

The FAIR Principles - an important challenge to variational linguistics

In scientific research, the question of correct research data management has become increasingly important in recent years. In the discussion about the correct management of research data, the FAIR principles formulated by Wilkinson et. al. in 2016 have been established as guidelines. According to these principles, research data have to be **F**indable, **A**ccessible, **I**nteroperable and **R**eusable.

What is special about the implementation of these criteria regarding variational linguistics? If we take the fundamental principle of variational linguistic analysis as a basis, in simplified terms we have to deal with a binarity where one side is to be examined in terms of variation (e.g. lexis) and another side which must necessarily be kept constant (e.g. word semantics). The constant side always requires standardisation, which allows data from different sources to be clearly referenced and thus also linked to each other. In order to achieve this, norm data are a suitable tool. The challenge, however, consists in creating these norm data for the different linguistic entities, in particular with regard to concepts but also regarding levels such as morphology. If linguistic atlases and dictionaries are checked for their FAIRness, it becomes clear that the interoperability of the individual resources, which aims to create persistent links between projects and data sets, is quite different. This is mainly because in most cases the linguistic data is not linked to norm data and thus the data is also not electronically addressable.

The research project VerbaAlpina of Munich University, which examines dialectal vocabulary from the entire Alpine region, attempts to address this problem by means of various approaches. In order to standardise the project's own data, the three core entities of the project - morpho-lexical types, concepts and political communities - are first linked with unique and persistent identifiers generated by VerbaAlpina. In this way, a clear identification of the project data is made possible. In order to enable the linking of own data with external data sets, in addition to the project-specific norm data, the following identifiers of external institutions are integrated: *Gemeinsame Normdatei* (GND, "Integrated Authority File") of the German National Library (for concepts), Q- and L-IDs of Wikidata (for concepts and morpho-lexical types), Geonames (for localities), language codes according to ISO 639-3 (for morpho-lexical and base types) and various reference dictionaries (for morpho-lexical types). In this way, the compliance with the principles F (findability) and I (interoperability) is ensured. By using Creative Commons Licences (CC) which are compatible with open access the data is also made accessible and thus generally reusable.

The talk will describe the concrete steps the project VerbaAlpina takes to approach the goal of being a FAIR research project. In particular, it will address the limits of the standardisation of the linguistic data stock and will show which linguistic sections still lack options for standardisation.

References

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>

Löffler, Heinrich (2005): Forschungsplanung bei Mikro- und Makrostudien / Research Planning in Micro- and Macro-Studies. In: Sociolinguistics / Soziolinguistik. An International Handbook of the Science of Language and Society. Vol 2, 1023 – 1033.

Lücke, S.: s.v. "FAIR-Prinzipien", in: VerbaAlpina-de 20/1 (Erstellt: 18/2), Methodologie,

<https://doi.org/10.5282/verba->

alpina?urlappend=%3Fpage_id%3D493%26db%3D201%26letter%3DF%23128

Lücke, S.: s.v. "Normdaten", in: VerbaAlpina-de 20/1 (Erstellt: 18/2, letzte Änderung: 19/2),

Methodologie, <https://doi.org/10.5282/verba->

alpina?urlappend=%3Fpage_id%3D493%26db%3D201%26letter%3DN%23114

Mutter, C.: s.v. "Wikidata", in: VerbaAlpina-de 20/1 (Erstellt: 18/1, letzte Änderung: 19/2),

Methodologie, <https://doi.org/10.5282/verba->

alpina?urlappend=%3Fpage_id%3D493%26db%3D201%26letter%3DW%23105

Wilkinson, Mark D. et al. (2016): "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data - a nature research journal*.

<https://www.nature.com/articles/sdata201618#_blank> (accessed October 31, 2020)