



Building a database on cryptocurrencies

Ilaria Mattei and Anamaria Illes, *Bank for International Settlements*

3rd IFC Workshop on "Data Science in Central Banking: Enhancing the access to and sharing of data"

17 October 2023

The views expressed herein are those of the authors and do not necessarily reflect the opinions of the BIS.

Database on cryptocurrencies

- **What is it?**
 - Overview, scope and definitions
- **How did we build it?**
 - Preparatory work
 - Data sources
 - Key challenges and measures taken to tackle these
 - Daily automated updates
 - Residual manual tasks (for stablecoins only)
- **Current applications**
- **What's next**

Overview, scope & definitions

- The cryptocurrencies database is a collection of:
 - **Market variables** (market capitalisation, prices, trading volumes, ...) for all cryptocurrencies coming from multiple data sources.
 - **Derived indicators**, mostly on *stablecoins*: market structure, price volatility, price deviations from peg, reserves.
- *Stablecoins* in scope: cryptocurrencies that aim (claim) to maintain a stable value relative to a specified asset, or a pool or basket of assets, and that have a positive market capitalisation.
 - A. Fiat-backed:** claim to be backed by assets denominated in a fiat currency.
 - B. Crypto-backed:** claim to be backed by other cryptocurrencies.
 - C. Commodity-backed:** claim to be backed by commodities.
 - D. Unbacked:** do not claim to be backed by any reserves and rather seek to maintain a stable value through, for instance, algorithms or protocols.

From the conceptual idea to its materialization

1. Definition of the **dimensions to monitor**:

- Size and features of the market
- Quality of the stabilization mechanism (for stablecoins only)
- Composition of the reserves (for stablecoins only)

2. Identification of the **variables**:

- Market capitalisation, volumes, prices
- Attestation reports (for stablecoins only)

3. Selection of the **data sources**: CoinGecko, CCData (CryptoCompare), websites of stablecoins

Data source and variables in details

Sources	Database of market variables	Database of derived indicators
CoinGecko	List of stablecoins	(1) Number of stablecoins
	Market capitalisation	(1) Levels of and changes in market capitalisation (2) Ranking (3) Market concentration (HHI) (4) Turnover
	(Off-chain) volume	(1) Trading volume (2) Turnover
	Closing prices	(1) Price (2) Price volatility (3) Deviations from peg
CCData (CryptoCompare)	OHLC prices	(1) Prices (2) Deviations from peg
Stablecoins' websites	Composition of reserves	(1) Breakdown and timeseries of reserves

Challenges

- Lack of *unique identifiers* across data sources
- Lack of consistent *definitions* of stablecoin types
- Lack of *reporting standards* for the reserves

- Lack of provision of data on "*dead*" cryptocurrencies
- Publicly available data being subject to *copyrights*

Solutions

- Creation of own **mapping rules**

- Creation of own **metadata**

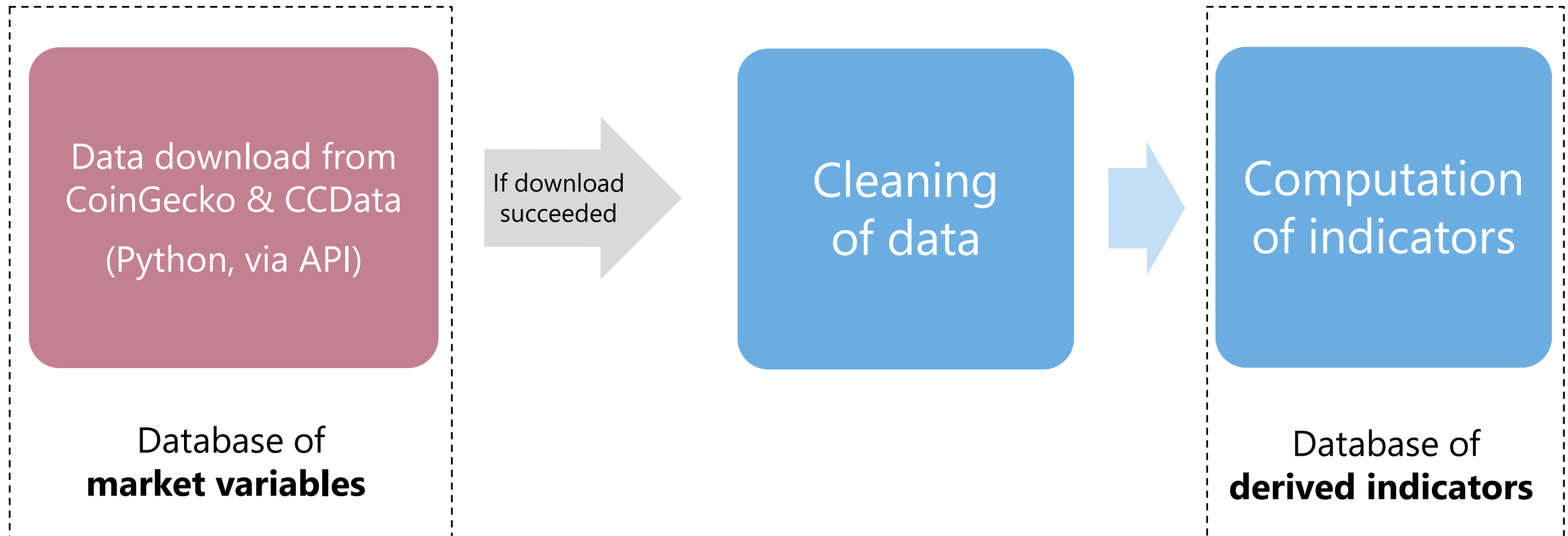
- Storing **vintages**

- Get in contact for **permissions**

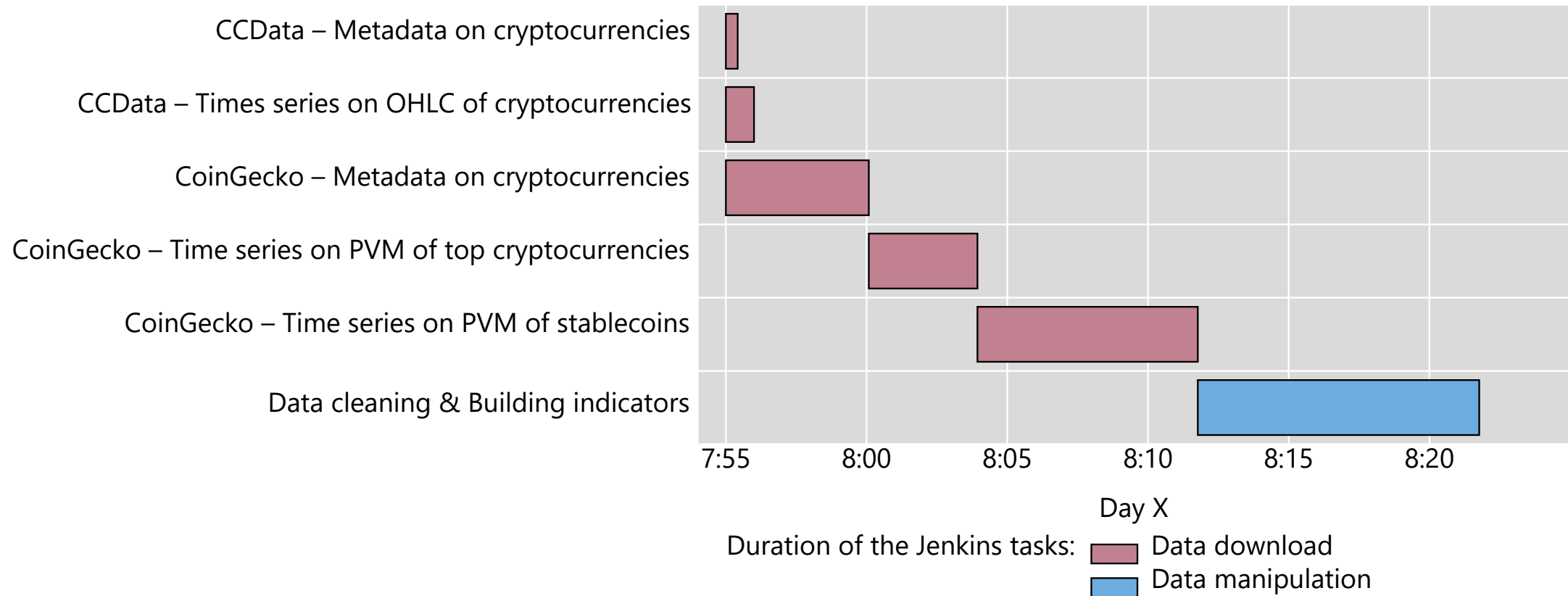
Result

Centralised database that is built in a structured way
to allow for a consistent comparison across data sources and time

The process runs automatically every day and takes less than 30 minutes



In a typical day...



OHLC= open, high, low and close prices; PVM = prices, volumes and market capitalisation.

Residuals manual tasks (for stablecoins only)

- **At the start and on an ad-hoc basis:**
 - Checking of (new) coins against our definition
 - Categorisation of (new) stablecoins by type
 - Mapping of (new) stablecoins among data sources
- **Troubleshooting codes which fail to run, eg due to:**
 - New (stable)coins
 - Temporary access restrictions to public API
 - Change in data structure (rare)
- **Update of breakdown of reserves on a monthly basis:**
 - Manually extract the information from the stablecoins' websites
 - Map the reserve-assets into our own high-level asset types

Current applications

- Daily (internal) newsletter on cryptocurrencies
- Stablecoins Dashboard (internal)
- Forthcoming paper

Ideas for future upgrades

- **Further automation:**

- String matching of cryptocurrencies' names across data sources
- Collection of information on reserves

- **Exploration of other sources, eg:**

- Coinmetrics
- IntoTheBlock
- Chainalysis
-

- **Addition of new indicators, eg:**

- Transactions
- Inflows to/outflows from exchanges
- DeFi
- ...



Questions?