|  |
| --- |
| MEMORANDUM FROM THE PARLIAMENTARY OFFICE |

**NATIONAL ASSEMBLY**

**FOR WRITTEN REPLY**

**QUESTION 997**

**DATE OF PUBLICATION OF INTERNAL QUESTION PAPER: 29/05/2020**

**INTERNAL QUESTION PAPER NO 18 OF 2020**

**Mr S L Ngcobo (IFP) to ask the Minister of Higher Education, Science and Technology:**

(1) Given that Monday, 25 May 2020, is marked Africa day, and seeing that indigenous African languages are faced with the unique challenge of adapting to a fast-changing technological era, what steps has his department taken to promote the ideal of a multilingual society as espoused in the Constitution of the Republic of South Africa, 1996;

(2) whether he has found that there are digitised efforts and carved-out spaces for indigenous languages within the digital space for them to not only survive, but also to thrive in the ever-changing technological era; if not, why not; if so, what are the full relevant details?

**NW1286E**

**REPLY:**

(1) The Department of Science and Innovation (DSI), through the South African Research Infrastructure Roadmap (SARIR), established the South African Centre for Digital Language Resources (SADiLaR). SADiLaR has an enabling function, with a focus on all official languages of South Africa, supporting research and development in the domains of language technologies and language-related studies in the humanities and social sciences. The Centre supports the creation, management and distribution of digital language resources, as well as applicable software, which are freely available for research purposes through its online repository.

The resources include language datasets (for all official South African languages, including the indigenous languages) as well as high-level resources, such as natural language processing tools that 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, and a variety of text-related applications, such as grammar and spelling checkers, online electronic dictionaries, and so forth.

SADiLaR plays a strategic role in ensuring the constitutional imperative is achieved in the long term to ensure that the historically diminished use and status of the indigenous languages of the people of South African are redressed and positive measures are taken to elevate the status and advance the use of these languages.

The Recognition of Prior Learning is an initiative of the DSI, which through the implementation of the Protection, Promotion, Development and Management of Indigenous Knowledge Act No 6 of 2019, aims to recognise the skills of indigenous practitioners in various IKS domains. The initiative focuses on the development of a competency-based qualification to be registered on the National Qualifications Framework. The Department is currently working with IK practitioners (Traditional Health Practice IK domain), to scope their

competencies of their various cultural settings, and has so far documented competencies in isiZulu, Setswana and TshiVenda languages. The workshops with IK practitioners are conducted in the vernacular languages of the IK practitioners. To this end, the IK occupations and accompanying competencies that are documented in the vernacular languages serve as a principle and as a means to promote and preserve the languages of the knowledge systems in its own context.

The National Recordal System (NRS) of the DSI supports the Protection, Promotion, Development and Management of Indigenous Knowledge Act No 6 of 2019 (herein after referred as the IK Act) through the registration of IK. The initiative promotes the recording of IK in vernacular languages using multimedia technology (recording of audio, video, images and transcriptions of each recorded IK story), as a means to preserve IK for future generations so that the context is not lost. Further hereto, the aim is to protect the IK from biopiracy and misappropriation, and to enable the sharing of benefits to the local and rural communities who have registered such IK in the system, should the knowledge be used by any 3rd party, following the various legal prescripts of the IK Act, No 6 of 2019. A key element of the NRS in the promotion of the vernacular languages is by having IK recorders from the participating communities to implement the documentation of IK. In this way, the youth are exposed to the value of their community IK, and through using their languages they are able to capture extensions of the very rich IK that are held by their own communities. The registered IK is held in a digital repository that stores, provide access to, transmit, manage and secure the registered indigenous knowledge via the digital platform.

(2) The South African Centre for Digital Language Resources provides a digital space for language resources and tools as part of its online repository available at https://repo.sadilar.org/. SADiLaR, through its nodes, focuses on ensuring African Languages are digitised, relevant text and speech processing technologies are developed, terminology development is supported through the creation of wordnets (which are large lexical databases containing nouns, verbs, etc. and their relationships) and language testing and training projects.

SADiLaR funds and supports a range of projects related to indigenous languages in collaboration with SADiLaR’s nodes (consisting of University of Pretoria (Department of African Languages); University of South Africa (Department of African Languages); CSIR (HLT Research Group); North-West University (Centre for Text Technology); and Inter-Institutional Centre for Language Development and Assessment (ICELDA). Projects relate to digitization, semantics and terminology, language development and teaching resources, speech resources, and text resources and technologies.

Collaboration between the North-West University, University of Pretoria and the CSIR in the area of Human Language Technologies predates the establishment of the SADiLaR. The development of a Human language technologies (HLT) speech-activated multilingual service delivery platform was funded from the European Union Government Budget Support programme, between 2014 and 2017. The platform is aimed at providing technology tools necessary for delivering information and services to South African citizens in their language of choice, in an affordable and sustainable manner. The focus was on the development of core technologies in automatic speech recognition (ASR) and text to speech (TTS) using mobile phones as the primary communication channel, furthermore, providing an HLT-enabled solution for website accessibility to print-disabled and low literate end-users.

The aim of the solution was to enable access to information and promote multilingualism. The solution involved the integration of TTS voices in South African English, Afrikaans and isiXhosa with the Non-Visual Desktop Access screen reader. Cape Access (CA) of the Western Cape Government was identified as a possible government partner following a need expressed to make their websites more accessible. CA identified 11 eCentres in which to pilot this technology. A demographics survey was conducted at these eCentres to determine who the typical visitors to these eCentres are and how they operate. After this, eCentre managers were trained on how to use the technology and the technology was subsequently installed and piloted at these eCentres.

The HLT-enabled solution which was also piloted at Kaleidoscope SA (Institute for the Blind). This pilot aimed at allowing blind students to use the Non-Visual Desktop Access (NVDA) screen reader with local languages as a basis for receiving training. Kaleidoscope SA offers formal qualifications (N4 & N5) in a number of fields to blind students.

Furthermore, an activity aimed at assessing communication practices and needs of multilingual persons using augmentative and alternative communication (AAC) was undertaken. The research was undertaken in collaboration with the Centre for Augmentative and Alternative Communication (CAAC) at the University of Pretoria and entailed the integration of CSIR Text-to-Speech (TTS) voices with AAC software. Two sets of evaluations were held and the local voices evaluated were South African English, isiXhosa, isiZulu, Afrikaans and Setswana.

The DSI is also currently funding the Centre for Artificial Intelligence and Research (CAIR), which has a node at North West University. This particular node’s area of expertise is led by a Multilingual Speech Technologies (MuST) research group focused at the creation and use of speech technologies in the less-resourced languages.

SADiLaR, through its involvement with the UNESCO Year of Indigenous languages, reached more than 850 participants directly through language celebration events. These events created a space for academics, lecturers, students (undergraduates – postgraduates), broader public as well as profound contributors in the various languages to interact, and were held across South Africa at various universities in cooperation with the National Lexicography Units of South Africa. These events culminated in SADiLaR taking part in the Language Technologies for All conference with a focus on Enabling Linguistic Diversity and Multilingualism Worldwide, creating awareness of how the South African Research Infrastructure Roadmap is directly contributing toward linguistic diversity and multilingualism through SADiLaR.

SADiLaR is also brainstorming its COVID-19 response, in particular to allow for “Rapid situational awareness in emerging situations like natural disasters or disease outbreaks”. This requires availability of Human Language Technology not only for the official languages of the country, but all languages spoken in South Africa.