# BIG DATA 4 PEACEBUILDING IN SOMALIA
RADIO CONTENT ANALYSIS TOOLKIT

The idea behind the **RADIO CONTENT ANALYSIS TOOLKIT** was to determine if radio data streaming and analysis can provide quality information in support of peace and state building processes in Somalia. UN Global Pulse rolled out new prototypes of the radio toolkit—first launched in Uganda—to analyse information from public radio broadcasts out of Mogadishu.
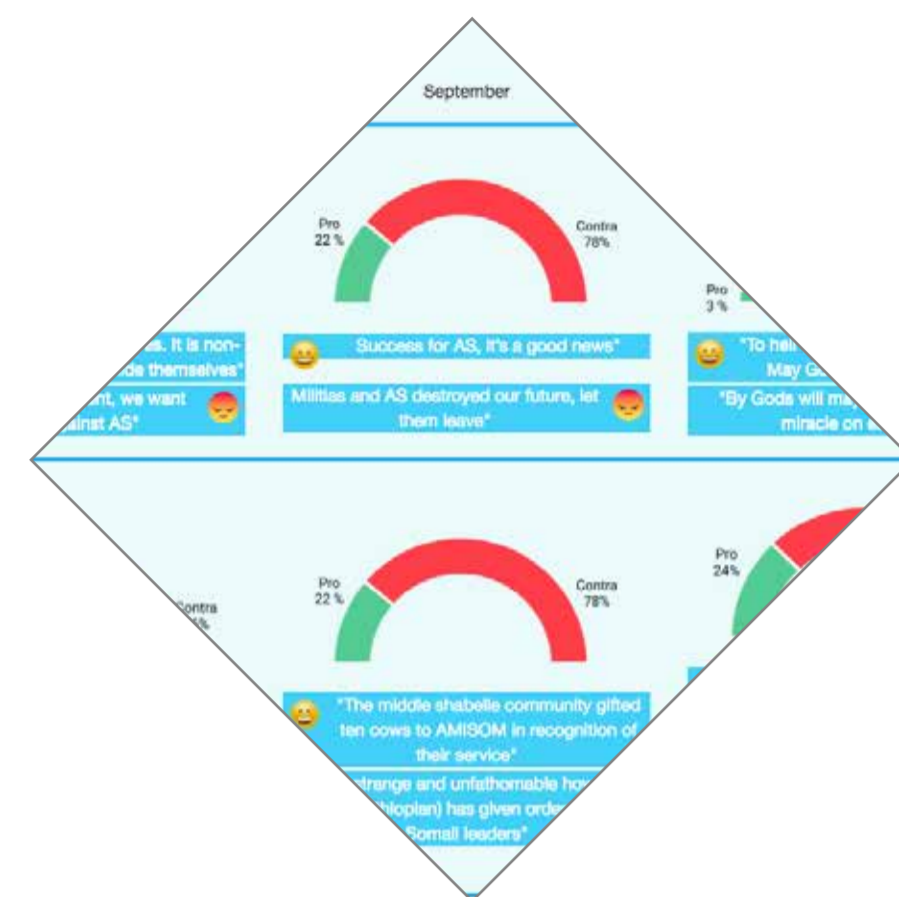
With support from the United Nations Peacebuilding and Support Office (PBSO), an automated technology for mining radio content was specifically designed for the Somali language. It analyses testimonies, rumors, opinions and reports shared on radio broadcasts. Once analysed, the data provides insight into the usefulness and effectiveness of governmental and humanitarian approaches across the country.

The prototype builds on previous UN Global Pulse efforts. The first version of the prototype, funded by the Government of Sweden, was tested in Uganda (using the Lugbara and Acholi languages). The success of the pilot, and potential of the software, led to new partnerships with PBSO, the United Nations Development Operations Coordination Office (UN DOCO). We are now in the midst of improving a second prototype.

Employing this digital innovation allows us to understand how big data can support efforts to monitor publicly-expressed perceptions and behaviors relevant to the ongoing peace and state building processes in Somalia, and doing so without exposing personnel to security-risks.

Global Pulse teamed up with Stellenbosch University of South Africa and PBSO, setting up hardware in Mogadishu and developing a speech recognition system for Somalia. The team streams content broadcast in Somali language from eight radio stations.

Roughly 800 radio clips (each five minutes in duration) are streamed through the cloud every day. Software filters music from audio content to zone in phone-in discussions that have been prioritized for analysis.

A beta version of the keyword spotting system for Somali language was developed using deep learning methods. This new version allows an easier and faster deployment for new languages. Where previous versions of the radio tool relied on large quantities of transcribed speech recordings, the next generation was built with far less training data.

Radio Content Analysis Toolkit developed for Somali language

- 3,906 hours/week in all radio stations in Mogadishu
- 1,008 hours/week in 8 targeted radio stations
- 653 hours/week only speech (no music)
- 251 hours/week targeted radio shows
- 40 hours/listening capacity of 1 analyst

Using filters, the technology prototype reduces 4,000 hours per of public radio content in Mogadishu each week to 40 hours of relevant content on a certain topic. This makes it easier for analysts to have a better understanding the information shared across the airwaves. To date, the software can filter content and automatically identify words programmed into the system with 40% accuracy.

A global innovation, the Radio Content Analysis Toolkit helps sustainable development, humanitarian practice and social change-makers in Somalia identify and filter discussions that previously existed in long, unfiltered and inaccessible formats. The next step is to improve the accuracy of the toolkit and deploy it outside of Mogadishu.

This advancement will help policy makers, government officials, and development partners remotely monitor Somali public perceptions and behaviours as they pertain to the ongoing peace, governance, and security situations in the country.