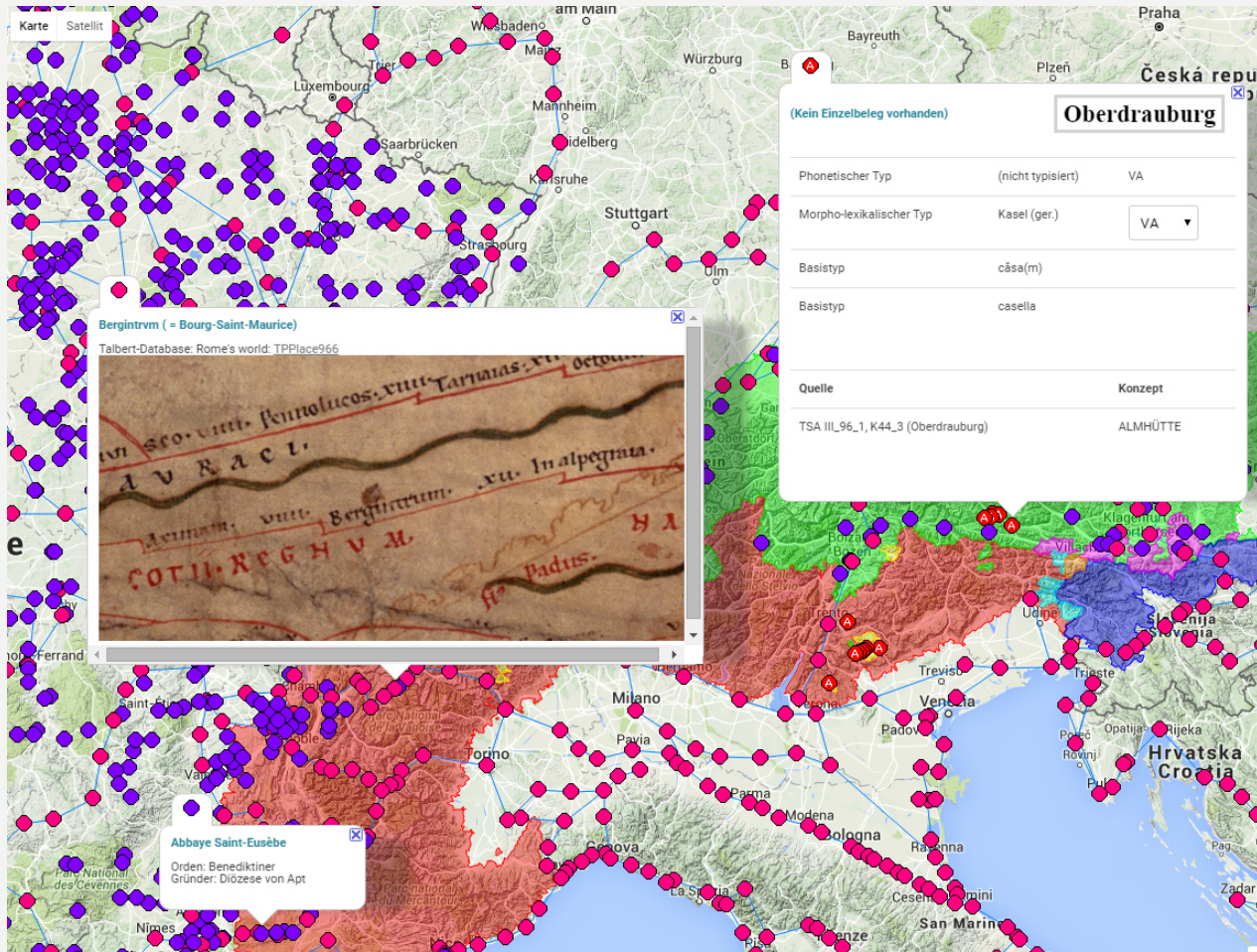
Mediathek (Sammlung von Referenten)

- Sammlung von Abbildungen zu KONZEPTEN von VerbaAlpina
- Systematische Verknüpfung der Bilder über Bezeichnungen mit KONZEPTEN





Erweiterter Kontext: Sprachexterne Daten (z.B. Geschichte)



- Antike
- Mittelalter
- ...

Beispiel Kombination

Sprache:

- Basistyp „casella“

Sprachextern:

- [Tabula Peutingeriana](#)
- Klöster

Team: 10 Personen

Projektleitung

- Prof. Dr. Thomas Krefeld (Institut für Romanische Philologie)
- Dr. Stephan Lücke (IT-Gruppe Geisteswissenschaften)

MitarbeiterInnen

- Susanne Oberholzer (wiss. Koordination, Germanistik/Romanistik)
- Markus Kunzmann (Germanistik)
- Aleksander Wiatr (Romanistik, Slowenisch)
- Florian Zacherl (Informatik)
- Mona Neumeier, Giorgia Grimaldi, Christina Mutter, Beatrice Colcuc (wissenschaftliche Hilfskräfte)

DFG-Förderung

- Förderung durch die DFG seit 2014 als “Langfristvorhaben”
- 1. Laufzeit: 2014–2017 (Perspektive bis 2025)
- Kombination von (Geo-)Linguistik mit *Digital Humanities*
- *Projektphasen:*
 - „Kultur“ (**Almwirtschaft**, volkstümliche Medizin, traditionelle Küche) (2014-2017)
 - „Natur“ (Landschaftsformationen, Wetter, Fauna, Flora) (2017-2020)
 - „Moderne Lebenswelt“ (Ökologie, Tourismus) (2020-2023)

VA-Projektpartner

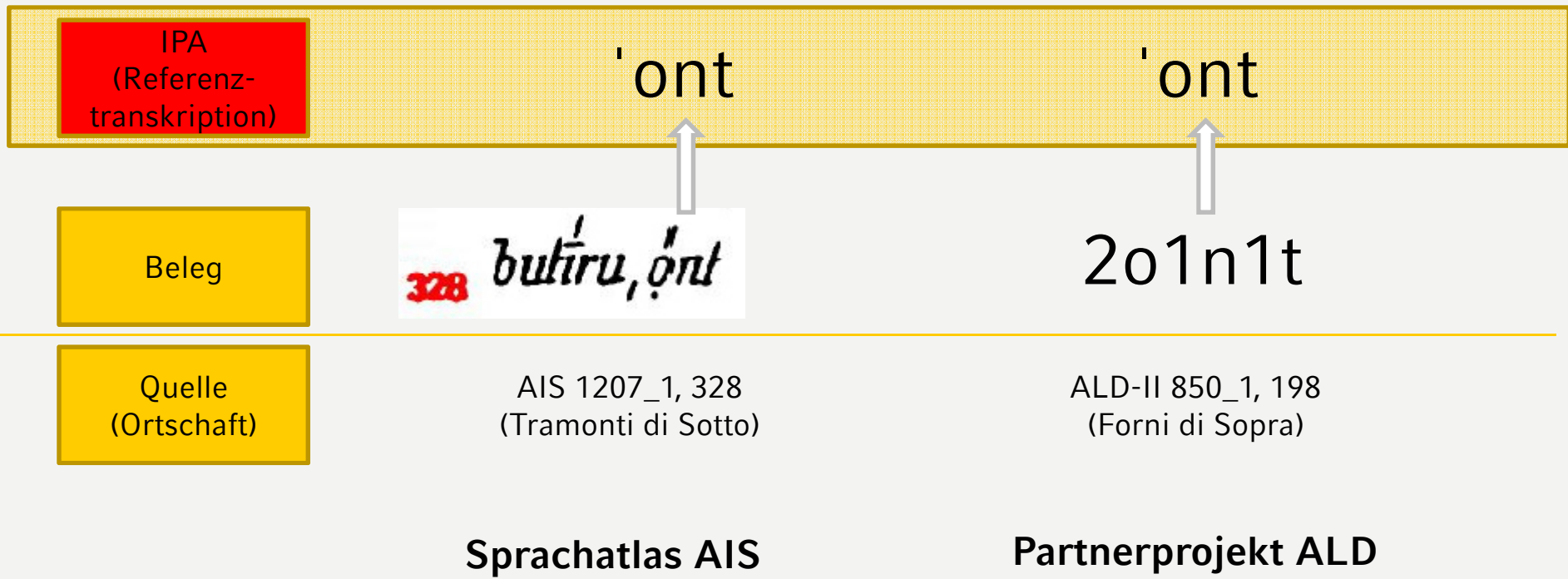
- Derzeit 30 Projektpartner im In- und Ausland
- Institutionen und Einzelpersonen
- Datenaustausch im Rahmen konkreter Kooperationsvereinbarungen
- Wissenschaftlicher Beirat



Arbeitsschritte

- Digitalisierung und strukturierte Erfassung von Sprachatlanten und Wörterbüchern mit georeferenziertem Sprachmaterial
- Georeferenzierung (Lat/Long)
- Abgleich und Übernahme von Sprachmaterial aus Partnerprojekten
- Ergänzung, Ausgleich und Validierung des gesammelten Materials durch die „Crowd“

Datenausgleich: Transkriptionssysteme (→ IPA)





Datenausgleich: Typisierung

Basistyp	tegia	tegia	tegia	tegia
Morph. Typ (Referenztyp) Lexikon	tegia	tegia	Teie(n)	
Phon. Typ			ta:iə	
Äußerung des Sprechers	t'i:dza	t'e:ɖza	?	
Quelle (Ortschaft)	AIS 1192_1, 27 (Latsch)	AIS 1192_1, 5 (Ems-Domat)	VALTS IV_73_1, T12 (St. Anton am Arlberg)	



Crowd-Sourcing I

- Zweck: Ergänzung, Ausgleich und Validierung des gesammelten Materials
- Wie erreichen wir die Crowd?
 - Netzwerk der Kooperationspartner
 - Kontakte zu Medien, Bildungseinrichtungen, Verbänden
- Welche Daten wollen wir einheben?
 - Text
 - Bilder
 - Kommentare, Diskussionen

Crowd-Sourcing II: Datenvalidierung

- Wie werden die Daten der Crowd validiert?
 - Intrinsisch nach dem Prinzip der unabhängigen Quellen („einmal ist keinmal und zweimal ist immer“)
 - Durch Kompetenztests nach dem Vorbild des Zooniverse-Projekts (<https://www.zooniverse.org/>)
 - Bildung einer „community“ durch Vergabe von Useraccounts



Crowd-Sourcing III: Frage aus onomasiologischer Perspektive

Entwurf

Helfen Sie uns!
Und der Wissenschaft!



Wie bezeichnen Sie den Begriff
ALMHÜTTE?

- **Webseite**
- **Smartphone App**

Almhütte

Tegia

Malga

Hütte

Andere Bezeichnung?



Crowd-Sourcing IV: Frage aus semasiologischer Perspektive


Entwurf



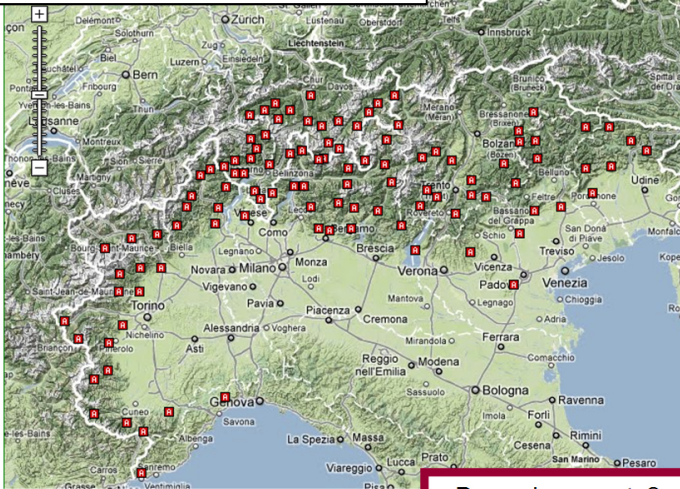
ALMGELÄNDE

Che cosa significa la parola
malga
?
Clicca sulla foto


Inserisci qui
un'altra foto,
se necessario!



ALMHÜTTTE



Dove viene usata?
Clicca sulla cartina!



HERDE

- parola sconosciuta
- parola conosciuta, ma non la uso

„Einmal ist kein Mal und zweimal ist immer!“

Crowd-Sourcing V: Datenvalidierung

user	timestamp	konzept	bedeutung	ort	relevanzindex
nutzer123	2013-10-21 13:17:22	malga	kuhherde	Tuenno	0

user	timestamp	konzept	bedeutung	ort	relevanzindex
nutzer123	2013-10-21 13:17:22	malga	kuhherde	Tuenno	3
nutzer11771	2014-02-12 22:50:43	malga	kuhherde	Tuenno	3
nutzer1822	2014-02-28 19:30:07	malga	kuhherde	Tuenno	3
usw.	usw.	malga	kuhherde	Tuenno	usw.





Crowd-Sourcing VI: Datenvalidierung - Kompetenztest

VerbaAlpina Kompetenzte x
www.verba-alpina.gwi.uni-n

Frage 3 von 5:

Ein 'Ziger' ist ein ...

- Rechen zur Verwendung bei der Heuernte
- Molkekäse
- Schlitten zum Holztransport

Verbleibende Zeit: 10 Sekunden

VerbaAlpina Kompetenzte x
www.verba-alpina.gwi.uni-muenchen.de/kc

Frage 5 von 5:

Welcher Vorgang ist hier abgebildet?



- Zerkleinerung des sog. Bruchs
- Brühen von Molke
- Butterherstellung

Verbleibende Zeit: 05 Sekunden



Idee: <http://www.galaxyzoo.org/>



Multifunktionales Portal

- Datensammlung
- Datenpräsentation
- Forum für Diskurs zwischen Wissenschaftlern und Laien gleichermaßen
- Publikation

Die Technik

- Ausschließlich **Webtechnologie**
- Backend: MySQL-Datenbank
- Frontend: Wordpress (PHP)
- Google Maps (Javascript)



Nachhaltigkeit (und Nachnutzbarkeit)

- Institutionelle Einbettung: ITG der LMU; UB; LRZ
- Datenversionierung
- Backups und Archivierung (z.B. Clarin-D, LRZ)
- Dokumentation (Rubrik „Methodologie“)
- Aufnahme in Kataloge, z.B der DNB (persistente Netz-Adressen)
- Probleme / Desiderate: langfristige Erhaltung der interaktiven Weboberfläche (Lebensdauer nach Abschluss der Entwicklungsarbeiten ca. 5 Jahre; abhängig von Fortentwicklung der Serversoftware – kein spezifisches VA-Problem)

Danke für Ihre Aufmerksamkeit!

www.verba-alpina.gwi.uni-muenchen.de

