



Contribution ID: 9

Type: **Talk**

What influences language resource reuse? The SADiLaR repository as a case study

The reuse of language resources is a cornerstone of responsible, sustainable, and FAIR-aligned research in linguistics and digital humanities. The South African Centre for Digital Language Resources, as a specialist repository for language-specific resources, serves as a case study for this presentation. We demonstrate that repository infrastructure alone, including persistent identifiers, rich metadata, and long-term preservation, does not automatically ensure visible or measurable reuse. Researchers and/or other possible users may be unaware of datasets, may cite secondary publications rather than the underlying resources, or may reuse data in ways that leave no explicit trace. As a result, valuable language resources can appear underutilised despite playing an important role in research processes.

To explore this further, we conducted a manual and exploratory search for references to SADiLaR-hosted resources in academic publications, as well as analysing web-based user activity. The labour-intensive and imperfect nature of this process highlights the absence of systematic mechanisms for identifying and documenting dataset reuse.

Improving the visibility of the reuse of language resources, especially of South(ern) African languages, is not only important for research transparency but also for demonstrating the broader societal value of language infrastructures. In multilingual contexts where accessible digital resources support language development, education, and inclusive knowledge production, this is even more important. Unfortunately, tracing the reuse of language resources remains a challenge and is often insufficiently supported by aspects like consistent citation practices. In this presentation, we will argue that current approaches to measuring language resource reuse underestimate the extent to which such resources are utilised by a broader user community.

Primary author: Dr TROLLIP, Benito (South African Centre for Digital Language Resources (SADiLaR), North-West University)

Co-author: Dr WHITE, Michelle (South African Centre for Digital Language Resources (SADiLaR), North-West University)

Presenter: Dr TROLLIP, Benito (South African Centre for Digital Language Resources (SADiLaR), North-West University)

Track Classification: DIRISA