Project: **Consolidation of Data and Application of Big Data Tools to Enhance National and Transboundary Data Sets in Southern Africa that Support Decision-Making for Security of Water Resources**

This research project was part of a series of four projects under the Big Data Analytics and Transboundary Water Collaboration for Southern Africa, bringing together key stakeholders in Water and Big Data sectors.

The overall goals were to:

- Enhance current understanding of shared groundwater resources.

- Provide big data skills development, capacity building and networking opportunities for Southern African researchers and their students.

- To promote innovative thinking and application of Big Data Analytics to the Transboundary Water sector for integrated decision-making.

There were four interconnected Themes that collaborated on this project:

Theme 1: Consolidation of data and application of big data tools to enhance national and transboundary data sets in Southern Africa that support decision-making for security of water resources.

Umvoto Africa and IGRAC, University of Botswana, GEMS/Water

Theme 2: Imagining solutions for extracting further value from existing datasets on surface and groundwater resources in Southern Africa

Wits University, Geological Services of Botswana, DWS

Theme 3: Localizing transboundary data sets in Southern African: A case study approach

University of the Western Cape, CSIR, L2K2 Consultants

**Theme 4: Groundwater secure transboundary systems**

Delta-H Groundwater Systems & Institute for Groundwater Studies

The study area focussed on the Ramotswa Transboundary Aquifer Area (RTBAA) located on the border between Botswana and South Africa. The RTBAA encompasses a region beyond the strict boundary of the Ramotswa aquifer to include the surface waters that interact directly with the aquifer. The aquifer is important for the upper part of the Limpopo River Basin, shared by Botswana, Mozambique, South Africa, and Zimbabwe. The Ramotswa aquifer makes important contributions to water security in the Gaborone Dam catchment and supplies water to communities on both sides of the border, primarily for domestic use and small-scale agriculture.

Effective management of groundwater resources requires ease of access to reliable data to support decisions. This is particularly important for transboundary systems where different environmental data are collected and managed by different institutions in different countries; thus, impeding the sustainable management of transboundary water resources. The data that was collected By Umvoto and IGRAC as part of this assessment was provided by four organisations: Department of Water and Sanitation (DWS) Botswana, DWS South Africa, IWMI South Africa, and SADC-GMI.

Almost all real-world data and datasets suffer from incompleteness, inconsistencies and errors, and the water resource data received as part of this project were no different. Often these data issues are solved in an ad-hoc and manual manner requiring familiarity with the data. The Umvoto and IGRAC project used a programmatic approach to effectively deal with these issues. The Big Data tools used focussed on statistics, visualization, and data comparisons using the Python programming language.

The project developed software tools to automate the collation and quality control of data from existing online databases into a format suitable for our primary repository, the existing SADC-GIP platform. The steps followed were data gathering, data assessment (e.g. detect outliers or any inconsistencies), data cleaning, and initial data quality control, all of which were undertaken programmatically using Python tools and documented in a Jupyter Notebook This included work done during an internship at IBM by an intern from the University of Botswana. This procedure is summarised as follows:
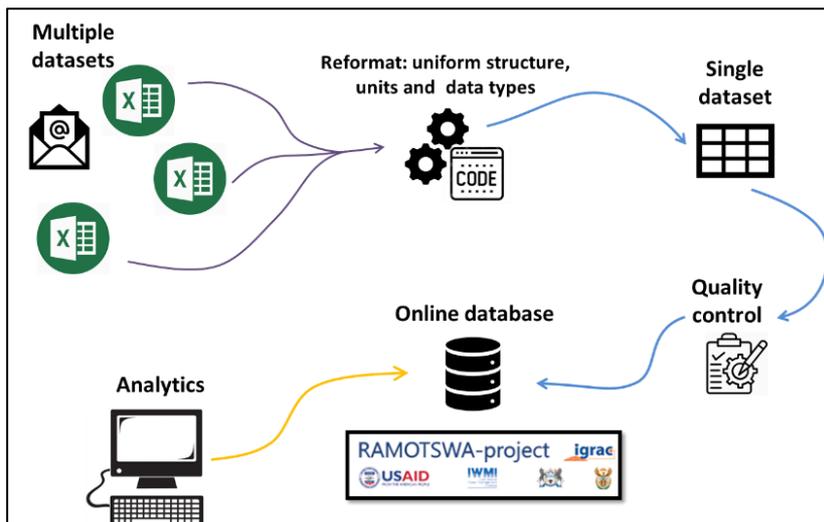
In addition, the data was shared to other online data platforms GGMN and GEMStat. The three platforms were selected as they each provide complimentary benefits.

As part of the project, a quality control system was formalised. This included protocols on designing a monitoring network, procedures for recording data (including metadata, logical and outlier checks), and a standard procedure for handling erroneous or dubious data.

To support groundwater data sharing between the DWS of Botswana and South Africa, protocols on groundwater data exchange were developed. The protocol could be taken up easily by other organisations undertaking projects on transboundary groundwater in the SADC region.

Umvoto Africa (Pty) Ltd

8 Beach Road, Muizenberg 7945
Cape Town, South Africa
P.O Box 61, Muizenberg 7950
Cape Town, South Africa.

Tel: +27 21 709 6700
Fax: +27 86 685 5725
Email: amanzi@umvoto.com
Website: www.umvoto.com