Automating workflows using EconData

Johannesburg R User Group 14 March 2023

CODERA
ANALYTICS

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

- 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

EconData

- Data Democratization: access data using variety of tools

- Definitional consistency of economic concepts

- Central data glossary of concepts and meta-data
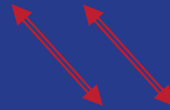
- Enabling data automation

CODERA
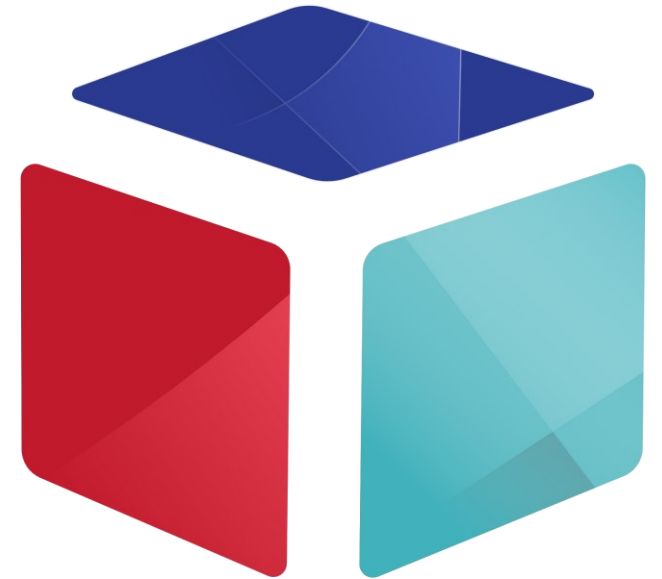
Essential Concepts

Definitional consistency
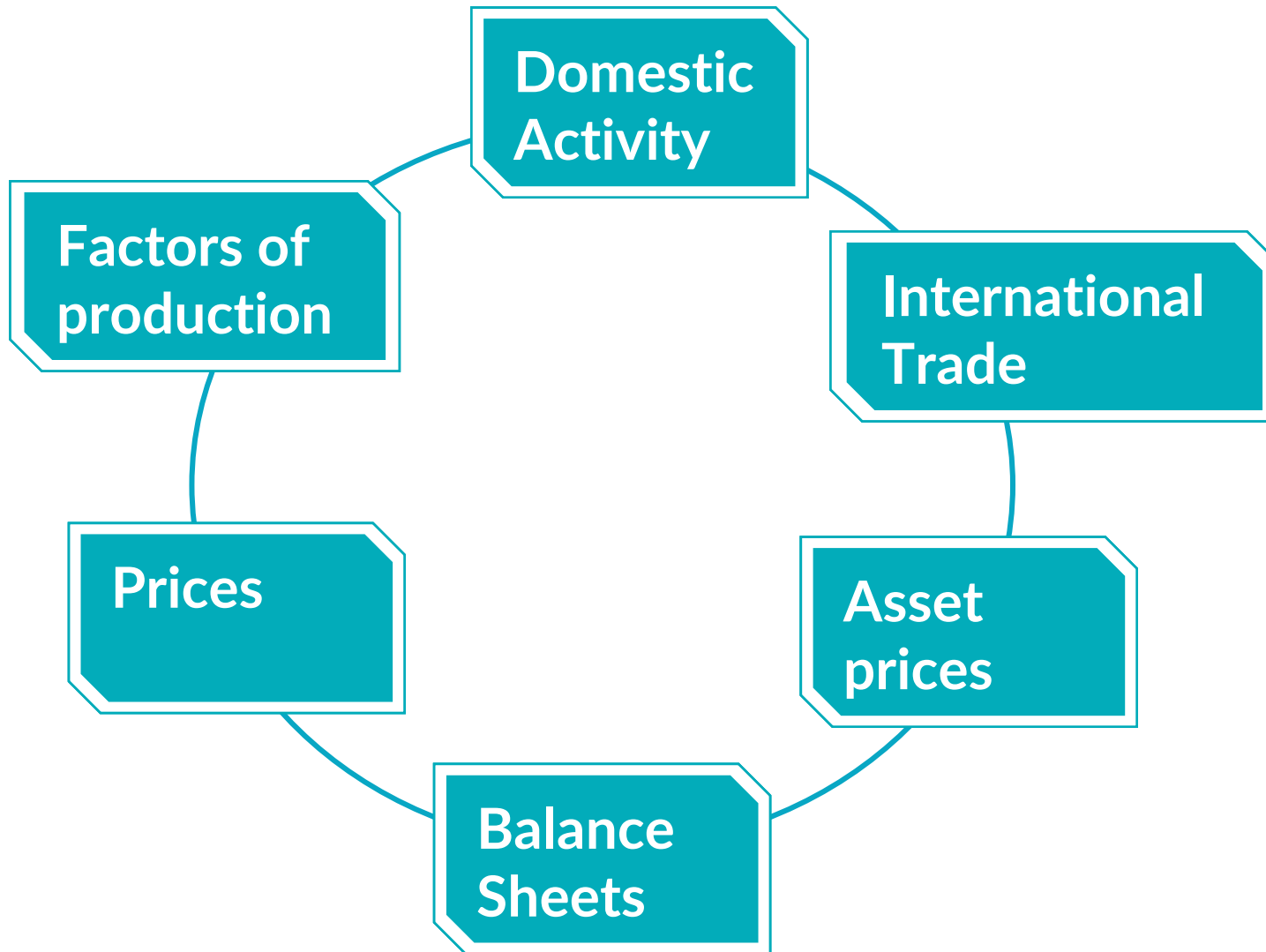
Definition
Sources
Ownership
Quality

SDMX

Standard data format
Metadata-driven governance
Process codification

Data Democratization

Data Catalogue
Meta-data
Sharing Data
Tools

Definitional consistency of economic concepts

- Domestic Activity
- International Trade
- Asset prices
- Balance Sheets
- Prices
- Factors of production

- 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
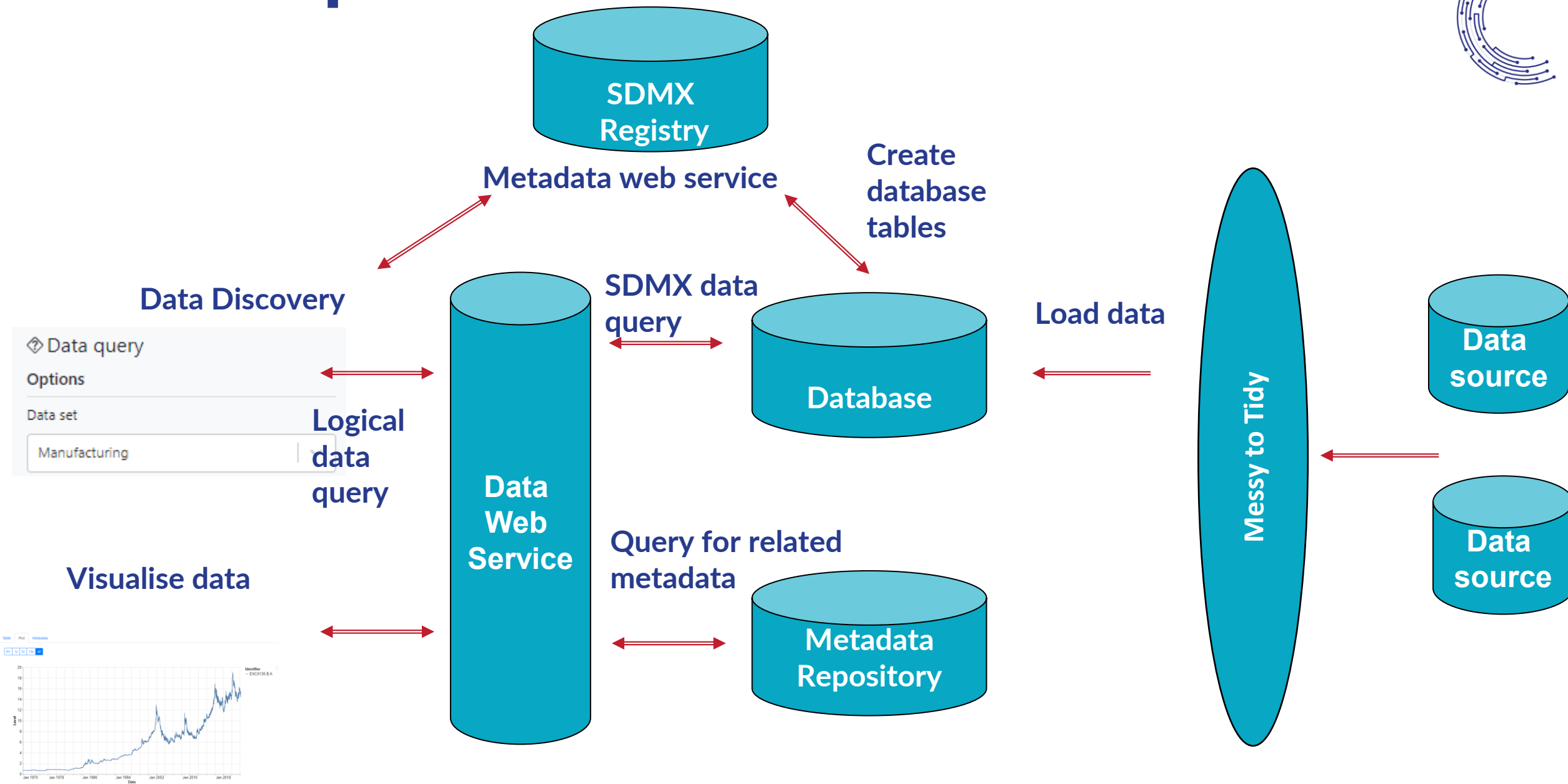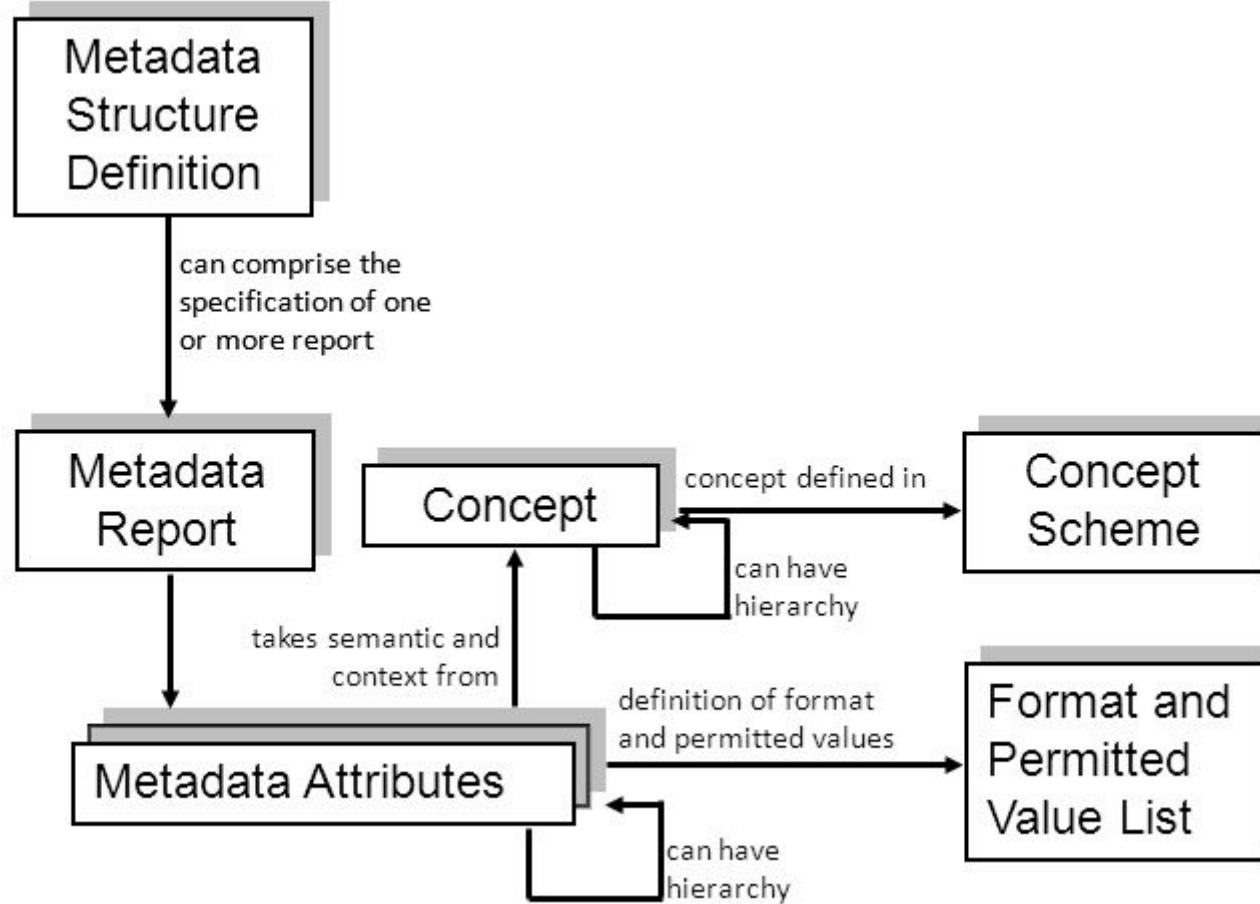
- Evolution of data model and tools

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 Registry

Metadata web service

Create database tables

Data Discovery

SDMX data query

⬥ Data query

Options

Data set

Manufacturing

Logical data query

Data Web Service

Database

Load data

Messy to Tidy

Data source

Query for related metadata

Visualise data

Metadata Repository

Data source

SDMX metadata

- Release calendar

- Provider

- Metadata from provider (i.e. Structure)

- Creation of reporting Taxonomy

Source: Arofan Gregory, Open Data Foundation

Data flows for controlling reporting & publishing
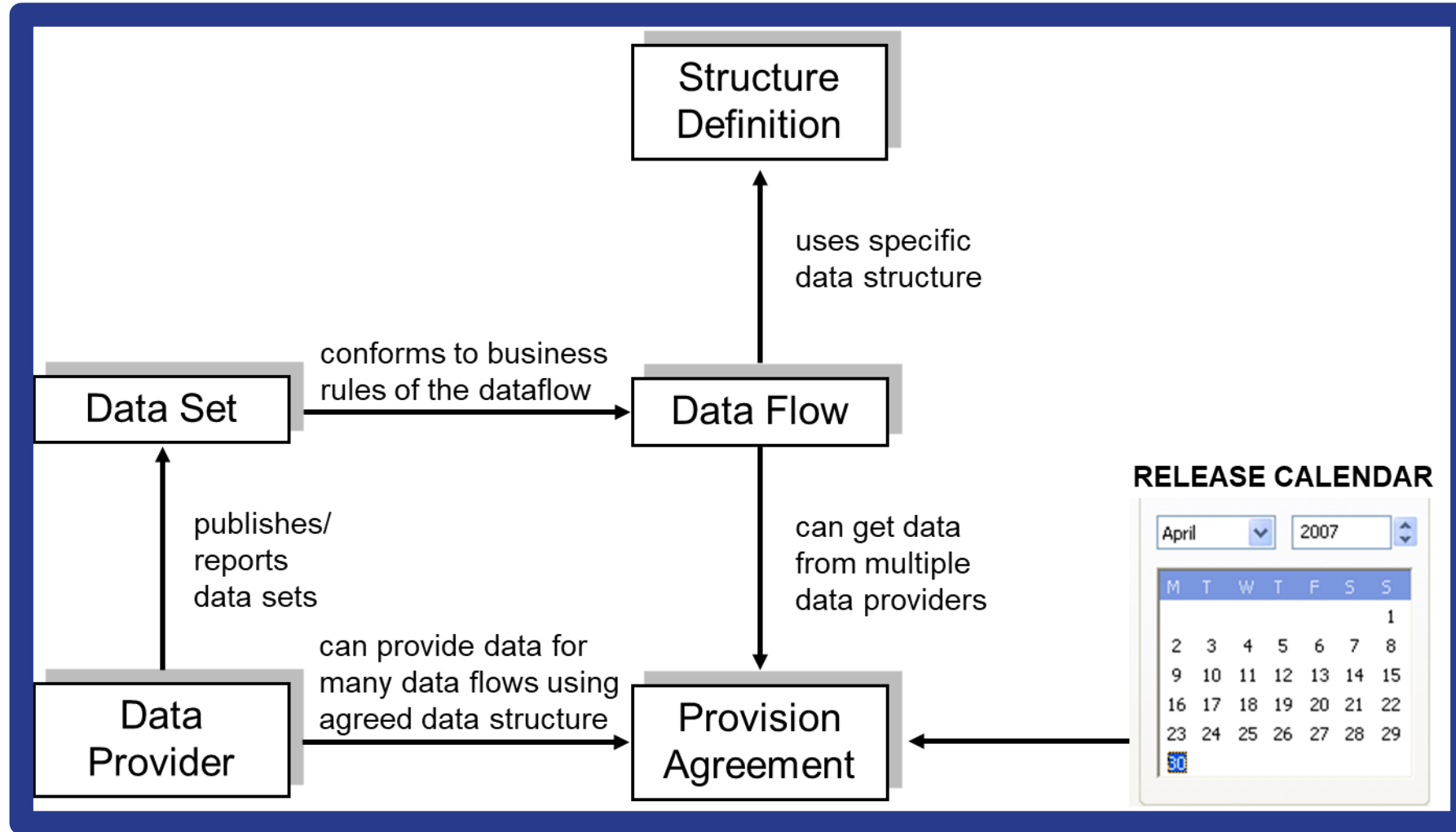
Source: Arofan Gregory, Open Data Foundation

EconData structures messy official statistics

Statistics South Africa

Annual Financial Statistics survey 2020[1] (AFS) preliminary estimates

Forestry, logging and related services

SIC 12	Carrying value at the beginning of the financial year (A)	Additions (capital expenditure) (B)	Disposal of assets (C)	Revaluation and other adjustments to carrying value (D)	Transfers-in and/or transfers-out (E)	Depreciation and amortisation (F)	Carrying value at the end of the financial year (G)[2]
				R million			
Tangible assets							
Land	1,745	48	134	243	0	5	1,897
Residential buildings	26	0	0	0	0	0	26
Total	11,793	1,971	321	254	3	1,231	12,469

Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing

SIC 13	Carrying value at the beginning of the financial year (A)	Additions (capital expenditure) (B)	Disposal of assets (C)	Revaluation and other adjustments to carrying value (D)	Transfers-in and/or transfers-out (E)	Depreciation and amortisation (F)	Carrying value at the end of the financial year (G)[2]
				R million			
Tangible assets							
Land	35	5	0	0	0	0	40
Residential buildings	63	2	0	0	0	0	65
Total	4,491	567	104	93	0	460	4,587

Forestry and fishing

Total - SIC 1	Carrying value at the beginning of the financial year (A)	Additions (capital expenditure) (B)	Disposal of assets (C)	Revaluation and other adjustments to carrying value (D)	Transfers-in and/or transfers-out (E)	Depreciation and amortisation (F)	Carrying value at the end of the financial year (G)[2]
				R million			
Tangible assets							
Land	1,780	53	134	243	0	5	1,937
Residential buildings	89	2	0	0	0	0	91
Total	16,284	2,538	425	347	3	1,691	17,056

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

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

Access EconData using R or direct data export

Data export options   X

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")
```
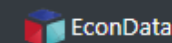
Citation for this query 📋

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

Select file for download

Choose format...   ⌄

Excel

OpenDocument

Comma separated values

HTML table

Text

Our Excel plugin is available for beta testing

Resources for automation
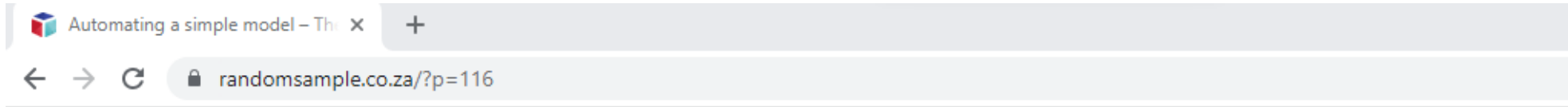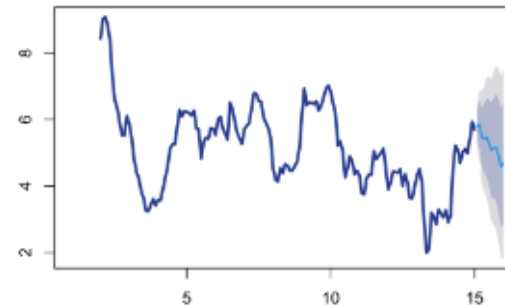
Our EconData Blog for code snippets

```r
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.

```r
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
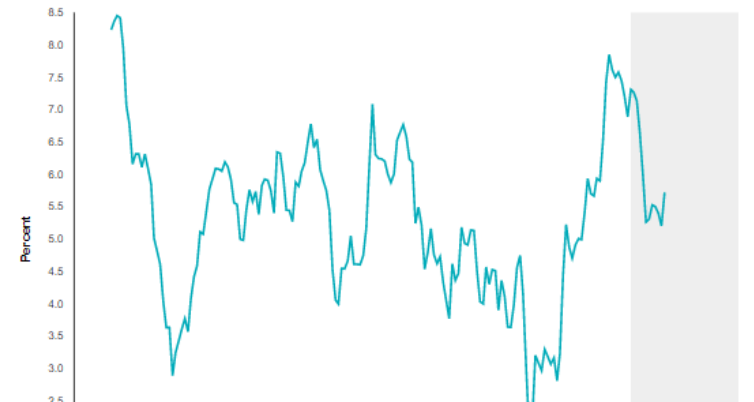
EconData Summary

WHY	WHAT	WHERE	WHEN	WHO	HOW
▸ Enabling automation and advanced analytics ▸ Centralise internal & external economic data ▸ Definitional consistency ▸ Store on premise or online ▸ Access data via variety of tools ▸ Process flows and quality assurance	▸ Centralised data ▸ Interoperability of tools ▸ SDMX for metadata-driven governance & process codification ▸ Answering policy, finance and economic questions	▸ Accessible across organisation from anywhere ▸ Central data glossary of concepts and meta-data	▸ Latest data available from single source ▸ Scheduled up-dating based on data suppliers	▸ User-level access security ▸ Data creators can push data to platform ▸ Provision agreements for custodian	▸ Interoperability ▸ Automation of models and dashboards ▸ Data quality management and policy enforcement

EconData Demo

System integration for advanced analytics

Data Integration
- Public data
- 3rd 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