



2023 UN WIDER Development Conference

**Revving up revenue for development
The role of domestic resource mobilization**

**Johnstone Makhubu
Deputy Commissioner for SARS**

6 to 8 September 2023
Oslo Norway

Tax matters: Data to policy

South African Revenue Service

- Tax administrative data
- Evolutionary steps taken by SARS
- Tax administrative data – external perspective
- Digital platforms
- Individual tax filing transformation
- Auto assessments
- Management of VAT returns
- Secure Data Research Lab
- Practical use of SA-TIED outputs
- SARS: The way forward
- Summary
- Conclusion

Tax administrative data

- ❑ Administrative tax data are important for many different reasons:
 - Tax policy design and reform
 - Study taxpayer responses to tax policy reform
 - Non-tax questions, such as intergenerational mobility, firm production networks, monitoring of various economic trends
- ❑ Tax administrative data are collected by tax administrations in the process of exercising its functions – collecting government revenue. There are five broad categories of tax administrative data
 - **Tax Register:** the list of all registered taxpayers at a point in time, with the unique identifier, name, geographic location, by sector and for firms the legal form
 - **Taxpayer self-assessment declarations:** declarations that taxpayers submit themselves such as individuals and firms submit income tax returns, employers submit PAYE declarations, and VAT vendors submit VAT declarations
 - **Informative declarations and withholding declarations:** forms submitted to the revenue authority by third parties that report on taxpayers' taxable transactions; these data are often at the transaction-level and are used for tax compliance purposes
 - **Customs data:** transaction-level records of imports and exports
 - **Process and HR data:** records of internal processes, e.g., audits; remuneration and bonuses of employees; administrative costs of revenue collections

Evolutionary steps – Internal Imperatives

- ❑ As part of its transformation journey, SARS started to make quantum leap advances in the fields of information and communication technology to improve on operational effectiveness and efficiency. SARS' Modernisation Programme, which began in 2007, has seen wide-scale migration to electronic channels, minimised the introduction of human errors which resulted in a high quality, data rich environment.
- ❑ The organisation continues to make use of the available information (and the information received from third parties) on taxpayers to inform risk assessments and reduce the administrative burden for the majority of compliant taxpayers.
- ❑ Through the pre-population of returns and declaration documents with data from own systems and third-party sources, SARS reduced the opportunity (risk) for false or inaccurate declarations. This process further reduced the compliance costs for taxpayers and results in a more efficient administration system for SARS.
- ❑ Most recently, as part of Vision 2024, SARS is committed to increase and expand the use of data within a comprehensive knowledge management framework to ensure integrity, drive insight and improve outcomes.

Tax administrative data – external perspective

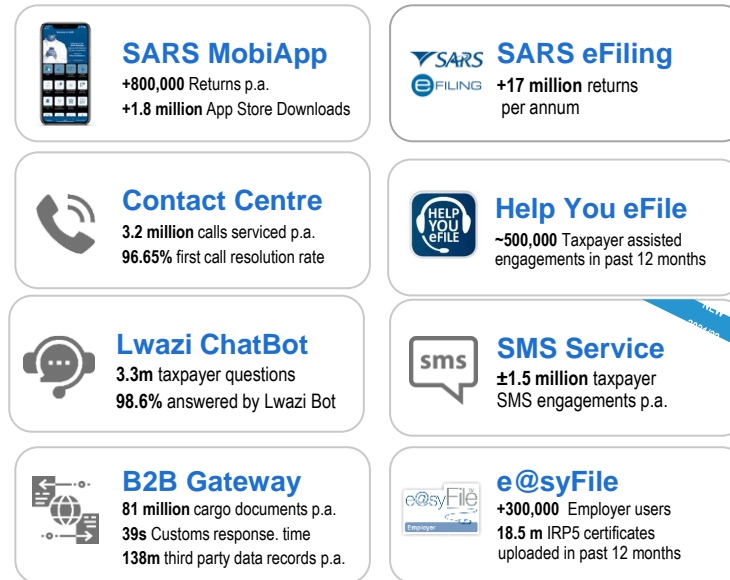
- ❑ Tax administration records find use outside the tax administration process
 - Tax administrative data in many countries underpins the production of official economic statistics.
 - As in many countries, VAT and CIT records are used by Statistics South Africa to create a business register and sampling frames for all official economic sample surveys.
- ❑ Tax records can be used for the direct derivation of statistics useful for planning and evidence-based decision-making.
- ❑ Tax administrative data, particularly at micro- level, provides the natural evidence base for tax policy analysis and policy-relevant research
- ❑ Increasingly tax data is being used in the evaluation of economic and social policy, and to assess the impact of incentives and targeted interventions.
- ❑ Currently collaboration projects with UNU-WIDER under the SA-TIED programme are underway to use tax administration data for research and the construction of micro simulation models based on tax return data.

Digital platforms

DIGITAL CHANNELS




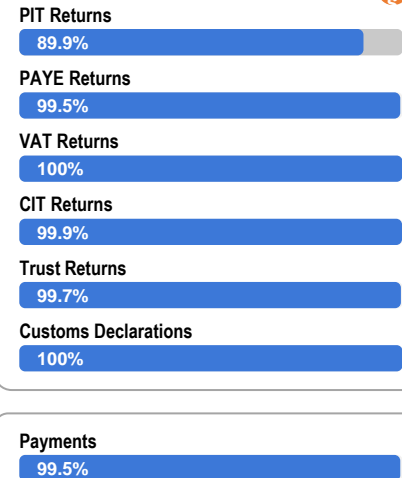
UPTAKE OF OUR SARS DIGITAL CHANNELS



TAXPAYER SUBMISSIONS

Majority of Tax Returns (19m), Customs Declarations (6m) and Payments (16m) submitted & processed digitally.

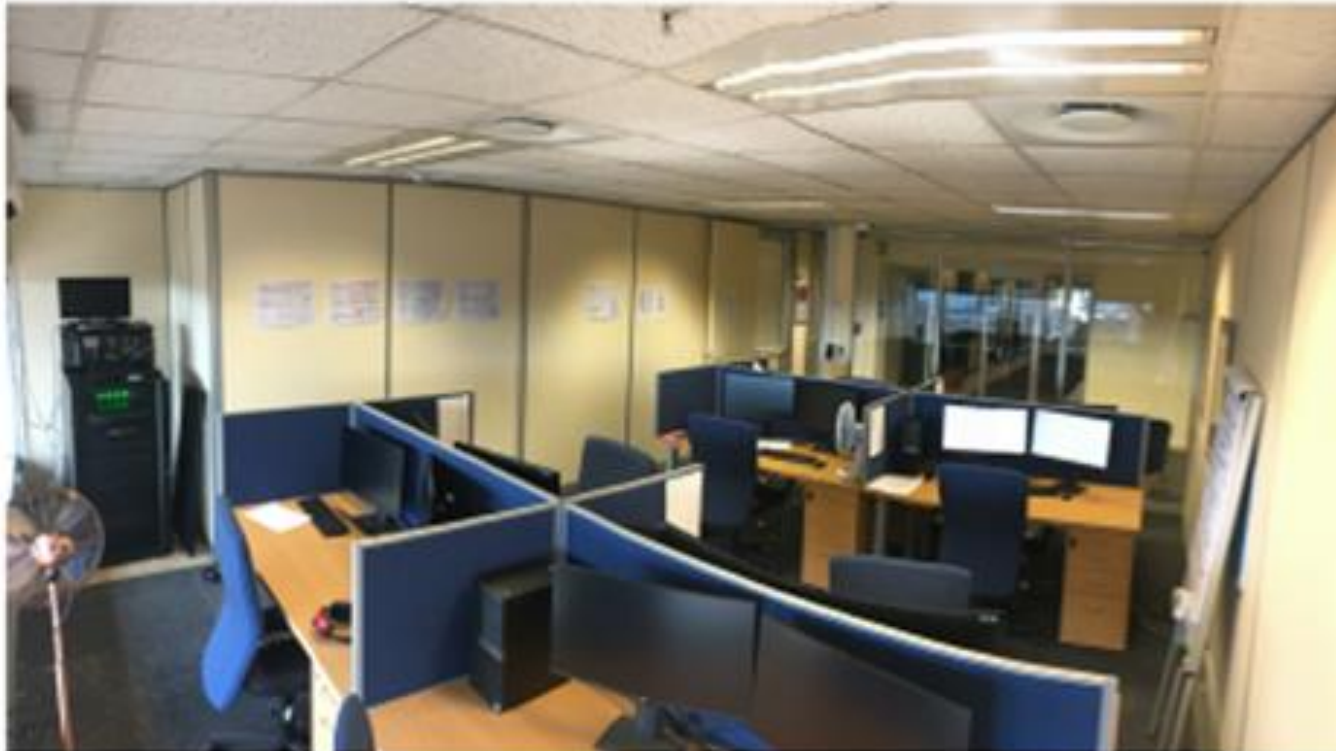
90% processed in less than 5 seconds

+85% Of all taxpayers, traders & intermediaries interact through our digital channels...

The Data Secured Research Lab

THE NATIONAL TREASURY SECURE DATA FACILITY (NT-SDF)



Practical use of SA-TIED outputs

- Southern Africa – Towards Inclusive Economic Development (SA-TIED) in collaboration between local and international research institutes and South African government
 - Build greater research capacity,
 - improve data management and analysis, and
 - provide research outputs that assist in evidence-based policy.

- Research paper: Simulating personal income tax in South Africa using administrative data and survey data: A comparison of PITMOD and SAMOD for tax year 2018
 - EUROMOD microsimulation platform
 - SAMOD simulates personal income tax and social benefit policies derived from National Income Dynamics Study survey (NITS)
 - PITMOD simulates the personal income tax system based on anonymised individual level tax administrative data

Research paper

- ❑ Growing interest to use tax administrative datasets as survey data tends to suffer from unit missing data especially on high-income earners
- ❑ PIT is the most important source of tax revenue in South Africa, contributing 38 per cent of total tax revenue in 2017/18
- ❑ Pressure on fiscal resources with highly skewed distribution of income
- ❑ PITMOD simulates almost all the elements of the policy rules based on all compliant taxpayers while SAMOD is limited to only a few policy rules but also simulates the social benefit options
- ❑ PITMOD underpinning dataset of 14.7 million individuals is derived from a central enterprise data warehouse combining information from tax certificates (IRP5 and IT3(a)) and income tax returns (ITR12) supplemented by third party data on medical scheme contributions

Constructing the administrative data file

- ❑ The Data Analytics team at the South African Revenue Service (SARS) constructed the required data file for the PITMOD
- ❑ The data was configured to have each individual case in the dataset relating to a single individual by creating a unique identifier in the order of completeness, tax reference number, ID number, passport number and certificate numbers that was merged with the ITR12 dataset and medical tax rebates information and finally with the assessed data
- ❑ Construction of the administrative data file was an iterative process by selecting small samples that were explored in detail which led to incremental improvements in the data file to accurately simulate final tax liability on an individual level
- ❑ A total of 1 390 variables were listed in the data file from the specified SARS source codes

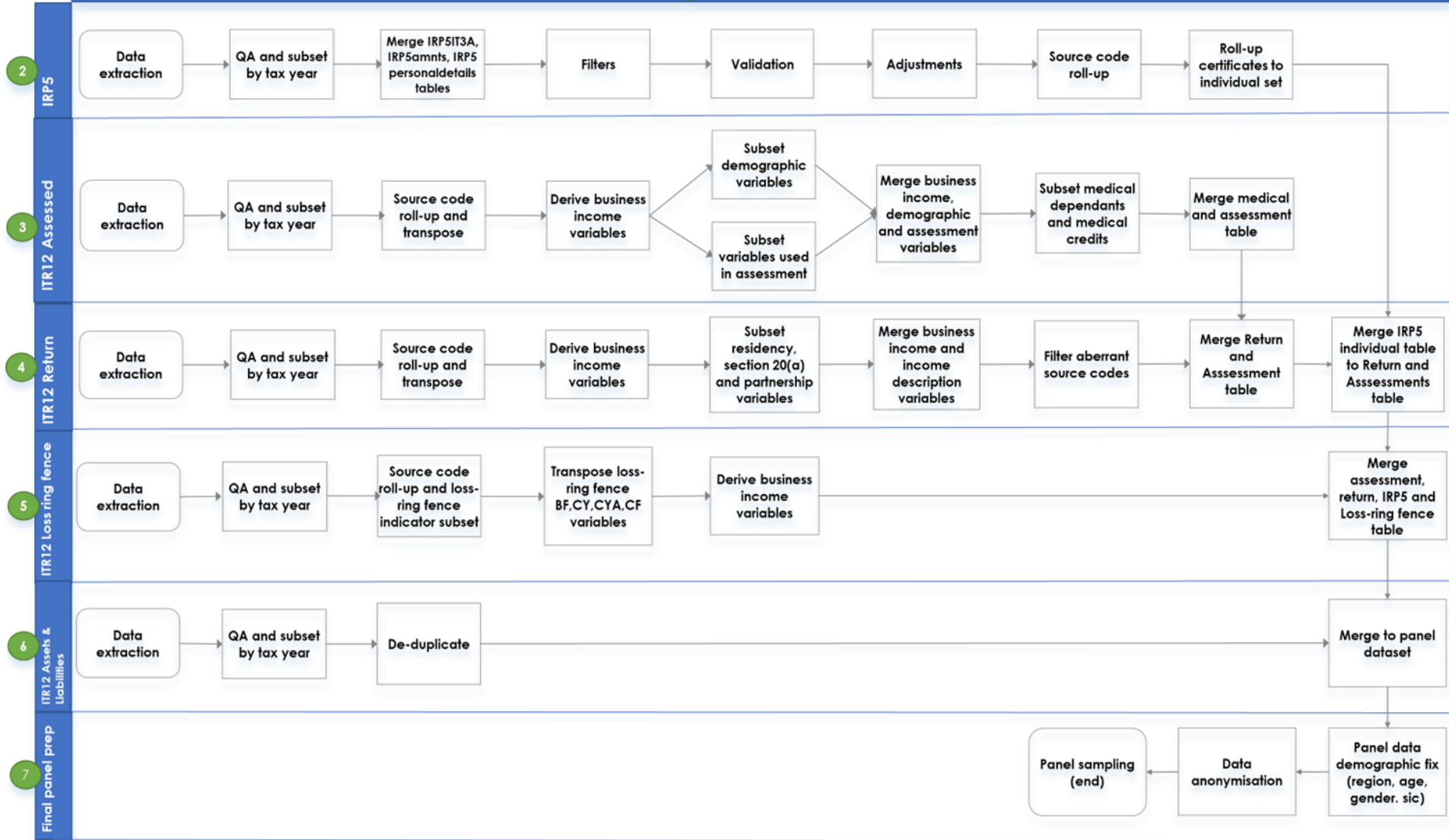
Administrative data process

- ❑ The extracted administrative dataset is anonymised, analysed and quality assured to ensure completeness, consistency, reliability and accuracy with SQL and SAS analytical tools
- ❑ A 10% sample is extracted from the original data set for the simulation of tax policy rules to calculate tax liability
- ❑ Datasets are large, 25GB for the full data set and 2.6GB for the 10% sample
- ❑ Once all the data management and governance protocols are finalised, the data is shared with the SARS PITMOD development team and SASPRI

PITMOD DATA ARCHITECTURE

PITMOD panel data construction process flow

1 Macros (start)



Tax administrative data facilitating research and tax administration

- ❑ Administrative data significantly enhance the quality of national statistics and produce synergies with tax administration and other government agencies, supporting better decision making, policy advice, and economic performance
- ❑ A robust system of data validation and quality checks are necessary to facilitate credible research results and identify non-compliance risks
- ❑ Administrative data sharing mechanisms and governance need to be formalized and strengthened with a memorandum of understanding to ensure confidentiality of taxpayer information
- ❑ Administrative tax data are used for many different reasons for instance to study taxpayer responses to tax policy reform and to institute or improve tax administration measures to improve taxpayer compliance

Categories of tax administrative data

- ❑ Five broad categories of tax administrative data can be identified
 - List of registered taxpayers with unique indicator, name, location
 - Taxpayer self-assessment declarations
 - Informative and withholding declarations by third parties reporting on taxable transactions
 - Customs data on transactional level records of imports and exports
 - Tax administration process data, e.g., compliance and investigative audits
- ❑ Modes of accessing taxpayer data
 - The data is available online (Scandinavian countries and Mexico)
 - Data extracts are shared with institutions under a MoU (Senegal and Pakistan)
 - De-identified data is shared to specific institutions under a MoU to a secure facility (Brazil, South Africa)
 - Remote access to the data via a secure server to selected or screened individuals or institutions
 - Access to data onsite (Datalab UK (HMRC)) External partners can work onsite or with a tax official who runs do-files or scripts

SARS – the way forward

- ❑ SARS is committed to be a data driven organisation with data products, analytics and visualisations that are delivered in a manner that provides one version of the truth and insights to drive optimization and strategic direction of the organisation.
- ❑ The emergence of new technologies such as Artificial Intelligence and Cloud Computing provide new possibilities for improving the efficiency and effectiveness of SARS' administration efforts.
- ❑ In an environment of big data and predictive analytics, the ability to identify risks, and drive evidence-based decisions has a huge impact on improving compliance.
- ❑ If implemented appropriately such technologies could yield savings in Information Technology (IT) infrastructure costs and data accessibility and usage for SARS, businesses and the public at large.
- ❑ The organisation plans to increase the data analytics capability, in both the collection and usage of data to enable better decision-making. The aim is to progress SARS through the analytics maturity curve to become a data and information driven organisation, to increase cost effectiveness and internal efficiencies.

Summary

- ❑ First, building a data centre is costly as sufficient processing power is required and it is important to get the technical configurations correct.
- ❑ Second, the South African experience has shown that it is possible to make sensitive tax data available for research without compromising the anonymity of firms or individuals. Many processes are in place at SARS and the NT-SDF to protect the data. These include non-disclosure agreements between the NT and the researcher and checking of output before sharing with researchers. This bodes well for the possibility of using other administrative datasets in the future and initiating similar tax projects in other countries.
- ❑ Third, development of tax data takes time.
- ❑ Fourth, building technical capacity across many individuals within a number of institutions will be critical to the sustainability of the NT-SDF. Academic capacity in the data science and economics fields are required to analyse the tax data but in addition to this, ICT capacity as well as efficient administrative skills are required to run and manage such a facility.
- ❑ Lastly, making administrative data available requires ‘buy-in’ from all relevant stakeholder including government departments and agencies.

Conclusion

- ❑ With global mature tax administrators, advanced analytics is now a core part of the decision-making toolset.
- ❑ Advanced analytics techniques can be applied wherever evidence-based predictions are needed and representative datasets are available.
- ❑ The analytics strategy helps with the evolutionary movement from the initial focus on case selection, to addressing problems in debt management, non-filer programmes, customer service, and policy evaluation.
- ❑ To enable this, the suite of techniques used should expand to include text mining, unsupervised learning, uplift modelling, and explanatory modelling.
- ❑ The key to successful application of these techniques is to carefully match the analytic approach used to the exact business problem to be addressed.
- ❑ In addition to solving the relevant analytical problems, SARS should also learn to address the major organisational challenges presented by advanced analytics:
 - How to achieve effective collaboration between analytics and business experts;
 - How to manage the uncertainty inherent in most analytics projects;
 - How to build datasets that meet analytical requirements; and, finally,
 - How to manage the process of change in order to achieve tangible results from new analytical models.
- ❑ To fully capitalise on the opportunities presented by advanced analytics, SARS should continue to collaborate actively with other countries to fully explore the analytical world.

Thank you
Re a leboha
Re a leboga
Ndza Khensa
Dankie
Ndi a livhuwa
Ngiyabonga
Enkosi
Ngiyathokoza