



**BANCA D'ITALIA**  
EUROSISTEMA

## **Data Science in Central Banking: Enhancing the access to and sharing of data**

Opening remarks by Bruno Tissot

Head of Statistics and Research Support, BIS, and Head of the IFC Secretariat

3<sup>rd</sup> IFC and Bank of Italy Workshop

Rome, 17 October 2023

Ladies and Gentlemen, dear colleagues,

I am very happy to welcome you to today's event on Data Science in Central Banking hosted by the Bank of Italy.

As you know, the Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS) decided a few years ago to organise recurrent workshops on "Data science in central banking". The aim has been to review the adoption of data analytics and business intelligence techniques and developments in the big data ecosystem. How can this be achieved? We are convinced that showcasing projects and sharing experience can provide a key opportunity to develop in-house knowledge and reduce the reliance on external services providers.

The most recent IFC workshop took place in 2022 and focused on the broad spectrum of data science applications/tools used in central banks. And indeed there are many projects currently going on in our community. A key supporting factor has been the sheer volume and complexity of financial data available in today's societies. It calls for more sophisticated techniques for data management and analysis, a trend reinforced by the new opportunities opened up by artificial intelligence and machine learning. Another factor has been the greater focus on real-time, evidence-based policymaking, which requires authorities to rely on better analytical and forecasting capacities to support their decisions. It also puts a premium on accessing data and disseminating them to interested users: international institutions, other domestic authorities, academia, the general public etc. These aspects are particularly relevant for central banks as both producers and users of official data – a quite unique perspective in fact.

But we should start by trying to define what data science really means for us. This concept generally refers to the study of data and therefore includes the various techniques for extracting insights from them. But data science is fundamentally different from traditional data analysis, as it typically applies to large, complex and/or unstructured information

sets. Hence the role of data scientists lies at the intersection of three areas: information technology (IT); mathematical and statistical methods; and business, or “subject-matter” expertise.

Let me discuss these three aspects in turn, noting that more details can be found in the IFC Bulletin no 59 just published by the BIS last week on “*Data science in central banking: applications and tools*”.

### **IT infrastructure and tools**

From the outset, IT has absorbed a great deal of attention and resources to support data science projects, as underscored by the IFC Report no 11 published in 2020 on “*Computing platforms for big data analytics and artificial intelligence*”. This report showed that central banks are increasingly aware that a modern IT architecture is crucial to reliably and securely deal with data. In addition, a key objective for them is to facilitate access to a large and diverse range of sources as well as relevant IT software and tools in a user-friendly way.

But implementing such an IT architecture can be challenging, calling as it does for careful implementation, clear governance frameworks, and the application of common standards. Exploratory work has sometimes barely started, and substantial staff and IT investment as well as business adjustments will continue to be needed. Looking ahead, the primary focus will thus be on adopting advanced IT tools and engineering practices – including cloud computing, software containers, automation tools etc.

### **Leveraging mathematical and statistical techniques**

Once a sufficient IT infrastructure is available, the next stage is to perform the various mathematical and statistical operations that are needed to deal with the raw data at hand. One needs not only to access very large and complex information sets, but also to derive statistics compiled via multiple sequential tasks (signal extraction, quality management, dissemination) before using them to extract relevant insights. Many different approaches can be leveraged for this purpose, including textual analysis, with the support of natural language processing (NLP) tools and large language models (LLMs).

This area is obviously key for us statisticians in central banks. Last year the Committee published the IFC Bulletin no 57 that took stock of how central banks are developing machine learning across a variety of use cases. It showed that the new techniques can help improving the efficiency and effectiveness their operations, including through the dealing of larger and new sources of information in a more automated way. This Bulletin also underlined the diversity and maturity of innovative approaches already developed and used.

In fact, the new techniques being contemplated can facilitate the gathering of more and better information, which is key for central banks that rely heavily on data. They can help respond to their information demand by enhancing statistical quality, providing richer context, and making sense of the abundance of data available to derive useful conclusions. This is particularly evident as regard the three main central bank policy areas, namely

monetary stability, micro supervision and macroprudential tasks... as well as for broader financial stability-related goals, including the functioning of the payment system, financial inclusion, consumer protection, and anti-money laundering as we will hear today.

The experience of central banks has highlighted several lessons. Let me highlight a few:

First, there is a wealth of alternative information sources that have barely been tapped and which can provide new and useful insights. Financial granular data, administrative registers as well as text-based information (a type of data that had been available well before the advent of big data techniques...) represent the new frontier for central bank statisticians.

Second, complementarity is essential: new methods can bring additional insights but have to be blended with other types of exercise.

Third, while data science can be instrumental in dealing with complexity, there are important challenges faced by users as the functionality of the new tools is rarely intuitive.

### **Subject-matter expertise**

A third key aspect of data science is to rely on a good understanding of the business cases, requiring therefore a close cooperation with subject-matter experts. This is because analysing data calls for an awareness of the way the statistics have been compiled as well as of the complex factors that drive them.

Moreover, subject-matter expertise is essential to support informed policy decisions: translating data insights into actionable recommendations for central banks cannot be communicated as a "black box" and requires transparent explanations to the various stakeholders involved – from other authorities to the general public.

From this perspective, the focus should be on:

Transparency ("metadata", as we statisticians usually refer to);

Ethics and governance (noting that central banks are in a rather good position, being able to build on the strong foundations provided by the international fundamental principles of official statistics); and

Public communication (putting a premium on enhancing economic and statistical literacy in society).

### **Accessing alternative information sources and sharing data**

Let me conclude by looking at the topic of today's event to deal, in particular, with issues related to the access to private and administrative sources and data-sharing.

As you know, one important development for central bank statisticians has been related to the international cooperation framework under the Data Gaps Initiative (DGI) endorsed

by the G20 and supported by many countries. This work was decided in response to the Great Financial Crisis (GFC) of 2007–09 to help to coordinate national work on financial data sets, including through enhanced data-sharing. It comprised several phases. The third one, initiated in 2022, calls for better data to understand today's key challenges and to make official statistics more detailed and timely. Main international financial organisations, including the BIS with the support of the IFC which represents more than 100 central banks globally, are actively involved.

The new five-year phase of the DGI covers priority policy needs in the areas of (i) climate change; (ii) distribution of household income and wealth; (iii) fintech and financial inclusion; and, particularly relevant for our event today, (iv) access to private and administrative sources of information and data-sharing. One needs, in this context, to emphasise the instrumental role played by the Bank of Italy in pushing this last topic on the agenda when chairing the G20 in 2021.

These international initiatives underscore that the topic of today's workshop, ie data access and sharing, is crucial. In fact, a key lesson of the recent Covid-19 crisis is that coping with the shifts in user demands triggered by unprecedented shocks such as pandemics requires tapping into new information sources, often held by private actors and government agencies. Another lesson is that the use of artificial intelligence in the compilation phase could transform the whole production chain of official statistics, making it faster and more resilient and user-friendly. Furthermore, it is also important to have standardised data publication and international data-sharing in the distribution phase. Let me remind you in this context that the BIS is an active sponsor of the Statistical Data and Metadata eXchange (SDMX) standard, and that these issues will be dealt extensively not only during this workshop but also on the occasion of the upcoming 9<sup>th</sup> SDMX Global Conference on "*Empowering Data Communities*" to be held in Manama, Bahrain, on 29 October-2 November 2023.

Looking ahead, we believe that the way forward should be based on cooperation, to spur knowledge-sharing and the exchange of best practice. This can be instrumental to facilitate **access** to privately-held data and administrative registers. For instance, recent pilot projects in central banks have showed that innovative text-mining techniques can greatly help to extract relevant information that policy makers are needing the most. Similarly, AI approaches can be used to better monitor and understand complex, granular developments in financial markets. Needless to say, substantial investments, especially in terms of IT hardware, software and staff skills will be required to replicate such initiatives more broadly.

Cooperation should also lead to better **data-sharing**, especially with other producers of official statistics, the private sector and academia – noting that several central banks are, together with a number of national statistical offices and international organisations, actively involved in INEXDA, the International Network for Exchanging Experience on Statistical Handling of Granular Data, to in particular facilitate the use of granular data for research purposes. Data-sharing across jurisdictions is also an essential element supporting global public goods. Fortunately, significant progress has been already achieved by the international community of central banks and supervisory authorities

in the sharing of financial and regulatory data, with the strong support of the BIS. Lastly, data-sharing can more generally help to ensure the broadest possible data availability in the most efficient way, while keeping the reporting burden under control.

Let me stop here. The fact that more than 100 institutions are represented in today's event is a clear indication of the relevance of the selected presentations. It also shows that the various issues covered are particularly topical for central banks and all statisticians more generally.

