



science & innovation

Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA



THE SOUTH AFRICAN CENTRE FOR DIGITAL LANGUAGE RESOURCES (SADILAR)

Strategic Plan

2021 - 2023

CONTENTS

| | |
|--|----------|
| 1. Executive summary of the strategic engagement at the HUB | 3 |
| 1.1. Digital Humanities (DH) | 3 |
| 1.2. Research | 3 |
| 1.3. Technical and Commercialisation | 3 |
| 1.4. Stakeholder engagement | 4 |
| 1.5. Communication | 4 |
| 1.6. General note regarding sustainable funding | 4 |
| 2. Foreword | 4 |
| 3. Vision | 5 |
| 4. Mission | 5 |
| 5. Objectives | 5 |
| 6. Stakeholders | 6 |
| 6.1. Internal Stakeholders | 6 |
| 6.2. External Stakeholders | 7 |
| 6.3. Funding Institutions | 8 |
| 7. Key Strategic Focus Areas | 8 |
| 7.1. The Digital Humanities Strategic Focus | 8 |
| 7.2. The Commercialisation Strategic Focus | 9 |
| 7.3. The Communication Strategic Focus | 10 |
| 7.4. The Research and Training Strategic Focus | 11 |
| 7.5. Stakeholder Engagement Strategic Focus | 12 |

1. EXECUTIVE SUMMARY OF THE STRATEGIC ENGAGEMENT AT THE HUB

The SADiLaR (draft) Strategic Plan 2021-2023 is premised on the work that is already underway, and also launches several new initiatives. While maintaining its strategic focus as articulated in the Department of Science and Innovation (DSI) contract, and excellent delivery and reporting, the imperative exists to create a strategic framework through which this DSI mandate can be achieved – and even exceeded. The objective is to strategically position SADiLaR as a locally engaged and impactful research infrastructure with a strong global presence. The strategic plan will steer SADiLaR through a period that is likely to demand that its programmes be responsive and impactful to society, while utilising innovative technology that predicts and addresses language-technology needs, and commercialisation initiatives that mitigate future government fiscal frugality as a result of economic challenges. The plan will also be reviewed and updated, where required, once the DSI's new Decadal Plan is finalised. The following strategic factors form the backdrop for the strategic choices underpinning the strategic plan.

In order for SADiLaR to strengthen the execution of its longer-term mandate, the following six strategic priority areas need to be supported and acted upon:

1.1. Digital Humanities (DH)

- DH is a relatively new research field within the South African context that offers many novel opportunities.
- Objective: To foster collaborations between computationally skilled groups and researchers from the Humanities and Social Sciences (HSS) as well as to capacitate with computational skills. The Escalator project has been initiated to address this imperative.

1.2. Research

- A lack of natural language processing (NLP) skills exists in South Africa.
- Objective: SADiLaR needs to stimulate and grow research skills within this domain. This could be done through the introduction of virtual labs with specific focus areas. The introduction of the BA in Language Technology Programme (NLP) by SADiLaR's CText node is a long-term strategic intervention.

1.3. Technical and Commercialisation

- There is a funding and sustainability risk for SADiLaR work beyond current DSI funding.
- Objective: SADiLaR needs to develop a commercialisation framework to mitigate this risk.

1.4. Stakeholder engagement

- Even though SADiLaR is known to stakeholders in the relevant fields, this network can be extended to increase SADiLaR's impact in society.
- Objective: SADiLaR needs to strengthen stakeholder relationships and build mutually beneficial, non-competing partnerships.

1.5. Communication

- Much of the work done by SADiLaR has been focused on the academic environment. However, the broader community may also benefit from the Centre's output.
- Objective: SADiLaR needs to translate research to the broader community to ensure that what it does is also clear to non-academics, and it needs to improve media outreach, e.g., radio, television, etc.

1.6. General note regarding sustainable funding

- As noted in the original DSI agreement, the Centre itself will be mainly reliant on government funding for sustained operations. However, alternative funding partnerships available through existing networks such as Common Language Resources and Technology Infrastructure (CLARIN) must be considered and exploited.
- Objective: Reach out and intensify collaborations with existing networks to achieve sustainable funding.

2. FOREWORD

From the Executive Director of the South African Centre for Digital Language Resources

The South African Centre for Digital Language Resources, also referred to as SADiLaR, is a research infrastructure (RI), whose strategic function is to create, manage and distribute digital resources, as well as applicable software, to all South Africa's official languages in order to stimulate and support research and development in the Humanities and Social Sciences (HHS). This strategic function is responsive to the imperatives of the Constitution of the Republic of South Africa, 1996, and the Use of Official Languages Act, 2012, pertaining to the equal development and use of the 11 official languages of the country.

SADiLaR is hosted at the North-West University (NWU) and is funded by the DSI. It is one of the RIs of the DSI's South African Research Infrastructure Programme (SARIR) programme. It is the first (and currently the only) infrastructure of its sort in Africa. SADiLaR's hub is located at the NWU and links to various nodes, located across South Africa, that assist in the development, research and support of all the official languages of South Africa. SADiLaR's core mandate is thus to drive the digitisation programme (natural language processing), which involves the creation of relevant text, speech and multimodal resources for the research and development of SA's 11 official languages, and also a DH programme that focuses on research capacity building.

This strategy seeks to articulate a vision that stimulates and enables digital research and development of all 11 official languages as articulated in the Constitution, 1996, read together with the Use of Official Languages Act, 2012. Through this strategy, SADIaR seeks to articulate the five focus areas that will operationalise the mandate of the hub and achieve the objectives of its contract as set out by DSI. The strategy will leverage the successes that have been achieved hitherto since the establishment of the Centre.

3. VISION

To stimulate and enable digital research and development in South Africa's 11 official languages, to cultivate, create awareness and grow DH scholarship in South Africa and the continent, while creating strategic synergies with local and international partners with similar or complementary infrastructure.

4. MISSION

To develop and promote all 11 official languages to be capable of expressing all forms of knowledge, and to drive their use and function in research and development, education, social transformation, trade, and economic and scientific development.

SADIaR aims to espouse the RICH Plus T (RICH^T) Principles:

R= SADIaR undertakes to promote mutual **respect**, courtesy and inclusivity in an environment driven by research and productivity.

I= SADIaR promises to inspire innovation that is **impactful** and contributes meaningfully to the community.

C= SADIaR commits to carefully and effectively **communicate** with all its strategic partners and stakeholders.

H= SADIaR undertakes to transform research and teaching in the **humanities and social sciences** through a high-impact DH programme.

T= SADIaR undertakes to drive with determination, with commitment to **teamwork and trust** between staff, management, strategic partners, and stakeholders.

5. OBJECTIVES

- a. To stimulate and advance the scholarship of DH across South African Higher Education (HE) sector through collective and collaborative engagements (e.g., enabled through the Escalator project and other incentives).
- b. To open new frontiers of research in the HSS in general by facilitating access to training, tools, and exposure to new methodologies as early as possible.

- c. To provide digital resources and tools for the development of a wide range of language-technology applications (e.g., in the fields of health services, education, social services and business, and support of other infrastructures and initiatives).
- d. To document the nature and use of local African languages, including cultural-heritage practices as part of a living archive (e.g., enabled through language laboratories/hubs and platforms).

6. STAKEHOLDERS

In order for SADiLaR to successfully execute its strategic plan, it has identified partner institutions that are crucial in supporting the Centre's work. The stakeholders are broadly divided into two groups, first those based in South African HE institutions (particularly where SADiLaR nodes are based) and, second, those based on the broader African continent and in Europe.

6.1. Internal Stakeholders

- Scope: within academic institutions
- African language and other Humanities schools/departments or Centres of Excellence
- Computer Science schools/departments
- University publishers (Wits University Press, UKZN Press, etc.)
- University executives and councils
- Public university language units/centres
- Research centres/institutes e.g., Indigenous Knowledge Systems (IKS)
- Technology Units/Centres
- University Innovation Hubs/Centres
- Museums for archives
- Corporate Relations/Public Relations divisions
- Students, especially mother-tongue enrolment

- SADIaR's nodes
- SADIaR governance committee members

6.2. External Stakeholders

- Scope: within Africa and broader
- The Department of Science and Innovation (Key Partner)
- The Pan South African Language Board (PanSALB), South Africa
- The Department of Sport, Arts and Culture (DSAC), South Africa
- Other government departments e.g., Department of Basic Education
- African Academy of Languages (ACALAN)
- Solomon Projects Language Academy (SPLA)
- GAFAM - Google, Apple, Facebook, Amazon, Microsoft
- Centre for Digital Humanities University of Lagos (CEDHUL) and other global DH centres
- The Southern African Development Community (SADC) and other regional communities
- Financial institutions and other private sector institutions, which could include corporations and organizations that have large amounts of text/data, e.g., Greenpeace/insurance companies
- Baidu, Alibaba, Tencent, Oracle, IBM, Accenture, SAP, Saigen, RI colleagues - CLARIN member countries
- The People of South Africa
- Language and culture organisations that are not affiliated with a centre or university department, e.g., The Zuid-Afrikahuis in Amsterdam, the Netherlands
- Researchers in countries similar to SA, for instance India, which also has many languages, many of them under-resourced. Several South American countries also fit this description

- Academic associations Southern African Linguistics and Applied Linguistics Society (SALALS), African Language Association of Southern Africa (ALASA), South African Association for Language Teaching (SAALT), Network for Digital Humanities in Africa etc.)
- GLAM Sector (Galleries, Libraries, Archives, and Museums)
- Summer Institution of Linguistics International

6.3. Funding Institutions

The Centre is fully funded by the DSI. The risk of relying on government funding is articulated in SADIaR's risk register (*attached*). The strategy to sustain the mission and mandate of the Centre will be articulated as part of SADIaR's **Commercialisation Strategy** that is currently under development.

In the medium term, partnering with similar institutions that share elements of SADIaR's mission and vision can help leverage limited funding. Investigating existing capacity and applying for external funding calls can be investigated.

7. KEY STRATEGIC FOCUS AREAS

7.1. The Digital Humanities Strategic Focus

This focus is consistent with **Objective a.** of this strategy: Objective a. to stimulate and advance the scholarship of DH across South Africa's HE sector). The imperative to cultivate and grow the DH scholarship across all South Africa's public universities has been articulated.

Currently, SADIaR is seen as linked with the DH research field. For instance, SADIaR organises monthly DH colloquia and is active in the local/regional Digital Humanities Association of Southern Africa (DHASA) as well as the world-wide Alliance of Digital Humanities Organizations (ADHO) organizations.

Given the structure of the organization, SADIaR is well placed to boost the field of DH in South Africa. A relatively large team with varied research backgrounds is available. Furthermore, researchers (mainly from HSS) in South Africa are interested in the new field of DH.

There are several challenges that need to be addressed. First, the visibility of DH in South Africa will need to be improved. This will happen through the organization of events such as the DH Colloquium series and DH conferences and workshops, both nationally as well as internationally. Second, researchers will need guidance to move into the field of DH. This requires reskilling and training, mostly in the more computational aspects of the field. Additionally, a support network is indispensable for researchers entering the DH field. Both DH education and the development of a community of practice is envisioned in the Escalator project. Specific attention will be paid to introducing DH scholarship and methods to Historically Disadvantaged Institutions (HDIs) and strategically partnering them with institutions where such work is already more advanced.

The unified approach of the Escalator project provides a central place of information (educational material), illustration of ongoing DH scholarship, and the building of a network in the form of a community of practice, which should also address the inherent interdisciplinary and collaborative nature of DH.

7.2. The Commercialisation Strategic Focus

This strategic focus is consistent with **Objective c**: To provide digital resources and tools for the development of a wide range of language-technology applications. It was identified at the RI's strategy meeting that SADiLaR was lagging behind other similar infrastructures world-wide, with respect to commercialisation and computational integration. The commercialisation strategic focus will articulate processes of leveraging current tools and resources in developing viable commercial outputs in line with DSI requirements. In order to develop this strategic focus area, certain baseline activities are required as well as the need to ascertain, from the RI and its governance structures, what commercialisation entails. The RI, where practically feasible, needs to be responsive to the imperative of the National System of Innovation (NSI), in order to contribute to the DSI's socio-economic transformation efforts.

Role of governance as part of SADiLaR's commercialisation strategy:

- Processes to be defined: Common terminology clarified and IP and revenue share discussed
- between different project partners
- Structure to be defined: Organisation and business-development processes to be confirmed
- SADiLaR needs to play the role of facilitator

Existing commercialisation possibilities:

- Start with existing areas of expertise and possible quick wins in new areas with low investment
- Basic text processing – in many ways non-revenue bearing
- Speech recognition – potentially revenue bearing

Areas to be developed:

- Child Language Capability Measure (CLN) potentially revenue bearing
- Analytics – revenue-bearing potential
- Service-orientated mindset that would enable sharing, use and sustainability of language data tools for research

Current commercialisation environment:

- Budget reduction of 10% makes introduction of commercialisation more critical^[1]_[SEP] but also more
- challenging, given constraints with regard to provision of seed funding, for example
- Economic power of small language groups is limited, meaning commercial return on investment is limited, and true commercial investment appears to be limited to larger language groups
- Non-revenue bearing commercialisation seen as an outcome of government funded RI

7.3. The Communication Strategic Focus

The imperative to develop a strong communication strategy has been articulated as a need more than once by stakeholders, as well as during internal analysis.

Communication is a vital factor in development efforts, whether in the form of information dissemination, offering guidelines and prescriptions, instigating advocacy, or prompting persuasion or dialogue. The strategy must also align with SADiLaR's mission and vision.

Aims and objectives of communication strategy:

- The communication strategy includes communication with all stakeholders (internal and external) to ensure that all the communities are reached and that SADiLaR plays a central role in enabling and facilitating access to digital research and development within the HSS.
- In reaching these objectives, the communication strategy is set out in accordance with the following specific RI aims, while also aligning it with the DSI Communication Strategy where synergies exist:
- Increase an understanding of the HSS – within communities as well as the broader public – of the nature of SADiLaR as an RI as well as its operations
- Building and fostering relationships with key SADiLaR stakeholders
- Strengthening internal communications to achieve acceptance and commitment from the core of the SADiLaR hub
- Strengthening communication and relations with stakeholders by reaching various target audiences via the right channels
- Disseminate outputs of SADiLaR projects to key stakeholders
- Get buy-in from stakeholders to facilitate commercialisation and deliver on end-user products/services

Key messages of SADiLaR:

- SADiLaR is an RI as part of the DSI SARIR programme
- SADiLaR enables research and development of South African languages to assist communities to receive information in their mother-tongue languages, to study in their languages, and to contribute to the knowledge community
- Support for open access to language resources and tools available to assist research and development within the HSS
- Open for collaboration opportunities – data storing/distributing, research projects, digitisation, terminology development, commercialization
- Boost mother-tongue education; provide centralised resources for several institutions and digital language resources for online learning

In 2019, SADiLaR had a strong community/direct drive, and the hub reached out to various institutions face-to-face with language celebrations. The Covid-19 pandemic led to restrictions in face-to-face communication and prompted a move to virtual engagements. It will be important to keep messaging personal.

The strategy of 2021-2023 will have a strong focus on reaching our audiences virtually, with a personal digital touch. The strategy will result in building the SADiLaR community, foster good and long-term relationships, and also have a strong outreach to external stakeholders for funding, collaboration and commercialisation.

7.4. The Research and Training Strategic Focus

This strategic focus is consistent with the **Objective b**: To open new frontiers of research in the HSS in general. This deals directly with the research performed and the outputs attained. It is also directly related to the DH focus, which deals with the development and advancement of DH as well as strengthening transdisciplinary transformation.

As a current starting point, a relatively large team of researchers is currently active at SADiLaR's hub. This means that a relatively large range of research topics can be addressed, encompassing all the official languages in South Africa. Additionally, SADiLaR hosts a repository that contains much of South Africa's natural language processing resources. This starting point aligns well with a research focus on (computational) linguistics in the field of DH. The repository will also be positioned to play a key role in providing data that can stimulate research in the fields of data science, language technology, and artificial intelligence.

Even though the field of DH and the corresponding research methodologies are still relatively unknown in South Africa, the DH's strategic focus aims to tackle this. However, given this situation, a large range of topics in this field are still unexplored. This provides a huge opportunity to boost research in this area.

On a practical note, SADiLaR aims to identify virtual labs that will be hosted by a researcher in the hub, but external researchers (e.g., from the nodes and other universities) are welcome to join. This will introduce several focus areas, allow for increased collaboration (internal and external), and will also boost training of (computational) skills and

methodologies. The virtual nature allows for quick and flexible reorganisation to tackle upcoming topics. The active creation and applicable curation of online training materials will position SADiLaR to become a service hub for easy and open access and long-term storage of training materials.

7.5. Stakeholder Engagement Strategic Focus

The imperative exists to identify and cultivate relations with relevant local and international stakeholders with similar and/or complementary RI and mandates, as it also links with relevant strategic DSI priorities. The aim is to be an actor at applicable national and international meetings, conferences and workshops for the advancement of SADiLaR's broad national mandate. To promote its DH research, innovations and digital language programmes through a robust communication that fosters and enhances national and international research cooperation.

Stakeholder engagement will be classified according to power/influence as well as general interest in the Centre's activities in order to engage in a context-specific manner, which is also in line with the communication strategy.

Key areas of engagement include:

- Boosting collaboration with DHET, DBE, DSAC and PanSALB regarding access to resources as well as collaboration on terminology development
- Engagements around Language in Education
- Mother-tongue education
- Digital language resources for online learning
- Centralised resources for several institutions in South Africa
- Historically Disadvantaged Institutions (HDIs)
- Strengthen and foster collaboration with international and African partners to, for example, leverage funding and training/education resources as well as promote knowledge sharing
- Strengthen engagement with different government departments to create awareness of SADiLaR's activities
- Monitor and explore connections to state-owned enterprises, industry, and the broader research-infrastructure landscape, nationally and internationally

- Strategy End -