

Swiss-AL: Language Data Platform for Applied Sciences

Language data is used not only by linguists, but also in many other research disciplines. In the Swiss-AL project, we aim to develop practices for a diverse community interested in exploring the potential of language data while implementing FAIR data management principles and integrating them into the Swiss-AL workbench.

Swiss-AL

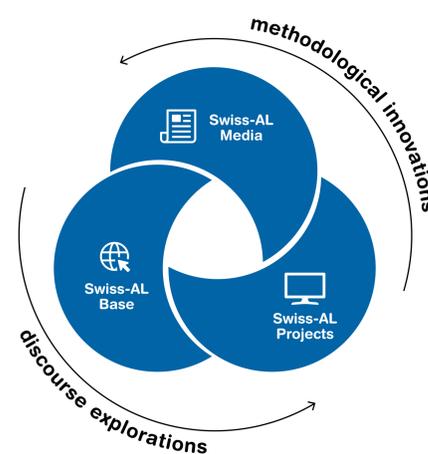


- Swiss-AL is a platform designed for the analysis of multilingual corpora derived from Swiss public communication.
- It contains Swiss journalistic media, media releases, news reports and blog posts from players in the worlds of politics and administration, industry, academia and civil society.
- A flexible processing pipeline enables the creation of tailor-made sub-corpora for analyzing and simulating public discourses.



Swiss-AL Workbench

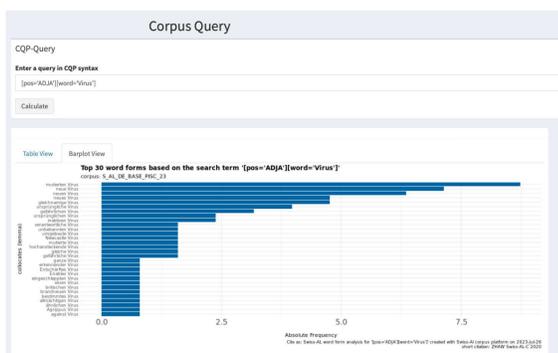
Swiss-AL Family of Corpora



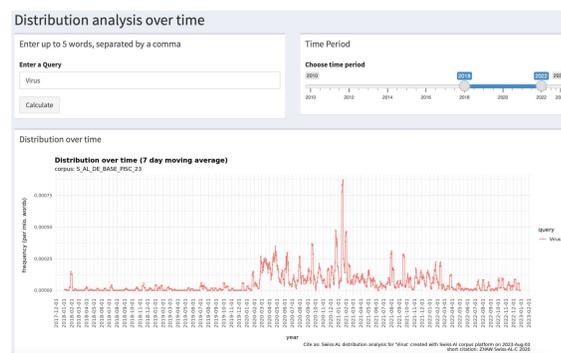
Swiss-AL Media
Journalistic Media
DE / IT / FR / RM
based on Swiss Media Database

Swiss-AL Base
politics & administration,
science, industry, civil society
and journalistic media
DE / IT / FR
based on websites and Swiss Media Database

Swiss-AL Projects
research projects
DE / IT / FR / EN
Based on Swiss-AL Base, Swiss-AL Media,
additional actors and sources



Adjectives that describe the word "Virus", sorted according to their frequency



Frequency distribution of the term "Virus" over time (2018-2022)



Pre-calculated topic bubbles (left) and frequencies of individual keywords (right)

Challenges Related to Linguistic Open Research Data

1. The use of language data by an interdisciplinary research community whose members use **heterogeneous practices** in dealing with such data (e.g., qualitative content analysis, distant reading methods)
2. The need to make **open research data** and associated infrastructures **flexible and expandable** and to ensure their continuous growth
3. Legal challenges related to providing and using language data posed by **the current data and copyright law**

Contact

Project Website

<https://www.zhaw.ch/en/linguistics/research/swiss-al-linguistic-open-research-data-practices-for-applied-sciences/>

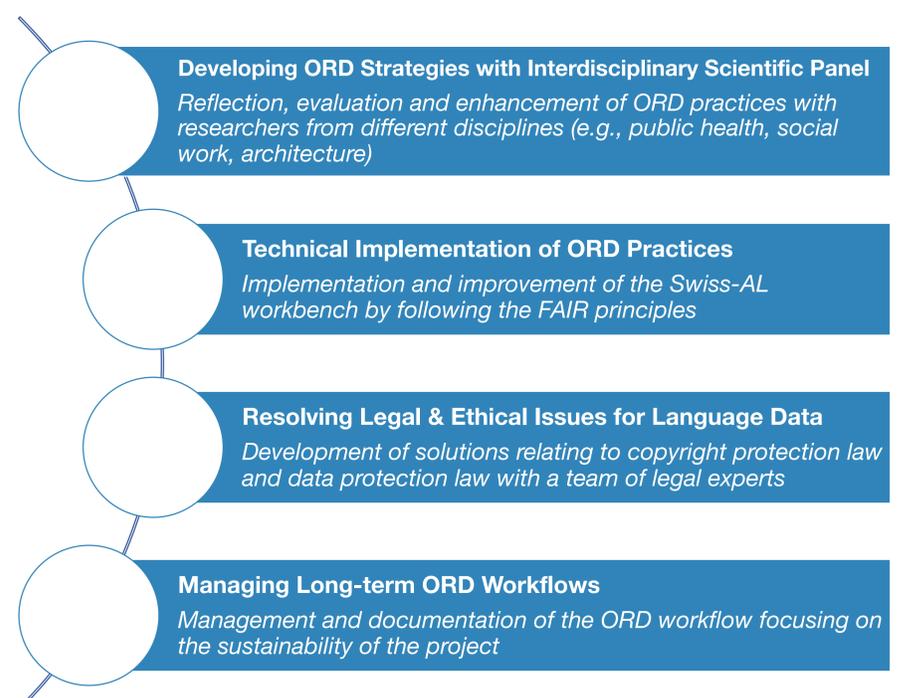


Project Team

Julia Krasselt, Philipp Dreesen, Peter Stücheli-Herlach, Matthias Fluor, Klaus Rothenhäusler, Dolores Lemmenmeier, Sooyeon Cho

E-mail: korpusinguistik@zhaw.ch

Steps towards Implementation of ORD Practices in Swiss-AL



Further Reading

Krasselt, J., Fluor, M., Rothenhäusler, K., & Dreesen, P. (2021). A workbench for corpus linguistic discourse analysis. In D. Gromann, G. Sérasset, T. Declerck, J. P. McCrae, J. Gracia, J. Bosque-Gil, F. Bobillo, & B. Heinisch (Hrsg.), 3rd conference on language, data and knowledge (LDK 2021) (Bd. 93, S. 26:1-26:9). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/OASlcs.LDK.2021.26>

Krasselt, J., Dreesen, P., Fluor, M., Mahlow, C., Rothenhäusler, K., & Runte, M. (2020). Swiss-AL: A Multilingual Swiss Web Corpus for Applied Linguistics. Proceedings of the 12th Language Resources and Evaluation Conference, 4138--4144. <https://aclanthology.org/2020.lrec-1.510/>