

2024

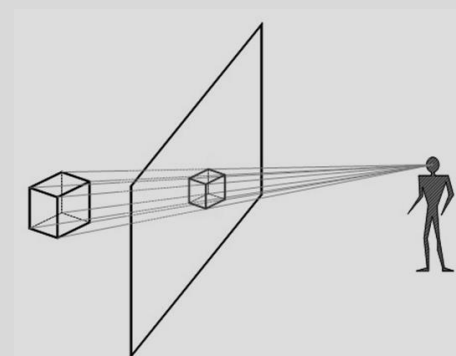


## 29<sup>th</sup> Eurofiling Conference

5<sup>th</sup> and 6<sup>th</sup> June 2024 | Central Bank of Ireland | Dublin, Ireland

# The digital reporting in the perspective of a national central bank

Laura Mellone  
Banca d'Italia



# AGENDA



Why continue the discussion on digital reporting?



The six pillars for the evolution of digital reporting



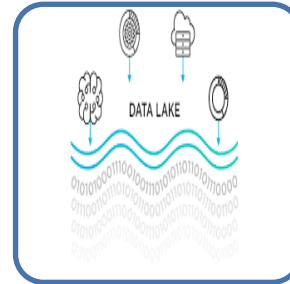
Standardization in reporting (LEI)



Cooperation with reporting entities and input-based approach



More efficient data production process with the use of AI/ML



Combined and shared use of structured reporting and unstructure data



Skills, knowledge, training



*Why continue the  
discussion on digital  
reporting?*

## Why continue the discussion on digital reporting? (1/2)

*In all areas affecting reporting (not just financial reporting), standards, taxonomies, practices and even regulations) have been introduced that, we might say, make **excel/paper/manual operations-based reporting a “thing of the past”***



## Why continue the discussion on digital reporting? (2/2)

... nonetheless, in order **to make digital reporting more effective and efficient**, there's still need of completing a **complex puzzle**, in terms for example of:

- ✓ deeply **redesigning the reporting requirements**
- ✓ fully **automatizing data processing** (central banks and at the reporting agents)
- ✓ associating **data integration** with digital reporting
- ✓ assessing carefully **changes of paradigm in data collection** (e.g. the real-time data-driven reporting)
- ✓ **growing the scope of reporting** in an orderly way (e.g. non-financial information)
- ✓ managing the **information overload** and the increasing data demand



## Our view

- ❑ **Why the digital reporting is still a key issue for Central Banks**
- ❑ **In our mind, the future of digital regulatory reporting (also in view of a reduced reporting burden) rests on six pillars:**
  - a. Standardization
  - b. Necessary data granularity but not maximum granularity
  - c. Machine-readable regulations and reporting instructions
  - d. Cooperation with reporting entities and input-based approach
  - e. More efficient data production process with the use of AI/ML
  - f. Combined and shared use of structured and unstructured data



*I will not touch  
b and c*

## STANDARDISATION



- ❑ The EU Commission “*Strategy on supervisory data in EU financial services*” (2021): a modern and integrated reporting system, that delivers accurate, consistent, and timely data to authorities at EU and national level while minimizing the aggregate reporting burden for all relevant parties, should include – among other things – standardized data.
- ❑ Standardisation includes many topics, of which the **identification codes (identifiers)** of the reporters and of the counterparties included in the reporting
- ❑ The **Legal Entity Identifier (LEI)** created for uniquely identifying entities engaged in financial transactions, may be a promising reference data to take into account in the evolution of reporting. In perspective, the LEI could be used as the **unique mandatory identifier** to be adopted by reporters or may be foreseen as an additional information very often accessible via public or official sources.



## Collaboration with reporting entities and input approach (1/2)

- ❑ **PUMA cooperation is a structured and voluntary collaboration between BoI and reporting entities** which has been producing tangible results for over thirty years.
- ❑ PUMA is aimed at supporting intermediaries in the automation and integration of their data systems and at optimising the overall reporting process in order to contain the reporting burden as well as to improve the consistency and the quality of the information transmitted.
- ❑ The basic assumption underlying PUMA approach is that the **quality and the timeliness of the data made available to end-users are strictly related to the properties of the data initially received from reporting agents, which, in turn, are influenced by the presence of a well-documented / highly standardized / automated process to produce the data.**



## Collaboration with reporting entities and input approach (2/2)

- ❑ This is why the BoI has an interest in investing also in the internal processes of reporting agents to produce the required statistical data, although such processes remain under the responsibility of the institutions themselves.
- ❑ The PUMA dictionary is the set of metadata guiding banks' internal systems. In order to assume the role of "active dictionary", the dictionary is described in a language as formal as possible, i.e. that it can be read by software more or less automatically; at the same time it is user-friendly enough to be managed independently by business analysts, i.e. not necessarily by IT experts.
- ❑ PUMA has been an example for other authorities in Europe: first, **Austria** and, a few years later, the ECB started a similar cooperation programme with their respective banking systems. Austria, although with some organizational differences, has a similar approach on a national scale. As a result of the progressive harmonization of reporting at the European level, the **ESCB** launched the BIRD cooperation initiative, involving a significant number of European NCBs and commercial banks.



## More efficient data production process with the use of AI/ML (1/2)

- ❑ new operational procedures to **accurately identify potential errors in a larger amount of data, but not only**
  - **Data validation:** AI/ML tools allow for a greater accuracy of the checks by reducing the so-called "false positives" and, as a consequence, they also contribute to limit the checking burden at the reporting agents side; they improve the quality of the data by quickly identifying anomalies which would otherwise remain unnoticed (so-called "false negatives")
  - **Process efficiency:** AI/ML tools significantly limit the area of manual intervention, as the machine takes charge of an increasingly large share of the decision-making process
  - **Data production/enrichment:** the outcome of the advanced checks or calculations may also be used to enrich the information asset already available by providing further information that may be not required to reporting institutions.



## More efficient data production process with the use of AI/ML (2/2)

### □ Our studies and use cases:

1. granular data in the statistical reporting of payment services: ML tools to **detect outliers** that cannot be detected using the standard/traditional procedures
2. granular datasets vs aggregated: ML tools to detect **potential outliers in granular AnaCredit data** compared firstly with BSI, and secondly with FinRep
3. banks' supervisory reports on loans to the private sector: **identification of errors**
4. ML tools used to predict **whether confirmations of data** (and relevant motivations) sent by banks to Bdl in response to warnings on data quality **should be accepted or rejected**
5. ML tools used as a decision-making rule to **evaluate and monitor the revisions** sent by intermediaries with regard to the national Non-Performing Loans reporting (very granular)
6. ML tools used to automatically **select the market notices issued by the Italian Stock Exchange** to update the Bank of Italy's Securities Database
7. ML tools used for the **classification by sector of economic activity of the Italian companies** recorded in the Bank of Italy Entities Register

## Combined and shared use of structured reporting and unstructured data (1/2)

- ❑ The need to jointly exploit the structured data collected from reporting entities and the unstructured information that is increasingly available to the community is a challenge particularly felt by data users and which in recent years has also become a priority at an international level. **Unstructured data are by their nature heterogeneous and fragmented and this makes their use complex, especially in conjunction with structured data.**
- ❑ An Bdl internal study on the evolution of our statistical data warehouse (February 2022), highlighted the need to create an **advanced technological environment** capable of responding effectively to the changing needs of the exploitation of statistical data.
- ❑ The conclusions of a subsequent feasibility study in November 2022 paved the way for starting a project for the creation of an **Enterprise Data Lake** oriented to statistical data, which allows the storage of large databases even with heterogeneous characteristics (e.g. structured data, texts, social data) and that facilitates integration and exploitation activities by users through methods and tools typical of data science.

## Combined and shared use of structured reporting and unstructured data (2/2)

- ❑ The EDL integrates a portfolio of diverse components, technologies, tools and services to support a broad and diverse set of use cases. The identified architecture aims to provide the user with **an integrated environment that allows access to data in a "transparent" way with respect to their location and the management systems that support their life cycle**. Data search will be facilitated by a new **data cataloging system** which will constitute a unifying layer placed "on top" of the current dictionaries and database census systems
- ❑ BoI has been always supported a **broad sharing of banking and financial data between the various Directorates**, without distinction between institutional functions, on the one hand, and headquarters or local branches, on the other
- ❑ At the European level, **the path towards the data sharing among "trusted" authorities is still very complicated and long**

## Knowledge, skills and continuous training



- Recruitment
- Rich portfolio of **internal introductory courses** (e.g. basis courses in Data Governance, Python, R, SAS, introductory statistical methods)
- Possibility to attend specific **post-graduate university-level courses**, selected from Data Science Masters organized by high rank Universities
- The **Data Science Academy**
- Possibility to attend a **one-year Master degree in Data Science**, which was recently organized in collaboration with a University located in Rome with a very high rank in data science topics
- Partnership with the Academy**

# CONCLUSIONS



- numerous challenging stages still to face
- IA/ML can address some of them challenges ... but taking care of issues such as privacy, bias, fairness, security, interpretability, transparency/accountability. ML may produce complex and opaque results that in turn may be hard to understand. It is essential to ensure that models are interpretable and transparent, providing clear explanations for findings.
- Very important for CBs (particularly those of the ESCB) to cooperate and move in a synergistic way. Strengthening the capacity for joint action is the only way to overcome the current challenges related to the reporting. Initiatives like the creation of the Joint Banks Reporting Committee, the IReF project, the DPM Alliance (and DPM Refit), BIRD, the ESCB Network on machine-learning, all go in that direction.



**Thank you for your attention!**

**[laura.mellone@bancaditalia.it](mailto:laura.mellone@bancaditalia.it)**

**Deputy Head of Statistical Data Collection and Processing Directorate**



## APPENDIX

- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - Quality checks on granular banking data: an experimental approach based on machine learning by Fabio Zambuto, Maria Rosaria Buzzi, Giuseppe Costanzo, Marco di Lucido, Barbara La Ganga, Pasquale Maddaloni, Fabio Papale and Emiliano Svezia, March 2020.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - Stacking machine-learning models for anomaly detection: comparing AnaCredit to other banking datasets, by Pasquale Maddaloni, Davide Nicola Continanza, Andrea del Monaco, Daniele Figoli, Marco di Lucido, Filippo Quarta and Giuseppe Turturiello, April 2022.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - Learning from revisions: a tool for detecting potential errors in banks' balance sheet statistical reporting, by by Francesco Cusano, Giuseppe Marinelli and Stefano Piermattei, March 2021.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - Application of classification algorithms for the assessment of confirmation to quality remarks, by Fabio Zambuto, Simona Arcuti, Roberto Sabatini and Daniele Zambuto, July 2021.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - A decision-making rule to detect insufficient data quality: an application of statistical learning techniques to the non-performing loans banking data? by Barbara La Ganga, Paolo Cimbali, Marco De Leonardis, Alessio Fiume, Luciana Meoli and Marco Orlandi, February 2022.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers), The market notices published by the Italian Stock Exchange: a machine learning approach for the selection of the relevant ones by Marta Bernardini, Paolo Massaro, Francesca Pepe and Francesco Tocco, July 2021.
- ❑ Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers) - Institutional sector classifier, a machine learning approach by Oliver Giudice, Paolo Massaro and Ilaria Vannini, March 2020.