Information Sheet

1. OVERVIEW

The South African Centre for Digital Language Resources (SADiLaR) is a research infrastructure set up by the Department of Science and Innovation (DSI) as part of the South African Research Infrastructure Roadmap (SARIR). SADiLaR is a multi-partner entity with the North-West University (NWU) functioning as host and hub of a network of linked nodes. The founding node members are the University of Pretoria (Department of African Languages), the University of South Africa (Department of African Languages), the Meraka Institute at the CSIR (Human Language Technology Research Group), the North-West University (Centre for Text Technology) and the Inter-Institutional Centre for Language Development and Assessment (ICELDA) which, in turn, includes the universities of the Free State, Stellenbosch, North-West, and Pretoria.

SADiLaR focuses on the creation, management, and distribution of digital language resources, as well as applicable software related to all official languages of South Africa.

SADiLaR runs two programs:

A **digitisation program** which entails the systematic creation of academically relevant digital text, speech and multi-modal resources related to all official languages of South Africa, as well as the development of appropriate natural language processing software tools for research and development purposes.

A **Digital Humanities program** which will facilitate research capacity building by promoting and supporting the use of digital data and innovative methodological approaches within the Humanities and Social Sciences. Pertinent fields of research include, for instance, language and literary studies, heritage studies, history, music, sociology, journalism, philosophy, psychology, media studies, visual and graphic art etc.

SADiLaR's enabling function impacts at least three domains:

Language technology domain

High-level resources and natural language processing tools are developed for use in applications, such as machine translation engines for local languages; automatic speech recognition systems; text-to-speech systems; speech-to-speech translation systems; interactive communication systems; as well as a variety of text-related applications such as grammar and spelling checkers, online electronic dictionaries, and so forth.

Humanities and social sciences domain

The building of research capacity is facilitated among scholars by promoting the use of digital data, innovative methods, and software tools that enhance research activities, and enable scholars to ask and pursue previously-unanswerable questions within their respective disciplines.

Socio-economic domain

Reusable digital language resources are licensed for use in interactive commercial applications in local languages. The nature and use of local African languages are documented as part of a living archive, including documentation of cultural heritage practices within different language communities.

SADiLaR clients include academic scholars and professionals in all domains of Humanities and Social Sciences, Language Technologies, Natural Language Processing, Computer Science, as well as potential end-users in education, business and industry.

SADiLaR is the first centre of its kind in Africa and intends to foster links with researchers and institutions in Sub-Saharan Africa. The Centre also promotes links with similar entities globally, especially with a major counterpart in Europe, CLARIN, which is also part of the European Strategic Framework for Research Infrastructures.