



DATA DRIVEN FINANCIAL STABILITY

#datadrivenstability
#financialstability

2–3.12.2019

OPPORTUNITIES
AND CHALLENGES
IN BIG DATA



Introductory speech

Olli Rehn

Governor

Bank of Finland



Introductory speech

Bruno Tissot

**Head of Statistics
and Research Support**

**Bank of International
Settlements**





Big data information and analytics: issues for central banks

Bruno Tissot

Head of Statistics and Research Support, BIS

Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC)

Conference on Data Driven Financial Stability: Opportunities and Challenges in Big Data

Bank of Finland, Helsinki, 2-3 December 2019

The views expressed are those of the author and do not necessarily reflect those of the BIS or the IFC.

Overview

- **Big data Information**
- **Working with big data: Artificial Intelligence potential**
- **Opportunities for central banks**
- **Organising big data work**
- **Challenges**
- **Policy issues with handling and using big data**

1. BD Information – Introduction

- **Three main sources of big data (UN)**

- 1. Social networks**

Human-sourced information

→eg blogs and searches

- 2. Traditional business systems**

Outcome of business processes

→eg files produced by commercial transactions, credit cards

- 3. The internet of things**

Machine-generated data

→eg by sensors, mobile phones, computer logs

1. BD Information - Introduction (cont'd)

- In practice, multiple types of heterogeneous datasets

- The three Vs

- Volume
- Velocity
- Variety

- ... and much more...

- 4Vs
- 5 Vs
- 7 Vs
- 12Vs
- ...
- The 42 V's of big data and data science
→Tom Shafer, Elder Research, Inc. (2017)



1. BD Information – Not just the Internet!

- **Designed versus organic data**

- Traditional statistics: **designed** data

- eg surveys, census

- **Organic** data: by-products

- of financial, commercial & administrative processes (and internet)

- Also interest in the financial industry for “**alternative data**”

- by-product of companies' operations

- Can be **large, granular, well-structured** datasets

- eg registers

1. BD – Not just the Internet! (cont'd)

- **Strong increase in granular data collected by central banks**

- Crisis impact
- Push for using administrative data

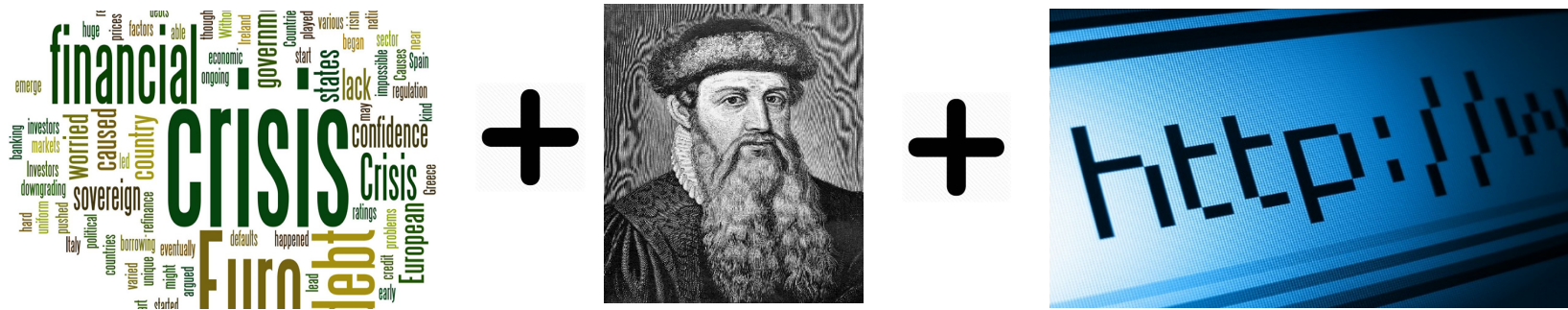


- **Importance of textual information**

- Generated by **internet...**
- Or "**traditional** text"
 - Digitised
 - Processed by big data tools



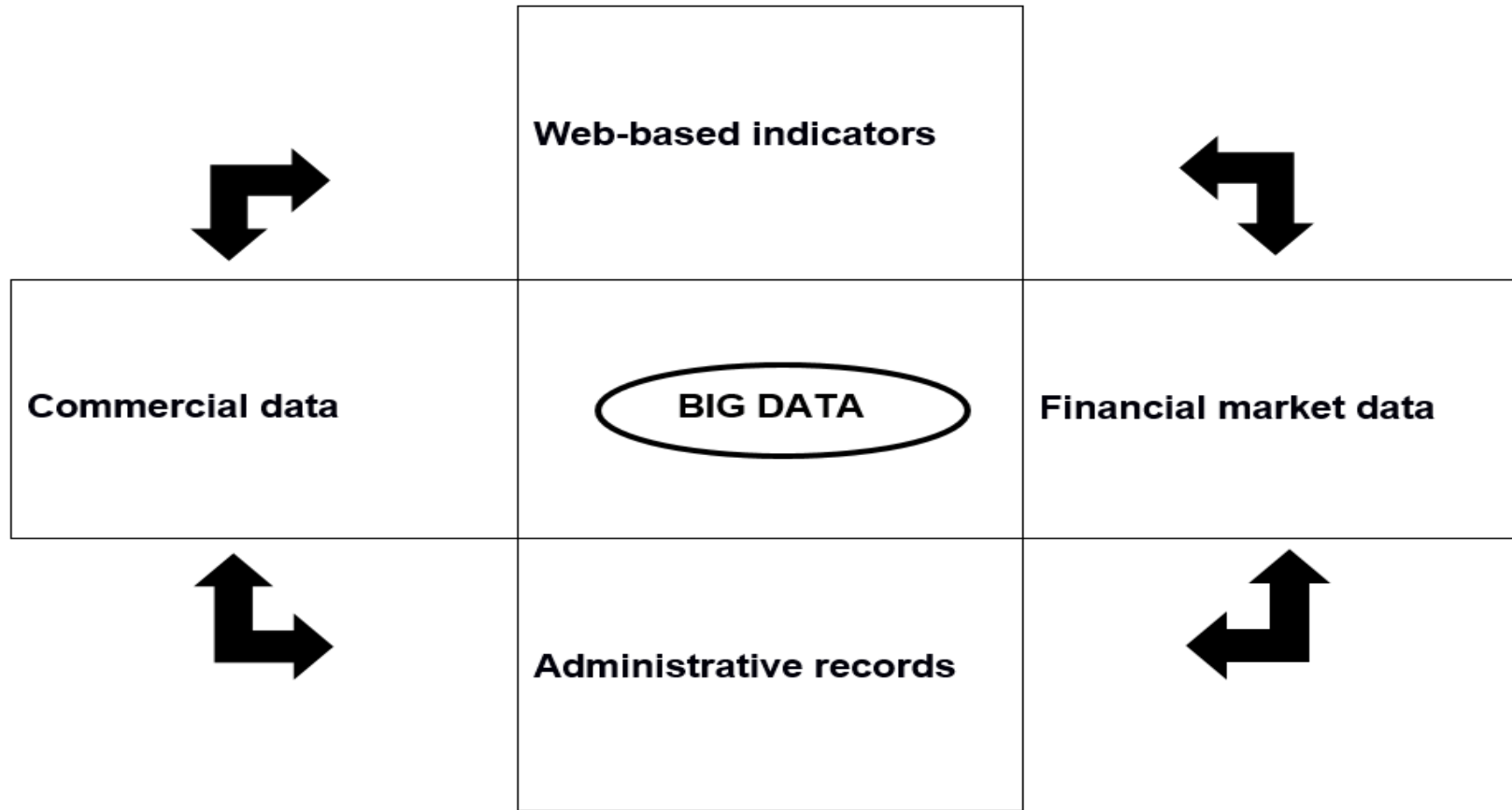
1. Big Data Information for Central Banks



Financial big data:

Crisis impact + Ability to process text information + Internet

1. Big Data: 4 types of “Financial Big Datasets”



1. Big Data: Central Banks' Interest / Pilots

*Active role of the **BIS Irving Fisher Committee on Central Bank Statistics (IFC)** in promoting the exchange of experiences on **Central Banks' pilot projects***

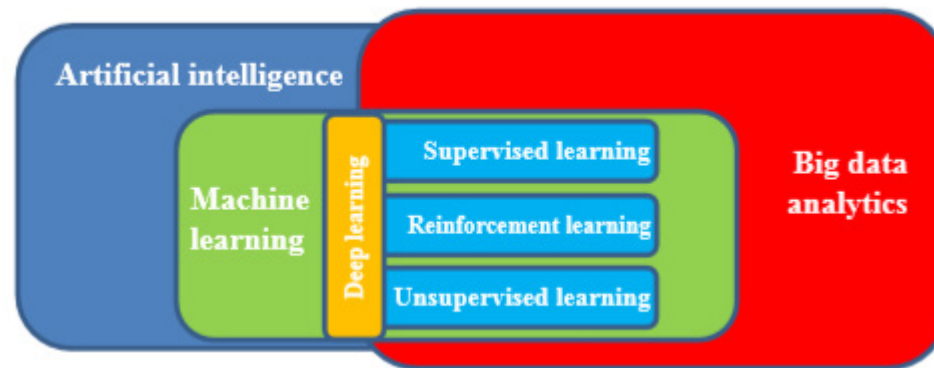
| Big data areas | Types of data-sets | Examples of projects |
|------------------------|--|---|
| Administrative records | Foreign trade operations / investment transactions | Balance of payments statistics eg tourism, exports |
| | Taxation / payroll / unemployment insurance | Employment, wages, business formation (SMEs) |
| | Central balance sheet offices | Performance vulnerabilities assessment |
| | Loans registers | Measurement of credit risk, FX exposures |
| | Financial market supervisors | Network analysis, exposures |
| | Public financial statements | Corporate balance sheet, group-level supervision |
| Web-based indicators | Financial market activity indicators | Payments systems, Trade repositories |
| | Internet clicks | Google searches |
| | social networks | Confidence & economic sentiment |
| | Digitalised content / text | Policy communication , analysis of expectations |
| | Websites' scraping | Various uses |
| | Job portals | Employment / activity |
| Commercial data-sets | Prices posted directly on websites | Measure specific components of the CPI, PPIs, Inflation nowcasting / forecasting, Pricing strategy analysis |
| | Real estate agencies | House price indices |
| | Credit card operations | Payments patterns, Tourism |
| Financial data-sets | Mobile operators | Mobile positioning data (eg travelers'), Financial inclusion |
| | Geo spatial information | National statistical system tasks |
| | Credit institutions | Balance sheet exposures, Investor behaviour/expectations |
| | Settlement operations | Operational risks, Market functioning |
| | Securities issuance | Security-by-security databases |
| | Market liquidity | Bid/ask spreads |
| | Custodians records | Securities holding statistics |
| | Tick-by-tick data | Real-time analysis of financial patterns |

Source: IFC.

2. Big Data Analytics: A bird's-eye view

- **New techniques to work with big data**
 - **“Big data analytics”**
 - General analysis of big datasets
 - **“Artificial Intelligence” (AI)**
 - Computer systems performing tasks traditionally requiring human intelligence

A schematic view of AI, machine learning and big data analytics

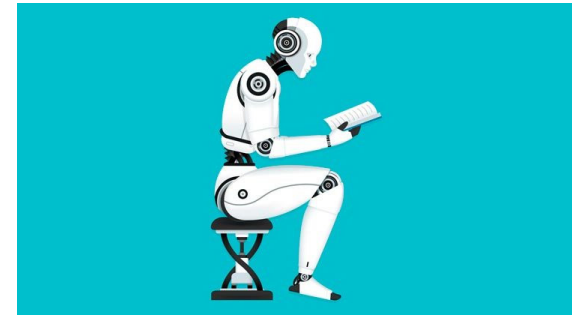


Source: Financial Stability Board: *Artificial intelligence and machine learning in financial services - Market developments and financial stability implications*, November 2017

2. BD Analytics: (i) Machine Learning (ML)

- **Design of actions to automatically solve calculations**

- Optimising algorithms, with limited human intervention
- **Classification**: finding similarities
- / **Clustering**: regrouping individual data points



- **Various types of ML techniques**

- **Supervised**: use of a “training dataset”
- **Unsupervised**: no identification available ex ante
- **Reinforcement** learning: additional feedback eg human expert
- **Deep learning**: based on artificial neural networks
 - Applied to eg unstructured data
 - Multiple layers of algorithms

2. BD Analytics: (i) ML (cont'd)

- **Multiple ML use cases for statisticians dealing with big datasets**

- Database **matching**
- **Outlier** detection
- **Imputation** of missing values
- **Forecasting** / nowcasting
- **Macro-micro** reconciliation

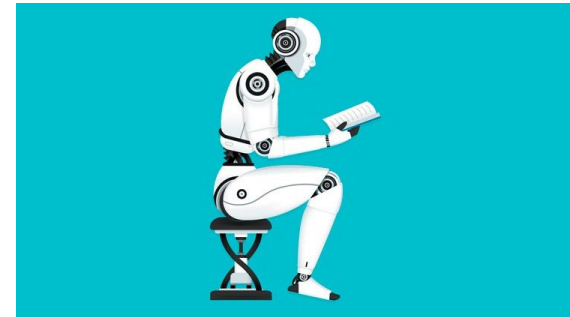
→eg to estimate granular patterns based on aggregate information

- **Classification**

→eg checking whether firms are correctly classified by economic sectors

- **Estimations** based on sub-samples

- Etc.



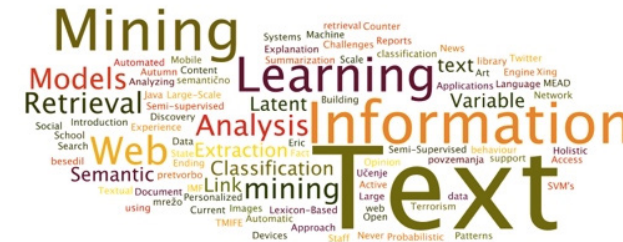
2. BD Analytics: (ii) Text Mining

- **Approach**

- Quantitative analysis of **semantic information**
- **"Unstructured"** (not quantitative)

- **Types of text**

- **Internet based-text**
 - eg blogs, social medias
- Also **"traditional"** text that can be digitised
 - eg Companies' financial reports / Authorities' statements / Press



2. BD Analytics: (ii) Text Mining (cont'd)

- **Process**

- Text **preparation**, cleaning etc
- Algorithms to derive a **vectoral representation**
 - eg topics characterised by key words

- **Multiple use cases**

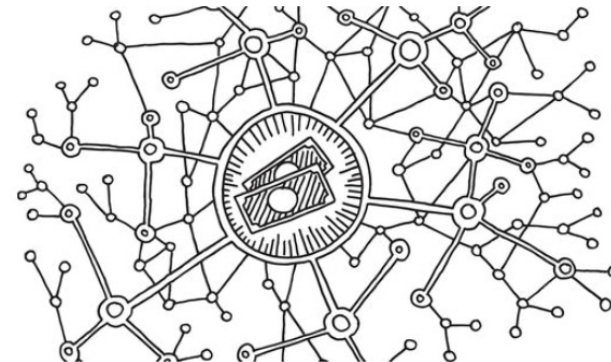
- Finding **similarities** between texts
- **Characterising** message content
 - Based on the appearance of specific terms
 - Taking into account the distance between semantic words
- **Highlighting specific text**
 - Detection of fraud/misconduct



2. BD Analytics: (iii) Network analysis

- **Approach**

- **Graphical techniques** to analyse relationships in the network
- Detection of **specific nodes** connected in the system
 - "Centrality"
- **Grouping** of "similar" nodes
 - "Community detection"
- Characterisation and analysis of the **connections**
- **Simulation** of the system functioning



2. BD Analytics: (iii) Network analysis (cont'd)

- **Multiple use cases**

- **Top-down** assessment of system-wide risk

- eg external shock

- **Bottom-up** assessment of systemically important nodes

- eg G-SIBs

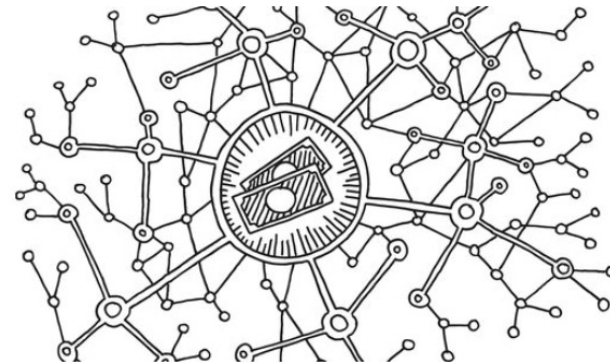
- **Network** features

- eg transmission channels

- Areas of risk concentration

- Role of **specific elements** in the system

- eg CCPs



3. Opportunities: a Central Banking Perspective

- **Multiple central bank tasks**

- Statistical production
- Monetary policy
- Micro prudential supervision, depending on countries
- Macro prudential supervision & financial stability analysis
- Payments systems, financial inclusion
- ...

- **4 main opportunity areas**

- Data
- Forecasts
- Markets monitoring
- Risk analysis



3. Opportunities: (i) More and Better Statistics

- Support of **current official statistical production**
 - Use of organic data that are a by-product of other activities
 - Available rapidly, ease of collection, low reporting burden
 - Improved statistics
 - Specific valuable information eg web scraping of certain prices
 - Substitute when data missing
- **Timelier publication**
 - Higher-frequency indicators
 - eg web searches, daily CPI
 - Rapid time release
 - Advanced compared to “hard”, lagging indicators



3. Opportunities: (i) Better Statistics (cont'd)

- **Complement “traditional” macroeconomic statistics**

- **Soft** indicators

- eg sentiment, uncertainty indicator

- **True** behaviours

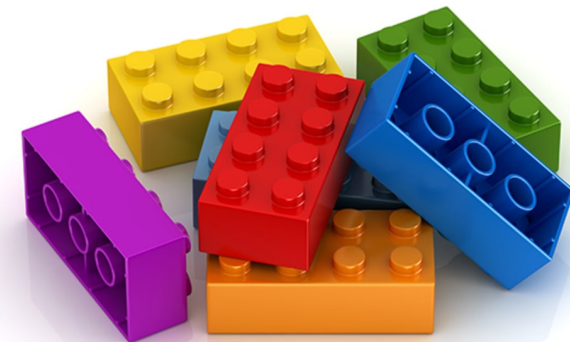
- Organic data can be less intrusive than surveys

- **Micro-macro** integration

- Macro aggregates compiled from granular “lego bricks”

- **Distribution** information

- Generally missing in the SNA framework



3. Opportunities: (ii) Forecasts

- **Support for various forecasting exercises**

- Use of **real-time** indicators for forecasting the past / present

- Nowcasting, advanced estimates

- Enhanced **short-term projections**

- Better lead time

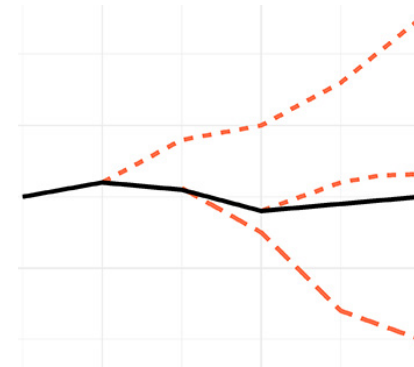
- Updating possibilities as incoming data arrive

- **Complement “traditional” forecasts** by detecting

- Non-linearities

- Abnormal patterns

- Turning points



3. Opportunities: (ii) Forecasts (cont'd)

- **Making use of more information**

- **Large variety**

- Web-based eg Google Trends, web searches
- Financial market indicators eg daily prices
- Qualitative information



- **Selection** of explanatory variables

- Greater choice, more potential data can be considered ex ante
- Flexibility, depending on the time horizon eg short- vs long-term

- **Use of ML algorithms** to

- Deal with / select (large) incoming data flows
- Adjust the forecast depending on the actual date of the exercise

3. Opportunities:(iii) Markets Monitoring

- Dealing with a **wealth of financial markets information**

- **BD analytics** to

- Process huge volume of quantitative data
- Test (potentially large) determinants of market prices



- Incorporation of **less structured information**

- Sentiment indicators eg uncertainty, optimism etc
- Role of expectations among market participants eg for policy rates

3. Opportunities:(iii) Markets Monitoring (cont'd)

- Better **understanding of markets' functioning**

- **"Animal spirits"**

- Factors driving sentiment / expectations

- Possibility to **"zoom in"** for specific events

- eg removal of the EUR/CHF floor in 2015

- Understanding of financial **markets' patterns**

- eg intra-day liquidity by using tick-by-tick data

- **Non-linearities**

- Large vs smaller market movements

- Risk-on vs risk-off moments



3. Opportunities: (iv) Risk Assessment

- **Core activity** of central banks, related to multiple areas
 - **Functioning** of financial markets
 - Monetary transmission mechanism
 - Collateral valuation
 - **Micro prudential** supervision
 - Individual firms' financial position
 - Likelihood of default
 - **Macro prudential** tasks and financial stability objectives
 - Systemic risk measurement & contagion effects
 - Incorporate non-quantitative information
 - Understand the functioning of the financial system



3. Opportunities: (iv) Risk Assessment (cont'd)

- **Technology and regulatory/supervisory processes**

- **Regtech**

- Technology support for financial institutions as regards regulatory monitoring, reporting, and compliance

- **Suptech**

- Provision of related insights for supervisors

- **Extracting information from data related to firms' activities**

- Dimensionality reduction algorithms to extract knowledge

- Classification algorithms to focus on specific groups

- Visualisation techniques to assess systemic risk / contagion

- **Detection of unusual data patterns**

- Text mining techniques to flag misconduct

- Odd patterns signal build-up of vulnerabilities



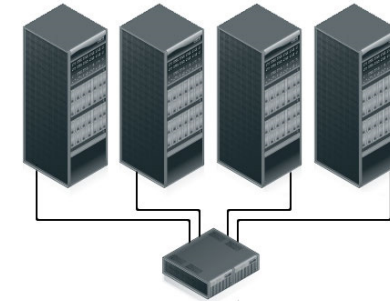
4. Organising Big Data Work

- How should central banks **organise themselves?**
 - Making the most from BD opportunities
- **Two areas of interest**
 - Organisation of the underlying data and IT infrastructure
 - Big Data vs High Performance Computing (HPC) **platforms**
 - On-premises vs cloud-based **IT services**
 - Use of appropriate techniques for further processing and analysing the new information collected
 - **Business Intelligence** (BI) systems
 - **Advanced Analytics** (AA) – “data science”



4. Organising Work: (i) Computing Platforms

- Two related needs:
 - BD Platforms for storing & processing of **large datasets**
 - Including unstructured data
 - **HPC infrastructure** for faster processing
 - In-depth statistical analysis and complex simulations



- Several ways **in practice for HPC**
 - Computing power aggregation (clusters of computing nodes)
 - Jobs to be scheduled vs flexible unified analytics engine
 - Quantum computing

4. Organising Work: (i) Computing platforms (cont'd)

- **What?**

- Various **layers**

- For data storage, processing, querying, accessing, analytics...

- **Dimensions** to be considered when adopting technologies

- eg technology trends, complexity, cost, performance, reliability, operating model, risks

- In response to clearly identified business needs

- **How?**

- Various components available from **open source community**

- Risk of **heterogeneous** software composition

- eg incompatibilities, performance issues, security risks

- **Packaged** Hadoop software available in the market

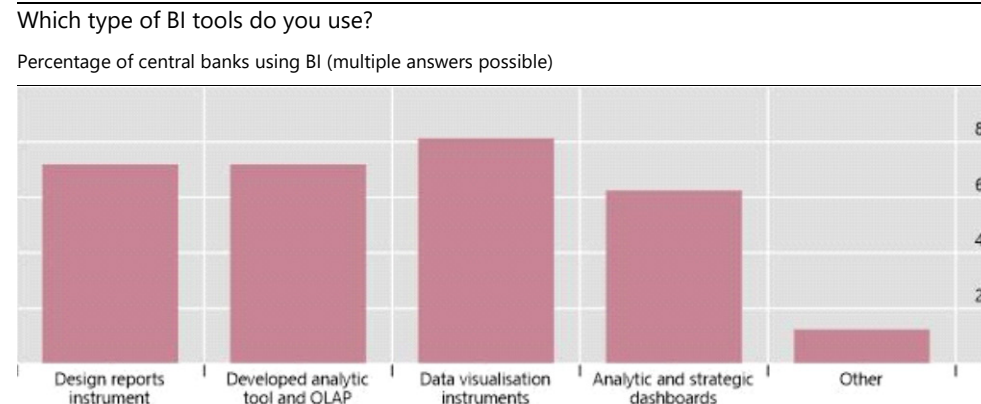
4. Organising Work: (ii) Data Analysis

- **Various needs**

- BI: technology-driven methods & techniques to manage/disseminate data
 - In order to inform business decisions
- AA: use of techniques & tools (eg ML) for examining data
 - Beyond BI tools, with “ad hoc” programming languages

- **IFC survey (2019) of central banks**

- Large range of BI tools used; High satisfaction



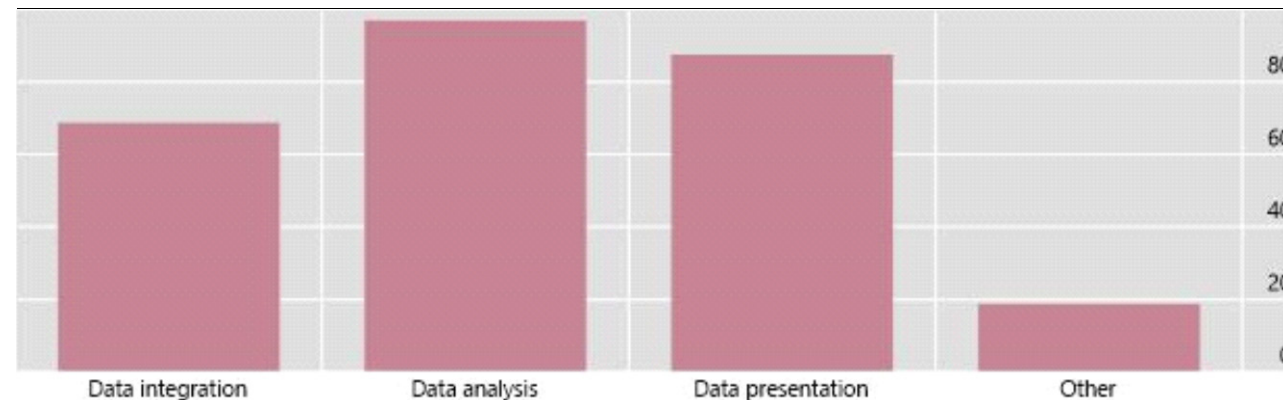
Source: IFC survey on BI systems in central bank statistics (2019).

4. Organising Work: (ii) Data Analysis (cont'd)

- BI key to support
 - **Statistical production** tasks
 - **Decision-making** processes
 - eg collecting, storing, integrating, searching and querying, reporting, analysing, presenting, publishing and visualising data
 - **Data analytics**
 - in practice complemented by “ad hoc” programming languages

For which function do you use BI tools?

Percentage of central banks using BI (multiple answers possible)



Source: IFC survey on BI systems in central bank statistics (2019).

4. Organising Work: (ii) Data Analysis (cont'd)

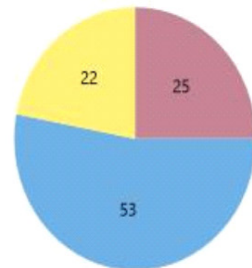
- Strong focus on tools for data visualisation
 - Key for central bank communication
 - Push for web/cloud-based dissemination strategy
 - To allow users to dynamically interact with publications
 - Move from monolithic to modular systems

Increasing push for web dissemination...

... and interactive data visualisation

What is the primary medium for the publication/dissemination process in your organisation?

Percentage of central banks using BI

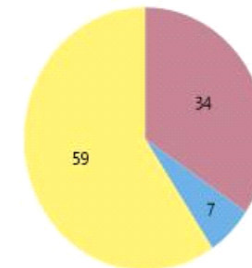


Print (pdf / Word content) Web (html content) Both web and print

Source: IFC survey on BI systems in central bank statistics (2019).

Do you disseminate static and/or dynamic graphical content on your corporate website?

Percentage of central banks using BI



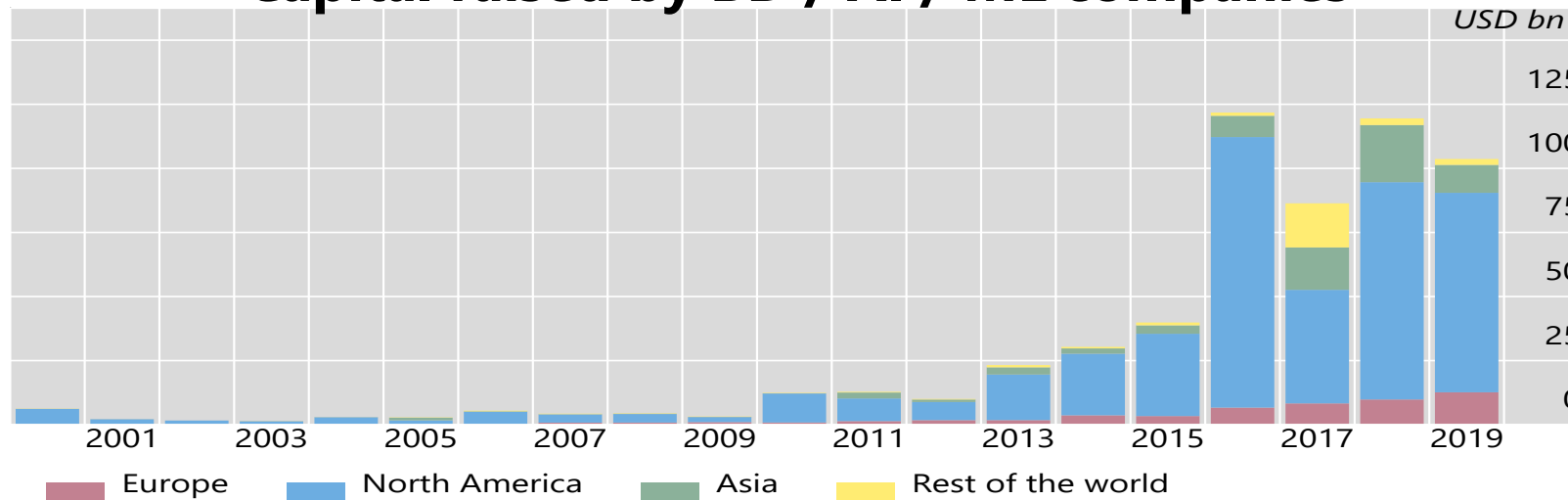
Static Other Both static and dynamic graphical content

Source: IFC survey on BI systems in central bank statistics (2019).

4. Organising Work: (iii) What's Next?

- Technology **landscape evolving** rapidly
 - Is it the **end**? Huge investment in BD firms in recent years
 - When will the technology **mature**?
 - **Risk** of investing too much / too early?

Capital raised by BD / AI / ML companies



Source: PitchBook Data Inc.; BIS calculations. Private Equity, Venture Capital and Mergers & Acquisitions excluding buyouts; 2019 figure annualised based on data available up to end Oct-19.



5. Key Big Data Challenges: (i) Variety

- In practice **various & heterogeneous “financial big data”**
 - Usually not designed for a direct statistical purpose
 - Indirectly, data exploited for addressing statistical needs
- Several **issues for statisticians**
 - Compilation (or acquisition) costs
 - Quality: attributes values / representativeness
 - Identifiers: connecting datasets / coverage of large entities



5. Key Big Data Challenges: (ii) Complexity

- Micro-level data universe is **complex and evolving**
 - Use of specific sources depends on policy questions
 - Example of payment systems :
of interest for supervision / tourism analysis
 - Interaction between data available, specific policy needs and actions (feedback loop)
- **Transforming data into relevant information** for policy
 - Public authorities at the beginning of making sense of these data:
connecting the dots, not just collecting them (Caruana, 2017)
 - “Smart data”: treatment of the raw, “organic” data is key



5. Key Big Data Challenges: (iii) Time Dependency

- **Information needs evolve over times**

- The **financial system changes**... not least **due to policy**
 - *Risk of “unknown unknowns” always there*
 - *Limitations of IA-type techniques based on past patterns*
- Assessment of how **fragilities are building up** typically rely on aggregated statistics to spot “abnormal patterns”
- In contrast, **resolution work** after a financial crisis will request much more timely and granular information
 - *Rough aggregates often OK to indicate rising fragilities*
 - *More granular data needed after a crisis (Carstens, 2018)*



6. Policy Issues: Handling Big Data...

- **Resources**

- IT, staff, security
- Proper arrangements for managing data

- **New statistical **production chain****

- Comprehensive information management process
- Data governance

- **Reputation risk** when handling the data

- Confidentiality and trust
- Ethical issues



6. Policy Issues: ... Using Big Data

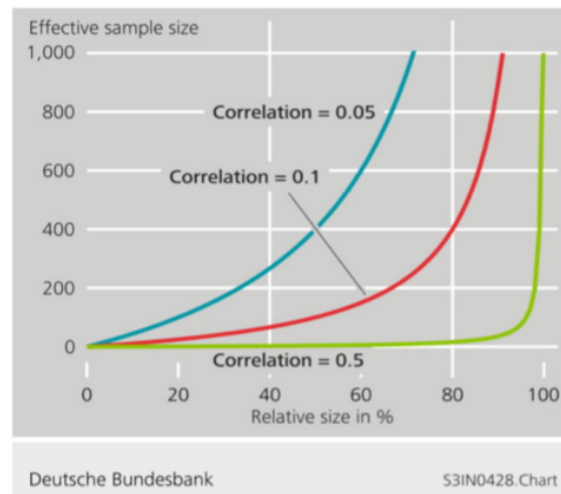
- **Accuracy?**

- **Unknown coverage bias**

- BD set (eg Twitter) can differ from the structure of the population
- Misperception that large amounts of data is a guarantee of accuracy
- Large self-selected samples can be of low quality vs probabilistic ones

- Challenges reinforced by the **opacity** of “black box” calculations

- For instance issue of ML algorithms trained on biased datasets



Depending on the statistical assumptions made, one may need to cover at least 50% of the entire US population to “beat” a simple random sample covering 400 individuals

Source: Mehrhoff J (2017): Central banks' use of and interest in "big data", IFC Bulletin no 44, September.

6. Policy Issues: ... Using Big Data (cont'd)

- **Reputation risk?**

- **Public use** of BD sources provides legitimacy...
- even if not governed by **Fundamental Principles of Official Statistics**

→ No 3: "(...) to present information according to scientific standards on the sources, methods and procedures of the statistics".

- **Social costs of misguided policy decisions?**

- Recent news of errors in Sweden official statistics
- Jobless rate revised by almost one percentage point

→ "What has happened is a catastrophe. Public statistics are the basis for economic, political decision making, and this is a sad story"

(Riksbank Deputy Governor H Ohlsson, Bloomberg 15/11/2019)

6. Policy Issues: ... Using Big Data (cont'd)

- **Altering decision-making?**

- Bias towards **responding to news?**

- Risk of encouraging shorter horizons for taking policy decisions?

- Incentives for **fine-tuning policy communication?**

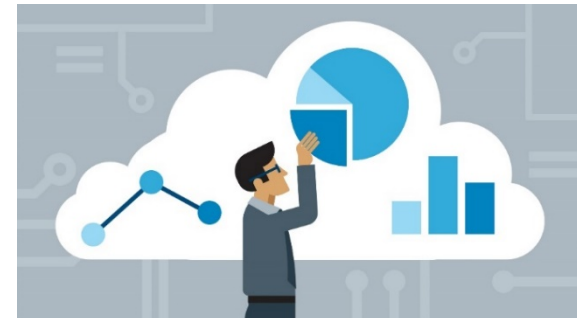
- Risk of a “monkey in the mirror” mimetic loop through BD (Coere, 2017)

- **Communicating on “black box”** calculations?

- A science fiction exercise

“Twitter messages collected since we last met in September indicate that the labor market remains strong. Our machine learning algorithms running on Google Trends data show that economic activity has been rising at a moderate rate. Consequently, the target range for the federal funds rate is adjusted to...”

(US Federal Open Market Committee, October **2050**)



Thank you!!

Questions?

bruno.tissot@bis.org

IFC.secretariat@bis.org

Selected Publications

- *IFC Report on business intelligence systems and central bank statistics, Oct 2019*
- *The use of big data analytics and artificial intelligence in central banking, IFC Bulletin No 50, May 2019*
- *Big Data, IFC Bulletin, No 44, Sept 2017*
- *IFC Report on central banks' use of and interest in "big data", Oct 2015*

Available at <https://www.bis.org/ifc/publications.htm?m=3%7C46%7C94>





DATA DRIVEN FINANCIAL STABILITY

#datadrivenstability
#financialstability

2–3.12.2019

OPPORTUNITIES
AND CHALLENGES
IN BIG DATA

