IFC-Bank of Italy Workshop on "Machine learning in central banking"

19-22 October 2021, Rome / virtual event

Monitoring at scale[1]

Enrico Apicella, Marco D'Errico and Pedro Marques, European Central Bank;
Antonio Ciullo, Deloitte; and Caroline Übelhör, Google

[1] This presentation was prepared for the Workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the central banks and other institutions represented at the event.

Monitoring at Scale

Enrico Apicella,[1] Antonio Ciullo,[3] Pedro Marques,[2] Caroline Übelhör,[4] and Marco D'Errico[1]

[1] European Central Bank (European Systemic Risk Board Secretariat)

[2] European Central Bank (Directorate General Information Systems)

[3] Deloitte

[4] Google

IFC and Bank of Italy Workshop on "Data science in central banking"

21 Oct 2021

Disclaimer and acknowledgements

The views expressed in this presentation are our own and do not necessarily reflect those of the European Central Bank, the European Systemic Risk Board or its member institutions.

Thanks to Roberto Stok, Grzegorz Skrzypczynski, Oana Furtuna, Marco Canu, Giulia Di Rienzo, Marco Tizzoni, Umberto Coda, Eva Abend, Sylwia Kret, Mario Picano and Teresa Felici for input, comments and suggestions.

Overview

- Background: the ESRB's monitoring mandate and analytical strategy
- The scaling problem in Systemic Risk Monitoring (SRM)
- Monitoring "at Scale":
  - developing a general SRM framework
  - building blocks and additional elements
- The framework at work:
  - the March 2020 market turmoil and the largest margin call in history
- Conclusions and way forward

The ESRB's mandate and strategy rely on data and technologies

Mandate

*"The ESRB's task should be **to monitor and assess systemic risk** in normal times for the purpose of mitigating the exposure of the system to the risk of failure of systemic components and enhancing the financial system's resilience to shocks."*

*"The **interconnectedness of financial institutions and markets** implies that the monitoring and assessment of potential systemic risks should be based **on a broad set of relevant macroeconomic and micro-financial data and indicators**"*

Regulation (EU) No 1092/2010 establishing the ESRB

Strategy

*"Monitoring an interconnected financial system involves the availability of **detailed and granular transactions data**. But **in order to have the full picture, it is vitally important to be able to link data across markets, instruments and counterparties**." "Only a holistic view of the system will allow potential contagion channels to be identified and modelled. And that requires **investing in new technologies for data analytics and enhancing the capacity for authorities to link and share data and technical knowledge**."*

Former ESRB Chair Mario Draghi

Welcome remarks at the third and fourth annual conference of the ESRB

27 September 2018 and 26 September 2019

"Data across markets, instruments and counterparties" (examples)

Several granular data collections stemming from post-crisis reforms...

Dataset	Topic	Frequency	Level	Size / Complexity
EMIR (ESMA)	Derivatives	daily	counterparty / transaction	300k entities reporting 100 mln / day approx 160 bln records in total
SFTR (ESMA)	Repos/SFTs/other	daily	counterparty / transaction	9-10 mln / day nested and complex data 15k+ entities
AIFMD (ESMA)	AIF managers	quarterly	holdings, risk profiles, counterparties	60k funds
STSS (ESMA)	Securitisation	variable	variable	TBA / In progress
AnaCredit (ECB)	Loans	monthly	loan level	~29 mln / month

The two dimensions of Systemic Risk Monitoring (SRM)

Analytical	Organisational
Descriptive / distributional (monitor risk distribution & clusters)	**Research and Development** (enable full cycle: from ideas to usable products)
Interconnectedness / contagion (monitor and quantify potential channels)	**User eXperience** (translation of results, dissemination, etc.)
What if / scenario / simulations (compare against potential developments in the system)	**Knowledge Management** (replicability, sharing, understand, knowledge transfer, training)

The scaling problem in SRM: the *analytical* dimension

Analytical scaling: dimensionality problem

the number of analyses, indicators, processes and aggregations increases exponentially

- **Complexity & size**:
  the financial system is inherently complex, interconnected and adaptive
- **Linking** datasets
- **Speed**: as markets shifts rapidly, analytical outputs must be produced at a faster pace
- **Aggregation layers**: seamlessly/iteratively move across several layers of aggregation
  - Micro ↔ meso ↔ macro
    counterparty, groups, sector, countries, areas
- *[Deal with the (alas substantial!) **data quality issues** in granular supervisory data (not in this talk)]*

The scaling problem in SRM: the *organisational* dimension

Organisational scaling: operationalising the full cycle in the organisation

- **Process:**
  - from research to production (full R&D cycle)
  - possibly language agnostic
  - reproducible, reusable
- **Policy**: development, calibration
- **Communication & collaboration**: across the whole organisation and at different technical levels (data scientist, researcher, expert, management)
- **Knowledge management:**
  - share, explain, reuse, replicate
  - skill set & education

Solution (step 1): generalising the SRM framework

A scalable data model…

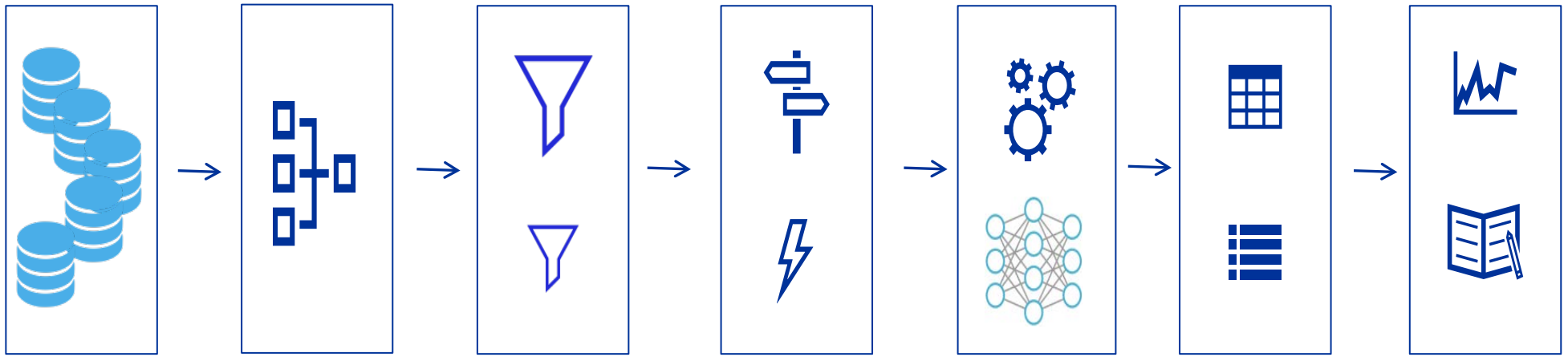
- model the **entire financial system at any level** as a *dynamic multilayer* network (multigraph)
- filters / aggregations / modifications ***preserve*** the same data model

Nodes	micro (counterparty sets), meso (groups), macro (sectors / countries)
Edges	(directed) balance sheet relationships (contracts, holdings, etc.)
Weights	any *measure* on the individual edges

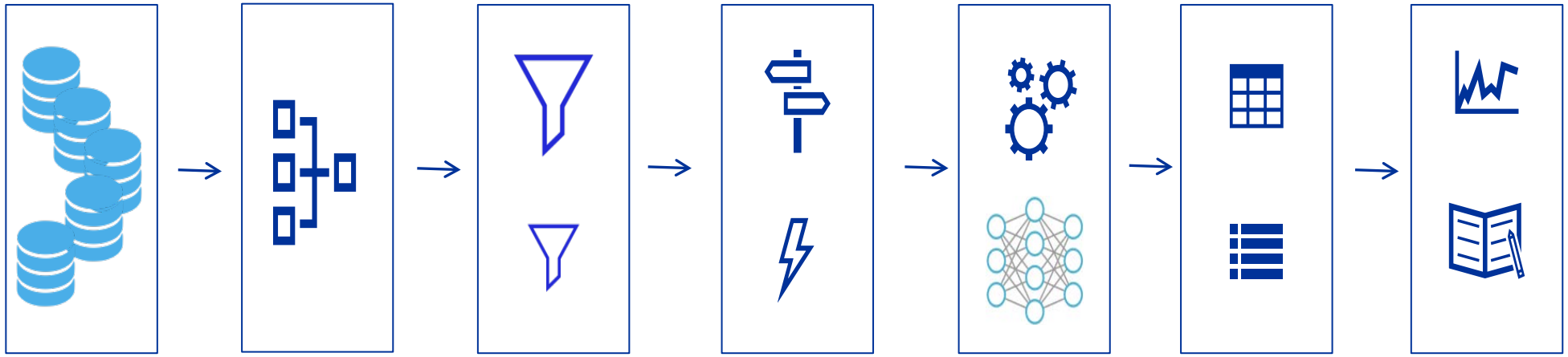
…enabling scalable operations on the data

Descriptive / distributional	descriptive statistics on the multilayer network
Interconnectedness / contagion	dynamics on the multilayer network
What if / scenario / simulations	operations on a transformed network

Solution (step 2): generalising the SRM framework



Solution (step 2): generalising the SRM framework



indicators	
conceptualisation	**implementation**
concepts	parameters
operations	functions
processes	compositions

→

Combining:
*- **operations (functions)** and*
*- **concepts (parameters)***

*we **obtain all indicators at every** **level***
with a precise explanation of the process

Solution (step 3): generalising the SRM framework

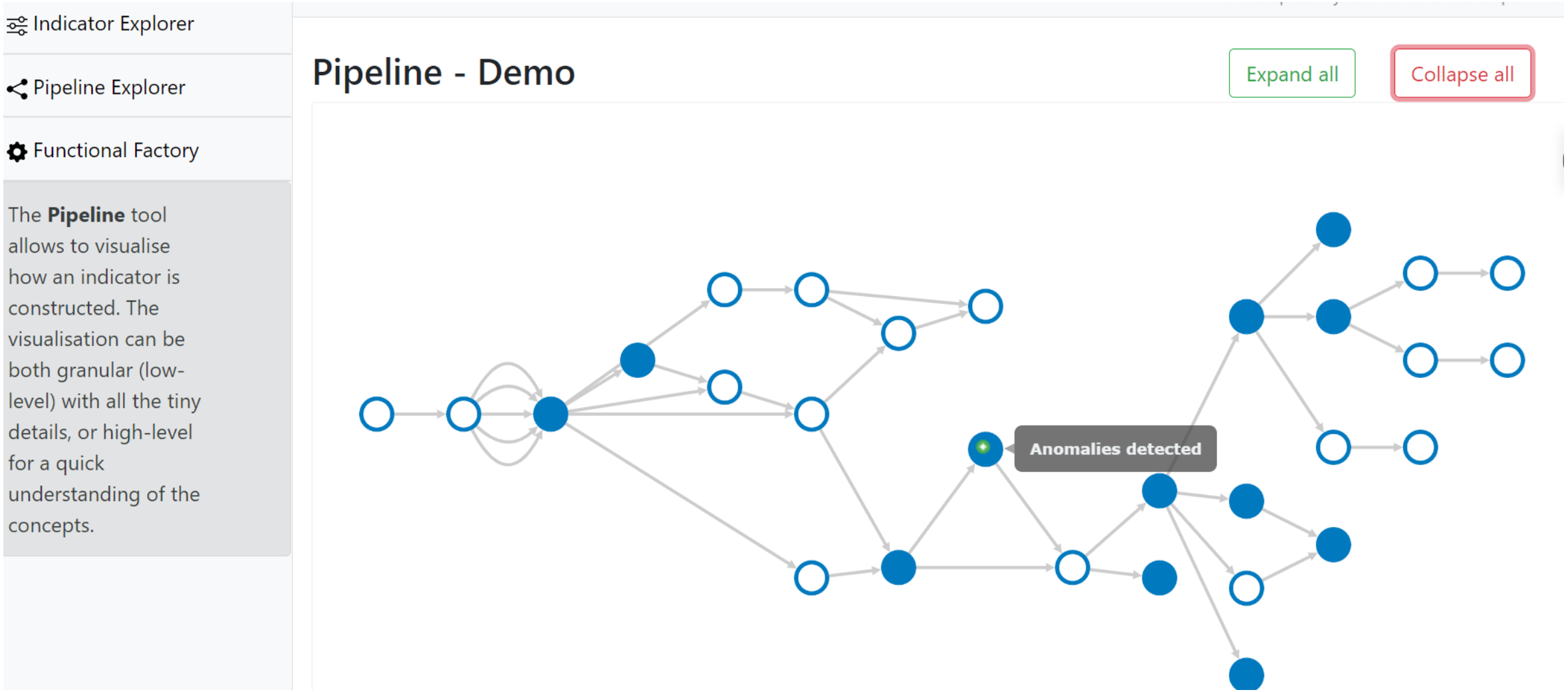
Building blocks (based on *functional* principles)

- **Functions** as operations defined for given **parameters** (concepts)
  - specialisation matters: first-class, higher-order, closures

- **Directed Acyclic Graphs** (DAG): encapsulating relationships between functions
  - sequential → composition (pipeline)
  - parallel → hierarchical / multivariable functions
    → adds expressivity via merging and branching, also increases complexity
    → counterfactuals / scenario analyses obtained via branching and merging
  - a DAG *becomes itself a function* → carries the description of the process (self documentation and rich semantics)
  - contagion models as fixed point iterations

- **Factories:** a function returning a (set of) function(s) from other ones or from DAGs
- **SRM indicators**: factory, pipelines or elementary functions bound to specified parameters

Solution (step 3): generalising the SRM framework



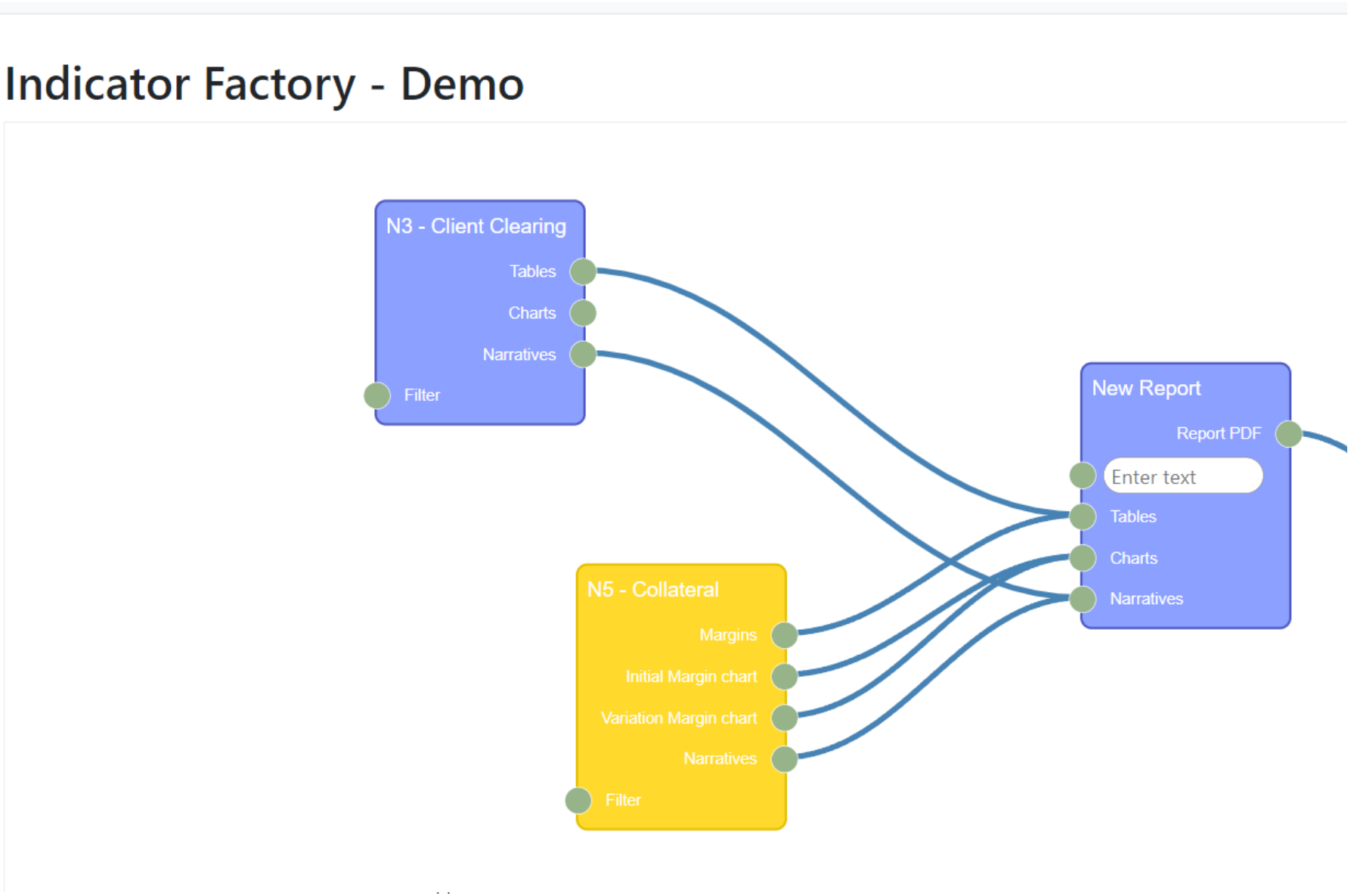
Solution (step 3): generalising the SRM framework

Indicator Explorer

Pipeline Explorer

Functional Factory

Indicator Factory: user can combine high-level concepts and/or predefined indicators to build a report which can be downloaded or automatically uploaded to a Darwin folder. This tool allows for some degree of customisation (e.g. filter some CCPs, choose a specific time window) without



Solution (step 4): generalising the SRM framework

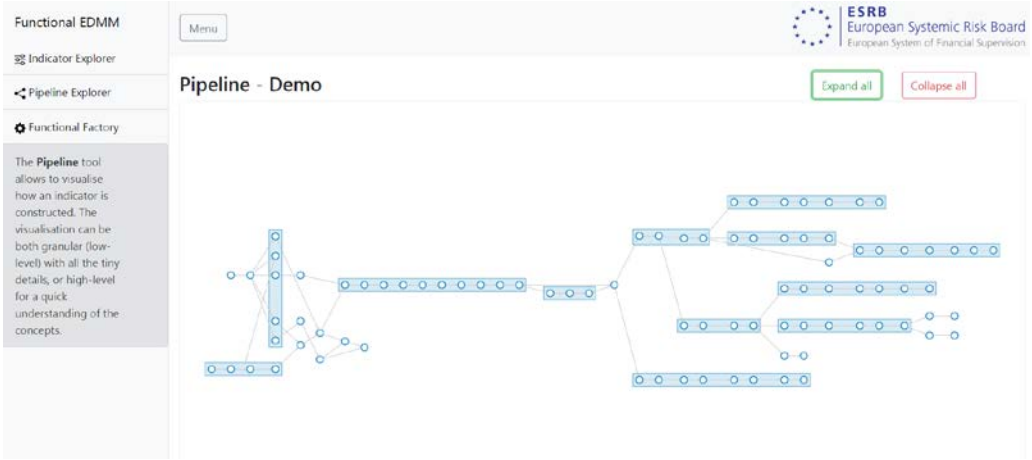
Building blocks - process

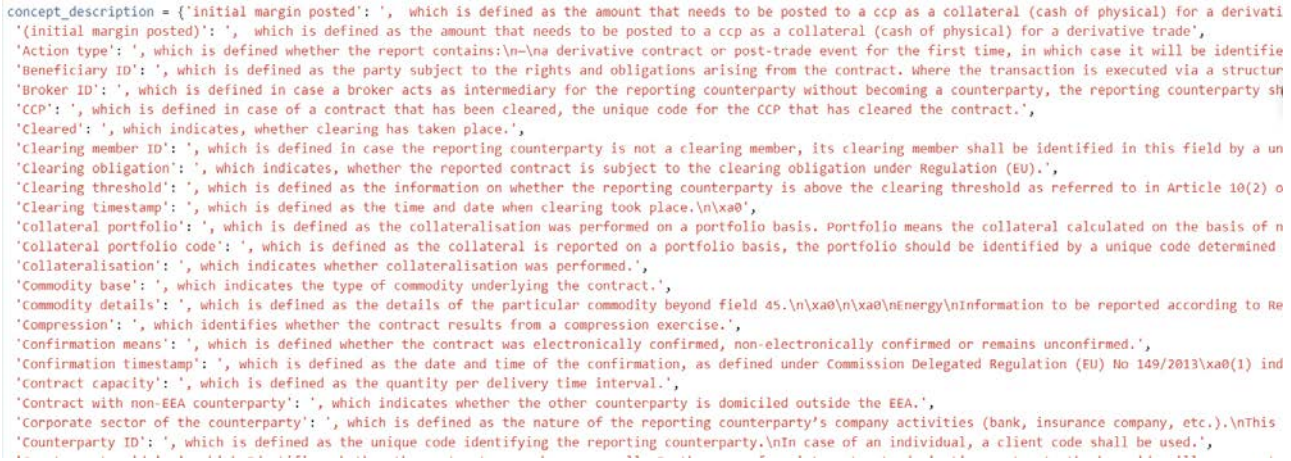
- Granular functions can be semantically framed by higher level ones and convey information about the hierarchical structure in the operations
  - ***filter*** *by* EU banks THEN ***group by*** bank country THEN ***compute*** max(margin call)
  - granularity tailored to each analysis

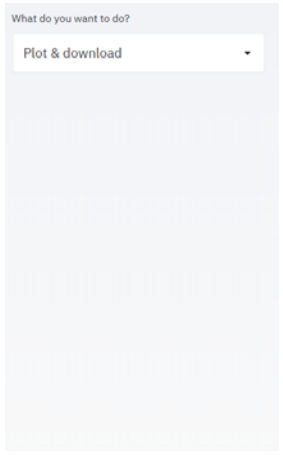
- Branching / merging conveys information on the reachability between operations:
  - compare two different scenarios (embeds causal relations)
  - reverse: find parameters at origin that would lead to given outcome (e.g. reverse stress testing)

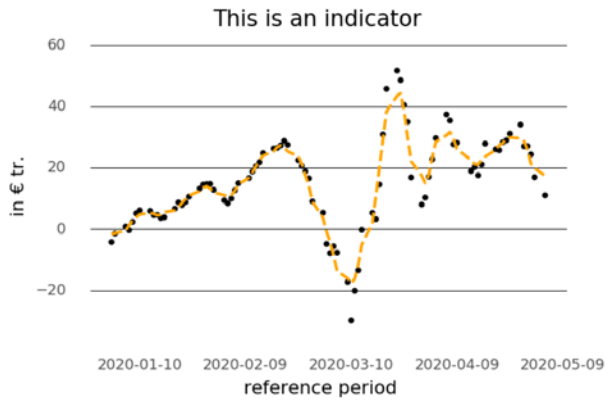
- Parameters can be linked to Knowledge Graphs: e.g. "*an* ***initial margin call*** *value of 10*, ***which is defined as***…"

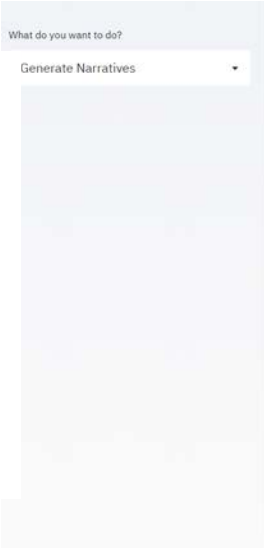
Solution (step 5): artefacts











Recent Value

The aggregated value of net_market_value (market value), which is defined as the current contract value of a contract, on the most recent date accounts for 11 trillion (exact value: 11000975800740.336).

Your data in a nutshell

The data frame consists of 8837 data points and 15 columns, which are the following: reference_period, ct, reporting_cpty_area, other_cpty_area, net_notional, net_market_value, net_to_gross_ratio, num_of_ctry1_ccps, num_of_ctry2_cpties, asset_class, contract_type, product, venue, gross_notional_eur_bn, gross_market_value.

The whole story

The data frame consists of 8837 data points and 15 columns, which are the following: reference_period, ct, reporting_cpty_area, other_cpty_area, net_notional, net_market_value, net_to_gross_ratio, num_of_ctry1_ccps, num_of_ctry2_cpties, asset_class, contract_type, product, venue, gross_notional_eur_bn, gross_market_value. The aggregated value of net_market_value (market value), which is defined as the current contract value of a contract, on the most recent date accounts for 11 trillion (exact value: 11000975800740.336).

Download Narratives

Solution (step 5): testing

Scaling issue: adding functions to the repository and/or an existing DAG → potential blind spots

Solution: key principle → tests as functions (can compose and added to a DAG)

Properties

- *Left to right inheritance*: a test on a given function is inherited by any other function that composes with it → a test on an indicator (factory) is inherited by any specification of the indicator
- *Right to left inheritance*: tests on a composed function hold (and do so recursively) if and only if all tests the argument functions hold
- Implication: integration testing corresponds to unit testing on composed functions / DAGs

Applications

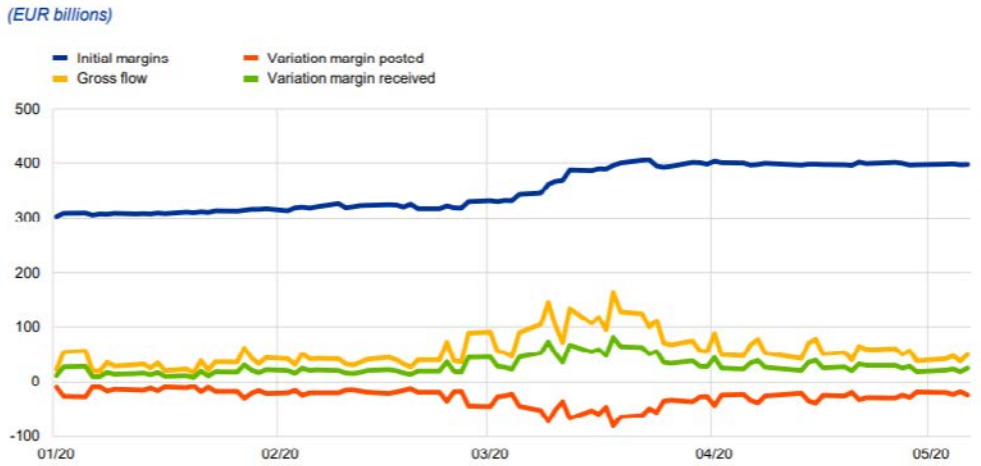
- A number of different test families can be applied as ***assertions on data***
  - Quality checks
  - Probabilistic testing (e.g. tests on the distribution or as r.v. transformations)
  - Business: testing on scenarios

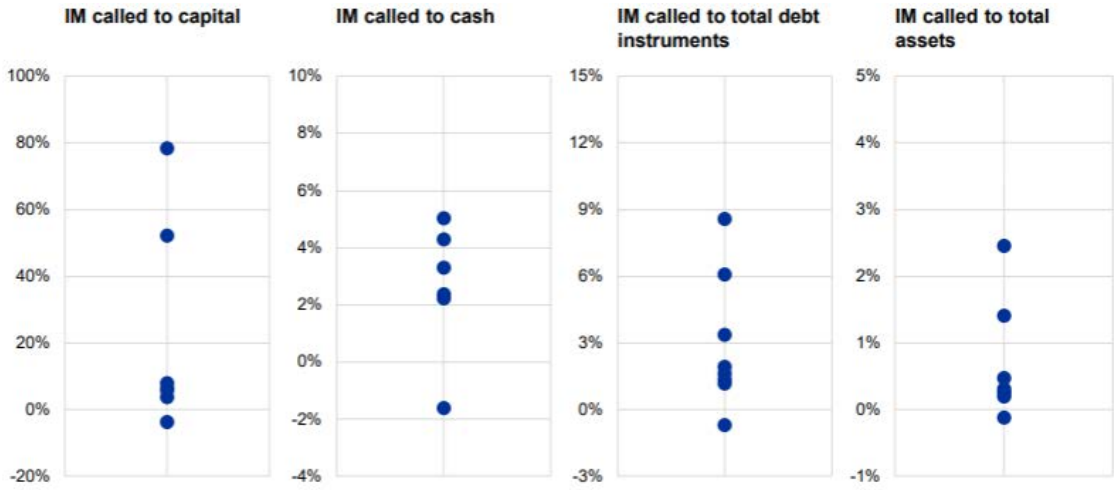
Monitoring at scale: the March 2020 margin call

- **March 2020**: sharp drop in asset prices and increased volatility, resulting in the **largest margin call in history**

- First ever experience of use of granular data in "**crisis mode**"

- ESRB tasks:
  - monitoring macro / meso / micro level
  - understanding the ***causes*** (what asset class? what dynamics?)
  - counterfactual analysis / reverse stress testing (depletion of liquid resources)
  - develop formal recommendation on liquidity risks





Conclusions and way forward

- Systemic Risk Monitoring as a core task of ESRB requires data, technologies and analytical frameworks
- Contribution to solving the scaling problem by developing a general SRM framework inspired by functional principles
- Such framework allows to “monitor at scale” the financial system and has been adopted in several instances (also for policy development)
- Next steps will focus on:
  - enhancing the indicators set (in particular: constrained simulations)
  - provide a richer user experience, linking with existing knowledge

References (a few…)

- *Functional principles in data science*:
  - filter-map-reduce / split-apply-combine (Wickham, 2011)
  - tidyverse (Wickham et al, 2019)
  - grammar of graphics (Wilkinson, 1999)
  - lambda calculus
- *Data engineering*:
  - Google Dataflow
  - Facebook Prophet (scaling issues in forecasting)
  - Airflow, Luigi, Dask
- *Systemic Risk Monitoring*:
  - Rethinking the financial network (Haldane, 2009)
  - Shedding light on dark markets (ESRB, 2016)
  - Mapping exposures (Abad et al., 2021)
  - Network Valuation (Barucca et al. 2020)
- *Statistics / data science*:
  - Causality (Pearl, 2000, 2009, 2018)
  - Functions of learning, Learning from data (Strang, 2019 and 2020)