

Have you thought about data governance? Think again. Think better.

Maria do Carmo Moreno. Banco de Portugal. Statistics Department. Lisboa, Portugal – mcmoreno@bportugal.pt

Abstract

This paper highlights the significance of data governance in central banks where data plays a crucial role in driving business outcomes. It outlines the essential components of a successful data governance framework, including data governance policy, data quality standards, data architecture, data management tools, data security and privacy, and data governance training.

The paper shares the experience of Banco de Portugal in developing a new data governance strategy based on a centralized coordination model with decentralized responsibilities. It highlights the creation of a data warehouse, reference data repository, data catalog, and data and SQL labs as key components of their data management framework. Challenges for the future include further data integration, integrated data quality procedures, managing the growing volume and diversity of data, and the need to develop new processes and technologies for analyzing and extracting more value from data.

The paper concludes by emphasizing that data governance is a complex and ongoing process requiring time, effort, and a cultural shift within an organization, and the need for dedicated teams passionate about data governance.

Keywords: data governance; data integration; data sharing; data management.

1 | Introduction

In today's digital age, data has become a critical asset for organizations across all businesses in general, and central banks are no exception. Consequently, data governance has emerged as a crucial aspect of business strategy, with the goal of ensuring that data is effectively managed, protected, and leveraged to drive business outcomes. The implementation of a data governance framework ensures that data is trusted and trustworthy. It implies that data is well-documented and easy to find and access within the whole organization, and that it is kept secure, compliant, and confidential when required.

A successful data governance framework must have several key components. These include a data governance policy, data governance processes, data quality standards, data architecture, reference data, data management tools, data security and privacy, and data governance training. Each of these components plays a crucial role in ensuring the effective management of data within a central bank.

Over the last five years, significant progress has been made in the development and implementation of data governance, along with other data management practices, reflecting the growing importance of data as a strategic asset. Looking ahead, data governance will continue to be a critical focus for central banks as they seek to efficiently manage and leverage

their data assets. However, several significant challenges need to be addressed to ensure the success of data governance initiatives. For instance, it is crucial to have a flexible data architecture that can handle new data requirements, such as large volumes of structured and unstructured data, cloud computing, new collaboration models, and the need for skilled personnel. Furthermore, central banks must develop comprehensive policies and procedures that ensure the privacy and security of data while also enabling effective use and collaboration with stakeholders. This requires a strategic approach to data governance that takes into account the evolving data needs of central banks.

Six years ago, Banco de Portugal embarked on a journey to develop a new data governance strategy, which has since evolved into a robust model with centralized coordination and decentralized responsibilities. In this paper we share our experience at Banco de Portugal in data management, including how we paved the way to set up a data governance framework, the challenges we faced and those that lie ahead, and a few lessons learned.

2 | Setting up a Data Governance Framework

One of the fundamental aspects of our data governance strategy is a clearly defined set of roles and responsibilities. We recognized the importance of having a **centralized coordination** through a partnership between the Statistics and IT departments to ensure alignment and suitability across the organization between business and technology. At the same time, we are demanding decentralized **responsibilities** that allow individual business teams to manage their data as they see fit. At the department level, **data stewards** were appointed to ensure compliance with data governance policies and standards, and to support data quality and management activities. This approach has helped us strike a **balance between standardization and flexibility**, ensuring that everyone in the organization understands their role in the data governance model.

The data governance strategy received strong **sponsorship** from the board, which was crucial for its success. Board sponsorship helped to promote a culture of data-driven decision-making and ensured that data governance was a priority within the organization. The board also provided resources and support for the implementation of this strategy, including funding for the development of an enterprise data warehouse and the creation of data science training programs.

In fact, a key component of our data governance strategy was the development of a **data warehouse** where all our organizational data that can be shared is gradually being stored. This data warehouse helps to balance the trade-off between data sharing and data security. On one hand, it is a single source of truth for all data, which ensures consistency and accuracy in data analytics and reporting across Banco de Portugal. On the other hand, it also provides a secure and scalable solution for storing large volumes of data that can be easily accessed by those who need it. By centralizing data in this way, the data warehouse is a powerful tool for sharing information across the organization, helping to break down silos and promote collaboration across different departments and teams.

To guarantee consistency in the way data is defined and used across the organization, we created a **reference data** repository. This repository includes master data, such as entities, and

common definitions for data elements which ensures that systems and applications in the organization share the same identifiers allowing to cross information from different datasets. A governance framework has been put in place for the reference data which clearly defines the roles and responsibilities of stakeholders involved in managing this repository, including data owners, stewards, custodians, and users. It also establishes processes for creating, updating, and using reference data and resolving any data quality issues or inconsistencies. Additionally, the governance framework addresses security and access controls for the repository, ensuring that only authorized users can access and modify reference data.

To make it easier for everyone in the organization to find the data they need, we developed a **data catalogue** which provides a comprehensive view of all the data stored in the data warehouse, along with metadata that describes the data's purpose, ownership, and usage. This approach has helped us to manage data assets more effectively. The catalogue is organized into four key pillars with relationships between them: data glossary (definition of data domains, terms and attributes), responsibilities (data owners, experts, stewards, custodians), systems (sources, databases, objects), and external relations (regulations, guidelines, protocols). Data cataloguing involves multiple parties, including the operational coordination and data stewards, data experts, and data custodians for the identification, mapping, and cataloguing of relevant technical objects with business terms.

We established both **data and SQL labs** as part of our data architecture. These labs provide a sandbox environment where our data science teams can test new technologies and experiment with different data management approaches. The data lab supports multiple languages (such as Python, R, and Julia) and tools and provides a secure environment for data scientists to test new models and algorithms, while the SQL lab provides a centralized repository for SQL scripts and queries. The labs also enable our teams to collaborate more effectively, sharing insights and learning from each other.

Recognizing the importance of having a skilled data science team, we created a **data science training program** called Data Science School. This program provides our employees with the necessary skills and knowledge to work with data effectively, enabling them to extract valuable insights from our data. The Data Science School also fosters communication and collaboration among the data science community in the bank through the promotion of talks by colleagues (called Spotlight sessions) and seminars with invited speakers (Trend time), or more challenging events (like a Datathon).

3 | Challenges for the Years Ahead

While our data governance journey has been successful, we recognize that the future brings new challenges.

One of the main challenges is the **integration of different sources and types** of information, namely statistical, prudential, and resolution. Data integration requires data harmonization. The most visible outcome of this integrated approach to data management is data cross-checking, which is critical to maintaining data quality and consistency at the source (banks). Banco de Portugal has identified the assessment and promotion of banks' reporting quality as a priority.

Opportunities to improve data quality, including integrity and accuracy, completeness, timeliness, and adaptability, will also be identified and pursued.

Another significant challenge will be the need to **effectively manage the growing volume and diversity of data**, particularly as the importance of analytics and artificial intelligence continues to grow. This will require the development of new processes and technologies for managing and analyzing data, as well as to continue investing in the appropriate skills and expertise in areas such as data science and machine learning to ensure that data is effectively leveraged to support decision-making.

The adoption of **cloud computing** creates new opportunities for managing and leveraging data more effectively, but it also brings new challenges for data governance. Cloud-based services offer a range of benefits, including improved scalability, flexibility, and cost-effectiveness. However, they also introduce new risks related to data privacy, security and auditability. As more data is stored in the cloud and processed by third-party providers, there will be growing concern about these issues. This will require the development of policies and procedures that are compliant with data protection regulations without undermining the advantages and benefits of cloud computing.

Banco de Portugal must collaborate with a range of stakeholders, including other central banks, financial institutions, and government agencies. The emergence of new **collaboration models**, such as open banking and platform-based ecosystems, presents significant challenges for data governance. This will, in turn, demand the development of policies and procedures that ensure the privacy and security of data while also enabling effective collaboration with these stakeholders.

Finally, the ongoing digital transformation will require the development of new strategies and approaches for managing data in a rapidly evolving business environment. This will require the development of **agile, flexible data governance frameworks** that can adapt to changing business needs and priorities.

4 | Lessons Learned and the Importance of Dedicated Teams

Throughout our journey in data governance, we have gained a wealth of valuable insights. These include the critical role of sponsorship, the significance of clearly defining roles and responsibilities, and the need to establish effective processes and technologies to manage data efficiently.

However, one of the most important lessons is that **data governance takes time**. The development of a comprehensive data governance framework is a time-consuming process that can take several years to complete.

Additionally, implementing such a framework involves a significant **cultural shift**, which can prove challenging to accomplish. Overcoming resistance to change, lack of understanding, and insufficient awareness of the benefits of data governance can all pose challenges. Therefore, it is essential to create a culture within the bank that values and prioritizes good data governance practices, both top-down and bottom-up.

Perhaps the most significant lesson we have learned is the importance of creating **dedicated teams** of people who are truly passionate about data governance. These teams play a critical role in the success of data governance initiatives, bringing specialized skills and expertise required to manage data assets effectively and ensure their optimal utilization. Establishing a dedicated data governance team requires not only technical expertise but also a deep understanding of the organization's business goals and strategic priorities. Moreover, there is also the need for ongoing collaboration between IT and business leaders and consistent investment in skills development and training.

A solid data governance model is particularly important when data sharing and integration are considered priorities within the organization.

data governance takes time

- Rationalization of collection processes

To ensure the rationalization of data collection processes, a data governance procedure was implemented which requires prior consultation with the data governance operational coordination before any new data requests are made. This procedure ensures that redundant data is not collected and that all business areas potentially interested in the requested data explicitly express their interest. Initially, some requesters did not follow the procedure, either because they were not totally aware of it or because they did not understand its benefits. Therefore, the Board had to emphasize its mandatory nature, because data governance policies and procedures are only effective if they are largely used. Fortunately, this is now a standard practice that not only avoids redundancy but also promotes data sharing.

- Data warehouse

During the initial year of implementing the data warehouse, only data related to loans was integrated. Over the following two years, four additional domains were integrated, and by the fourth year, the integration of eight domains was achieved. The significant leap in progress during the fourth year can be attributed to a new governance strategy, which involved a dedicated team creating an integration framework that streamlined the analysis phase with the business team, enabling them to work more efficiently with the IT team during the implementation phase.

data governance involves culture shift

- Data stewards' network

When the head of departments appointed the data stewards (two per department), many of them were initially unsure about their roles and responsibilities. During the monthly meetings, they primarily listened passively. Over time, the situation has improved, and at the end of the second year, when we conducted a survey asking if they would prefer to stop the meetings, everyone expressed their desire to continue attending. Their level of active participation began to increase, and they began to present new data developments occurring in their departments to other data stewards. Nowadays, this network facilitates genuine collaboration and

communication among different departments. They are the main point of contact in each department in what refers to data management and play an essential role in sharing knowledge and good data governance practices.

- Data science school

The training program began in June 2020 through Coursera, a MOOC platform, and in-house blended courses designed to cover internal subjects, and is accessible to all departments. It was further reinforced by a Community of Practice (DATA.go) with over 480 members on MS Teams which is an excellent field for multiple chats and discussions on data science topics. To date, over 350 participants have completed around 400 distinct courses, resulting in more than 10 000 hours of training. More than 30 events (talks, seminars, and master classes) have been organized over time in the DATA.go community and these are opportunities to share experiences and acquire new knowledge. In total, 28% of the employees from 18 out of 21 departments have attended courses and 57% have participated in the DATA.go initiatives. Last November, the Data Science School organized a 24-hour Datathon, which was a huge success. A total of 53 participants from 14 departments were grouped into 12 teams, tasked with driving insights from Payment's data. The results were presented to a jury that selected three winners. The event received tremendous feedback from participants, and it was a fantastic opportunity to promote collaboration among departments.

data governance requires dedicated teams

- Data Management Unit

The Data Integration Program in 2017 led to the creation of a data management unit within the Statistics Department, which supported data management and data governance activities. As a central hub for data governance issues, the team has various responsibilities, including managing reference data, coordinating data catalogue and data warehouse activities, and promoting collaboration among data stewards. The team serves as a facilitator for implementing data governance guidelines, such as adopting harmonized classifications, breaking down silos, integrating data from different business areas, and sharing best practices. Additionally, the team contributes to developing data literacy and raising awareness of data governance principles.

One of the main benefits of having dedicated teams is the willingness to work collaboratively with other departments to develop and implement strategies that improve data sharing and integration. In the long run, the success of data governance initiatives depends on the effective management and utilization of data, and the promotion of a culture that truly values and prioritizes data sharing. To achieve these goals, it is critical to foster collaboration and communication among stakeholders.

Data governance is ultimately about people - the people who create, manage, and use data.