

Using machine learning to improve the science-policy-society interface on the SDGs in South Africa

Willem Fourie, Teresa Kroesen and Jurgens de Lange, South African SDG Hub, University of Pretoria, South Africa
(willem.fourie@up.ac.za; teresa.kroesen@up.ac.za; jurgens.delange@up.ac.za)

Abstract

The [South African SDG Hub](#), hosted by the University of Pretoria (UP) in South Africa, uses machine learning to increase the uptake of research relevant to the Sustainable Development Goals (SDGs) into national policymaking, as also reported in the journal *Nature* (Fourie et al., 2019). Leveraging a three-way partnership between UP, the Presidency of South Africa and the German government through the GIZ, the South African SDG Hub provides policymakers with cost-free access to more than 150 000 research items tagged in terms of one or more of the SDGs. In February 2023 it launched an innovative tool aimed at making it significantly easier for policymakers to access relevant research.

Technology undoubtedly has a major role to play in enabling evidence-informed policymaking in Africa. The experience gained from building and expanding the South African SDG Hub helps to highlight some of the opportunities but also limitations inherent to public digital infrastructure used for this purpose. One of the key technologies of the SA SDG Hub is its classification algorithm, which utilizes the abstract and title of harvested records to automatically classify articles in terms of the 17 SDGs. Another is its Intelligent Search tool, which recommends the most relevant research from its database based on PDFs provided by users.

Context: Barriers to using research in SDG-relevant policymaking

In theory, African universities' increasing research output has the potential to improve policymaking significantly (e.g. Chiware and Becker, 2018). By using this growing reservoir of local research, policymakers can develop and implement more impactful policies.

In practice, however, the growth in African research productivity will not lead to better policies as a matter of course. There are several reasons for this.

The *complexity of evidence* is often misunderstood, as policymakers rely on a variety of evidence types beyond peer-reviewed journal articles, including practical experience and political know-how (Hunsmann, 2012). However, using just one type of evidence presents challenges, as peer-reviewed research can come from different sources with varying methodologies, perspectives, and ideologies (e.g. Juntti et al., 2009).

The *absence of personal relationships* between researchers and policymakers is a key barrier to the uptake of research evidence (Innvaer et al., 2002 and Lorenc et al., 2014). Weak relationships can impact the relevance and timeliness of research, and technical

terminology can exclude non-academic partners from the knowledge-creation process.

Peer-reviewed research can be *time-consuming*, which can create a disconnect between research timeframes and policymakers' need for urgent responses (e.g. Innvaer et al., 2002). Additionally, the perceived irrelevance of research is more about how it is communicated than its content, as academic writing can obscure its practical relevance. Developing actionable messages for decision-makers can help overcome this barrier.

The *lack of analytical capacity* among government workers can also hinder the uptake of research evidence, particularly when facing time pressures and a high volume of evidence (e.g. Howlett 2009). Finally, *budgetary constraints* present another challenge, as evidence-based policy interventions can be costly.

Project: South African SDG Hub

Using machine learning, the South African SDG Hub contributes to strengthening Africa's digital policymaking infrastructure. It provides policymakers with free access to peer-reviewed research, classified in terms of one or more of the SDGs.

The platform currently addresses some of the challenges outlined in the previous section, focusing on the following:

- *Access.* By harvesting university repositories, the South African SDG Hub offers policymakers access to research at no cost.
- *Relevance.* The platform's machine-learning-based classification tool addresses the problem of relevance by classifying research in terms of the SDGs. In addition, the platform's advanced search-relevance tools can provide a more relevant ranking of search results than traditional keyword-based search engines.

- *Time.* The platform saves policymakers time by offering them centralised access to research produced at South Africa's public universities.

Technical features: Using machine learning to ease policymaker access to research

The South African SDG Hub uses institutional repository metadata to collect and classify research from South African and selected non-South African universities in terms of the SDGs and redirect traffic back to the originating repository.

The platform benefits from the Open Access Initiative for Public Metadata Harvesting (OAI-PMH) protocol to gather this metadata. The OAI-PMH helps to create and maintain standards for exposing data and content in a way that allows it to be accessed, received, and consumed by a variety of end-users or software, with the ultimate goal of open access. This enables the Hub to harvest relevant data from universities' repositories. After collecting the data, it is cleaned and several filters are applied to ensure that only high-quality, relevant, and complete data is provided.

One of the key technologies of the South African SDG Hub is its classification algorithm, which utilizes the abstract and title of harvested records to classify articles in terms of the 17 SDGs. The algorithm is built on a pre-trained neural network using Google AI's Bidirectional Encoder Representations from Transformers (BERT) architecture (Devlin *et al* 2019), which is then further trained on over 60,000 documents pre-classified according to SDGs. This has resulted in the creation of a new model, sdgBERT, which includes an additional classification layer that converts the embedding layer into a probability layer, providing a probability of the text belonging to each SDG.

To facilitate user accessibility, the South African SDG Hub offers three search options.

Recent innovation: Intelligent Search

In February 2023, the South African SDG Hub [launched a new tool](#), aimed at revolutionising how policymakers access research.

Called 'Intelligent Search' this tool provides an alternative to the platform's current keyword and semantic searches.

The new tool allows users to upload any PDF, based on which the most relevant articles from the platform are recommended. This is done through a topic analysis of both the uploaded PDF and the platform's entire database. Any records from the platform's database that

share the same underlying semantic structure and discusses similar topics are returned as relevant articles. This algorithm allows policymakers to search for relevant research that might not be easily encapsulated into a set of keywords or key phrases, as is done in traditional search routines.

Figure 1. The South African SDG Hub's three search tools, highlighting Intelligent Search

Data source: Authors

Lessons learned and recommendations

At a high level, the South African SDG Hub and its partners have learned that *technology on its own is not the full solution*. Whereas technology can play a major role in enabling better access to policy-relevant research, it needs to be supplemented by human engagement and intervention. This is why the South African SDG Policy Support Initiative was launched as a complement to the research platform. The South African SDG Policy Support Initiative connects policymakers to experts based at South Africa's 26 public universities.

Regarding users, it is clear that *end-users should be involved continuously* in determining both the type of tools that are developed and how they are presented on the platform. Conducting interviews and user experience testing to understand the needs and

usability can help to ensure the platform is user-friendly and meets the needs of its intended audience.

At the level of data, it was found that *not enough attention is given to quality metadata*. There are standards (e.g. OAI-PMH) that help to access data, but standards are not effective if they are not upheld. They must be adhered to ensure the reliability of data and information being shared. It is important to check and verify the data quality.

Practically, this means that institutional repositories – the source of the South African SDG Hub’s data – should be incentivized to devote more time and energy to ensure metadata, such as publication date, authors, source publication, etc. are more accurate.

References

- Chiwere, E. R. T. and Becker, D. A., 2018, Research trends and collaborations by applied science researchers in South African universities of technology: 2007–2017, *The Journal of Academic Librarianship* 44(4).
- Fourie, W., Van der Walt, I., Strydom, H. and Marais, C., 2019, South African system tracks SDG research, *Nature* 573.
- Howlett, M. 2009, Policy analytical capacity and evidence-based policy-making: Lessons from Canada, *Canadian Public Administration*, 52(2).
- Hunsmann, M., 2012, Limits to evidence-based health policymaking: Policy hurdles to structural HIV prevention in Tanzania, *Social Science & Medicine* 74.
- Innvaer, S., Vist, G., Trommald, M., Oxman, A., 2002, Health policy-makers’ perceptions of their use of evidence: a systematic review, *Journal of Health Services Research & Policy* 7(4).
- Juntti, M, Russel, D., Turnpenny, J. ,2009, Evidence, politics and power in public policy for the environment, *Environmental Science and Policy* 12.
- Lorenc, T., Tyner, E. F., Petticrew, M., Duffy, S., Martineau, F. P., Phillips, G., Lock, K., 2014, Cultures of evidence across policy sectors: systematic review of qualitative evidence. *European Journal of Public Health* 24(6).
- Devlin, J., Chang, M-W., Lee, K., Toutanova, K. 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of NAACL-HLT 2019*, 4171(15)