

# Automating workflows using EconData

Johannesburg R User Group 14 March 2023



[www.codera.co.za](http://www.codera.co.za)



# Data challenges for data professionals/analysts



Data  
Availability

Data cleaning  
and  
wrangling

Data  
comparability

Automation

Monitoring  
data value

Data  
Analytics

Data risk  
management

Data  
Governance

# Our EconData platform



The screenshot shows the EconData website homepage. At the top left is the EconData logo, which consists of a 3D cube with red, blue, and green faces. To the right of the logo is the text 'EconData'. In the top right corner, there are three navigation links: 'Home' (which is underlined), 'About', and 'Blog'. The main heading is 'Automatable Economic Data' in a large, bold, white font. Below this is a subtitle 'A single source of South African Macroeconomic Data' in a smaller, white font. The body text explains that there is a wealth of economic data in the public domain and that EconData centralizes it. It also mentions an open API and an R package. At the bottom center is a white button with the text 'Explore'.

**EconData**

Home About Blog

## Automatable Economic Data

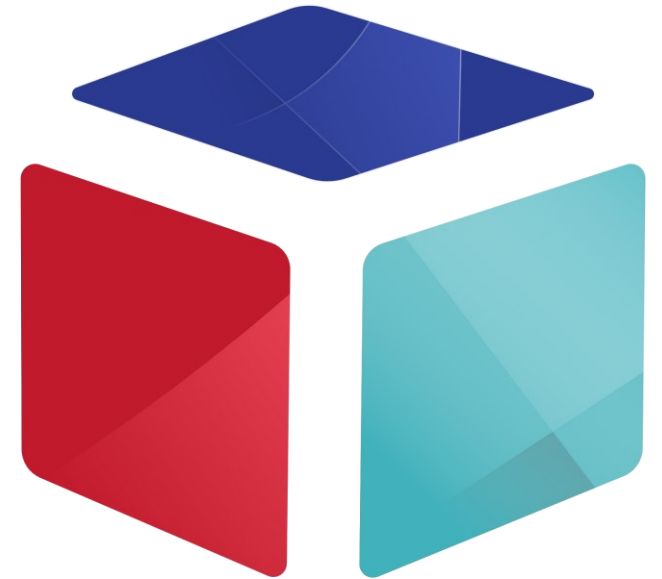
A single source of South African Macroeconomic Data

There exists a wealth of economic data spread over the public domain. EconData brings these disparate sources of data together into a single centralised repository. An open API and a package for R are also provided to allow for automatable workflows.

**Explore**

- Provides access to a growing collection of South Africa data sources (currently over 80k series)
- Enables users to auto-update their charts and models of South African data
- For real-time forecast comparisons and model evaluation, EconData provides historical vintages of all curated data
- To create a user account, sign up at [www.econdata.co.za](http://www.econdata.co.za)

- Data Democratization: access data using variety of tools
- Definitional consistency of economic concepts
- Central data glossary of concepts and meta-data
- Enabling data automation

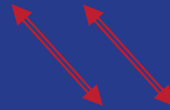


# Essential Concepts



## Definitional consistency

Definition  
Sources  
Ownership  
Quality



## SDMX

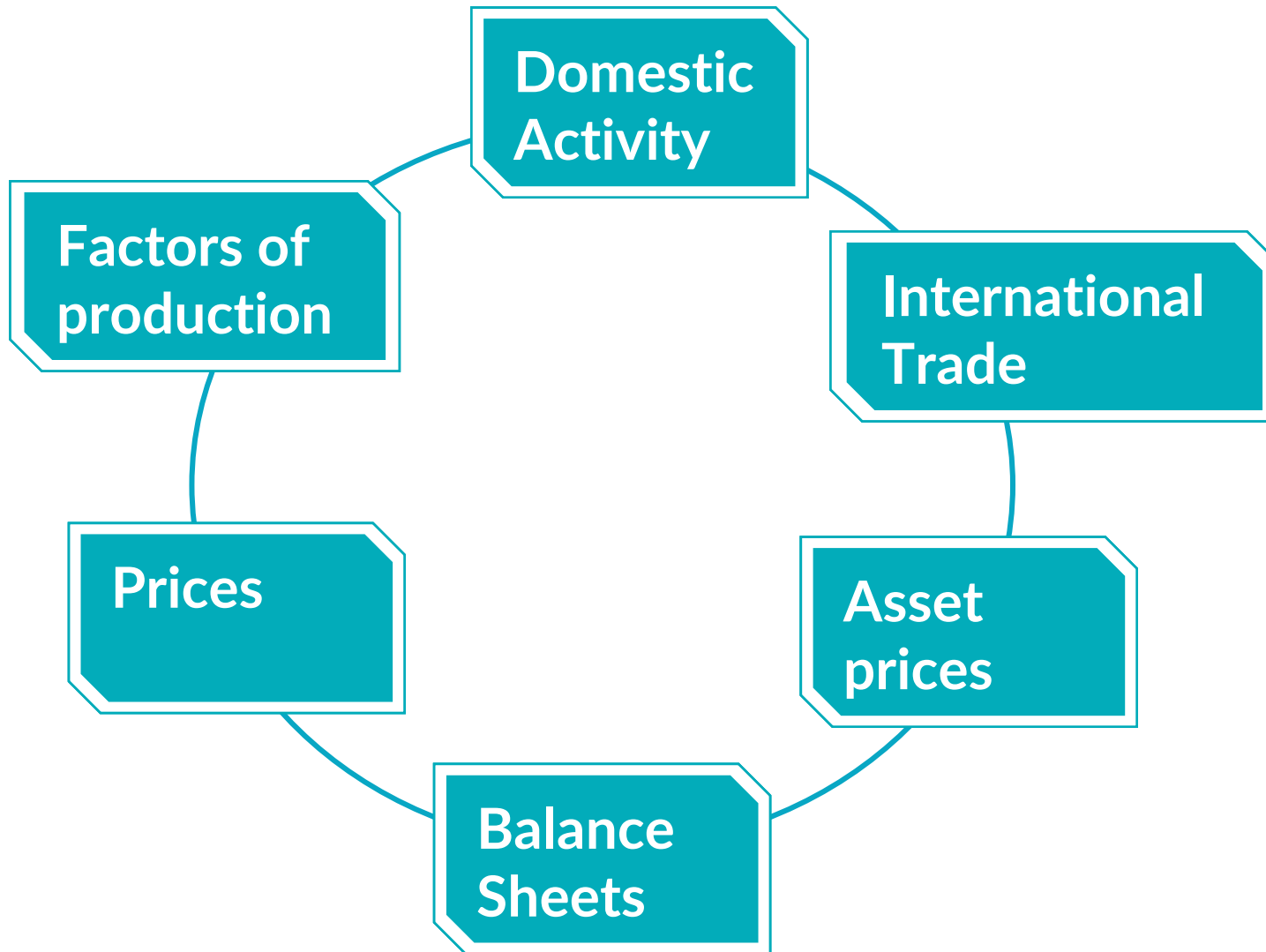
Standard data format  
Metadata-driven governance  
Process codification



## Data Democratization

Data Catalogue  
Meta-data  
Sharing Data  
Tools

# Definitional consistency of economic concepts



- Common definitions
- Clear relationships
- Authoritative sources
- Frequencies
- Transformations
- Ownership
- Security classification
- Quality assessment
- Vintages

# Data Democratization



- Data Catalogue for discovery and meta-data visibility
- Build trust through sharing data & ensuring consistency
- Self service access & analytics via variety of tools
- Evolution of data model and tools

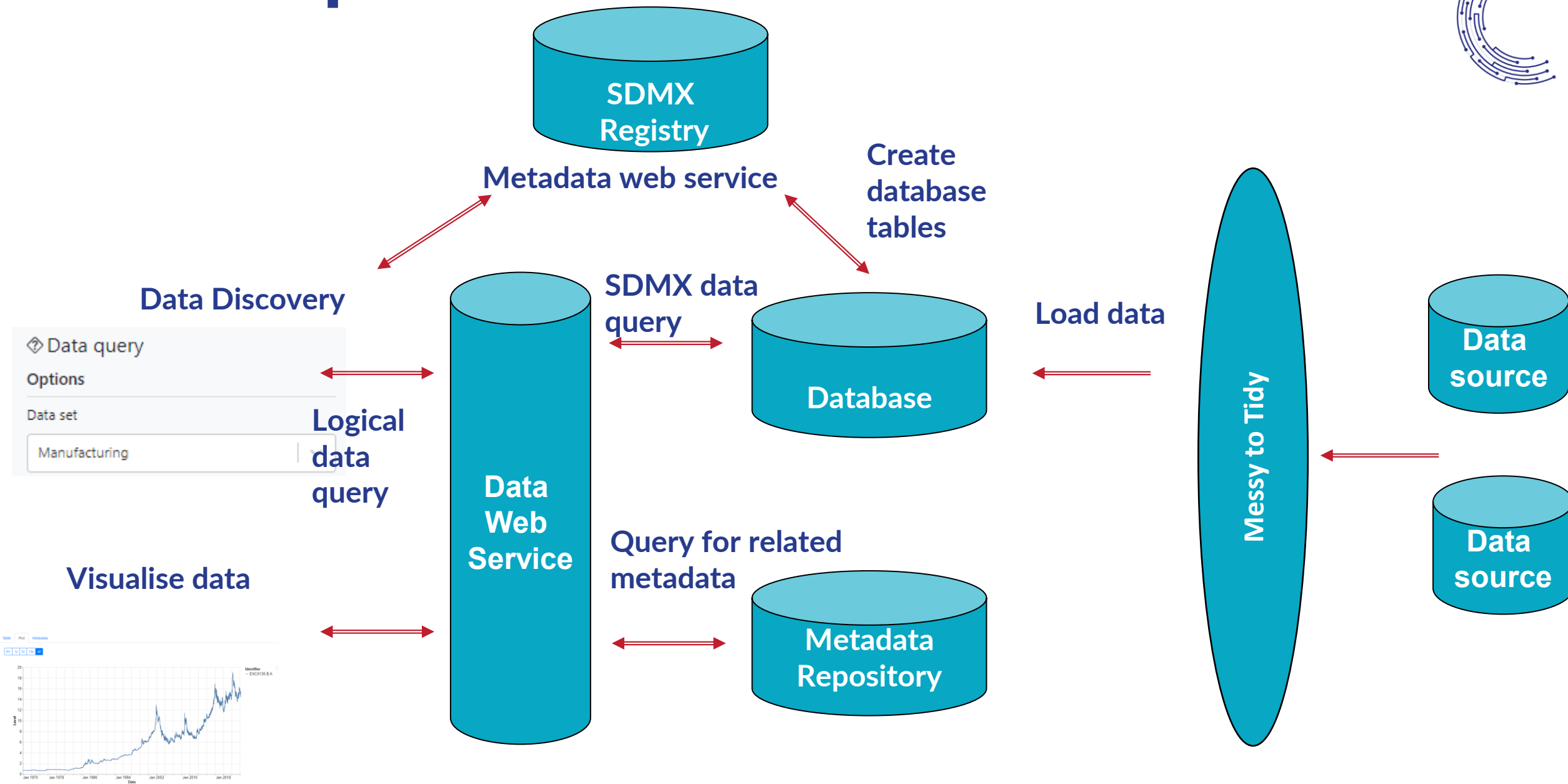
# Why SDMX?



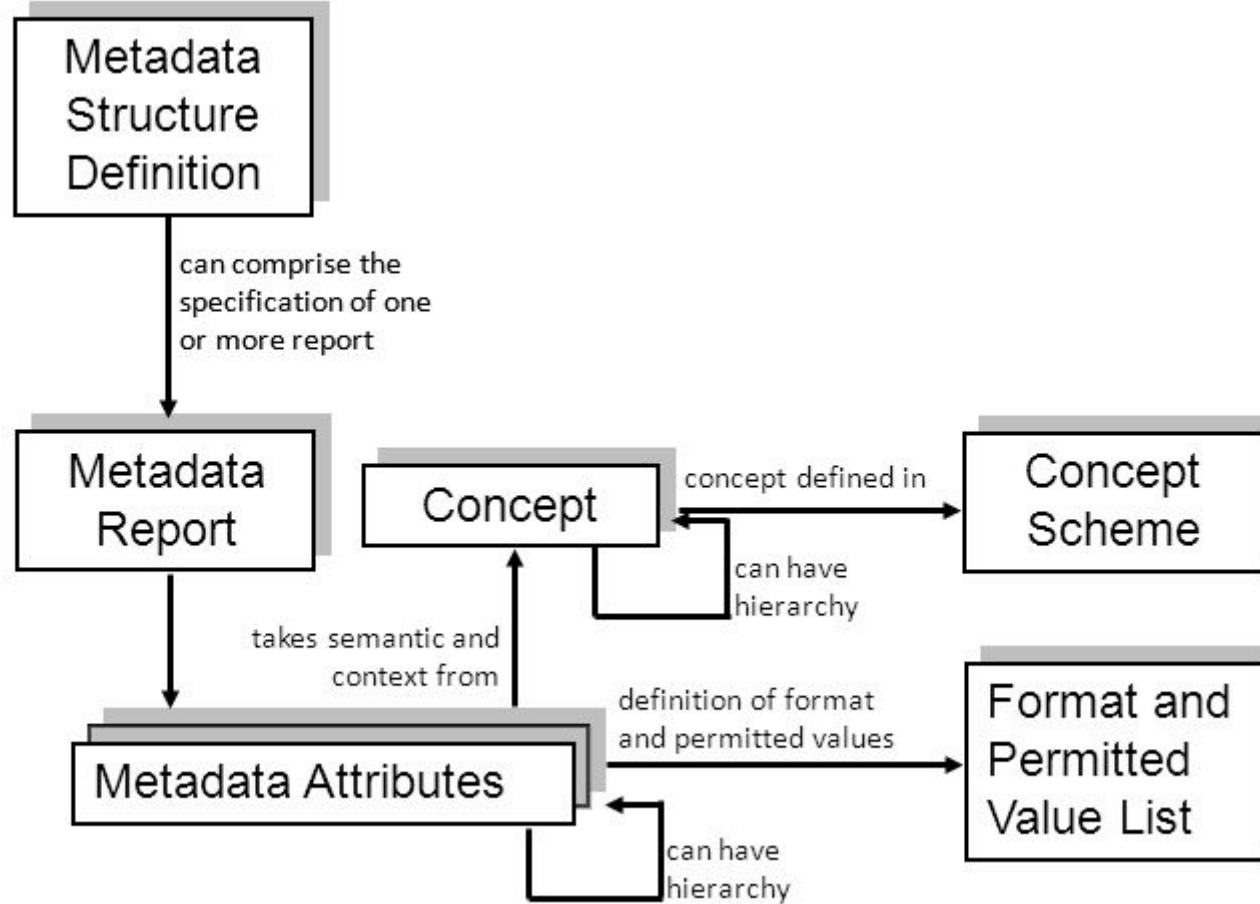
- Statistical Data and Metadata Exchange (SDMX) format is an ISO standard widely adopted (Eurostat, ECB, BIS, IMF, OECD etc.)
- Single, **metadata-rich** reporting format for each type of data
- **Integrating** data from variety of sources
- Supports **data-sharing** across databases and within institutions
- Simplified reporting **taxonomies** for data management and governance
- Document steps of data management **process flows**



# SDMX in practice



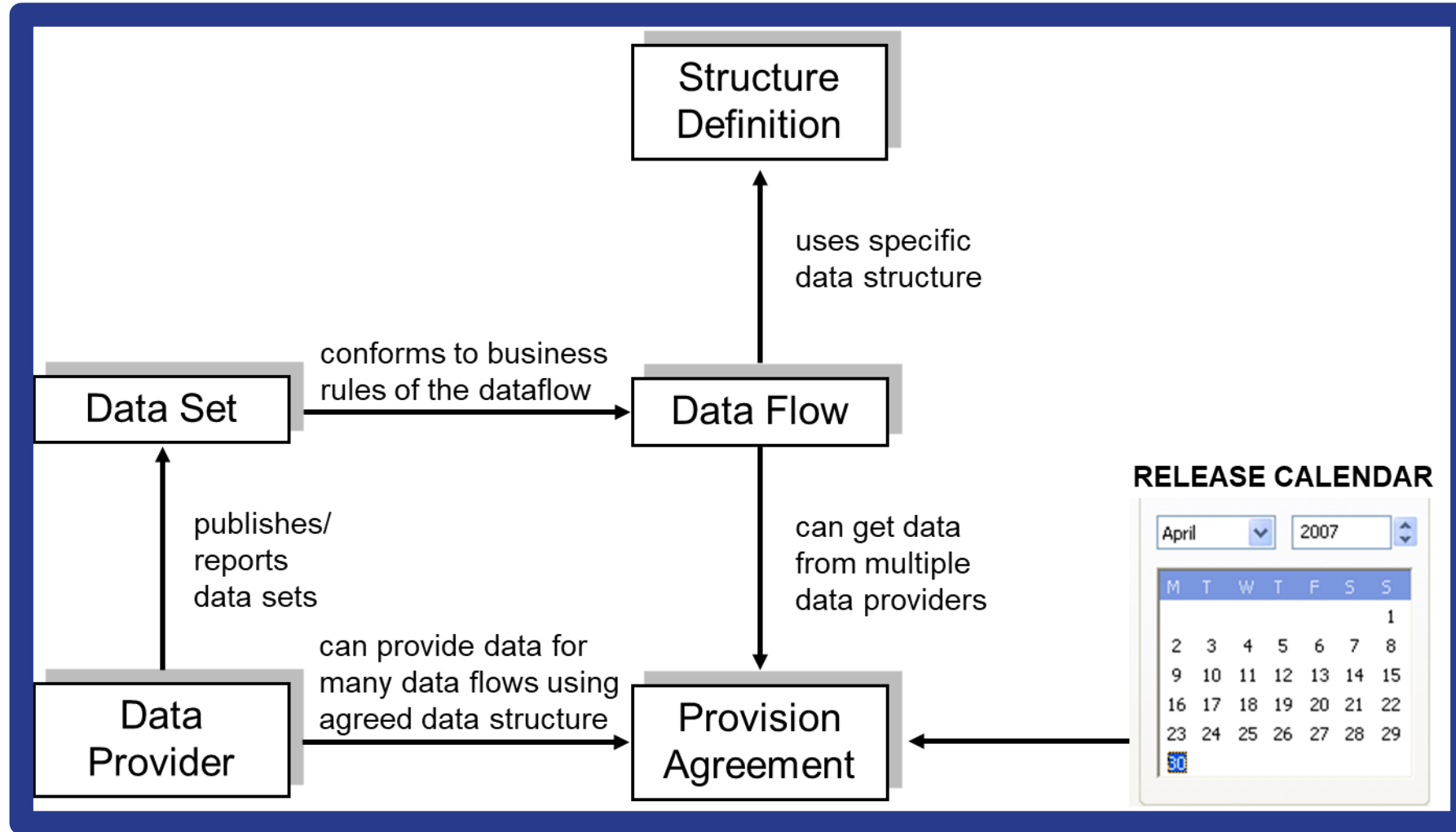
# SDMX metadata



Source: Arofan Gregory, Open Data Foundation

- Release calendar
- Provider
- Metadata from provider (i.e. Structure)
- Creation of reporting Taxonomy

# Data flows for controlling reporting & publishing



Source: Arofan Gregory, Open Data Foundation

# EconData structures messy official statistics



<b>Statistics South Africa</b> <b>Annual Financial Statistics survey</b> <b>2020<sup>1</sup> (AFS) preliminary estimates</b>							
SIC 12	<b>Forestry, logging and related services</b>						
	Carrying value at the beginning of the financial year	Additions (capital expenditure)	Disposal of assets	Revaluation and other adjustments to carrying value	Transfers-in and/or transfers-out	Depreciation and amortisation	Carrying value at the end of the financial year
	(A)	(B)	(C)	(D)	(E)	(F)	(G) <sup>2</sup>
	R million						
<b>Tangible assets</b>							
Land	1,745	48	134	243	0	5	1,897
Residential buildings	26	0	0	0	0	0	26
<b>Total</b>	<b>11,793</b>	<b>1,971</b>	<b>321</b>	<b>254</b>	<b>3</b>	<b>1,231</b>	<b>12,469</b>
SIC 13	<b>Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing</b>						
	Carrying value at the beginning of the financial year	Additions (capital expenditure)	Disposal of assets	Revaluation and other adjustments to carrying value	Transfers-in and/or transfers-out	Depreciation and amortisation	Carrying value at the end of the financial year
	(A)	(B)	(C)	(D)	(E)	(F)	(G) <sup>2</sup>
	R million						
<b>Tangible assets</b>							
Land	35	5	0	0	0	0	40
Residential buildings	63	2	0	0	0	0	65
<b>Total</b>	<b>4,491</b>	<b>567</b>	<b>104</b>	<b>93</b>	<b>0</b>	<b>460</b>	<b>4,587</b>
Total = SIC 1	<b>Forestry and fishing</b>						
	Carrying value at the beginning of the financial year	Additions (capital expenditure)	Disposal of assets	Revaluation and other adjustments to carrying value	Transfers-in and/or transfers-out	Depreciation and amortisation	Carrying value at the end of the financial year
	(A)	(B)	(C)	(D)	(E)	(F)	(G) <sup>2</sup>
	R million						
<b>Tangible assets</b>							
Land	1,780	53	134	243	0	5	1,937
Residential buildings	89	2	0	0	0	0	91
<b>Total</b>	<b>16,284</b>	<b>2,538</b>	<b>425</b>	<b>347</b>	<b>3</b>	<b>1,691</b>	<b>17,056</b>

# EconData metadata



## Dimensions

[BANK] Bank

[AGG] Category/Aggregation

[LINE] Line

[TIME\_PERIOD] Time period

## Primary Measure

[OBS\_VALUE] Observation value

## Dataset Attributes

(M) [RELEASE] Release description

## Series Attributes

(M) [UNIT\_MEASURE] Unit of measure

(M) [UNIT\_MULT] Unit multiplier

(M) [FREQ] Frequency

## Observation Attributes

- n/a -

## Group Attributes

- n/a -

Table

Plot

Metadata

datakey	Bank	Category/Aggregation	Line	Frequency	Unit multiplier	Unit of measure
TOT.A1.L001	Total	Liabilities: Deposits: Cheque (T1)	Deposits (L001)	Monthly	Thousands	Rand
TOT.A1.L002	Total	Liabilities: Deposits: Cheque (T1)	Deposits denominated in Rand (L002)	Monthly	Thousands	Rand
TOT.A1.L003	Total	Liabilities: Deposits: Cheque (T1)	SA Bank (L003)	Monthly	Thousands	Rand
TOT.A1.L005	Total	Liabilities: Deposits: Cheque (T1)	Other deposits (L005)	Monthly	Thousands	Rand

# EconData Registry



EconData Registry

Home

Organisations

Data

Data Definitions

Data Structures

Dataflows

Data Reporting

Provision Agreements

Reporting Constraints

Reporting Templates

Validation Schemes

Convert Data

Convert Data

Data Dissemination

Pre-Defined Queries

Items

Metadata Structures

Structure Maps

Web Service

Export Structures

Structure References

Search

Data Structure Definitions

All

Id

Name

ECONDATA

BA100

Bank Balance Sheet (IFRS)

ECONDATA

BA120

Bank Income Statement (IFRS)

ECONDATA

BA200

Bank Credit Risk Exposure

ECONDATA

BA325

Bank Selected Risk Exposure

ECONDATA

BA700

Bank Capital Adequacy

ECONDATA

BA900

Bank Balance Sheet (institutional)

ECONDATA

BUSINESS\_CYCLES

Business cycles

ECONDATA

CIVIL\_CASES

Civil Cases

ECONDATA

CPD\_RATES

CPD rates

ECONDATA

CPI\_ANL\_SERIES

CPI Analytical Series

ECONDATA

CPI\_COICOP\_5

CPI COICOP 5-digits

ECONDATA

ELECTRICITY

Electricity generated and available for distribution

ECONDATA

EXTERNAL\_SECTOR

External sector

ECONDATA

FINANCIAL\_SECTOR

Financial sector

ECONDATA

FISCAL\_SECTOR

Fiscal sector


Showing 1 to 15 of 30 entries

Data Structure Definition Details

- Data catalogue
  - Detailed meta-data
  - Taxonomy for definitional consistency
  - Quality checks
  - Process codification
  - Enables automation
  - User-level access control
  - Provisioning agreements
  - Interoperability
  - Vintage storage

# EconData R package




 Search or jump to... / Pull requests Issues Codespaces Marketplace Explore

coderaanalytics / econdatar Public

<> Code Issues 5 Pull requests Actions Projects Wiki Security Insights

master 3 branches 3 tags Go to file Add file <> Code

 byrongsby Update README.md e3d145f last week 42 commits

R	If condition bug fix	last week
inst	Fixed bug in write_econdatar	last year
man	Added returnrange query parameter to read release	10 months ago
.gitignore	Initial commit	3 years ago
DESCRIPTION	If condition bug fix	last week
LICENSE	Initial commit	3 years ago
NAMESPACE	Clean up the namespace	last week
README.md	Update README.md	last week

README.md

## Quick start

```
install.packages(c("remotes", "tcltk"), repos = "https://cran.mirror.ac.za")
library("remotes")
install_github("coderaanalytics/econdatar", ref = "1.1.6")
```

Install from disk

```
install.packages("path/to/econdatar", repos = NULL, type = "source")
```

Re-install (upgrade when new versions are available) - start a new session after running

```
library("remotes")
remove.packages("econdatar")
install_github("coderaanalytics/econdatar")
```

### About

R package for uploading and downloading data to/from [www.econdatar.co.za](http://www.econdatar.co.za)

- Readme
- EPL-2.0 license
- 2 stars
- 2 watching
- 0 forks

### Releases

3 tags

[Create a new release](#)

### Packages

No packages published

[Publish your first package](#)

### Languages

R 86.9%	PowerShell 13.1%
---------	------------------

# Access EconData using R or direct data export



## Data export options

**R command** (use with the [econdatar package](#))

```
read_econdata(  
  agencyid = "ECONDATA",  
  id = "NATL_ACC",  
  version = "1.4",  
  key = "KBP6006.R.",  
  provideragencyid = "ECONDATA",  
  providerid = "SARB",  
  releasedescription = "2023-03-01")
```

**Citation for this query**

SARB. NATL\_ACC. Version 1.4. Pretoria: SARB [producer], 2023-03-01. Johannesburg: EconData [distributor], 2023.

**Select file for download**

- Excel
- OpenDocument
- Comma separated values
- HTML table
- Text



# Our Excel plugin is available for beta testing



T23

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ECONDATA:NATL_ACC(1	KBP6006.N.S	ECONDATA:NATL_ACC(1	KBP6006.R.S																
2	.4)		.4)																	
3	Label	GDP at market prices	Label	GDP at market prices																
4	Frequency	Quarterly	Frequency	Quarterly																
5	Transformation	None	Transformation	None																
6	Start	3/31/1960	Start	3/31/1960																
7	End	3/30/2022	End	3/30/2022																
8	Base period	2015	Base period	2015																
9	Comment	Gross domestic product at market prices. Current prices. Seasonally adjusted at annual rate.	Comment	Gross domestic product at market prices. Constant 2015 prices. Seasonally adjusted at annual rate.																
10	Price concept	Nominal	Price concept	Real																
11	Seasonal adjustment	Seasonally adjusted data, not calendar adjusted	Seasonal adjustment	Seasonally adjusted data, not calendar adjusted																
12	Series name	GDP	Series name	GDP																
13	Source ID	NRI6006L	Source ID	NRI6006D																
14	Unit multiplier	Millions	Unit multiplier	Millions																
15	Unit of measure	Rand	Unit of measure	Rand																
16		3/31/1960		3/31/1960																
17		5295.34		565040.3567																
18		6/30/1960		6/30/1960																
19		5352.5122		574254.6957																
20		3/30/1960		3/30/1960																
21		5424.1419		584835.9301																
22		12/31/1960		12/31/1960																
23		5572.2046		530608.622																
24		3/31/1961		3/31/1961																
25		5602.5434		594892.3743																
26		6/30/1961		6/30/1961																
27		5585.2621		592292.8323																
28		3/30/1961		3/30/1961																
29		5768.8003		605522.4264																
30		12/31/1961		12/31/1961																
31		5824.362		611026.621																
32		3/31/1962		3/31/1962																
33		5932.9797		626831.3482																

## Data export options

R command (use with the [econdatar](#) package)

```
read_econdatar(agencyid = "ECONDATA", id = "NATL_ACC", version = "1.4", key = "KBP6006.R.", provideragencyid = "ECONDATA")
```

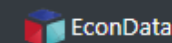
## Citation for this query

.. Version . [producer], 2022-05-07. Johannesburg: EconData [distributor], 2022.

## Select file for download

Choose format...

## EconData Add-In



### Data query

#### Options

#### Data set

National Accounts

#### Data release

2023-03-01

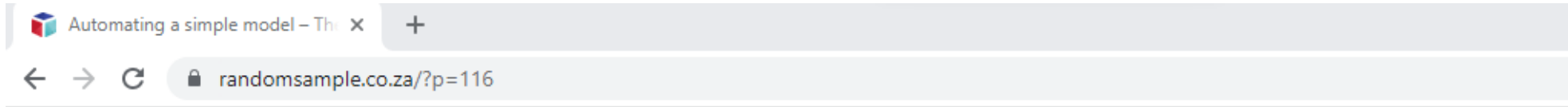
☒ Include metadata

#### Data key (filter)

#### Series name

GDP

# Resources for automation



THE ECONDATA BLOG  
Tutorials and other resources

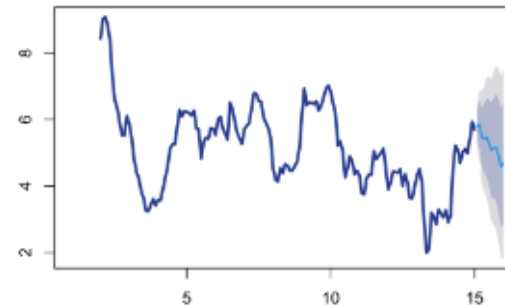
[Home](#) [Contact](#)

## Automating a simple model

---

A part from automating your charts, EconData makes it possible for you to automate your models.

Forecast of 12-month ahead y-o-y headline inflation



# Our EconData Blog for code snippets



```
library(econdatar)
library(dplyr)
library(ggplot2)
library(readr)
library(tidyr)
```

Next, we create a helper function that converts the output of the EconData R package to *dplyr tibble* format.

```
format_gdp <- function(x) {
  as.yearqtr <- function(y)
    paste0(format(y, format = "%Y"), quarters(y))
  tibble(Value = x$KBP6006.R.S$OBS_VALUE) %>%
    mutate(Date = x$KBP6006.R.S %>%
      rownames() %>%
      as.Date() %>%
      as.yearqtr()) %>%
    mutate(Vintage = release$Description)
}
```

# Auto updated models and dashboards



Consumer Price Inflation  
Disaggregated Inflation Model

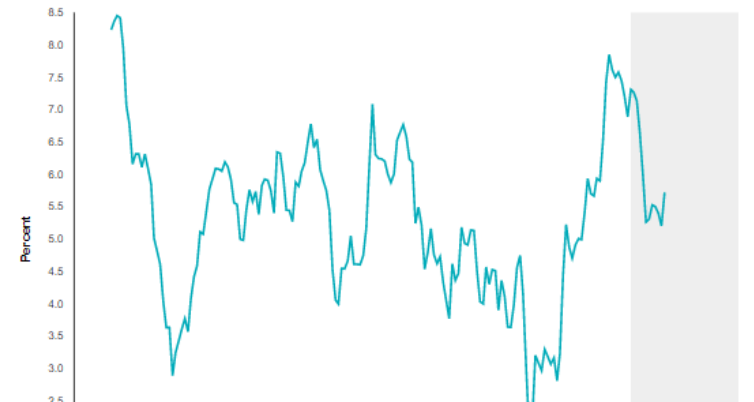
:

Scenario Charts Table Error Analysis Tutorial

## Baseline scenario



## Overview



## All Items (100%)



## Food and non alcoholic beverages (19.07%)



## Food (17.1%)



# EconData Summary



WHY	WHAT	WHERE	WHEN	WHO	HOW
<ul style="list-style-type: none"><li>▶ Enabling automation and advanced analytics</li><li>▶ Centralise internal &amp; external economic data</li><li>▶ Definitional consistency</li><li>▶ Store on premise or online</li><li>▶ Access data via variety of tools</li><li>▶ Process flows and quality assurance</li></ul>	<ul style="list-style-type: none"><li>▶ Centralised data</li><li>▶ Interoperability of tools</li><li>▶ SDMX for metadata-driven governance &amp; process codification</li><li>▶ Answering policy, finance and economic questions</li></ul>	<ul style="list-style-type: none"><li>▶ Accessible across organisation from anywhere</li><li>▶ Central data glossary of concepts and meta-data</li></ul>	<ul style="list-style-type: none"><li>▶ Latest data available from single source</li><li>▶ Scheduled up-dating based on data suppliers</li></ul>	<ul style="list-style-type: none"><li>▶ User-level access security</li><li>▶ Data creators can push data to platform</li><li>▶ Provision agreements for custodian</li></ul>	<ul style="list-style-type: none"><li>▶ Interoperability</li><li>▶ Automation of models and dashboards</li><li>▶ Data quality management and policy enforcement</li></ul>



# EconData Demo

# System integration for advanced analytics

## Data Integration

- Public data
- 3<sup>rd</sup> Party data
- Internal data



## Data Model

- Meta-data
- Provisioning agreements
- Registry/Catalogue
- Taxonomy
- User-level access



## Data Automation

- API for automation of dashboards/models
- Web interface for exploration and download



## Publication

- Auto-upload reports / dashboards
- Disseminate via emails/blogs



## Model Refresh

- Data and model update
- Forecasts and analysis
- Diagnostics



## Value added tools

- Probabilistic assessment
- User defined dashboards
- Judgement addition
- Scenario analysis





**[info@codera.co.za](mailto:info@codera.co.za)**