Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

Ninth IFC Conference on "Are post-crisis statistical initiatives completed?"

Basel, 30-31 August 2018

ECB data for analysis and decision-making: data governance and technology[1]

Emily Witt and Jannick Blaschke,
European Central Bank

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

ECB data for analysis and decision-making: Data Governance and technology

Emily Witt, Jannick Blaschke

Abstract

Data are necessary to enable analysis and support decision-making. But the use of data raises many questions: How can we best exploit the wealth of data available within the ECB whilst safeguarding confidentiality? And how can we best leverage the data expertise available across business areas? This paper outlines how the Data Intelligence Service Centre (DISC) project will provide a high performant technical platform that supports users to easily access and experiment with data and deliver new analysis in a speedy manner. But even more important is to have a good data governance structure, which fosters strategic alignment, standardisation, as well as collaboration, data and knowledge exchange across business areas.

Keywords: data governance, data-driven decision making, central banking, data access, data sharing,

JEL classification: C80, D70, E50, M14, O30

Contents

ECB data for analysis and decision-making: Data Governance and technology 1

Exponentially increasing availability of data calls for a holistic approach 3

DISC facilitates data usage, collaboration, automation 4

Strong ECB data governance ensures alignment 5

- ECB Data Committee steers data management 5
- Data Steward Group fosters collaboration and operational alignment 5
- Directorate General Statistics adjusts its organisation to new needs 6

Standardisation is essential for efficiency and integration 7

- Global identifiers will facilitate data integration and reduce reporting burden 7
- Eurosystem initiatives foster standards implementation 7
  - ESCB Integrated Reporting Framework (IReF) 8
  - Banks' Integrated Reporting Dictionary (BIRD) 8
  - The Single Data Dictionary (SDD) 9
  - Register of Institutions and Affiliates Data (RIAD) 9
  - Centralised Securities Data Base (CSDB) 10

A shift in working culture with closer collaboration is required to leverage the data potential 10
Make data fit for use whilst protecting confidentiality 10
Clear guidance for publications and reports ensures consistency 11
Create multidisciplinary teams and share knowledge 11
Example: Shared data services pilot on funding and assets of banks 12
Conclusion 12
Abbreviations used 12

Exponentially increasing availability of data calls for a holistic approach

Benoît Cœuré, Member of the Executive Board of the European Central Bank (ECB), outlined many examples how granular data can support policy making in his speech "Policy analysis with big data"[1] on 24 November 2017. He concluded that "*Central banks have made considerable progress in recent years in integrating big new datasets into their policy analysis and decision-making. Granular data collected by central banks themselves have, in particular, become an indispensable source of information for policymakers. … The potential of such data to enrich central bank analysis in the future is considerable, however, as are the challenges that come along with it.*"

This paper describes how the ECB addresses these data related challenges with a focus on technology and organisational themes.

The ECB has a 20 year history in preparing jointly with the National Central Banks (NCBs) and Eurostat macroeconomic statistics to prepare policy measures and assess their impact.

The financial crisis, and the euro area sovereign debt crisis that followed, were characterised by periods of increased heterogeneity, market fragmentation and sudden turns in economic activity. Recent data-related ECB and EU regulations paved the way for the ECB to get an encompassing view on the developments in EU financial market and instruments like loans (AnaCredit), money markets (MMSR)[2], security holdings (SHSDB), derivatives (EMIR), secure financed transactions (SFT-DS) and banks (COREP/FINREP). Such data will help to analyse diverse economic signals in a timely manner. They can be used to assess the underlying forces driving economic behaviour and understand the interconnectedness between financial institutions to better assess risks and to calibrate policy measures, e.g. by analysing how changes in the volume and price of central bank money impacted the monetary policy transmission mechanism or understanding how decisions in the asset purchasing programme impact the credit and funding behaviour of banks.

The exploitation of granular data poses various challenges to statisticians and users:

- The increasing volume and speed of the data
- The heterogeneity and complexity of the data
- Data confidentiality.

To address these challenges, the ECB applies a holistic approach, covering technology, governance, standardisation and collaboration.

1 Speech given at the conference on "Economic and Financial Regulation in the Era of Big Data", organised by the Banque de France, Paris, 24 November 2017.

2 For example, MMSR and AnaCredit are based respectively on regulations (EU) No 1333/2014 and No 2016/867 of the ECB. MMSR data contain confidential daily information on the individual euro-denominated loans in the euro money market from the 52 largest euro area banks, accounting for approximately 80-85% of the total balance sheet of euro area banks. AnaCredit will deliver loan-by-loan information, mostly on a monthly basis, on credit to around 8 million companies and other legal entities extended by about 4,500 euro area banks, comprising almost 100 different attributes.

DISC facilitates data usage, collaboration, automation

The ECB uses a Forecasting Analysis and Modeling Environment (FAME) for the production of time series and macroeconomic statistics and the Statistical Data Warehouse (SDW) for their dissemination to internal and external users.

However, FAME is not suitable for the production of granular data sets. For this reason, several IT systems were developed next to each other to fulfil specific data processing and business reporting needs resulting from the recent data-related ECB and EU regulations. The requirements of these systems were focused on answering particular business questions. Each of these systems has its own distribution channel and separate areas where users could get access to these granular data sources. Thus, the need emerged for data analysts to have a common environment to work with data in a secure and performant fashion.

Complementing the efforts by the ECB Directorate General (DG) Statistics to collect micro data and establish sound processes with the NCBs to ensure an appropriate level of quality, the ECB's IT department (DG Information Systems) established the Data Intelligence Service Centre (DISC). The DISC data platform project was initiated in March 2016 to provide a single shared platform for (i) business centric analytical capabilities, (ii) data integration and ETL services[3], and (iii) metadata management based on the Single Data Dictionary (SDD) and Data Inventory (see explanation in section *The Single Data Dictionary (SDD)*).

The DISC data platform is the central secure place for organising, storing, analysing data and related collaboration within the ECB. It aims to make large data volumes and the variety of data structures usable by employing the latest big data technologies available, such as Hadoop, Oracle Exadata, complemented with products from Cloudera and Informatica. It offers (connections to) tools such as Matlab, Python, R, SAS, Stata, Tableau for statistical analysis and visualisation purposes.

The central place for users to access data will be the DISC corporate store, which will contain the "golden copy" of relevant data from the statistical production processes and from other systems or sources for further usage by the end users. Currently, several micro data, such as statistics on the securities issued (CSDB) and the securities held (SHSDB), as well as the Register of Institutions and Affiliates Data (RIAD) are being on-boarded into the DISC corporate store. In a next phase, data sets from the SDW will be made available via DISC.

Business areas can request so-called data labs, which are self-service spaces within the DISC platform in which users can independently store and analyse data. They can manage the access rights to their data labs and thereby decide with whom to collaborate or share results.

DISC also provides data management components, which will technically enable activities to control, protect, deliver and enhance the value of data and information assets within the ECB.

[3] The term Extract, transform, load (ETL) refers to a process where data is first processed and then disseminated (usually via a data warehouse).

The service offering is complemented with a data factory, which allows automating the loading of data and make it available in a ready-to-use format.

DISC thus offers a highly performant environment, which has great potential to help exploiting large volumes of data sets and fostering collaboration between business areas.

But like any technology, the technical platform and services can only be successful if they are complemented with sound governance and the organisational willingness to share data, knowledge and results and to prioritise related work.

Strong ECB data governance ensures alignment

"*Support[ing] analysis and decision-making through high quality, timely, integrated and relevant data*" is part of the ECB Business Strategy 2018-2020.

In October 2016, the Executive Board approved a new ECB-wide data governance structure of a federalised type: data management is steered by a single Data Committee, and is jointly performed by a dedicated central team (DG Statistics) and dedicated roles in the business areas (data stewards and data owners) with a clear split of responsibilities. This approach shall ensure that data is governed and managed centrally whenever it is effective and efficient, without slowing down business areas' local data management and core business.

ECB Data Committee steers data management

The Data Committee is the overarching steering and coordination body for the management of data at the ECB. It is chaired by the Chief Services Officer and comprises area heads from the data providing and data using business areas as well as the technology provider.

The Data Committee proposes the data management strategy, data related policies and organisational changes to the Executive Board, decides on requirements for new internal data collections and steers the purchasing and provision of market data. It coordinates ECB-wide data standardisation activities and steers the implementation of the data management strategy and roadmap.

Data Steward Group fosters collaboration and operational alignment

The Data Steward Group comprises data representatives at senior expert level per business area to ensure cross-business area and conceptual alignment regarding data management. The data stewards act as sounding board for the data integration section in DG Statistics and contribute to the development of the data strategy, data policies, standards and prioritisation representing their business area specific activities and needs. The data stewards support the maintenance of the Data Inventory and the SDD. They strongly foster collaboration and the overcoming of silos, by sharing related experiences, issues and needs of their business area in the data steward group. They brief their business area representatives in the Data Committee and support the local implementation of the data management strategy and policies in their business area.

Directorate General Statistics adjusts its organisation to new needs

DG Statistics develops, collects, compiles and disseminates data, master data, statistics, statistical indicators, and metadata, and provides related user support services required for monetary policy, financial stability, banking supervision, the other ECB and the European System of Central Banks (ESCB) tasks, and for the support of the European Systemic Risk Board (ESRB). It thereby defines the concepts and methods and determines the master data for entities and transactions. DG Statistics also provides centralised market data services and develops financial market databases, e.g. for yields and credit risk. It makes non-confidential data available to market participants and the public, and shares confidential data in line with legal and contractual provisions. DG Statistics contributes in all fields to European and international standards related to statistics and data.

In June 2017 the DG Statistics adjusted its organisation to the new needs of the various internal and external stakeholders. It now comprises three divisions focussing on micro data, two divisions focussing on macro data, and two horizontal functions, namely one division responsible for statistical applications and tools, and a data integration and services section.

Micro data divisions deal with analytical credit data, financial market data, banking supervision data and centralised market data services. They strengthen the process-oriented clustering of micro data, with a focus on database development, data integration and supervisory services and enable innovative solutions. They also develop and manage RIAD and the SDD and work towards the reconciliation of statistical reporting requirements of credit institutions in particular through the Banks' Integrated Reporting Dictionary (BIRD) and Integrated European Reporting Framework (IReF) for reducing banks' reporting burden (see related section *Standardisation is essential for efficiency and integration*).

The two macro data divisions focus on monetary and general economic statistics, and macroeconomic statistics such as balance of payments and sector accounts including government finance statistics. They cluster well-established statistics with a strategic focus on a closer connection between external statistics and sector accounts capturing better the phenomenon of globalisation.

The statistical applications and tools division develops and maintains the ECB's statistical production environment (e.g. FAME) for collecting, producing and managing the ECB's macro-economic statistical databases, and the supervisory banking database. It is in charge of the SDW for disseminating macro-economic statistics and coordinates the on-boarding of data sets to the DISC corporate store. It contributes to the definition and implementation of analytical tools for accessing supervisory and statistical data and implements best practices and optimises data exchange processes and standards (SDMX, XBRL). It produces statistical publications and supports internal and external stakeholders with the visualisation of macro and micro data.

The newly established Data Integration and Services Section provides a focal point for the data integration within the ECB and offers ECB shared data services coordinated by the Data Committee. The section supports the Data Committee in defining and implementing the data management strategy and work program, ensuring existence and implementation of respective ECB-internal data policies, standards and processes (in particular data access and quality). It also coordinates the design and maintenance of a comprehensive data inventory covering all data

sets available within the ECB to increase transparency for users and foster multi-purpose usage of data. The section closely collaborates with the Data Stewards, with other DG Statistics functions as main data provider and with DG Information Systems who is responsible for the implementation of technical solutions, such as DISC. In addition, the section offers a shared data services pilot, which is co-funded by business areas (see below).

Overall, DG Statistics contributes to high data quality, efficiencies and less operational risks by promoting best practices across the ECB (quality and methodology), ensuring optimal (re-)use of existing data, methods and code, exploiting synergies between similar business area needs.

Standardisation is essential for efficiency and integration

Global identifiers will facilitate data integration and reduce reporting burden

Whilst DISC provides a common platform, data inefficiencies and inconsistencies continue to be caused by a lack of harmonisation and standardisation. Standardisation efforts are critical to ensure national, regional and global aggregation where needed and for data integration to gain a consistent overview of inter linkages and concentration risks in the banking sector and other parts of the financial system. They will also help to consolidate reporting and lighten the burden on banks.

Standardisation and harmonisation efforts have been stepped up, as demonstrated by the ongoing work of the Committee on Payments and Market Infrastructures (CPMI) and the International Organization of Securities Commissions (IOSCO). Progress has been made in developing globally harmonised identifiers, such as the Legal Entity Identifier (LEI) for banks and their counterparts, Unique Transaction Identifiers (UTIs) and Unique Product Identifiers (UPIs), relying wherever possible, on data standards that already exist (e.g. ISO standards).

Eurosystem initiatives foster standards implementation

However, where harmonised identifiers exist, they have not yet been globally implemented. Therefore, in addition to promoting data standardisation at international and European level, the ECB pursues three further initiatives to facilitate the integration of different data sets for analytical purposes and to reduce the reporting burden for banks[4]: IReF, BIRD and SDD. Moreover, it has established RIAD, an important reference data bases for entities, and a Centralised Securities Database (CSDB).

[4] See also '*The ESCB's long-term approach to banks' data reporting*': http://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html

ESCB Integrated Reporting Framework (IReF)[5]

The long-term approach of the ESCB and its Statistics Committee (STC) to data collection from banks aims at standardising and integrating the existing ESCB statistical frameworks, as far as possible, across domains and countries[6]. The main objective is increasing the efficiency of the reporting and reducing the burden for banks, while continuing to provide users with high quality data.

One element of the approach is the ESCB IReF, which is intended to integrate banks' statistical reporting requirements. The other element is the Banks' Integrated Reporting Dictionary (BIRD), which aims at supporting reporting agents in optimising the organisation of the information stored in their internal systems to fulfil reporting requirements.

The IReF aims at integrating the existing ESCB statistical data requirements related to banks, as far as possible, into a unique and standardised reporting framework that would be applicable across the euro area and might also be adopted by other European countries. The main focus of the project is on the requirements of the ECB regulations on balance sheet items (BSI) and interest rates (MIR) statistics of monetary financial institutions (MFIs), the securities holdings statistics (SHSDB) of MFIs, and granular credit and credit risk data (AnaCredit).

The ESCB has initiated a cost-benefit analysis to assess the impact of the IReF on the supply and demand sides, in close cooperation with the banking industry. This will help to identify the most appropriate approach for the banking industry and the ESCB to take.

Banks' Integrated Reporting Dictionary (BIRD)[7]

The BIRD is a recommended tool that helps banks to organise their data by providing a harmonised data model (so called *input layer*), describing the data to be extracted from banks' internal IT systems, and a common set of transformation rules to derive the specific final regulatory figures. The BIRD offers a transposition of the legal requirements at a more operational level. It facilitates the implementation, in a uniform manner, of reporting requirements by banks and software companies in their internal operational systems (e.g. for accounting, risk management, securities or deposits), also helping to report data to the required level of granularity.

The BIRD is being developed thanks to the joint efforts and close collaboration by banking institutions, NCBs and the ECB. Based on harmonised concepts and using clear classifications (i.e. a data dictionary), the input layer provides an accurate, standardised and unique means of defining and identifying individual business positions and transactions, together with their corresponding attributes.

5 See also The ESCB Integrated Reporting Framework (IReF) – An overview http://www.ecb.europa.eu/pub/pdf/other/ecb.escb_integrated_reporting_framework201804.en.pdf

6 See also Par. 2.12 of the "Medium term work programme of the ESCB Statistics Committee": https://www.ecb.europa.eu/stats/pdf/stcworkprogramme2019.en.pdf.

7 See also the BIRD website at: http://banks-integrated-reporting-dictionary.eu/.

The Single Data Dictionary (SDD)

The SDD contains metadata with definitions and concepts for describing ECB datasets and their content, which aims at enhancing the harmonisation of the description of datasets content and, thus, data integration within the ECB[8]. The final goal of the SDD is to have a unique metadata dictionary that can serve the ECB in defining its reporting requirements, as well as in fully exploiting already available data. The SDD started as a home-made stand-alone application and is now being implemented in the DISC infrastructure.

Users will benefit by having all data sets described following the same standard. In addition, producers will benefit when using different data sets in compilation procedures or for checking the consistency between different data sets. The SDD also serves BIRD.

A dedicated team in DG Statistics defines and maintains the reference dictionary and the process for direct semantic integration. Other concepts, which are defined using terminologies different from the SDD (e.g. under responsibility of entities outside the ECB) can be mapped to the SDD terminology.

The decision of which data frameworks are to be included in the SDD in the next years will depend on requests by ECB business areas coordinated by the Data Committee. DG Statistics will propose to the Data Committee a migration plan of the current metadata dictionaries to the SDD, taking into consideration also the input provided by the Data Stewards ensuring that datasets managed by all ECB business areas are properly taken into consideration, and following a cost-benefit assessment. The migration plan will describe the "whether", "how", and "when" existing datasets get included in the SDD.

In case a new reporting framework is defined or an existing reporting framework is amended, DG Statistics will propose an integration strategy to the Data Committee following the same approach described above.

Register of Institutions and Affiliates Data (RIAD)

RIAD is an ESCB-wide unique master dataset covering reference data on legal and other statistical institutional units and the relationships between them, relevant for statistical and several other business processes in the ESCB and the Single Supervisory Mechanism (SSM).

RIAD is pivotal for the collection, joint management and provision of reference data needed for all ESCB granular data collections and enables the integration of a variety of datasets, via a shared identification of the counterparties. This holistic approach avoids maintaining parallel reference datasets in different business areas and allows for a harmonised view of counterparties within the ECB and in associated institutions (NCBs, National Competent Authorities (NCAs)). In RIAD group structures are derived using a wide range of attributes on individual entities and relationships between them. This approach offers a wide flexibility in modelling different group structures according to different business needs.

Different stakeholders update and enrich the RIAD dataset, perform the necessary data quality management and deliver the up-to-date information to end-

[8] See also http://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/smcube/html/index.en.html.

users (which may be individuals or other client systems). For this, a network of "hubs" has been set up for some years. Hubs are in place in NCBs (or NCAs) to coordinate the work at country level and ensure that high quality standards are met. The national hubs meet in the "RIAD Hub Network" under the auspices of the STC, which sets out principles on how to proceed, prepares instructions – including via guidelines and agreements with stakeholders –, and monitors all new requests and developments to ensure a high quality of service and data.

Centralised Securities Data Base (CSDB)

The CSDB is jointly operated by the ESCB and contains timely and high-quality security-by-security reference data on around 7 million debt securities, equity shares and investment fund units issued worldwide. This includes securities issued by EU residents; securities likely to be held and transacted in by EU residents; and securities denominated in euro, whoever the issuer is and wherever they are held.

CSDB processes input from multiple data sources (several commercial data providers and NCBs) with overlapping and even conflicting information. Based on this input, the system compounds a consistent 'golden copy' data set in a fully automated way, taking into account all information available and relying on pre-set compounding priorities and statistical compounding algorithms. The system benefits from an efficient shared data quality management by NCBs and the ECB.

A shift in working culture with closer collaboration is required to leverage the data potential

Make data fit for use whilst protecting confidentiality

The Data Inventory creates transparency of all data sets that exist in the ECB, documents the confidentiality of the data and the access request procedures to be followed. The data integration section in DG Statistics has defined the requirements for the Data Inventory, which will be migrated from an interim access database to DISC. It is also performing a quality control on the content. The Data Stewards coordinate the population of the inventory for their business areas.

The ECB aims to maximise the use of available data whilst safeguarding the confidentiality. It has established clear access approval and reconciliation processes. Users are trained in awareness sessions to understand their responsibilities and apply appropriate measures to ensure data protection. The current access approach is rather restrictive and user needs are assessed on a case-by-case basis. The combination of granular data sets creates new challenges for the protection of confidential information. Work is ongoing at technical and organisational level to reduce the complexity and the effort spent to manage and reconcile access rights and perform output controls.

International initiatives, such as the G-20 Data Gaps Initiative (DGI-2)[9] and the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) are expected to contribute to these endeavours[10].

Clear guidance for publications and reports ensures consistency

The efforts by the DISC team and DG Statistics will give users more flexibility to use the granular data, apply own models and methods and easily create reports using tools like Tableau.

It is therefore essential that users fully document which data they used, which methods and codes they applied to ensure the traceability and reproducibility of their results. This is a prerequisite to ensure the soundness of decision making and understand the potentially different results based on the same data. Likewise, it is important that users, often coming from different business areas, share their knowledge, codes and methods to avoid duplication of efforts and reach results even faster. DISC will therefore provide functionalities to easily document the data related processes, provide versioning and historisation of data and of codes.

The Statistical Application and Tools Division has created a Tableau user guide jointly with the Communications and the IT department. They take into account existing practices and needs in the business areas, and provide corporate standards to ensure a harmonised look and feel for dashboards and reports produced across the organisation.

Create multidisciplinary teams and share knowledge

Multidisciplinary teams are best suited to exploit the potential of the granular data, new technologies and new methods and techniques. It is impossible for a single expert, even for a single division to master the complexity of the data. Therefore, the ECB started a few projects analysing data combining expertise from different fields, e.g. economists, financial stability experts, statisticians, data scientists and IT experts to jointly work on concrete questions.

These teams have demonstrated creativity by leveraging on the different skill sets and expertise, experimenting with machine learning and exchanging knowledge, code and output. Especially the co-ordinated feedback to data producers will help both data users and data producers.

[9] As stated in the report "Update on the Data Gaps Initiative and the Outcome of the Workshop on Data Sharing" by the Inter-Agency Group on Economic and Financial Statistics, dated March 2017 "The G-20 economies are also encouraged to increase the sharing and accessibility of granular data, if needed by revisiting existing confidentiality constraints" … "Every overhaul of existing or introduction of new legislations (or legal texts to the extent possible) which may have implications for data collection (including for administrative uses) should address data sharing and accessibility at national and potentially regional levels to prevent duplicated information requests by different authorities."

[10] INEXDA was launched by five central banks in 2017 (Banca d'Italia, Banco de Portugal, Bank of England, Banque de France and Deutsche Bundesbank), and the ECB and Banco de España joined in 2018.

Example: Shared data services pilot on funding and assets of banks

One of such projects is the shared data services project on the funding and assets of banks. Colleagues from various business areas are collaborating with the help of the DISC environment to combine security data with the objective to create a high quality integrated data set and conduct further analysis. They provide feedback to the DISC team on performance and user friendliness. They discuss with the data producers data quality questions and they work closely with user areas to provide reports containing answers to specific policy relevant questions.

The project combines data from RIAD, CSDB, SHS, several commercial data providers and CEPH[11]. With this concrete business case, it is performing a proof of concept of DISC. It started with the on-boarding of the data sets into the corporate store. This included the optimisation of table structures, which dramatically improved the performance of queries, which took hours in the CSDB environment and then was tuned from minutes to seconds in the DISC environment. The team is testing a new solution to ensure secure access. It is harmonising definitions and codes via the SDD and using a rule based approach to integrate the different data sets using the ISIN and LEI and other identifiers where available.

Conclusion

We are in the middle of a journey to exploit the potential of granular data for analytical and policy making purposes. The compass has been set right, but we will require persistence and continued collaboration at global, European and business area level to further progress on standardisation, optimise the use of technology, improve the legal framework for sharing data and develop appropriate methods to preserve confidentiality without unduly impacting the usability of the data. New business processes will emerge and statisticians and users will enter into new modes of cooperation, for which the rules need to be fine-tuned along the process.

Abbreviations used

BIRD	Banks' Integrated Reporting Dictionary
CEPH	Common Eurosystem Pricing Hub
CSDB	Centralised Securities Database
DG	Directorate General
DISC	Data Intelligence Service Centre

[11] CSDB contains reference and price information for around 10 mio securities at monthly frequency. The commercial suppliers cover 4,000 to 5,000 securities from different number of issuers in different ways, either from large banks or EU sovereign bonds, daily or tick-by-tick. The Common Eurosystem Pricing Hub (CEPH) provides daily data for around 40,000 securities, and provides all NCBs on a daily basis with a unique price "Final Eurosystem Price" for each marketable asset eligible as collateral for Eurosystem monetary policy operations.

ECB	European Central Bank
ESCB	European System of Central Banks
IReF	ESCB Integrated Reporting Framework
MFIs	monetary financial institutions
NCA	National Competent Authority
NCB	National Central Bank,
RIAD	Register of Institutions and Affiliates Data
SDD	Single Data Dictionary
SHSDB	Securities Holdings Statistics Data Base
SSM	Single Supervisory Mechanism
STC	Statistics Committee
MMSR	Money Market Statistical Reporting
FAME	Forecasting Analysis and Modeling Environment

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

Ninth IFC Conference on "Are post-crisis statistical initiatives completed?"

Basel, 30-31 August 2018

ECB data for analysis and decision-making: data governance and technology[1]

Emily Witt and Jannick Blaschke,
European Central Bank

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

EUROPEAN CENTRAL BANK
EUROSYSTEM

E Witt
Senior Advisor Statistics
European Central Bank

J Blaschke
Research Analyst Statistics
European Central Bank

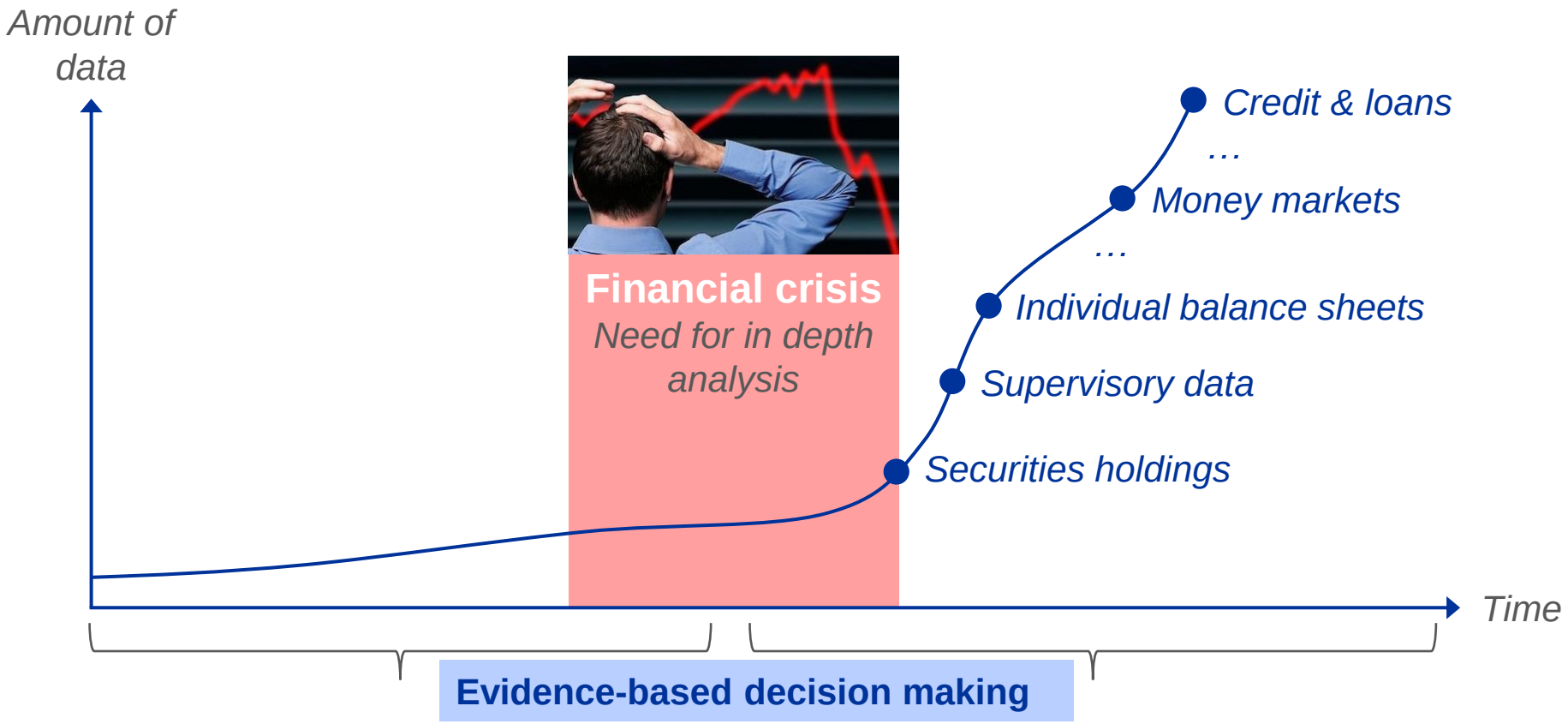
Governance and dissemination

ECB data for analysis and decision-making: Data Governance and technology

9th IFC Conference
"Are post-crisis statistical initiatives completed?"

Basel, 31 August 2018

Exponentially increasing availability of data



Macroeconomic statistics
to analyse
- economic signals
- macroeconomic forecasts
- economic linkages

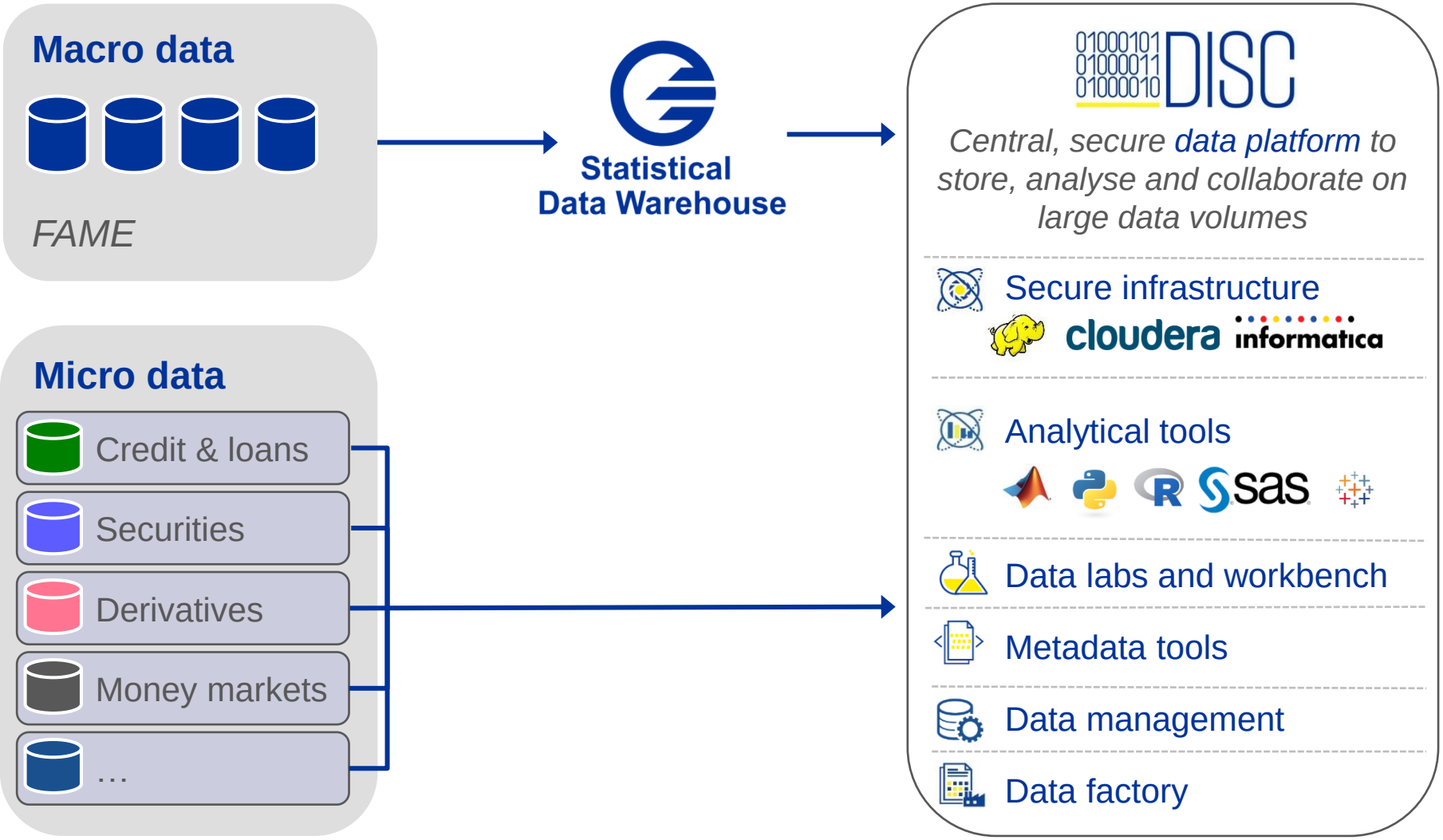
Micro data (in addition to macroeconomic statistics) to analyse
- timely and diverse economic signals
- risk concentration and distributional effects
- flexibly new questions

New challenges call for a holistic approach

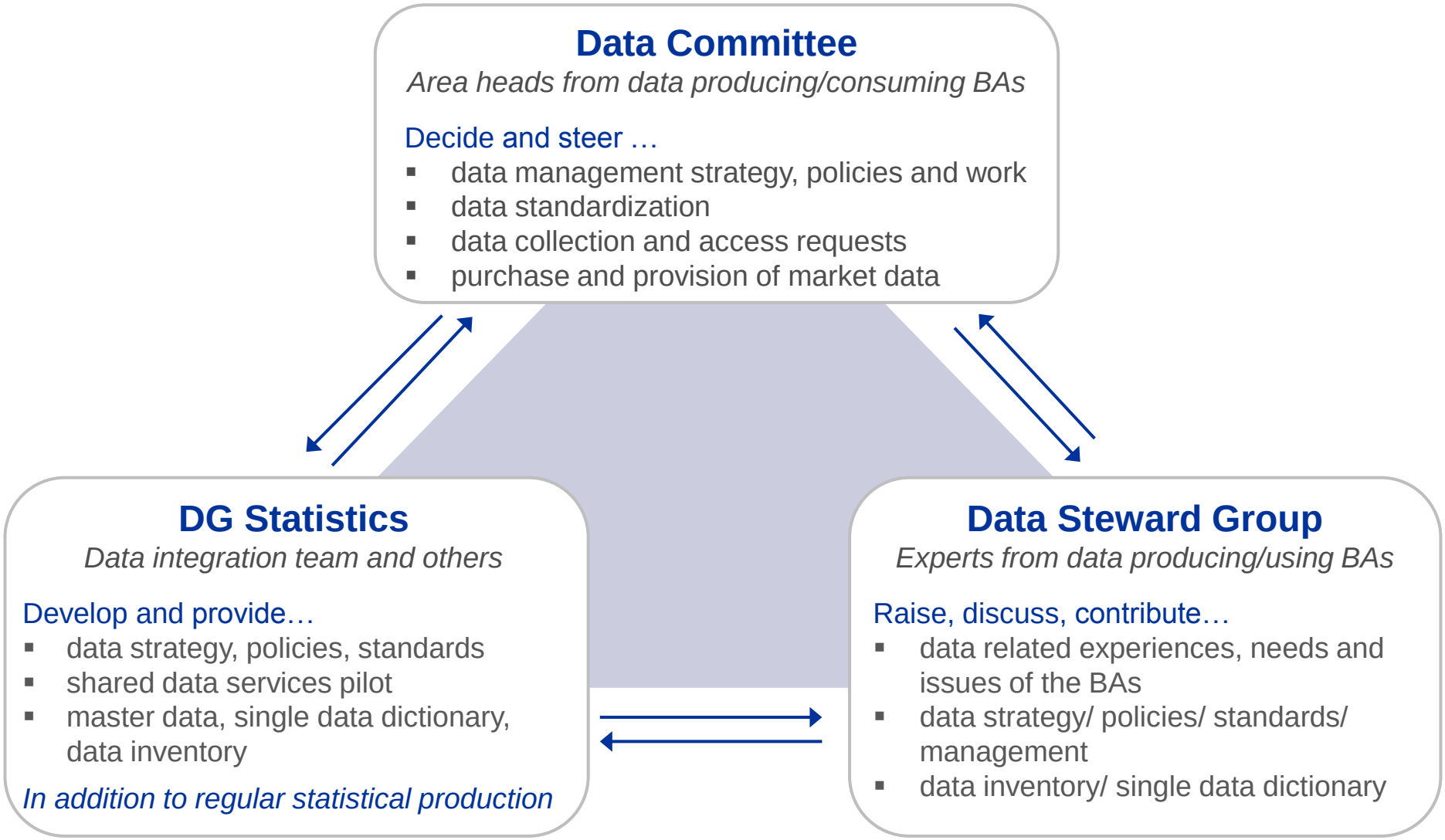
To exploit the increasing availability of micro data challenges have to be addressed:

- Increasing volume and speed
- Heterogeneity of data
- Complexity of data
- Data confidentiality

DISC facilitates data usage, collaboration, automation



Strong ECB data governance ensures alignment



Standardisation for efficiency gains and integration

LEI

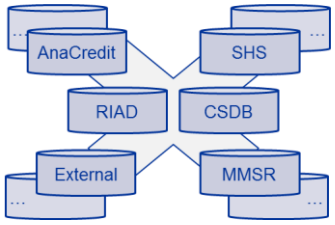
ISIN ISO

- **Global standardisation of identifiers**
  - Entity
  - product
  - transaction

BIRD IReF

SDD

- **Joint initiatives with European banks**
  - Banks' Integrated Reporting Dictionary (BIRD)
  - Integrated Reporting Framework (IReF)
- **ECB Single Data Dictionary**



- **Reference data for entities and securities**
  - Register of Institutes and Affiliates Data (RIAD)
  - Centralised Securities DataBase (CSDB)

Shift in working culture to optimise data usage

Make data fit for use whilst protecting confidentiality

- Create transparency of data and data needs
- Make data access clear and efficient within legal boundaries
- Apply appropriate measures to protect data
- Adhere to clear rules for data usage and output control

Document guides, methodologies and code

- Use metadata and document methods and code to reproduce results
- Apply corporate standards for visualisation (Tableau user guide)

Create multidisciplinary teams and share knowledge

- Combine expertise from different fields e.g. economics, statistics, data science, legal, IT
- Define concrete projects to jointly analyse data for multiple purposes
- Exchange knowledge, code and output
- Provide quality feedback for data producers
- Learn new skills and experiment with e.g. machine learning, AI

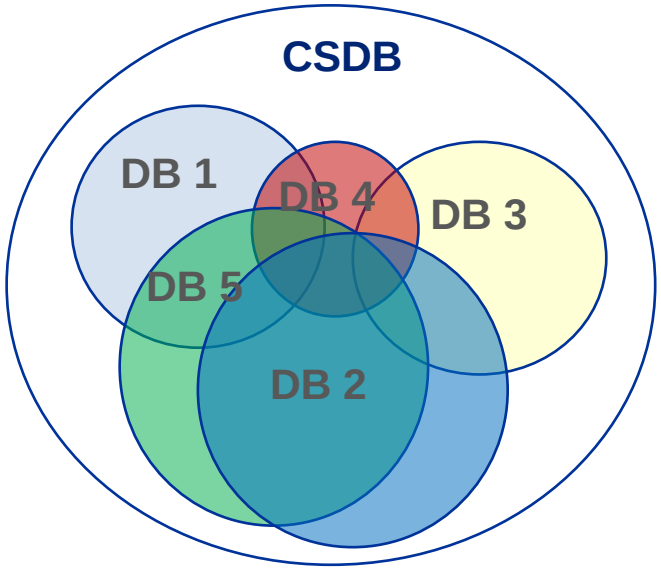
Joint project to leverage expertise and data

High quality integrated security data on assets and liabilities of banks

Joint team

- Economics
- Statistics
- Financial Stability
- IT

Security data



- **On-board** all relevant data sets to DISC, optimise table structures
- Harmonise definitions and codes via **SDD, comparable identifier**, etc.
- Ensure **secure access** (Jumphost)
- Use **DISC workbench** for storing, distributing and analysing data
- Manage and collaborate on code (BitBucket)
- Use **Tableau** for data visualisation
- Interact with **production teams** to clarify quality issues

7. Time to conclude and ask questions

- ❑ The journey is going into the right direction
- ❑ Persistence and collaboration are required

Time for questions and comments