EUROPEAN CENTRAL BANK
EUROSYSTEM

Aurel Schubert
Director-General Statistics

Data as a core central banking asset – The strategy of the ECB

Big data workshop

Sveriges Riksbank, 9 September 2015

Overview

1. Big Data and Central Banking
2. Data: a core central bank asset
3. The strategy of the ECB: Integrating and leveraging available data
4. Integration of granular data collections

* The views expressed are those of the author and do not necessarily reflect the position of the ECB.

EUROPEAN CENTRAL BANK

EUROSYSTEM

TÜRKİYE CUMHURİYET MERKEZ BANKASI

Irving Fisher Committee on Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Survey on Big Data

Preliminary survey results on

"Central Banks' use of and interest in big data"

Should central banks statisticians play a role, contribute and develop the concept of "Big data" or is it only a temporary phenomenon ?

IFC Survey on Big Data

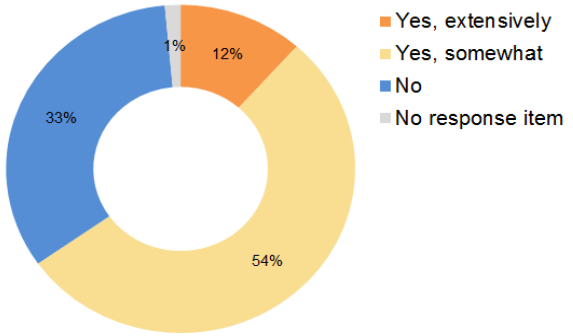
- IFC on-line survey was launched in January 2015
- Survey was answered by **69** IFC members and monetary authorities (**83%** response rate)
- Aim was to assess central banks' experiences and interest in exploring big data related to financial and economic topics of interest to central banks
- The concept is not clearly defined. Different understanding and perception
  - While some central banks may consider single sourced data, such as granular "administrative data" (credit registers) or micro "financial information data" (security-by-security datasets) as "big data"; others
  - may take a more holistic approach of complexity of combining size, formats and sources mainly focused on private web-sources
- **Big data is not just about large data sets.** The 4 Vs (IBM) relates to Volume, Velocity, Variety and Veracity.

*Draft paper "Big data – The hunt for timely insights and decision certainty. Central banking reflections on the use of big data for policy purposes" P. Nymand- Andersen (2015)

Central banking findings

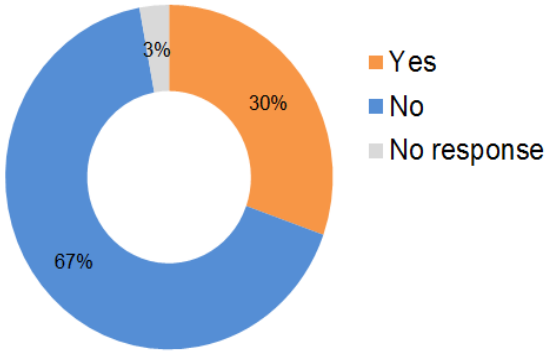
1. At senior policy level, there is **significant interest** in big data within the central banking community

66%



2. Despite the interest, central banks have **limited experience** in use of big data

30%



3. Central banks are interested in **cooperating together** on specific topics to explore the usefulness of big data

71%

Address key statistics topics

- Relevance of sources
- Quality
- New indicators/statistics
- Statistics methods
- Sampling & representativeness

Central banking findings

4. Big data can be useful for central banking purposes and is perceived as useful for supporting central banking policies

→ Macro-economics
Forecasting/nowcasting
Financial stability
Business cycle analysis
Supervisory purposes (micro-economics)
Sentiments and behaviour indexes
Improve quality

5. Central banks are interested in cooperating together in a structural approach

→ Setting up
- a road map
- identify joint pilot projects
- sharing experience

6. Explore synergies to overcome barriers and challenges

→ Barriers and challenges
- Resources and costs
- Skilled human capital
- IT constraints

Examples of future central banking projects

- **Databases within central banks and national statistics offices** → **Micro-economic behaviour, modelling of SME, Network analysis, improve quality controls**
- **Public authorities sources** → **Analysis of micro-statistics on loans, derivatives, government balance sheet Transactional data for network analysis**
- **Internet based data** → **Macro-economic indicators (such as unemployment), Economic sentiments House indexes, consumer price dynamics, Improve web-based quality**
- **Databases of financial institutions** → **Credit data and risks, loans to loans, securities holdings, mutual funds data, investor behaviour/expectations, Financial markets liquidity and patterns**
- **Media and social networks** → **Perception of central banks**

IFC way forward

a) **to define and contribute to a central bank "big data" roadmap**

b) **Share and contribute to selected big data pilot projects**

i. **administrative dataset** (e.g. corporate balance sheet data)

ii. **web search data set** (e.g. Google type search info)

iii. **commercial dataset** (e.g. credit card operations)

iv. **financial market data** (e.g. high frequency trading, bid-offer spreads)

ECB & Google search data

- ECB receives weekly data from Google search machines in a CSV file
- The data is an index of weekly volume changes of Google queries by geographic location and category
- Google search data is more accurate and uses much larger samples than Google Trends
- Google search data includes the following 14 countries: Austria, France, Italy, Slovenia, USA, Belgium, Germany, Netherlands, Spain, United Kingdom, Denmark, Ireland, Portugal, Sweden
- Google search data includes 26 categories and 269 subcategories
  E.g. Finance is a category and Banking is a subcategory
- The data are normalised starting at 1, one can see the relative change in Google searches by category but nothing can be said about the absolute search volumes

ECB & uses of Google search data/big data

- Findings of the ECB Statistics Paper Series released on this topic
  - "Nowcasting GDP with electronic payments data" by John W. Galbraith and Greg Tkacz
    - Electronic payment transactions and cheques can be used to formulate nowcasts of current gross domestic product growth
    - Assesses this technique and finds that debit card transactions contribute most to forecast accuracy
  - "Social media sentiment and consumer confidence" by Piet J. H. Daas and Marco J. H. Puts
    - What is the relationship between the changes in Dutch consumer confidence and the Dutch public social media?
    - The changes in social media sentiment have the same underlying phenomenon as Dutch consumer confidence
    - Could be used as an indicator for changes in consumer confidence and as an early indicator

ECB & uses of Google search data/big data (cont'd)

- Forthcoming publications by the ECB staff

  - "Big data – the hunt for timely insights and decision certainty: Central banking reflections on the use of big data for policy purposes" by Per Nymand-Andersen

  – Big data might lead to new economic theories with statistical algorithms applied to multiple big data sources from various disciplines finding new causations
  – Big data as opportunity for the central banks to apply expertise in testing existing and new models, data sets and theories; to explore new data sources and to obtain new, timely knowledge from the feedback loop between monetary policy and market reactions
  – Central banks need to start by taking a structural approach to systematically testing the use of non-official big data sources

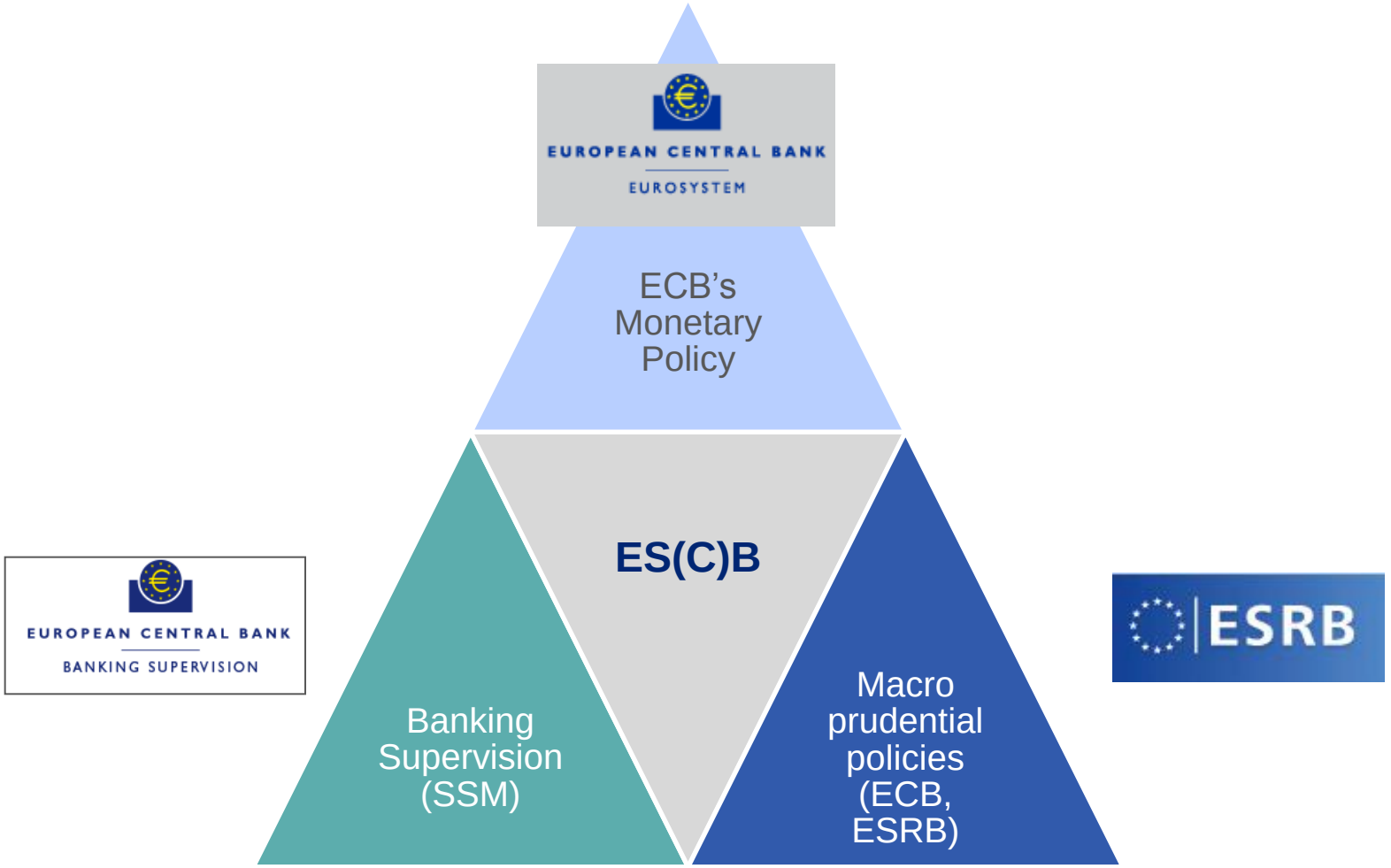
  - "Predicting euro area unemployment rate using Google data: Central banks interest and use of big data" by Per Nymand-Andersen and Heikki Koivupalo

  – explores how Google search data has been used for macro-economic and financial purposes within the literature
  – tests how Google search data can be used for predicting the euro area unemployment rate in advance of the official statistics
  – demonstrates that applying Google data within a simple model can improve the predictability of the euro area employment rates
  – Google search data used as part of central banking statistical and analytical toolkit?

Overview

1. Big Data and Central Banking
2. Data: a core central bank asset
3. The strategy of the ECB: Integrating and leveraging available data
4. Integration of granular data collections

Rapid expansion of needs for policy-making



Challenges

Reports and data collection systems have been developed

- At different times by different teams
- Following different concepts and legal foundations
- Serving different requirements, which itself changed over time

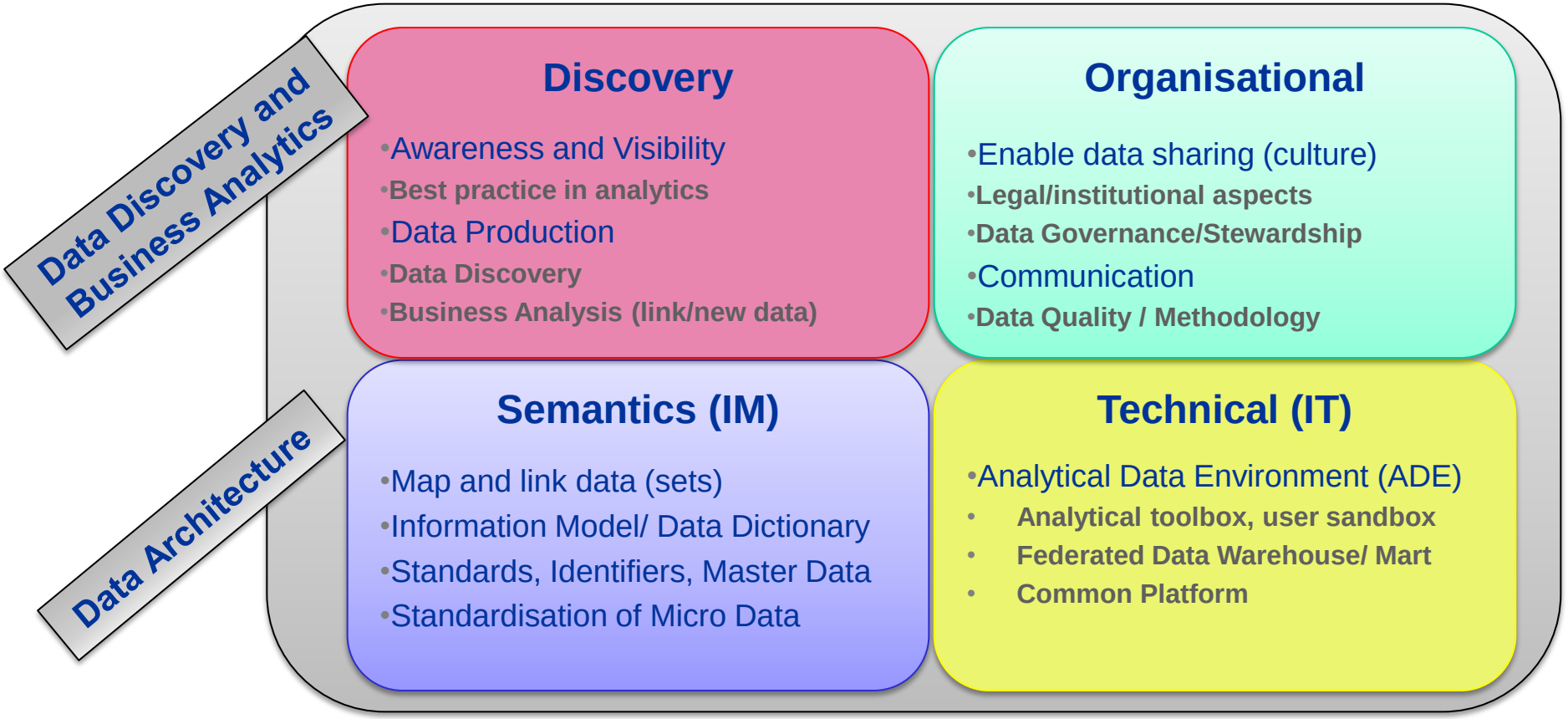
For historical reasons, information about the same phenomenon is collected several times in different reports referring to *slightly different*

- definitions
- filtering and aggregation rules
- derivation algorithms and valuation principles

The system presently is in place is compared with **stove pipes** or **silos**

The strategy to get through this situation is **integration** and **long term harmonisation**

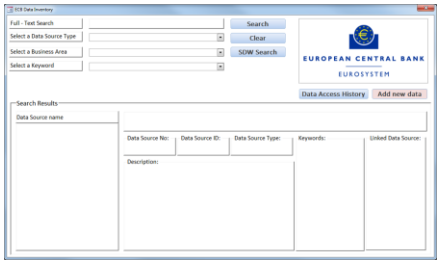
ECB integrated Data Management (EiDM)



One integrated View on ECB data

Data Inventory: ECB wide access via the ECB intranet

- ***Description* of ≈ 500 ECB datasets**
- **Accessed every two business days on average**



- **Benefits:**
  - Enables usage and leveraging existing data
  - Search for ECB data in both ***user terms and keywords***
  - **Finding *"new" data*** and support in access
  - Enables ***new analyses*** **and research**
  - Supports ***policy making*** **and supervision**
    - ABS, SSM Data Sharing, ESMA cooperation
  - Little effort **to maintain**

Common Data Dictionary and Semantics

Interoperability of Information and Data → Common processing of information (regardless of its origin)

What is required?

Common semantics → "Central Banks" = "the central bank"

AU = AUT, ECB = EZB = BCE

What is a non-performing loan?

"People can't share knowledge if they don't speak a common language" [1]

1. Big Data and Central Banking
2. Data: a core central bank asset
3. The strategy of the ECB: Integrating and leveraging available data
4. Integration of granular data collections

Data Quality requires holistic approach: from *Source* to *Final use*

"The ECB has every interest to facilitate and promote *integration and standardisation* also on the *"input side",* in the *internal systems of the banks*, for only this will ensure *coherent* information."

Mario Draghi, Seventh ECB Statistics Conference (15/10/ 2014)

1. Banks' Integrated Reporting Dictionary (BIRD): Common language with the industry

- A logical description of data and transformation rules
- Common reference, to which reporting requirements can refer

2. ECB Single Data Dictionary (SDD): Common language within the ECB

3. European Reporting Framework (ERF)

- **Vision**: single and integrated ERF for the ECB and EBA (ITS), ultimately for the EU
- "one" report replacing the multiple, overlapping reports
- BIRD will allow "short" term reports as the data and data structure hold is known

BIRD description

- The BIRD aims to provide a standardised model for organising the banks' internal data warehouses in an integrated way, and the transformations to the reports that the banks need to transmit
- The BIRD consists of **documentation**, it is **not an IT tool**
- The application of the BIRD by the banks is strictly **voluntary**
- **Responsibility** for the correctness of the data remains with the banks
- The BIRD is **not adding reporting requirements**
- Advantages:
  - better quality data at source
  - more efficient, and, in the longer run, less costly report production
  - more consistent and harmonised data
  - univocal interpretation and clarity of regulations

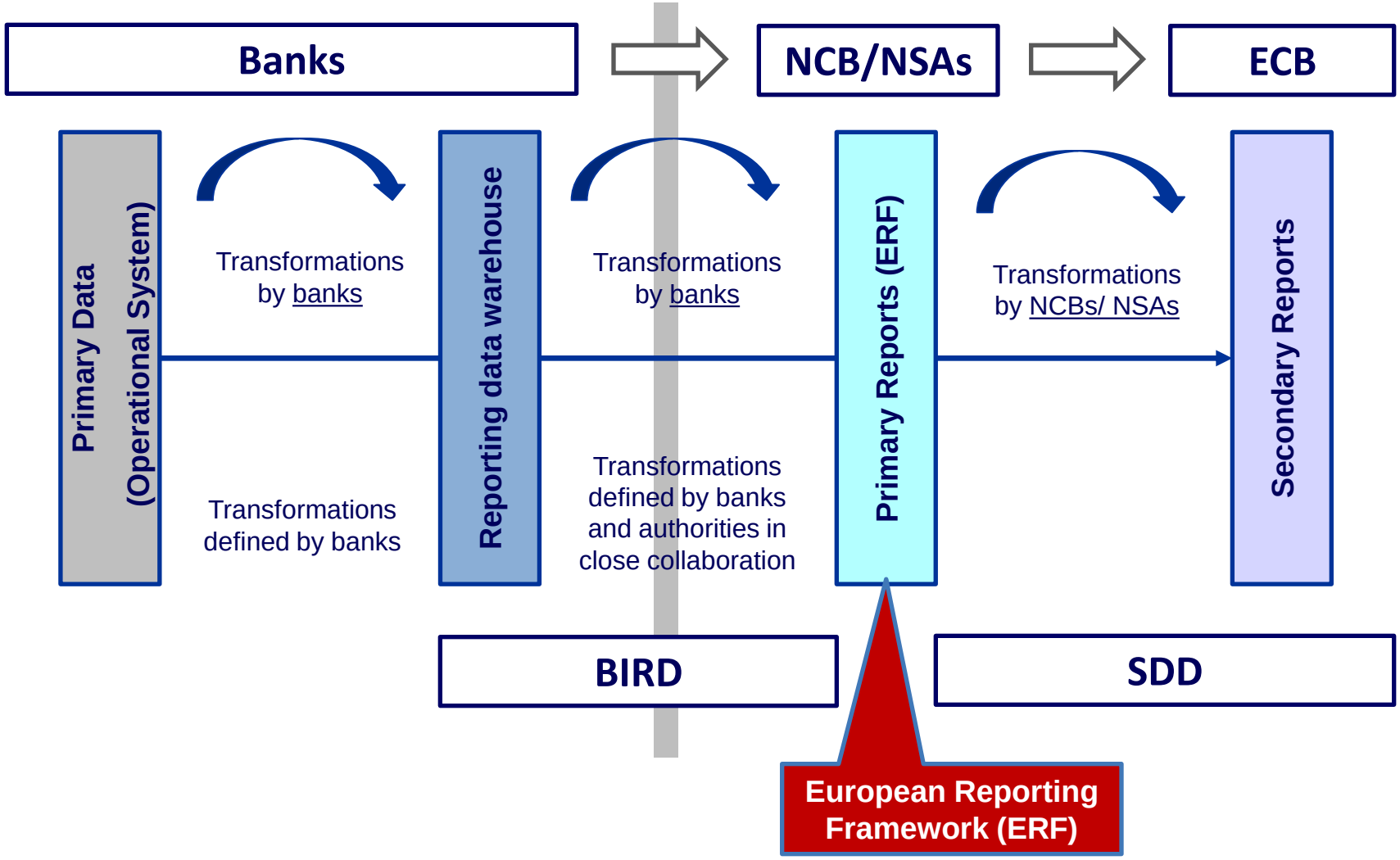
Single data dictionary (SDD)

- The SDD is a single data dictionary, developed by the ECB/DG-S
- Goal: methodological and semantic integration of existing European reporting frameworks
- How:
  - by creating a **common information model and methodology** (SMCube), able to read and process all relevant file exchange formats;
  - by creating **clear, non-overlapping definitions of the data** (SDD) - the meaning of the data will be reconciled across several regulatory frameworks.
- Benefits:
  - reconciliation done once and available for all users;
  - enables the creation of an integrated statistical and analytical system;
  - unified data definition, collection, processing, dissemination, and its final use.

European Reporting Framework (ERF)

- Integrated and harmonised cross-county reporting scheme for banks, covering most reporting requirements of ECB and EBA.
- Best practice for collecting banks' data for different purposes, **avoiding duplication** across various reports.
- Basis for **harmonised production of secondary statistics** via unique transformation rules.
- Project in its early stages: design, adoption and timeline of eventual implementation are "work in progress".
- ERF key facts and information (June 2015) available at:
  http://www.ecb.europa.eu/pub/pdf/other/europeanreportingframeworkkeyfactsandinformation062015.en.pdf

The Role of BIRD, ERF and SDD



1. Big Data and Central Banking
2. Data: a core central bank asset
3. The strategy of the ECB: Integrating and leveraging available data
4. Integration of granular data collections

The example of Money Market Statistical Reporting

- The Money Market Statistical Reporting (MMSR) aims to provide a standardised reporting of daily transactions on the money market (unsecured, secured and derivatives) in an automated way
- The MMSR has been prepared with the reporting agents via multiple contacts, in particular to prepare Reporting Instructions and XML schemas
- The Reporting Instructions have been approved as an ISO 20022 standard
- The MMSR abides with the statistical data dictionary and uses standard referential: e.g. Register (RIAD) for the identification of reporting agents and their counterparts (using the Legal Entity Identifiers)
- <u>Advantages:</u>
  - better quality data at source (for large daily volumes: ca 50,000 reports per day)
  - efficient data transmission (after initial teething problems)
  - timely production and delivery to users (8:30 a.m. the next day)
  - standard/harmonised data, shareable with other systems

Analytical credit datasets

- Several National Central Banks traditionally hold “Central Credit Registers”, e.g. for supervisory purposes or to assist the banking industry in their assessment of credit risk
- Such Credit Registers are rather heterogeneous due to their history and main functions
- Harmonisation in coverage, definitions and concepts and standardisation of transmission features (e.g. frequency, timeliness, instruments covered) help increase the relevance, also for benchmarking
- The Analytical credit datasets will be a distributed system
  - Allowing NCBs to keep with national specificities and needs; also
  - Abiding with the statistical data dictionary and using standard referential: e.g. Register (RIAD) for the identification of reporting agents and their counterparts
- <u>Advantages:</u>
  - better quality data at source, consistent and harmonised big dataset (Terabytes)
  - efficient and stable system over time (lower production cost)
  - Supporting a wide range of user needs, in an agile way

Conclusion

Data: **a core central bank asset**

- Statistics/Data Management to keep up with the ever increasing inflow of data
- From stove pipes to **data integration** from source to final use
- Integrated granular data provides new insights, enables new macroeconomic forecasts and more **informed (economic) policy making**.
- **Big Data (r)evolution:** Data collection and analysis done in a single place, with common standards - just one keystroke away for the users

With **data integration**, such seemingly (r)evolutional ideas might become a reality …

Thank you for your attention

Any questions?