The Virtual Research Environment of VerbaAlpina and its Lexicographic Function

Christina Mutter, Aleksander Wiatr
Ludwig-Maximilians-Universität München
E-mail: christina.mutter@lmu.de, aleksander.wiatr@lmu.de

Abstract

This paper describes the long-term research project VerbaAlpina of Munich University, which has been funded by the German Research Foundation (DFG) (http://gepris.dfg.de/gepris/projekt/253900505) since October 2014. The project investigates the Alpine lexis of three conceptual domains in the Alpine region where dialects and languages belonging to three large language families (Germanic, Romance and Slavonic) are spoken. This paper emphasizes one of the project's main functional areas, its lexicographic function, which serves to gather, process, access and visualize lexical data. To this end, data from traditional linguistic atlases and dictionaries as well as recent data gathered via the project's crowdsourcing tool first have to undergo a process of systematic data processing to fit the unified structure of the relational database (MySQL). This process can be subdivided into three major steps: transcription, tokenization and typification. Apart from the multi-directionality of the project which collects, documents and disseminates structured linguistic and ethnographic data, VerbaAlpina also provides an innovative online publishing platform that will prove sustainable and can be easily cited.

Keywords: digitalization; crowdsourcing; interlingual geolinguistics

1   Project Description

1.1   Area Under Investigation: The Alpine Region

The project "VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit" (VerbaAlpina. The Alpine cultural region reflected through its multilingualism) seeks to investigate the linguistic and cultural area of the entire Alpine region from a transnational perspective through selective analyses. The geographical region of investigation is limited to the territorial borders defined by the Alpine convention[1]. This covers a surface area of 190,600 km^2 and encompasses parts of six different countries (Austria, Italy, France, Switzerland, Germany and Slovenia) as well as two entire countries (Liechtenstein and Monaco). The Alpine region is characterized by its ethnographic and topographic homogeneity, and at the same time by its strong linguistic heterogeneity. This linguistic heterogeneity, which includes three large language families (Germanic, Romance and Slavonic), has made the region a topic of interest for linguists. Accordingly, VerbaAlpina (in the following text often abbreviated as VA) focuses on the following languages and their respective dialects: German, French, Italian, Slovenian, Franco-Provençal, Romansh, Ladin, Friulian and Occitan (c.f. Krefeld/Lücke 2014b: 189).

---

[1]   http://www.alpconv.org/en/convention/default.aspx, [last access: 23.03.2018]: "The Alpine Convention is an international treaty between the Alpine Countries (Austria, France, Germany, Italy, Liechtenstein, Monaco, Slovenia and Switzerland) as well as the EU, for the sustainable development and protection of the Alps."
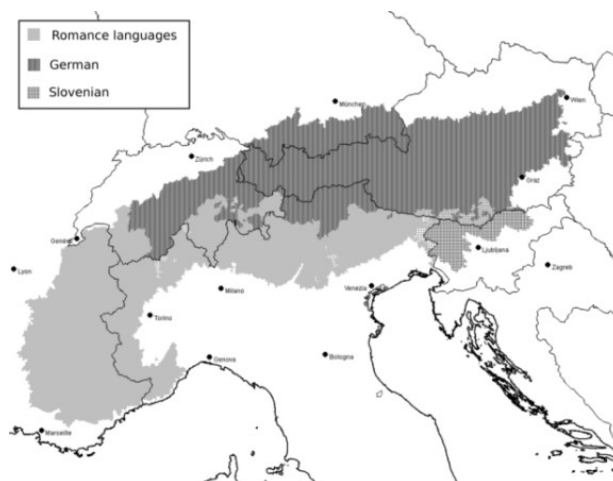
Figure 1: The area under investigation in VerbaAlpina corresponding to the Alpine region as defined by the Alpine Convention. The data collection of the three conceptional domains is broken down into three stages. Stage one (from October 2014 to October 2017) focused on vocabulary related to Alpine pasture farming, in particular, milk processing. In the current phase (from November 2017 to November 2020) the project is concerned with the lexis of the domains fauna, flora, landscape formation and weather. The last stage (from December 2020 to December 2023) will focus on the vocabulary of modern Alpine life (ecology, tourism).

1.2   Data and Methodology[2]

VerbaAlpina gathers and analyzes linguistic data derived, on the one hand, from linguistic atlases and geo-referenced dictionaries from the past one hundred years. Figure 2 gives an overview of the large number of traditional atlases and lexica in which the relevant vocabulary of the Alpine region is recorded. Yet these resources lack a multilingual perspective, and only cover parts of the Alpine vocabulary. They were created at different times and document diverse concepts. This project's on-line crowdsourcing tool (c.f. Wiatr 2016; Krefeld/Lücke 2017b) thus helps to even out, complete and correct this inhomogeneous data stock (www.lmu.de/verbaalpina).
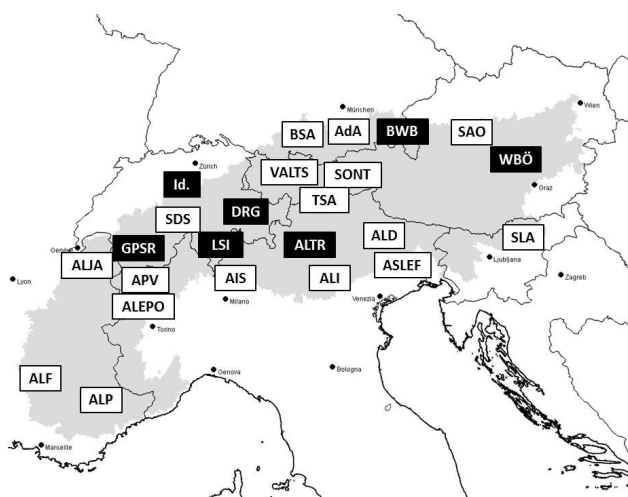

Figure 2: Overview of traditional atlases (white boxes) and dictionaries (black boxes) of the Alpine region (map created by VerbaAlpina)[3].

2   All relevant processes are documented in the menu item "methodology" on the project platform (https://www.verba-alpina.gwi. uni-muenchen.de/?page_id=493&db=172).

3   See the reference section for the abbreviations of the different atlases and dictionaries.

VerbaAlpina thus combines and further develops three different approaches to digital geolinguistics: (1) atlases which are digitally published but whose data has been gathered through traditional methods (e.g., ALD); (2) atlases which document diverse languages and language families (e.g., WALS); and (3) web-based atlases which use the internet to create data sets (e.g., AdA). (c.f. Krefeld 2017f)

The project addresses the major challenge posed by the lack of uniformity of the data from the individual data sources (linguistic atlases, dictionaries, crowdsourcing) which are not structured in the same way. VerbaAlpina must first unify the different transcription systems of the atlases and dictionaries. For this purpose, the already existing data in both digital and analogue form undergo a process of systematic data processing to fit the unified structure of the relational database (MySQL) in which all project data is stored (for more details c.f. Oberholzer/Kunzmann in press). This process can be subdivided into three major steps: transcription, tokenization and typification. These sub-processes will be described in more detail in section 2.

1.3   Research Aims

VerbaAlpina seeks, first of all, to investigate the Alpine region which is characterized by numerous languages (cf. section 1.1.) and their corresponding dialects in its historico-cultural and historical linguistic unity. As such, VerbaAlpina also overcomes the traditional limitation of geolinguistic investigation to nation-states.

Of primary importance is the recognition of connections regarding the etymology of the individual dialectical words. In this way, differences and similarities between the individual language groups of the Alpine region can be found. Many words share a common etymology, even if this cannot be seen anymore at first sight (c.f. Krefeld 2017c). For example, the German word *Butter*, French *beurre* and Italian *burro* all go back to the Greco-Latin word *butyru(m). Two additional examples are: the* Swiss-German *Staffel*, German *Stadel*, French *étable* and Italian *stabbio* are all based on the Latin word *stabulum*; the Swiss-German *Schotte(n)*, Italian *scotta* and Slovenian *skuta* all derive from Latin **excocta*.

The cooperation with other projects is fundamental for VerbaAlpina, as reflected by numerous cooperation agreements with international partners from the entire Alpine region. Each cooperation is based on a formal agreement which guarantees the project partners their own database to upload their data. Each project partner's data is then available to all partners in a structured form. The cooperation is not limited to data exchange, as all partners are also invited to use and further develop all the functional areas offered by VerbaAlpina (c.f. Krefeld 2017d).

2   Lexicographic Function

One of the core functions of the research environment of VerbaAlpina is its lexicographic function. This enables researchers to explore the data from two points of view: onomasiological and semasiological. The data from investigations from either perspective can be accessed through the interactive map on the project platform or via direct database query (see the detailed description in section 2.4.). For this purpose, the raw source data must be implemented and processed to fit the structure of the relational database. Since the sources of the input data are both analogue and digital, VerbaAlpina must adapt and handle them individually. Regardless of the type of source data there are three steps that always take place to process linguistic material: (1) transcription/transfer, (2) tokenization and (3) typification. This section provides an overview of these steps. Moreover, we will also discussed the challenges we encounter and how we deal with them in order to standardize the data to make it comparable (c.f. Lücke 2017c).

2.1 Transcription of the Input Data

2.1.1 Analogue Data

The main sources of our data stock are diverse language atlases from the Alpine region. The project's two biggest challenges regarding data processing are:

- the conception of the language atlases; and
- the transcription system used.

Most linguistic atlases differ in conception. This often depends on the type and objective of each atlas. In most cases one single map shows linguistic attestations for a single concept. It is also common for other kinds of information to be displayed: attestations, types, other concepts, and explorer's remarks. Furthermore, the linguistic material on the maps is not always presented in the same way. Whereas in the Romance (e.g. AIS, ALF) and Slavic (e.g. SLA) tradition the atlases show the actual expression that stands for a given concept in a given place, the Germanic atlases (e.g. VALTS) provide maps that show types (phonetical or morphological ones) as well as mixed forms, namely forms with types and expressions. Moreover, there are also atlases which show types referring to a certain area (e.g. ALP) (see Figure 3 below).
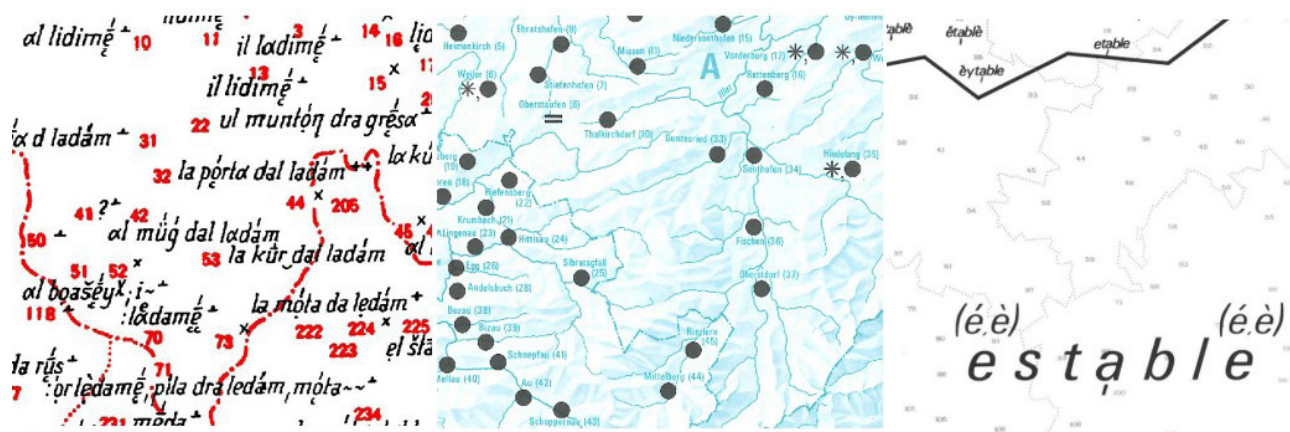

Figure 3: The picture illustrates the differences between language atlases regarding the presentation of linguistic material. On the left: AIS with linguistic attestations for each recording location; in the middle: VALTS with phonetical types; on the right: ALP showing a mixed form of attestations and phonetical types referring to a certain area.

Second, the transcription systems vary strongly depending on the individual scientific tradition. For example, AIS uses the transcription system Böhmer-Ascoli; ALF and ALJA use the Gilliéron-Edmont system; VALTS uses Teuthonista. When looking at the data in the context of a specific atlas and a specific region, the transcription problem may not be obvious at first. However, the use of different transcription systems turns out to be quite problematic when investigating language phenomena that cross national and linguistic borders. Since VerbaAlpina aims to examine the Alpine region from a transnational perspective, we have developed methods and tools to extract, handle and save all information (linguistic and non-linguistic one) to make the data structured and comparable. VerbaAlpina's transcription tool is essential when transferring analogue data from traditional atlases into digital data (c.f. Krefeld 2017g).

The main window of the tool shows the transcription object, in this case a single map. The attestations that have not yet been transcribed are suggested automatically. At this point, the transcriber must decide whether the individual attestation is defined as a real expression or as a phonetical or a morphological type. Before entering the transcription into the database, the corresponding concept

must be chosen. This is particularly helpful when dealing with maps containing multiple concepts, such as the maps of the AIS. We use Beta Code to transcribe the data from the different atlases in order to preserve all phonetic information and have access to the original transcription at any time. Beta Code is a system of unequivocal signs that are independent of any computer system and font type, and which can be used for the transcription of diverse phonetic systems. Since the transcription takes place on a graphematic and not a phonetic level, the Beta Code can be applied to every kind of atlas because it only symbolizes graphemes and not the phonetic values. Thus, only one code is necessary for the transcription of all data, allowing the transcription work to be more efficient and delegated to non-specialized staff. The Beta Code differentiates between basic signs (normal letters) and diacritics. First, the transcriber writes down the basic sign followed by diacritics from left to right and from bottom to top (see Figure 4) and then moves on to the next basic sign (c.f. Krefeld/Lücke 2017a).

Figure 4: Use of Beta Code in the transcription tool of VerbaAlpina.

Unique combinations of Beta Code signs and the corresponding IPA symbols have been listed in a specific code page for every atlas. The transcribed linguistic material is converted into IPA without losing the original transcription at any stage of the process. Table 1 illustrates how the original transcription is transcribed with the help of the Beta Code and its conversion into IPA.

Table 1: Using Beta Code to transcribe different transcription systems.

Source attestation	Beta Code transcription
	una1 mu:g/a1 da1 va/c)/

IPA: [una myʤa dɑ vˈatɕ]

Source	Beta Code	IPA	
AIS	a1	ɑ	&#x0251
AIS	c)/	tɕ	&#x02a8
AIS	d	D	&#x0064
AIS	g/	ʤ	&#x02a4
AIS	m	M	&#x006D
AIS	n	N	&#x006E
AIS	u	U	&#x007
AIS	u:	y	&#x0079
AIS	v	V	&#x0076

2.1.2 Digital Data

Digital data sources present a different set of challenges compared with analogue data. To cope with these, we first create unique databases for every cooperation partner of VerbaAlpina in which the unaltered source data is stored. Although in most cases the data is provided as a database dump or Excel-data, we must adapt it to the structure of our relational database. All steps regarding the data processing of every specific source are documented in detail for consultation at a later stage.

The second and most challenging part of the implementation of digital data consists in the coding of signs or, more precisely, the font types used by the individual partner projects. In many cases the coding does not respect the standards of Unicode, which leads to problems when the data is transferred into a different system. The non-standard coding of signs is unproblematic as long as the coding is restricted to the project itself. It becomes problematic when two or more projects using different codings are brought together. For that reason, we replace the incorrect coding and add the corresponding IPA to the code page (c.f. Lücke 2017a).

2.2   Tokenization

All transcribed and transferred data is first stored in the table *aeusserungen* (utterances) in our database. Single words as well as entire utterances can be found there. Every single record is related to a geographic point of reference and to the corresponding concept and stimulus. During the process of tokenization we subdivide utterances into their single components and we separate the lexical information from the grammatical one, such as the article. At this point all data is uploaded into the table 'tokens'. In case of single word attestations, the concept is inherited from the table *aeusserungen*. Multi-word lexical units additionally have to be split up into their single components (tokens) and then assigned to new concepts. This step is quite challenging, since it is not always obvious which concept is the correct one. However, it is always possible to reconstruct the data before the transformation. The conversion into IPA also takes place during this step.

Table 2: Illustration of the multi-word lexical unit and its components and the corresponding concepts.

Attestation in Beta Code	Attestation in IPA	Concept
una1 mu:g/a1 da1 va/c)/	una myʤɑ dɑ vˈatɕ	HERD OF COWS
TOKENIZATION		
una1		ARTICLE
mu:g/a1	myʤɑ	HERD
da1	dɑ	PREPOSITION
va/c)/	vˈatɕ	COW

2.3   Typification

The third process of the lexicographic function is the typification. In this all single tokens as well as whole utterances are assigned to a so-called morpho-lexical type. A morpho-lexical type is defined by the following categories: language family, part of speech (PoS), affix and gender. For example, for the tokens *barga, barg, margun,* and *bargun* we would have three different morpho-lexical types, as shown in Table 3 (c.f. Krefeld/Lücke 2017e).

Table 3: Creating morpho-lexical types for existing tokens.

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

With regard to orthography, we first search for an equivalent in a standard language and use it as an orthographic form. If standard language dictionaries do not provide an appropriate form, we search for one in regional and dialectical dictionaries. For example, for a series of single attestations one morpho-lexical type is created:

[lˈatʃ], [lˈatɕ], [lˈetʃ], [lˈetʂ], [lˈatɕ], [lˈatɕ], [lˈatɕ], [lˈat], [lˈetʃ], [lˈetʃ], [lˈetʃ], [lˈatɕ], [lˈæjt], ß 'lait/latte rom. m.'

Since we only work with information we have, it is also possible that morpho-lexical types without gender marking are created:

(1) Butter (ger. m.)
(2) Butter (ger. f.)
(3) Butter (ger. n.)
(4) Butter (ger.)

In the next step of the typification process the morpho-lexical types are assigned to a so-called basic type, which is only defined by the source language. The basic type is not to be confused with the etymon of a word, which can but is not necessarily be the same word as the basic type. As in the case of morpho-lexical types, we use dictionaries as a reference (c.f. Krefeld/Lücke 2017d).

2.4 Accessing the Processed Data in Two Ways: Via an Interactive Map and the Database Interface

Having completed the process of typification the data is made accessible through two different user interfaces. The first is the interactive map[4] which allows the user to choose elements from linguistic core data as well as data from the linguistic periphery, also called extra-linguistic data, such as toponyms, demographical and other information. For exploring the linguistic data, one can choose from concepts (semasiological point of view), morpho-lexical types and basic types (onomasiological point of view). In addition, different kinds of data can be easily grouped on a single map. The interactive legend which is updated dynamically according to the presentation of the data gives the user the possibility to switch on and off preselected data. (c.f. Krefeld 2017a) The interactive map offers two different views: the physical and the hexagonal one. The latter, due to its algebraic simplification of the geographic dimension to hexagons, is very useful when using the quantifying feature of the interactive map, enabling the user to quantify existing data in relation to a chosen territorial unit (language areas, NUTS-3) of the Alpine region (see Figure 5 below) (c.f. Lücke 2017e). Registered users have the possibility to save their selected data in a so-called synoptic map and to reuse or modify it later (c.f. Krefeld 2017e).

---

4 To understand how the interactive map works and what possibilities it offers, please visit our website and experiment with it: https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=133&db=172 [31.03.2018].
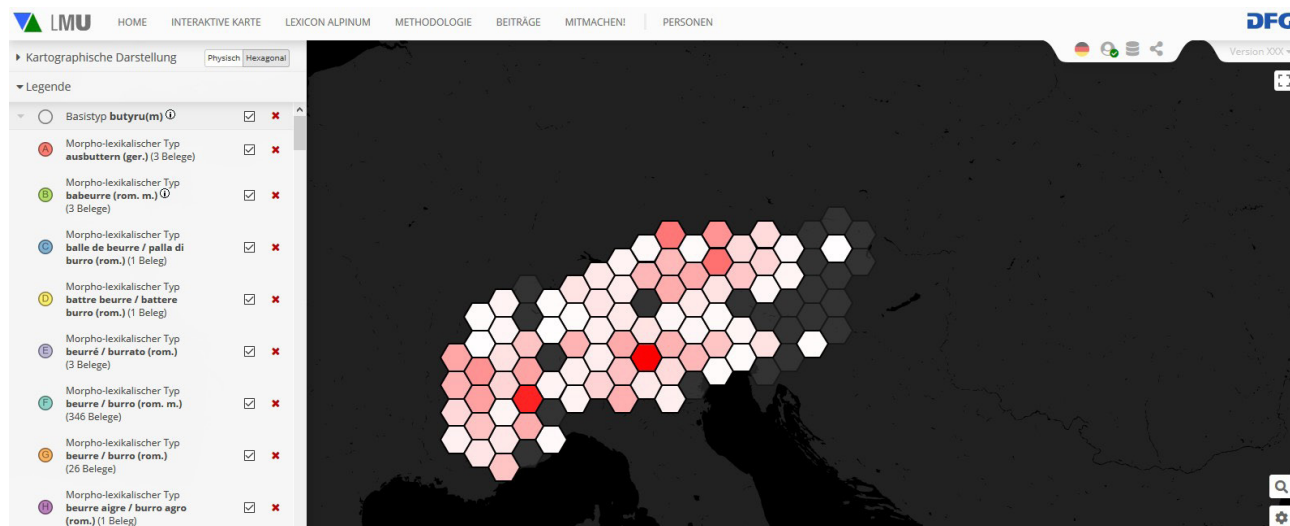
Figure 5: Demonstration of the quantification tool of VerbaAlpina in use (in the hexagonal mode). This screenshot illustrates the frequency of different morpho-lexical types related to the basic type *butyru(m)*. The redder the polygon the more tokens are present in this area. The legend on the left side of the screenshot shows the morpho-lexical types that matched with the basic type *butyrum*.

The other and more precise way to access the data is via the database interface. Whereas the interactive map is open access for everybody, access to the database of VerbaAlpina is currently granted to project partners only. It is, however, possible that other users are also granted access as soon as they register and are accepted by the VerbaAlpina team[5]. The database interface[6], which is a MySQL view, presents the user with the data in a structured and comparable form. Provided they have some MySQL knowledge, the user can select, filter, group and organize the data in an unlimited number of ways. For example, one database query could ask for all tokens from AIS, ASLEF and SLA that have the basic type *butyrum* in common, that are all feminine and have no suffixes. It is necessary for users to have some knowledge of MySQL, in particular the structure of the relational databases in MySQL and of the query language. However, VerbaAlpina offers some options to download parts or all of the data and edit it with the help of spreadsheet software such as Microsoft Excel. Accessing data through our interfaces gives the user the possibility to work with recent data. Since the team of VerbaAlpina works continuously on the data, the access interface is updated every day (c.f. Lücke 2017b).

2.4.1 Data from the Linguistic Periphery

VerbaAlpina does not only use and process linguistic data. In fact, every kind of data which is geo-referenced can be inserted into the structure of VerbaAlpina. The additional types of data help with the interpretation of the created maps. Thus, some linguistic phenomena can be understood better when combined with other data, such as demographic or infrastructural data. Currently we dispose of the following extra-linguistic and geo-referenced information: data about settlements, infrastructure, castles, Latin inscriptions, early medieval sites, churches and toponymy (c.f. Krefeld 2017b).

---

5   You can register using the following link: https://www.verba-alpina.gwi.uni-muenchen.de/wp-signup.php.

6   The database interface is aaccessible in all working languages of VerbaAlpina, at present in German, French, Italian, Slovenian and Rhaeto-Romance. We are also working on an English version to make it accessible to a broader audience.

3   VerbaAlpina as an Innovative and Sustainable Online Publishing Platform

One of the major and most debated problems of digitalization consists in the sustainability of digitized data (for further information see for example: Maron, Smith, Loy 2009, Bradley 2007, Krefeld/Lücke 2017c). To cope with this issue VerbaAlpina has developed an innovative approach of versioning and of citability.

3.1   Versioning

VerbaAlpina consists of the following modules: VA_DB, VA_WEB, VA_MT (see Figure 6 below). The module VA_DB comprises the data stock which is stored in the project's MySQL database (va_xxx). VA_WEB encompasses the program code of the web interface of the project platform www.verba-alpina.gwi.uni-muenchen.de including the corresponding WordPress database (va_wp). The module VA_MT contains media files (photos, videos, text and sound documents) which are stored in the media library of the web interface. All three modules build a consistent whole that is interrelated and interdependent, and therefore they cannot be separated from each other. During the duration of the project, the current status of the modules VA_DB and VA_WEB is "frozen" simultaneously in the form of an electronic copy every six months, on 15 June and 15 December each year. These frozen copies receive versioning numbers according to the scheme [year]/[sequence number] (e.g. 15/1). Every productive VA version is named XXX. Due to the large size of media files it is impossible to produce copies of the VA-media library (VA_MT). For that reason, no copy of this module is produced in the course of a versioning process. Media files which have been stored there once cannot be removed from the VA media library as soon as just a single VA version relates to them. On the project platform it is possible to switch between the "productive" VA version—subject to constant changes—to the archived, "frozen" versions. Appropriate coloring of the background and of certain control elements makes clear whether the productive or an archived version of VerbaAlpina is displayed (c.f. Lücke 2017f).
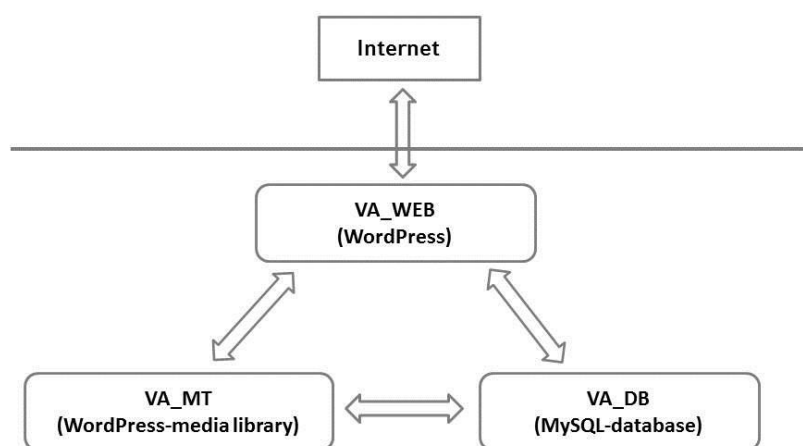

Figure 6: The modules of VerbaAlpina

3.2   Citability

The versioning process also makes it possible for the contents of VerbaAlpina to be accurately cited. Unlike common online sources, the date of last access is not necessary since the cited versions (unlike the productive version XXX) are stable and will not be changed. Thus, VerbaAlpina as an Internet-based digitized source becomes viable for academic writing. As part of the bibliography VerbaAlpina can be cited in the following way:

> VerbaAlpina (VA), http://www.verba-alpina.gwi.uni-muenchen.de,[version]
> e.g.: VerbaAlpina (VA), http://www.verba-alpina.gwi.uni-muenchen.de, 15/1

For citations of contributions that belong to the menu "methodology" on the project platform, the following scheme is recommended:

> [author/s]: s.v.* "[lemma]", in: VA-[language code according to ISO 639-1] [version], methodology, [URL]
> e.g.: Krefeld, T. / Lücke, S.: s.v. "Tipizzazione", in: VA-it 15/1, Metodologia, http://www.verba-alpina.gwi.uni-muenchen.de/it/?page_id=21&letter=T#tipizzazione

The authors, marked with "auct.", at the end of every contribution, always need to be mentioned. Besides textual contents created by VerbaAlpina, graphic ones may also be cited. Every view of the interactive map can easily be shared and cited, since an individual URL is produced for every page and every pop-up window (c.f. Lücke/Oberholzer 2017).

3.3   Long-term Archiving

The project takes several steps to guarantee the sustainable use of all data. First, we take care to entrust several institutions with the long-term archiving of the data. Second, we document the data structuring, the logical relationships between data and data categories, and the applied character encoding. There are several different options to archive the project data by third parties. We plan to have multiple copies of the project data stored by several institutions. Currently, our data is saved on a regular basis by the IT-Gruppe Geisteswissenschaften of Munich University (ITG, IT group of humanities, http://www.itg.uni-muenchen.de/index.html) on backup servers of the Leibniz Rechenzentrum (LRZ, the Leibniz Computing Centre). At the same time that the data is archived, the different versions of VerbaAlpina are also created. At random intervals the module VA_WEB is also stored in the Internet Archive (https://archive.org). In addition to the automatic archiving by archive.org through their wayback crawler, VerbaAlpina also actively archives the data (since 2018 on a regular basis in the course of the versioning every six months). We also intend to store further backup copies at other appropriate institutions like CLARIN-D. We hope in the long term to have the archiving conducted by the University Library of Munich so that project contents are also accessible via the electronic catalogues (c.f. Lücke 2017d).

References

AdA = Elspaß, S., Möller, R. (2003 ff.). *Atlas zur deutschen Alltagssprache*. Universität Salzburg.

AIS = Jaberg, K., Jud, J. (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen.

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

ALD-II = Goebl, H. (2012). *Atlant linguistich dl ladin dolomitich y di dialec vejins*. Editions de Linguistique et de Philologie, vol. 1-5.

ALJA = Martin, J.-B., Tuaillon, G. (1971, 1978, 1981). *Atlas linguistique et ethnographique du Jura et des Alpes du nord*. Paris: Éd. du Centre National de la Recherche Scientifique.

ALF = Gilliéron, J., Edmont, E. (1897-1900). *L'Atlas linguistique de la France*. Paris: Champion.

ALP = Bouvier, Jean-Claude (1975, 1979, 1986). *Atlas linguistique et ethnographique de la Provence*. Paris: Éd. du Centre National de la Recherche Scientifique, vol. 1, 2, 3.

Bradley, K. (2007). Defining digital sustainability. In *Library Trends*, (56)1, pp. 148-163.

Krefeld, T. (2017a). s.v. "Dokumentation", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=D#21.

Krefeld, T. (2017b). s.v. "Ergänzende Daten", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=E#3.

Krefeld, T. (2017c). s.v. "Interlinguale Geolinguistik", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=I#32.

Krefeld, T. (2017d). s.v. "Kooperation", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=K#22.

Krefeld, T. (2017e). s.v. "Synoptische Karte", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=S#56.

Krefeld, T. (2017f). s.v. "Sprachatlanten im Alpenraum", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=S#52.

Krefeld, T. (2017g). s.v. "Transkription", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=T#57.

Krefeld, T., Lücke, S. (2017a). s.v. "Betacode", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=B#7.

Krefeld, T./Lücke, S. (2017b). s.v. "Crowdsourcing", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=C#12.

Krefeld, T./Lücke, S. (2017c). Nachhaltigkeit – aus der Sicht virtueller Forschungsumgebungen. Korpus im Text. Version 7 (10.03.2017, 12:27). url: http://www.kit.gwi.uni-muenchen.de/?p=5773&v=7.

Krefeld, T., Lücke, S. (2017d). s.v. "Referenzwörterbücher", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=R#51.

Krefeld, T., Lücke, S. (2017e). s.v. "Typisierung", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=T#58.

Krefeld, T./Lücke, S. (2014b). VerbaAlpina - Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit. In *Ladinia*, XXXVIII, pp. 189-211.

Lücke, S. (2017a). s.v. "Codepage", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=C#11.

Lücke, S. (2017b). s.v. "Datenzugriffsschicht", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=D#70.

Lücke, S. (2017c). s.v. "Digitalisierung", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=D#15.

Lücke, S. (2017d). s.v. "Langzeitarchivierung", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=L#40.

Lücke, S. (2017e). s.v. "Quantifizierende Darstellungen", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=Q#88.

Lücke, S. (2017f). s.v. "Versionierung", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=V#61.

Lücke, S./Oberholzer, S. (2017). s.v. "Zitierweise", in: VA-de 17/2, Methodologie, https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=493&db=172&letter=Z#64.

Maron, N., Smith, K. K. & Loy, M. (2009). *Sustaining digital resources: an on-the-ground view of projects today: Ithaka case studies in sustainability*. Ithaka: S+R.

Oberholzer, S., Kunzmann, M. (2017). Geolinguistic Documentation of Multilingual Areas – VerbaAlpina and the Challenges of Digital Humanities. In I. Buchstaller, B. Siebenhaar (eds.). *Language Variation - European Perspectives VI. Selected papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8)*. Leipzig: Benjamins, pp. 199-214.

SLA = Skofic, J. (eds.) (2011ff.). *Slovenski lingvistični atlas 1. Človek – telo, bolezni, družina*. Ljubljana: ZRC SAZU.

VALTS = Gabriel, E. (1985-2004). *Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus, vol. 1-5*. Bregenz: Vorarlberger Landesbibliothek.

WALS = Haspelmath, M. (2005ff.). *World Atlas of Language Structures*. Leipzig: Max Planck Institut für evolutionäre Anthropologie.

Wiatr, A. (2016). Bedeutung und Funktion von Crowdsourcing im Projekt VerbaAlpina. In *Jezikovni zapiski*, (22)2, pp. 161-175.