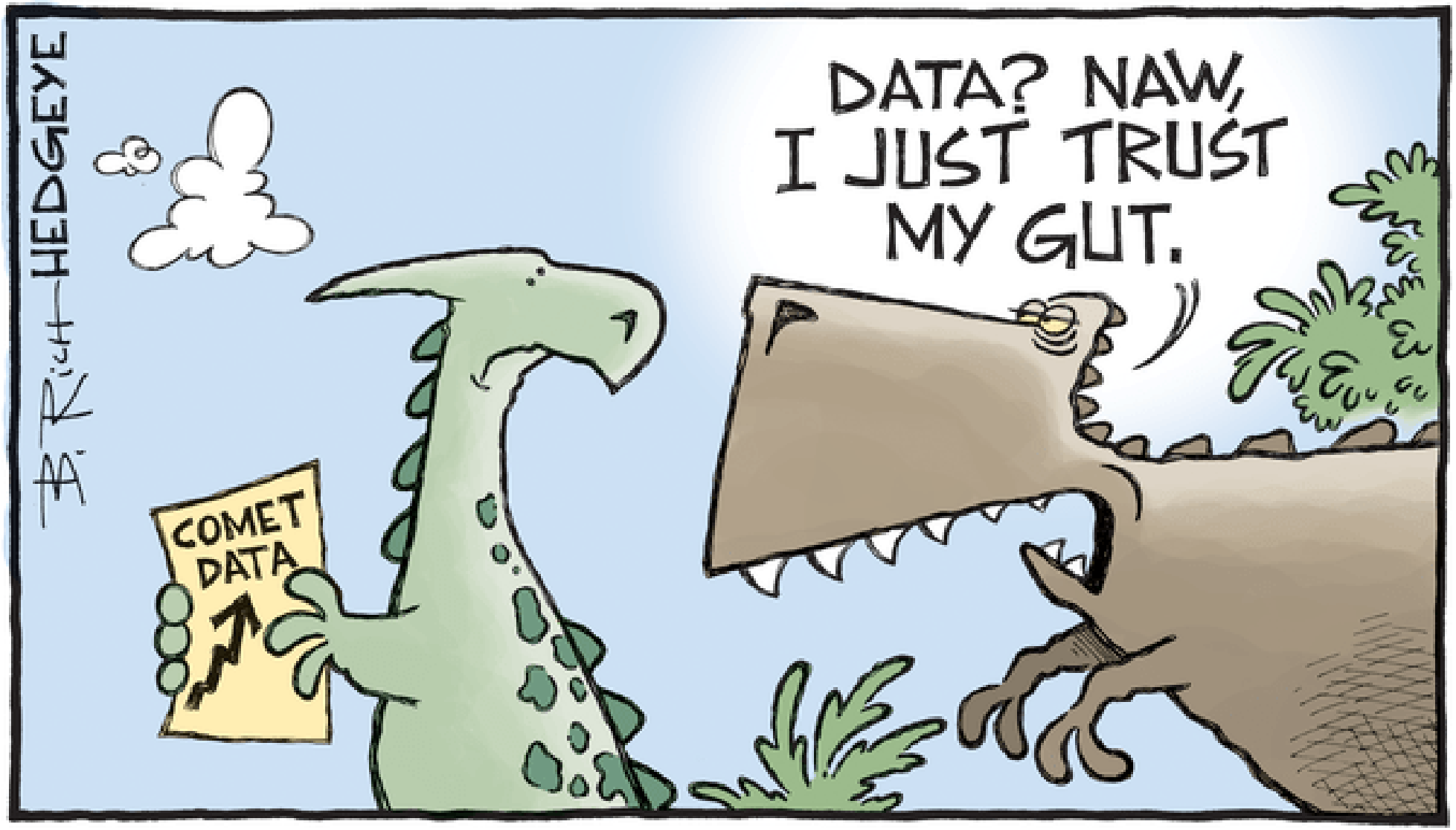
Working with Data in South Africa: Challenges and Opportunities

SARB-UCT Financial Stability Research Conference 2023

Dr Daan Steenkamp

www.codera.co.za


HEDGEYE
DATA? NAW, I JUST TRUST MY GUT.
COMET DATA

Data Matters

- Reliable and timely data important for sound policy & decision-making
- Data access & sharing promotes trust & technological adoption
  - Social & economic benefits between 1%-2.5% of GDP (OECD)
- Data transparency strengthens government credibility
  - May even lower borrowing cost (Gonzalez-Garcia, IMF, 2022)

Overview

Data

- Availability
- Quality
- Regulation

Challenges and Opportunities

FAIR Principles for data governance & sharing

F

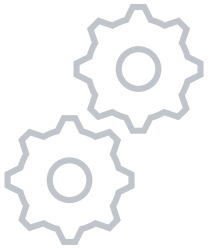
Findable

A

Accessible

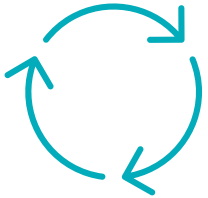
I

Interoperable



R

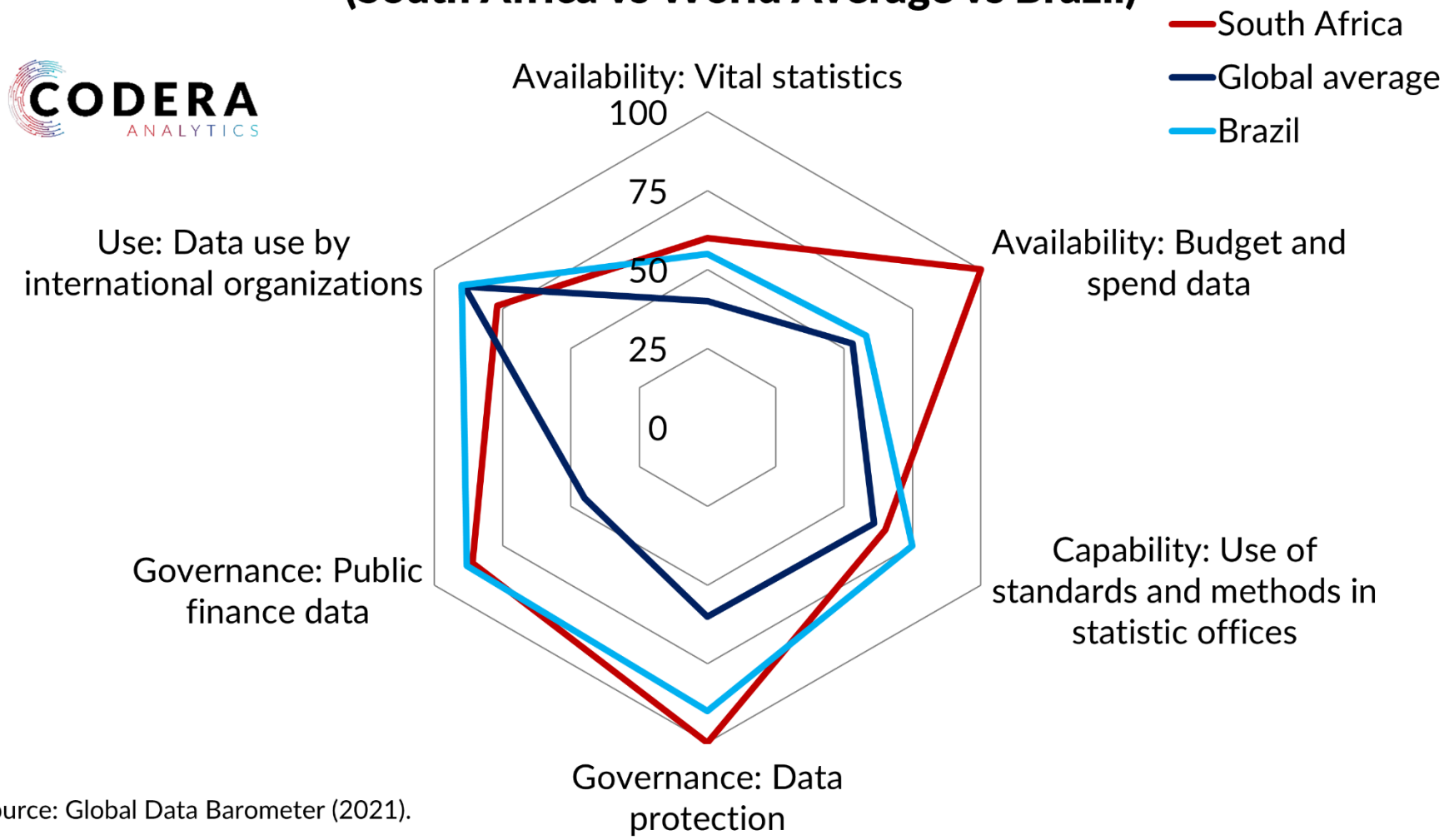
Reusable



SA's data ecosystem: Comparative Strength

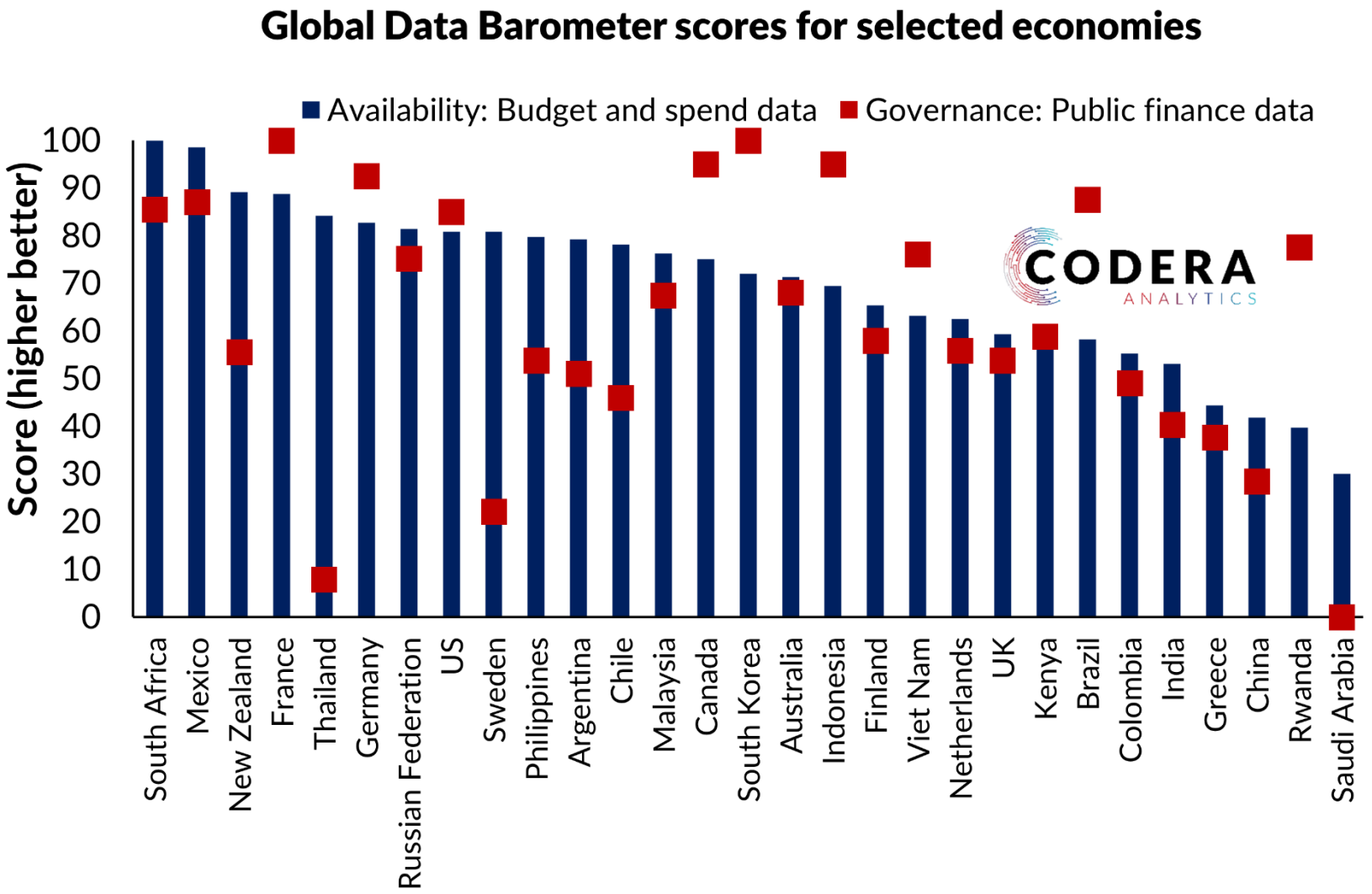
We do well on vital statistics & data protection

Global Data Barometer scores for selected economies (South Africa vs World Average vs Brazil)



Source: Global Data Barometer (2021).

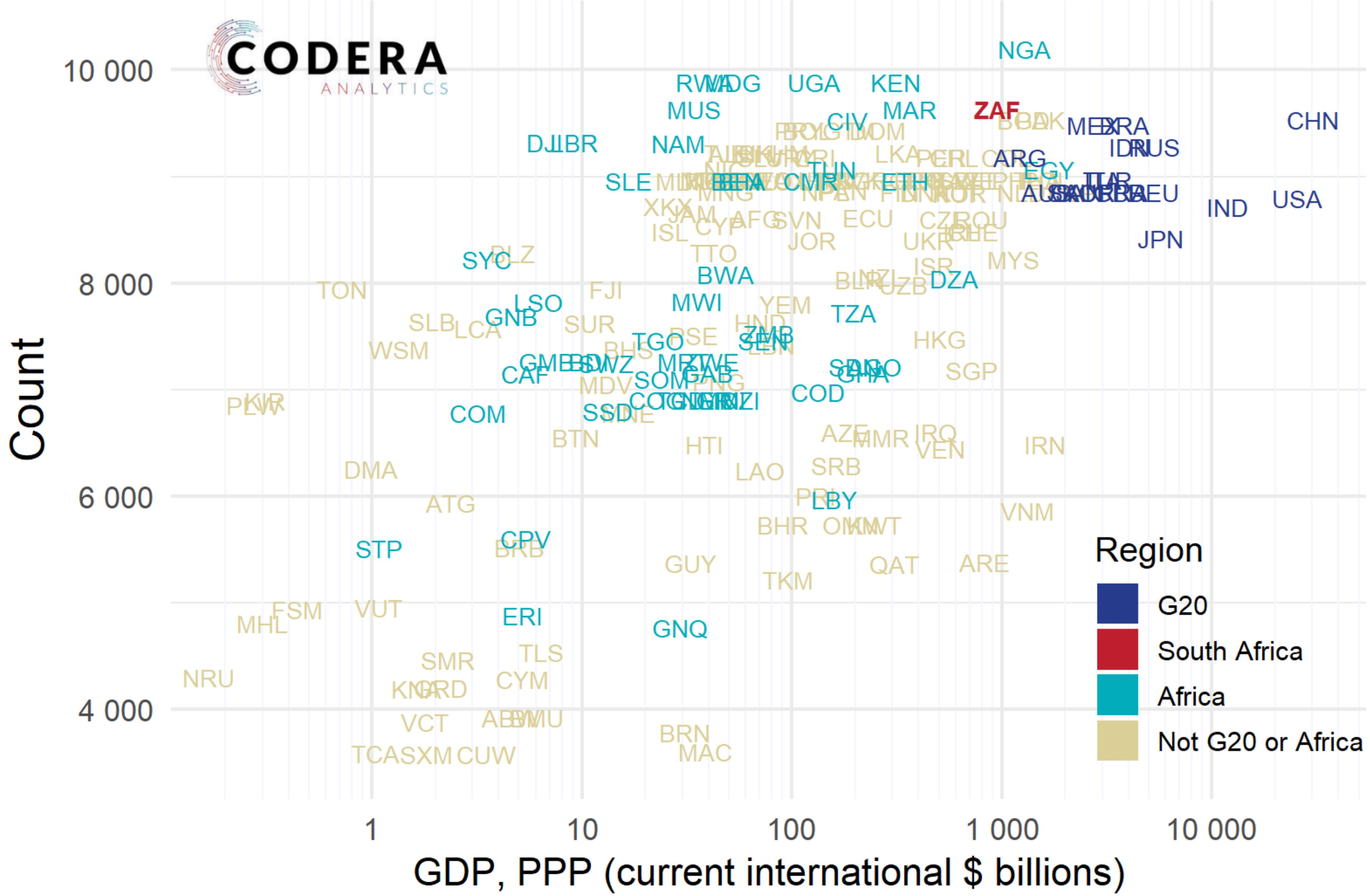
Particularly public finance data



Source: Global Data Barometer (2021).

Number of indicators available from the World Bank

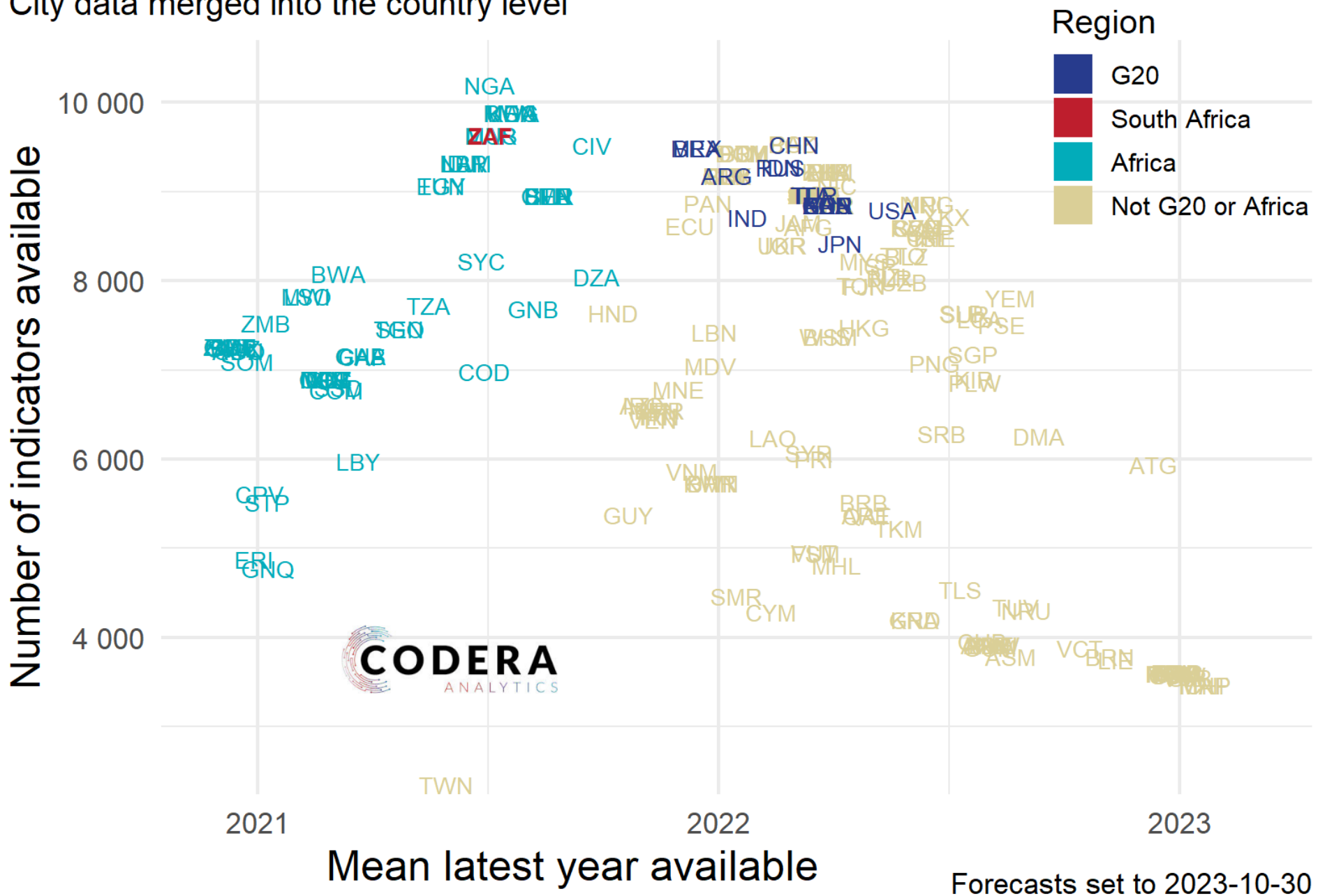
City data merged into the country level



We stand out on availability

Average latest year available in World Bank indicators, across countries

City data merged into the country level

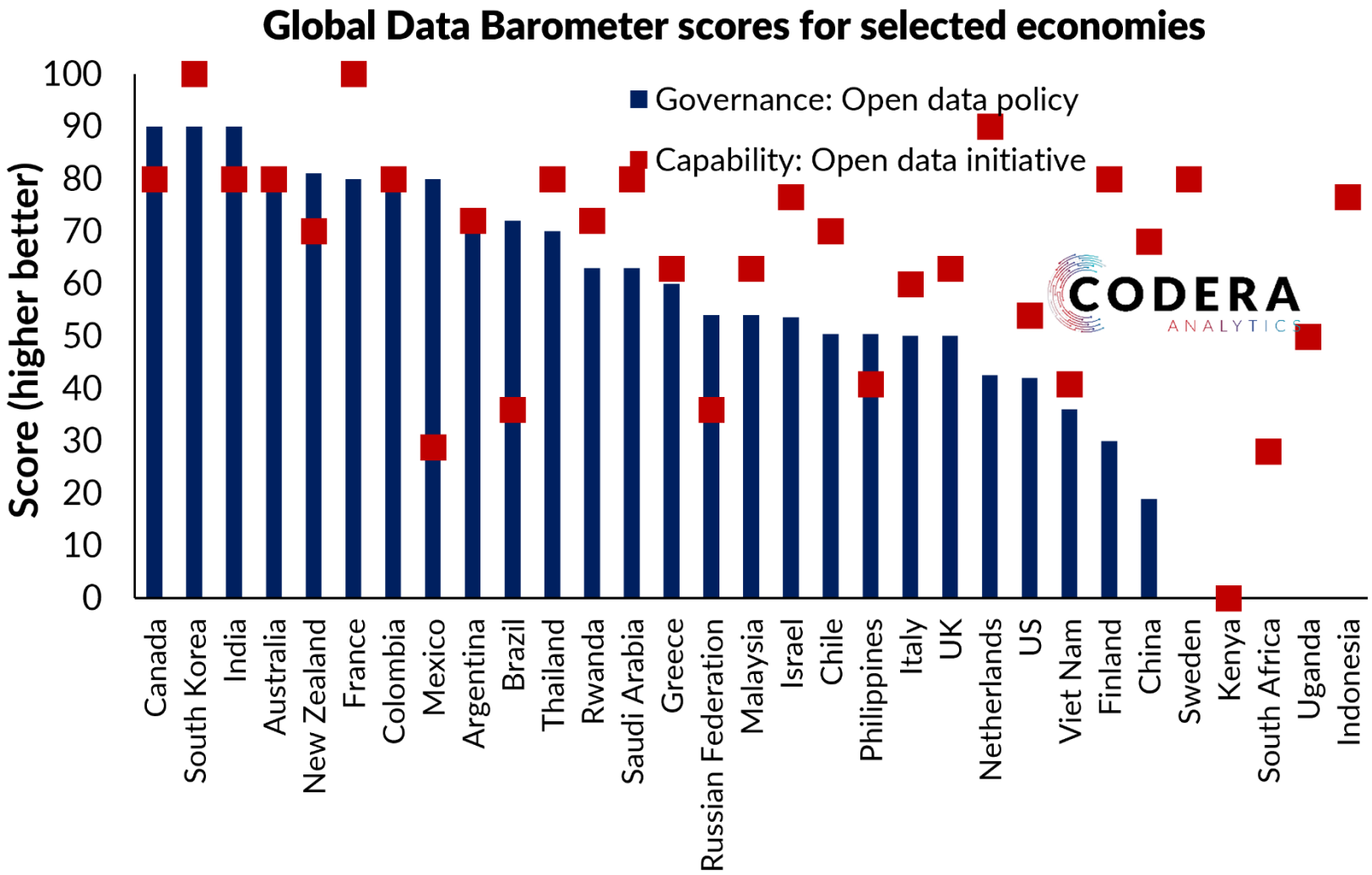


Forecasts set to 2023-10-30

Though less up-to-date

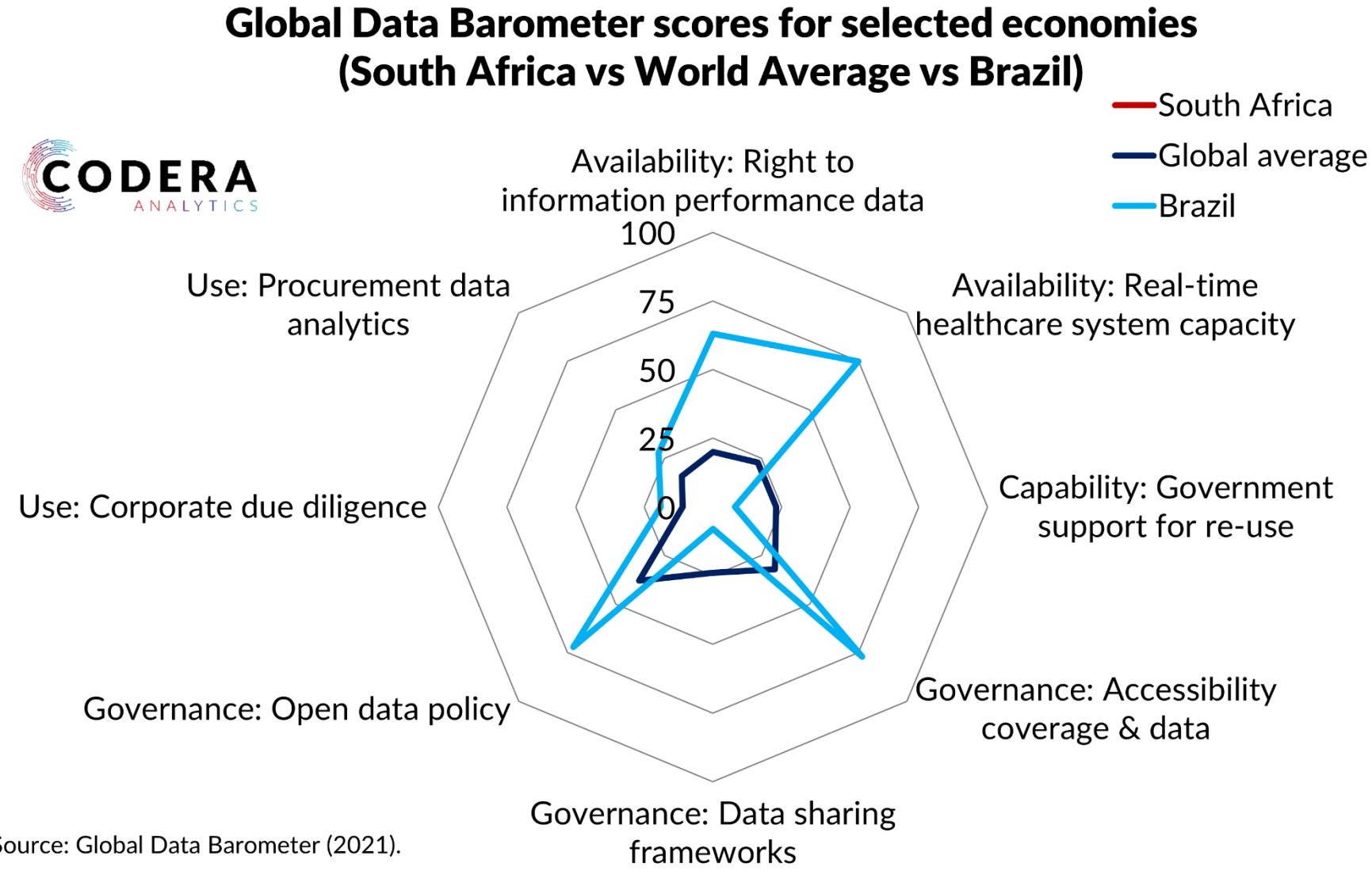
SA's data ecosystem: Comparative Weakness

Weak on open data



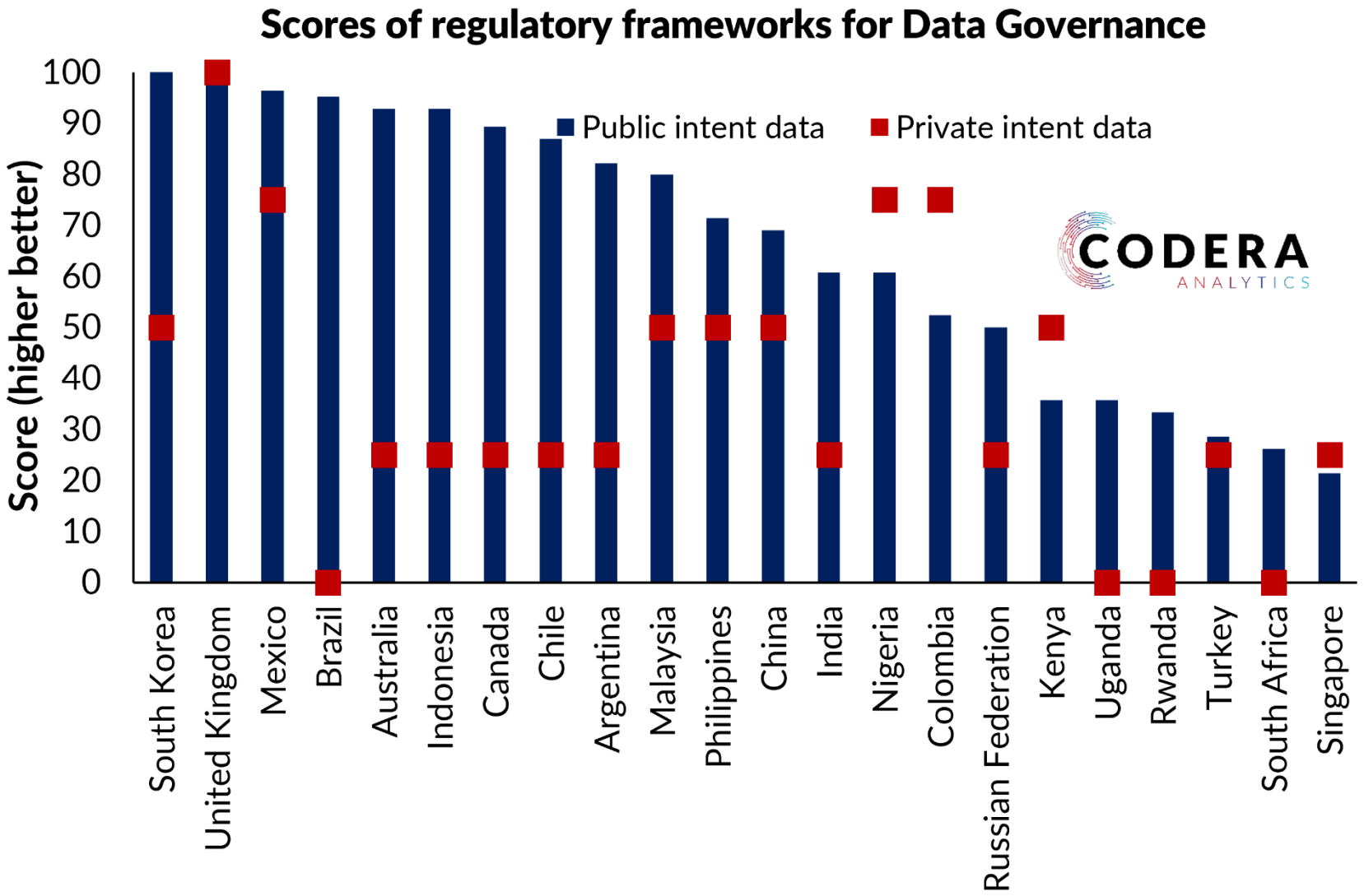
Source: Global Data Barometer (2021).

And frameworks and accessibility



Source: Global Data Barometer (2021).

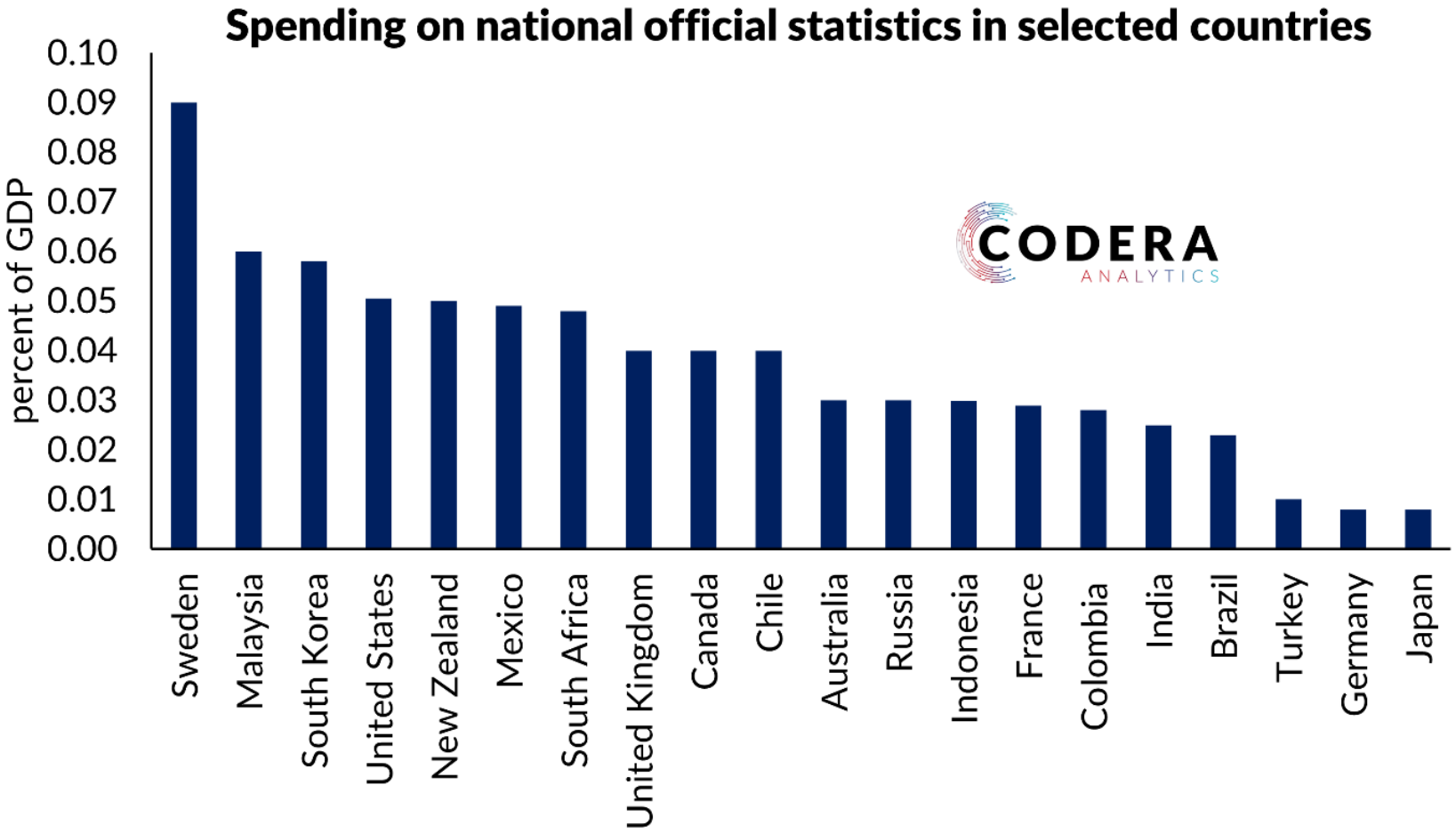
Especially on regulatory frameworks



Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

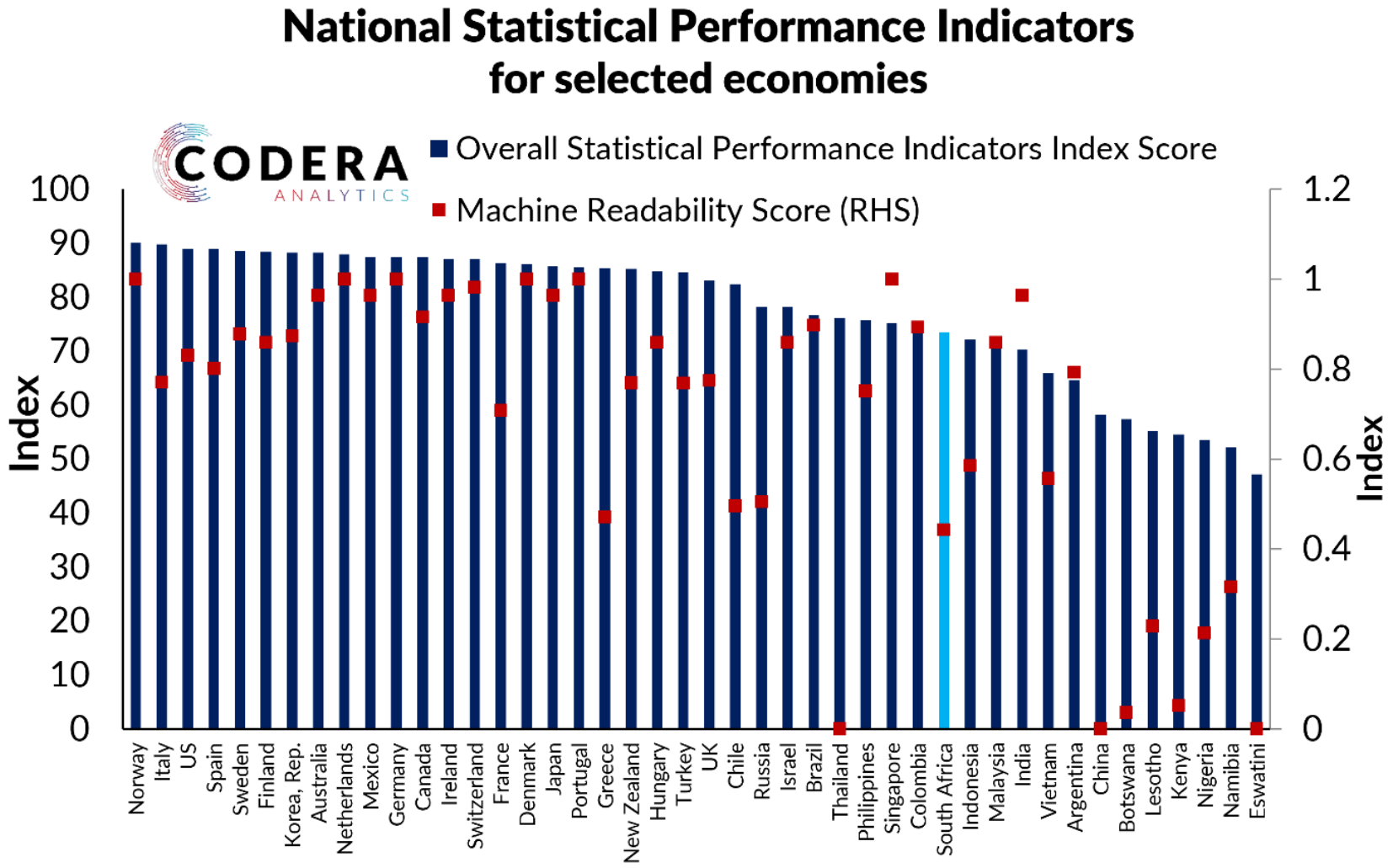
Why is SA falling behind?

It does not reflect a lack of prioritisation



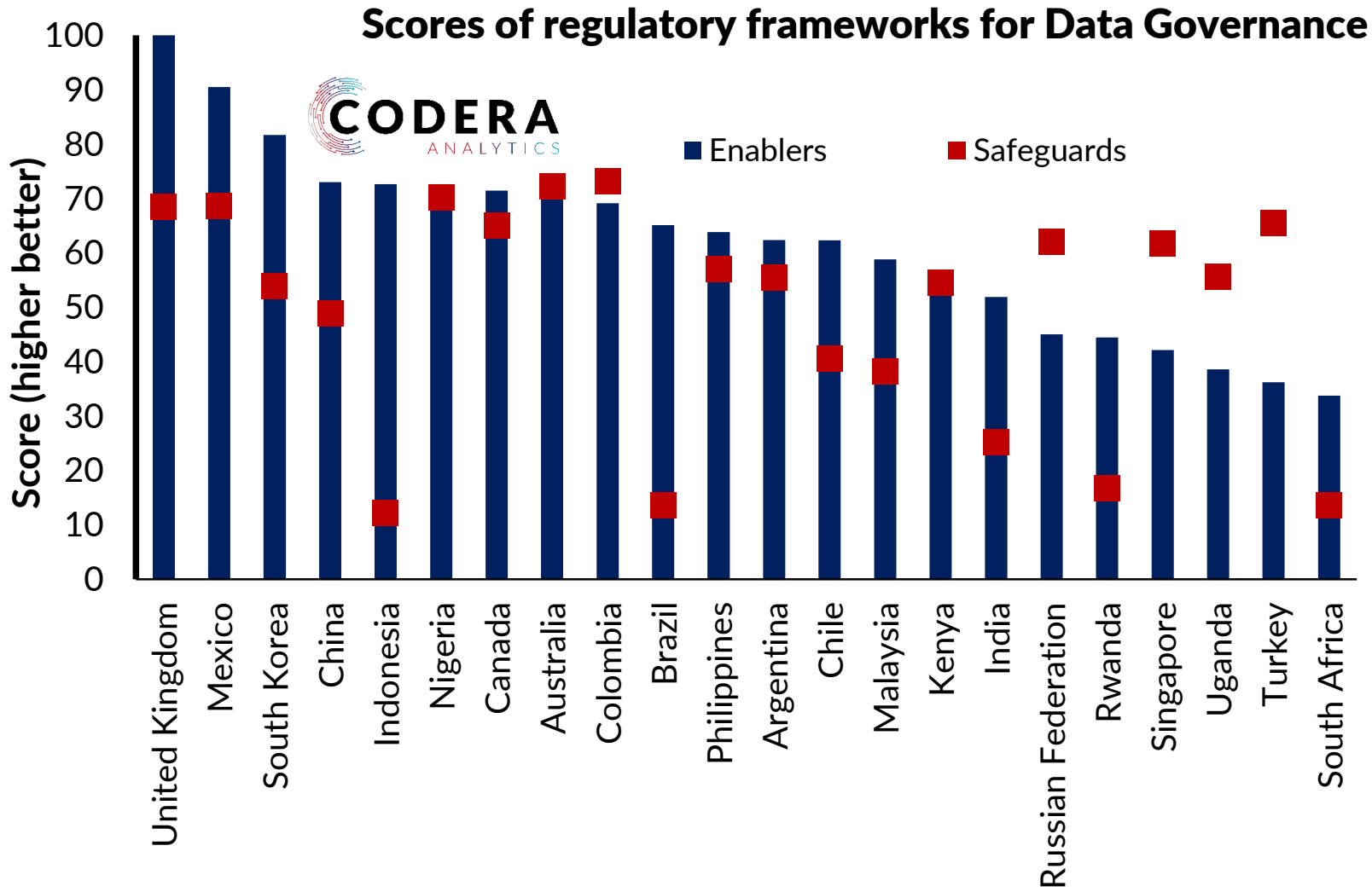
Source: National Statistical Agencies, World Bank, Codera Analytics. Latest data (generally 2021). Note that it is difficult to compare such figures across countries as some countries report data in fiscal year and others in calendar year and irregular census expenditures may be included in budgets. Some countries have multiple national statistical agencies so figures may underestimate total expenditure for some countries. For the US, the figures include Bureau of Economic Analysis, Bureau of Labour Statistics, the Census Bureau and National Center for Health Statistics.

… but are behind in terms of accessibility



Source: World Bank, estimates for 2019. The Statistical Performance score contains indicators along five dimensions: data use, data services, data products, data sources, and data infrastructure. Each dimension of the five pillars incorporates several indicators.

… we lag on ‘safeguards’ & ‘enablers’



Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

SAFEGUARD
SAFEGUARD
ENABLER
ENABLER

Safeguards

Template data use agreements
Data protection standards
Data security measures
Information regulator

Enablers

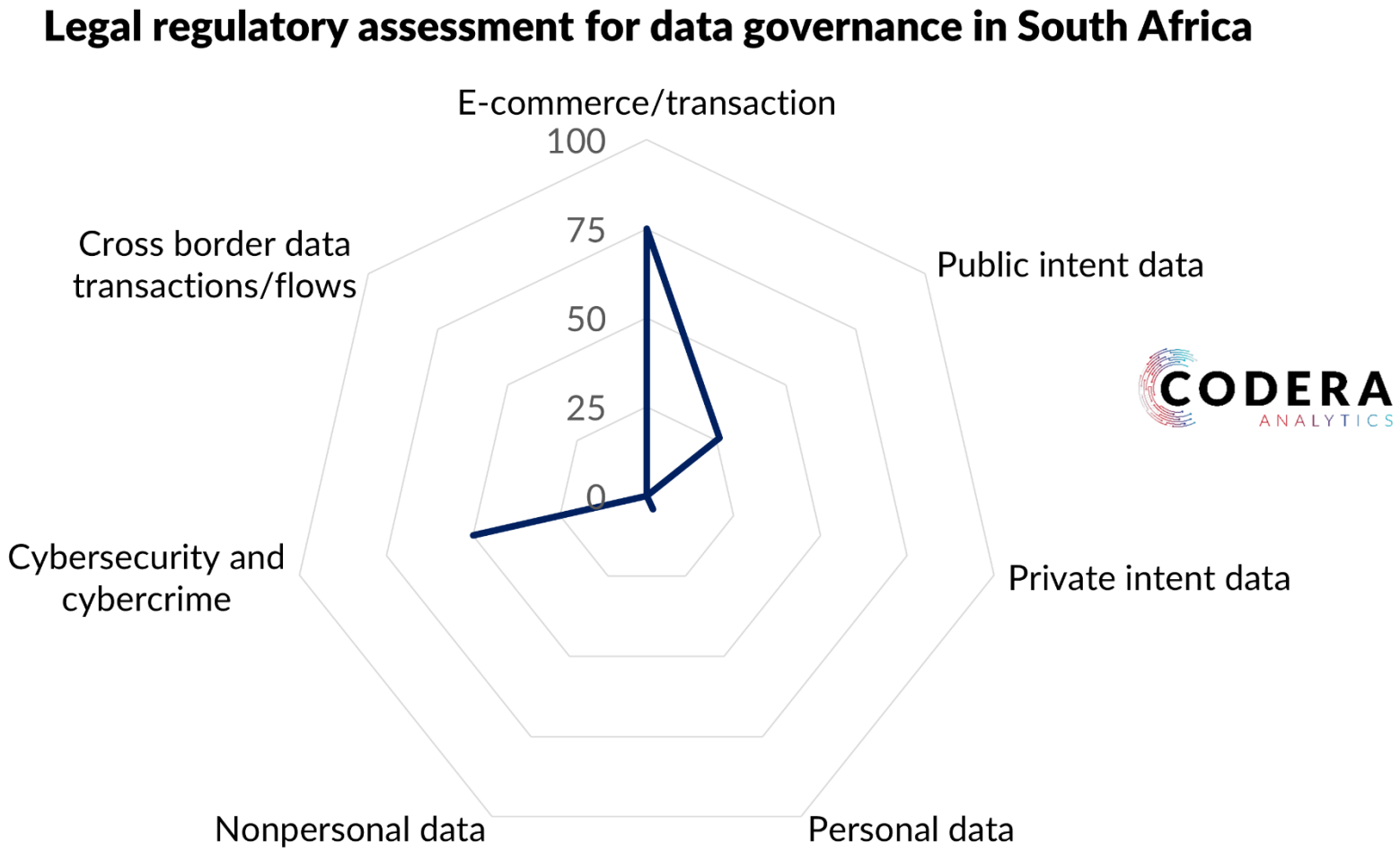
PUBLIC INTENT DATA

Open data laws | Open licensing regimes | Data classification policy | Interoperability | Data Portals | Metadata catalogues | Data exchange platforms | Data classification policy | Freedom of information

PRIVATE INTENT DATA

Data portability rights | Digital verification| Voluntary licensing access

We lack frameworks for data sharing & access



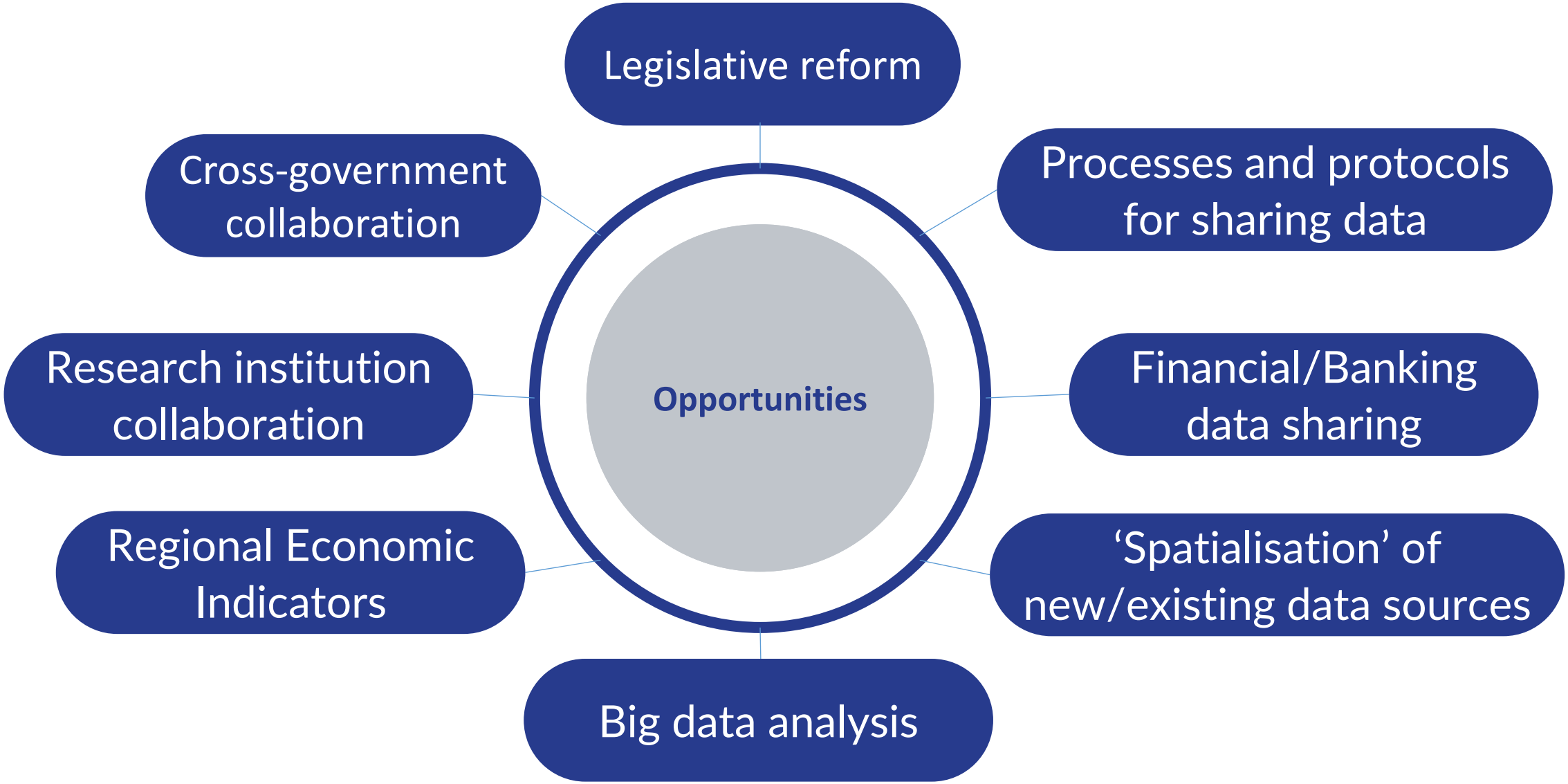
Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

What do we need to do?

Time to democratise data

Security	Governance	Literacy	Infrastructure	Culture
Standards for encryption, access controls, data masking Authorisation procedures	Governance policies & procedures Data quality standards Privacy policies Data management practices Data use agreements & templates Applications from researchers Standards & processes for data management, use & accessibility within and between organisations	Use of analysis tools Data visualisation Statistical concepts Training of public sector data stewards in standards	Enable self-service access Security Cost effective Integration of data & indicator creation Data Definitions Codebooks Interoperability Data portals Visualisation Self-service	Open Data Transparency Data access Sharing insights Collaboration

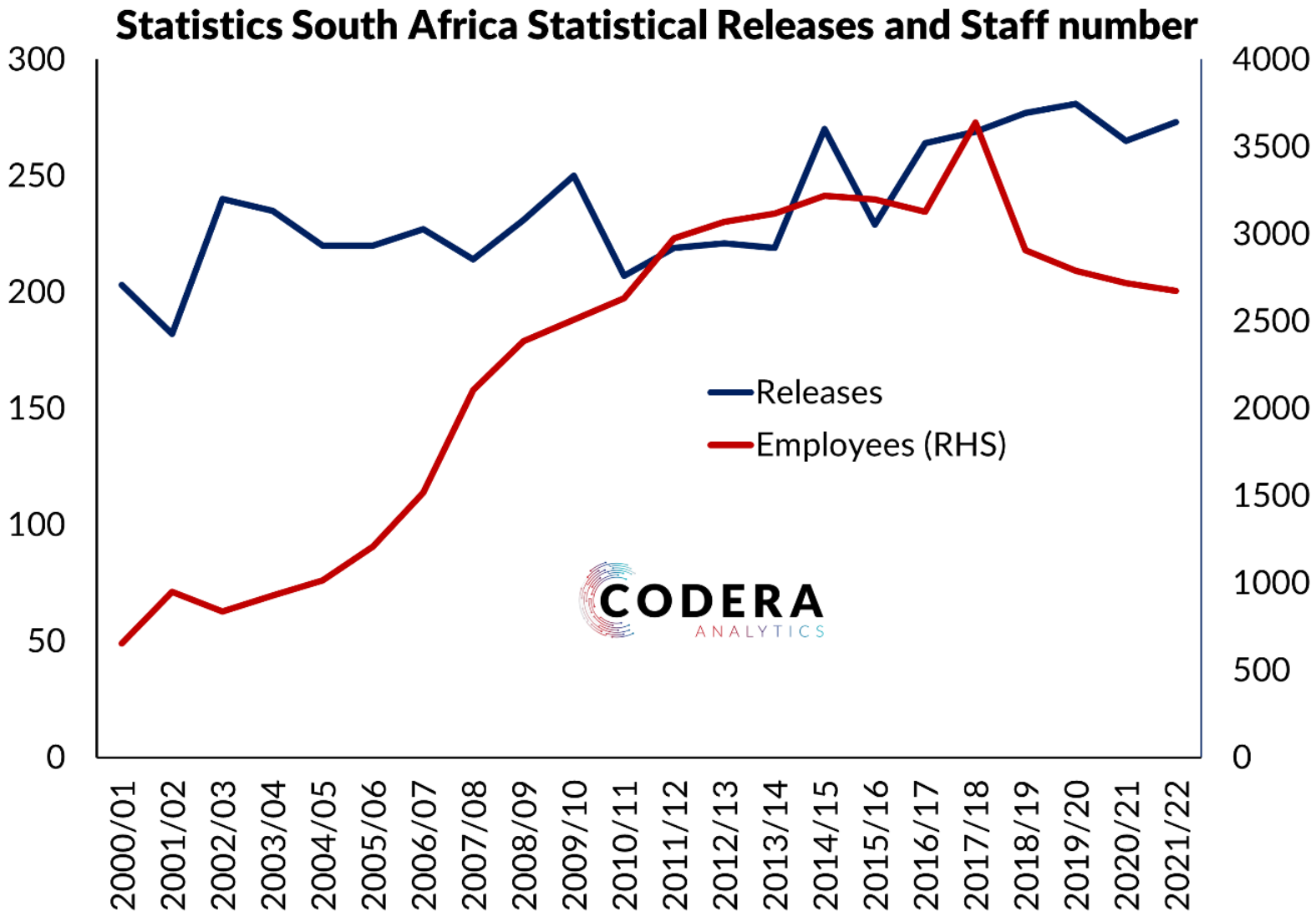
Opportunities



daan@codera.co.za

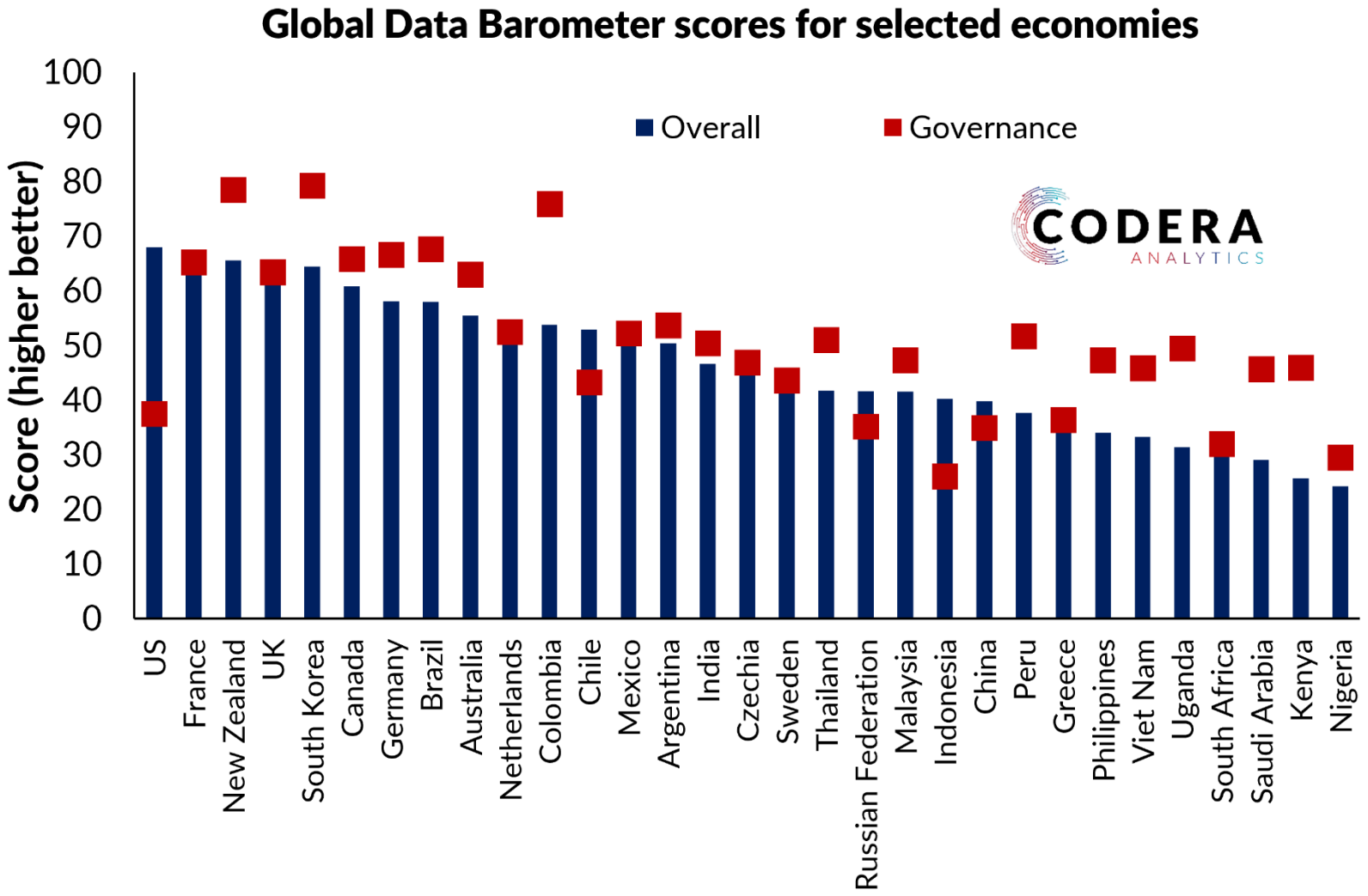
Extra Slides

Increase in official releases



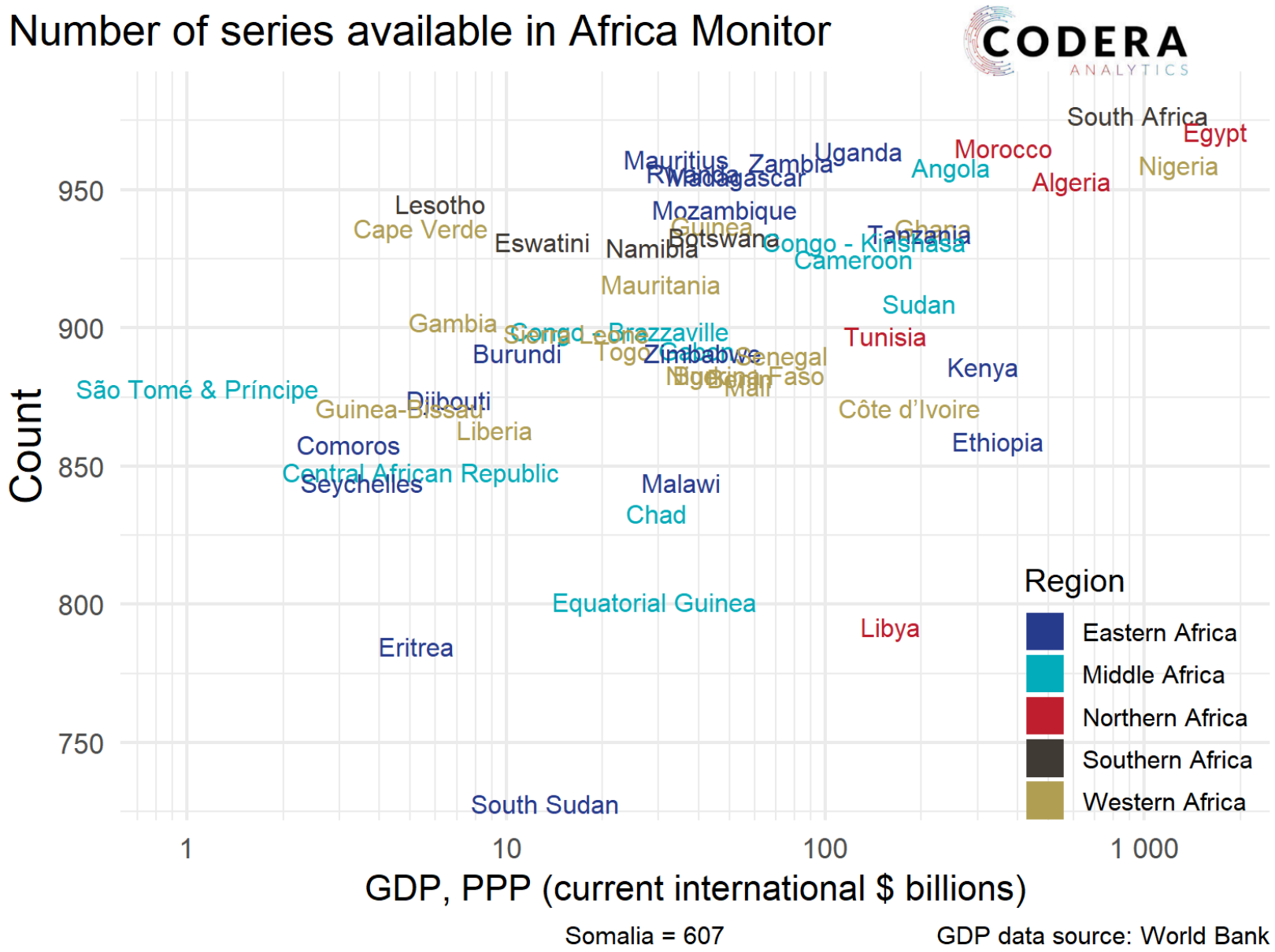
Source: Statistics South Africa Annual Reports.

Overview: SA's data ecosystem

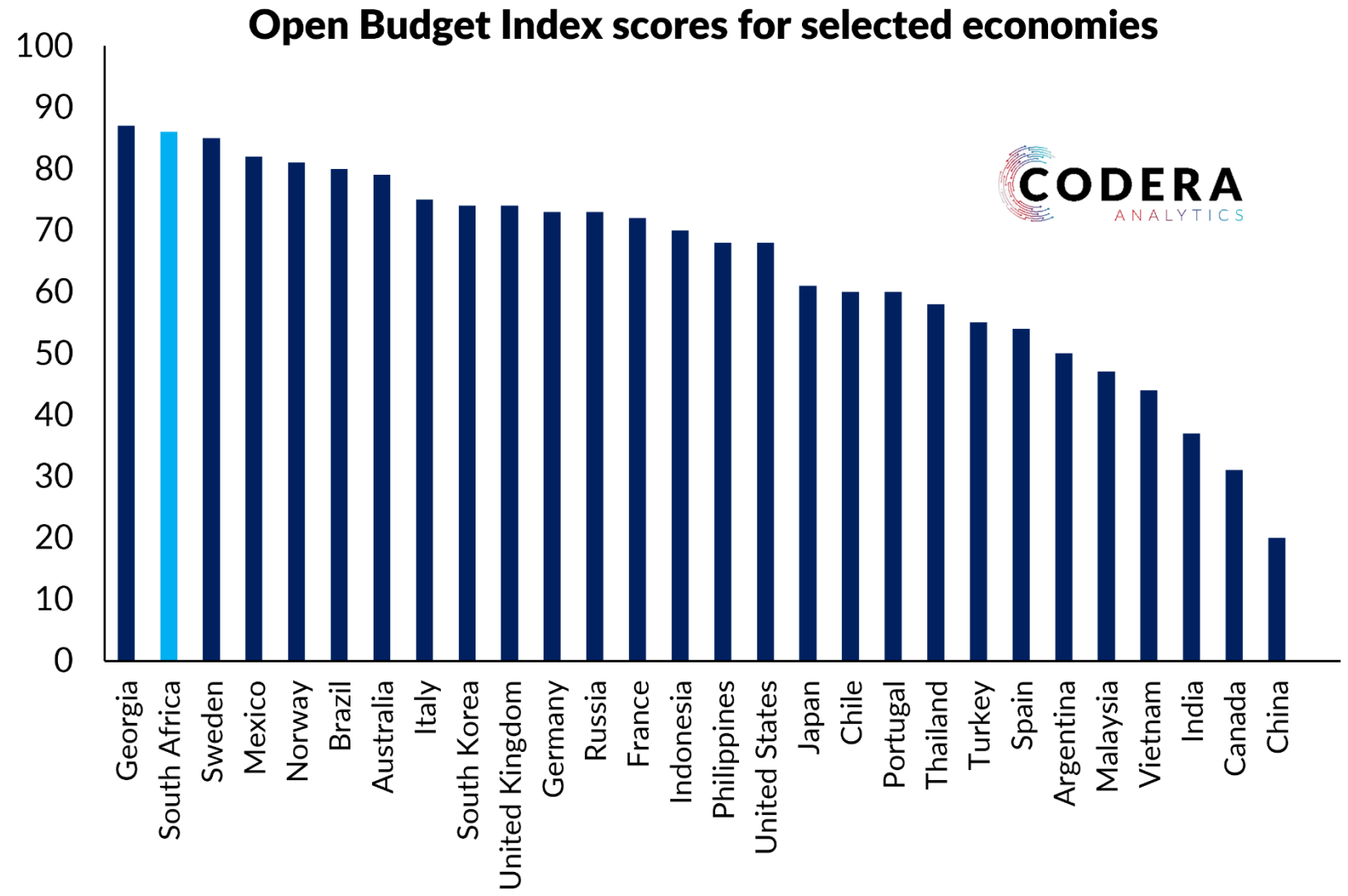


Source: Global Data Barometer (2021).

Indicator Availability

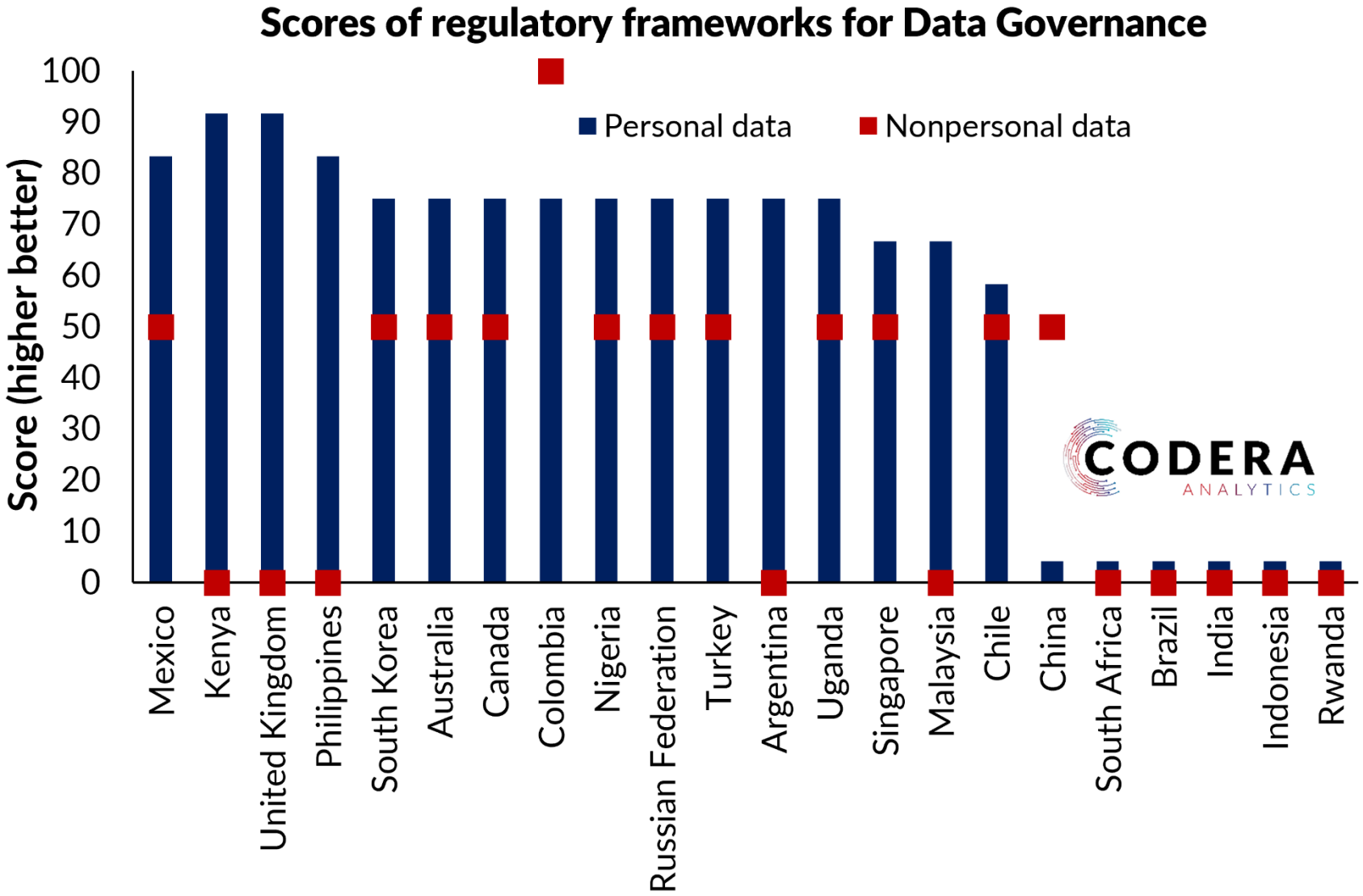


Budget Transparency



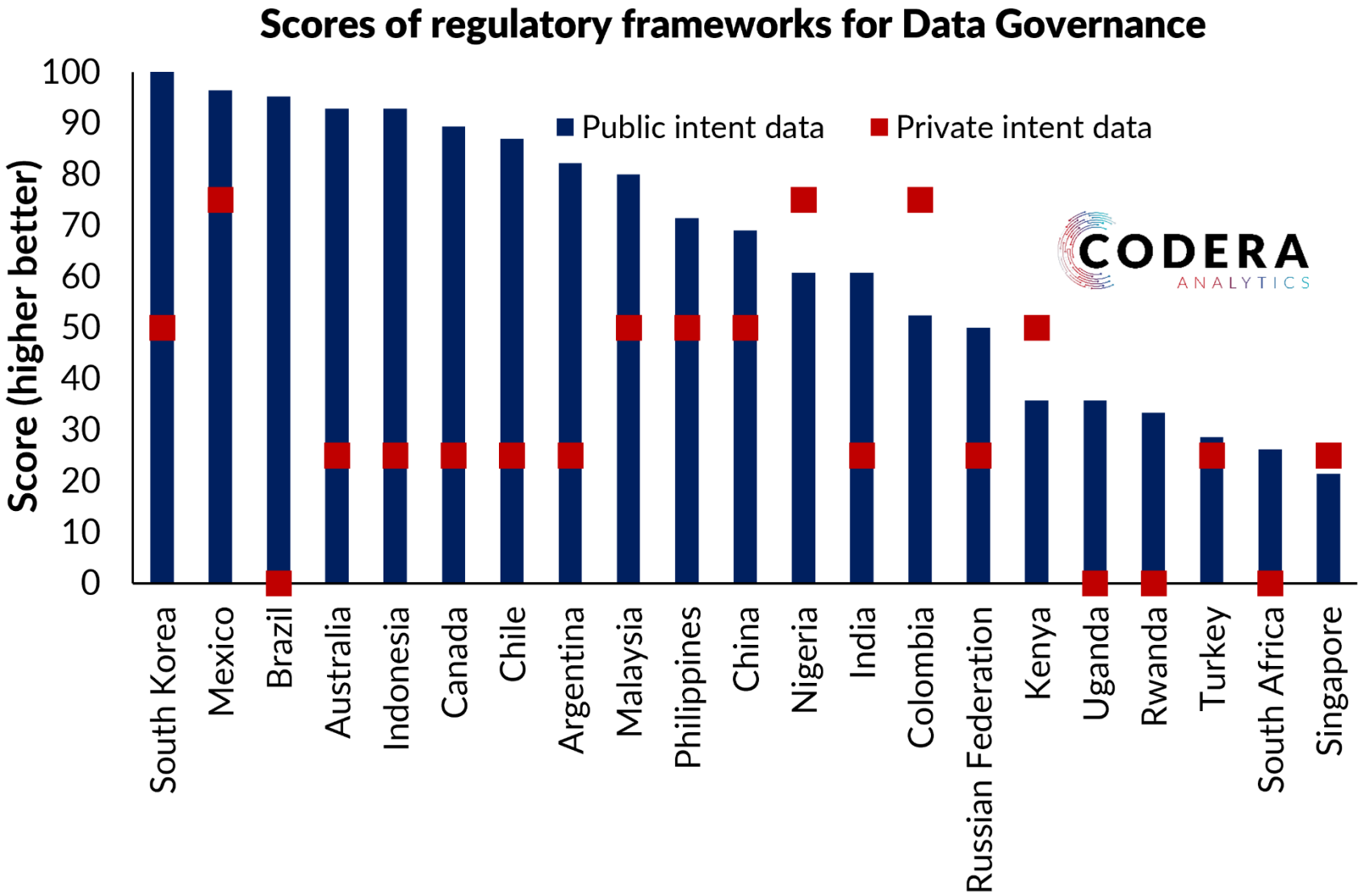
Source: Open Budget Survey 2021.

Best practice for data sharing & access



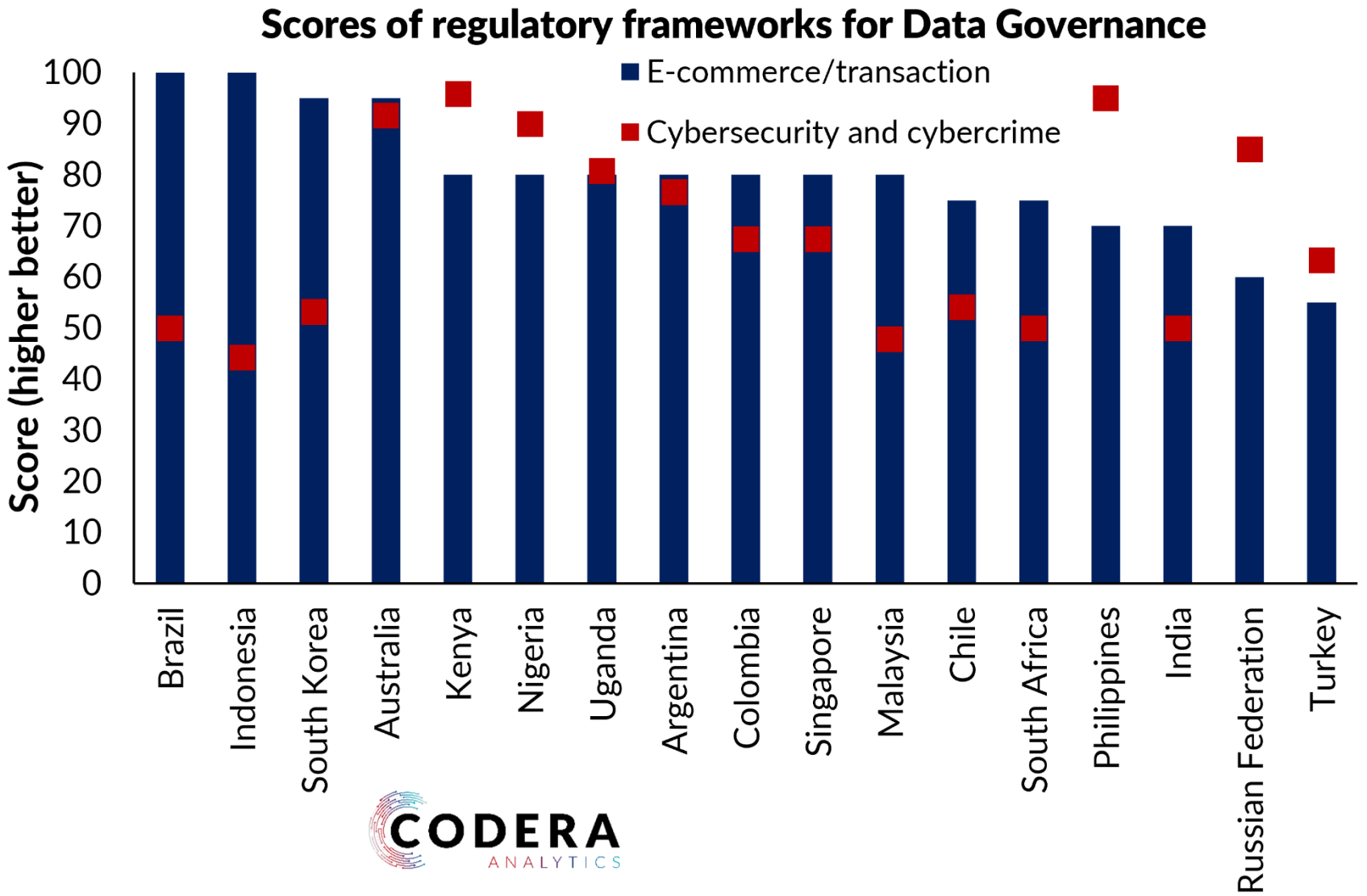
Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

Best practice for data sharing & access



Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

Best practice for data sharing & access



Source: Chen,Rong - DECIG, 2021. "Mapping Data Governance Legal Frameworks Around the World : Findings from the Global Data Regulation Diagnostic," Policy Research Working Paper Series 9615, The World Bank.

South African data gaps

Availability:

- 4/63 G20 Data Gaps Initiative targets/Intermediate targets not met
  - Dissemination of stocks of non-financial assets
  - Non-Bank Financial Intermediation
  - Currency breakdown of international investment position
  - Commercial Property Prices
- Spatial data (e.g. regional breakdowns of official statistics)
- Administrative data
- Non-traditional data (e.g. mobility, payment system)
- Regulatory compliance (e.g. bank-level Basel III)
- Input-output tables & KLEMS
- State Owned Enterprise data
- Municipality-level data
- Mobile Money – SA not included in IMF MFS database

South African data gaps

Quality & Accessibility

- No self-assessment / peer review with OECD Council on Good Statistical Practice
- Infrastructure (e.g. APIs)

Regulation

- Legislative & Policy
  - Procedures for construction and use of ‘Big Data’ and private data

Enhancing data access and sharing

- OECD recommends:
  - Whole-of-government approach to data access and sharing
  - Templates for data access and sharing agreements
  - Open data arrangements & APIs

- **OECD Data Sharing Recommendations**
  - Enhancing Access to and Sharing of Data
  - Access to Research Data from Public Funding
  - Digital Government Strategies
  - Open Government
  - Protection of Privacy and Transborder Flows of Personal Data

Data Democratisation

- Data Catalogues for discovery and meta-data visibility
- Build trust through sharing data & ensuring consistency
- Self service access & analytics via variety of tools
- Evolution of data models and tools
- Data-driven decisions & crowd-sourcing innovation
- Data as a service

Noteworthy initiatives

National Treasury Secure Data Facility

Spatial Economic Activity Data: South Africa

- NT/HSRC/Cities Support Programme

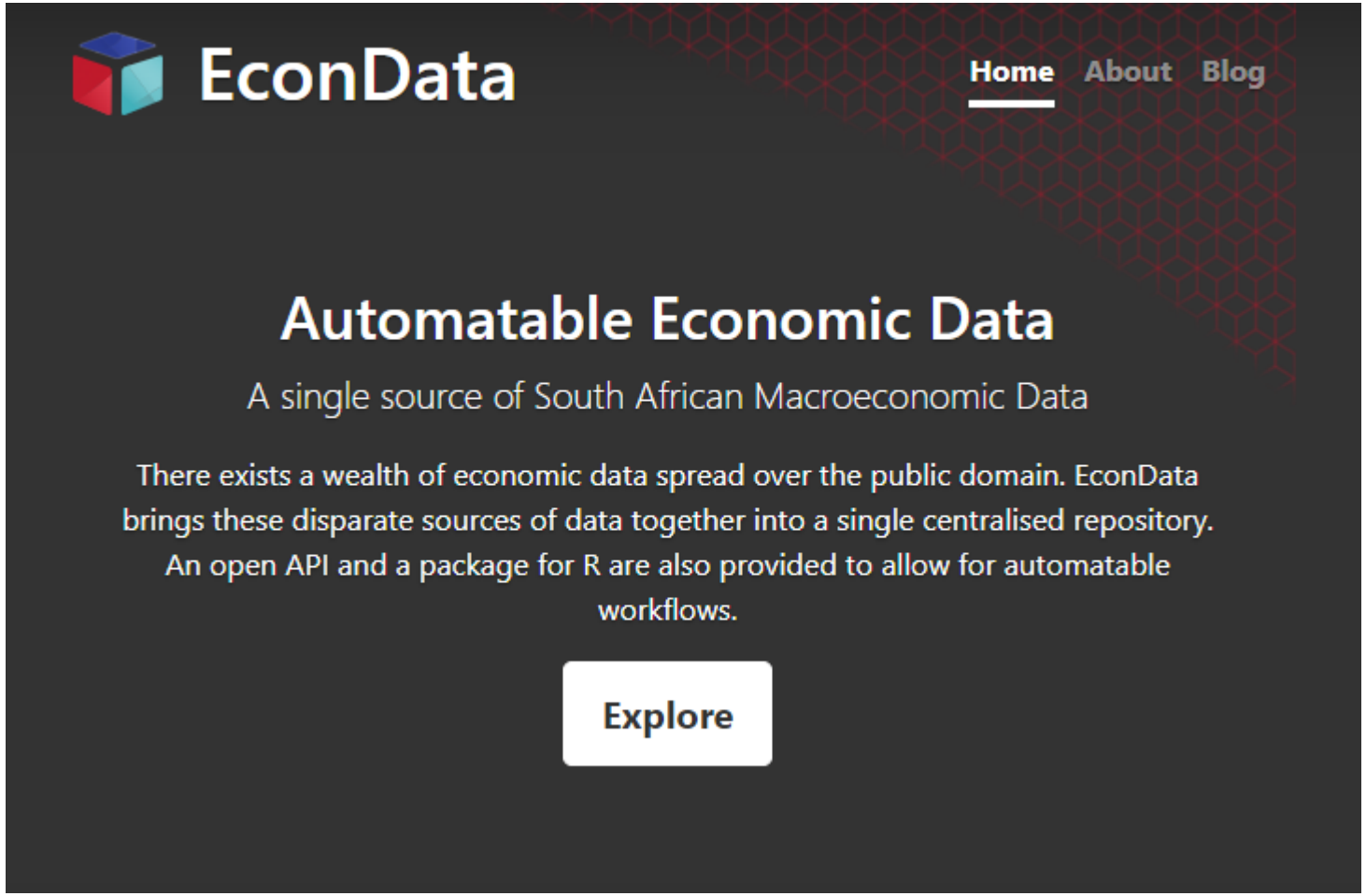
G20 New Data Gaps Initiative

Eurostat Data Sharing mechanisms

OpenData Initiative

Open Data South Africa Toolkit

Our EconData econometric data service



- Data Democratization: access data using variety of tools (excel, API, website)
- SDMX-based with best practice approach to metadata, data governance, data catalogue
- Access to a growing collection of data sources (over 80k series)
- Enables users to auto-update their charts and models of SA data
- For real-time forecast comparisons and model evaluation, EconData provides historical vintages of all curated data
- To create a free research user account, sign up at www.econdata.co.za

South Africa Macroeconomic Database and Nowcast

- Over 10k time series for South Africa collected from the SARB Quarterly Bulletin, EconData and STATSSA, with weekly updates (every Thursday)
- SA Nowcast: An automated weekly nowcast of GDP and Unemployment (updated every Friday) based on a mixed-frequency dynamic factor model and 57 core macroeconomic time series that receive real-time updates through EconData
- https://codera.co.za/research/

Opportunities

Legislative reform	Legislative constraints on release of unit-level enterprise data as part of current amendments to Statistics Act, National data, cloud policy
Cross-government collaboration	Administrative datasets for research and policy making, Data gaps stocktake, Public sector data science Labs (or scaling up NT Data Lab), Shared IT infrastructure, data quality standards, National Data Catalogue and Public list of Data Owners
Research institution collaboration	Mechanisms for data sharing and dataset development
Processes and protocols for sharing data	Centralised coordinating body to develop of guidelines and templates for data access and sharing agreements, promoting data sharing partnerships, and standardised protocols for legal, IT, governance, including definitions of confidential data & data reuse modalities, Universities or key stakeholders to act as honest data brokers
Regional Economic Indicators	Public-private collaboration to use third party data to create macroeconomic and regional indicators (e.g. SOE, public institution and member organisation data such as SANRAL, Transnet, SARS, CIPC, UIF, ENatis, BankServ, Metros, Research institutions, SASSA)
'Spatialisation' of new/existing data sources	UIF, CIPC, StatsSA data
Financial/Banking data sharing	Standardised data standards and dictionaries for regulators, institutions and researchers
Big data analysis	Access to SOE and department data, regulated entities, listed companies, financial market, web scraping, payments system, Financial Regulatory Data (EconData)

Incorporating recommendations from City Support Programme Report (2021)