Eurofiling Conference 2023

Keynote speech by Silke Stapel-Weber, Director General Statistics and Chief Data Officer, ECB

The future of digital reporting

Starting from the past

My speech today will be about the future of digital regulatory reporting. Let me very briefly remind you of the importance of regulatory reporting and the key role it plays in ensuring a sound financial system. Regulatory reporting is the main way in which authorities collect data from financial institutions, and it is the basis for informed decision-making to fulfil the authorities' mandates in the areas of monetary policy, financial stability and supervision.

During the next 20 to 30 minutes, I will try to convey how the ECB is contributing to shaping the future of digital regulatory reporting and the vision I have for that future. And, as always when talking about the future, the past is a good place to start.

Around the end of the 20th century, the first moves towards digital reporting were taken as authorities started to digitise their paper-based collections. As some of you may remember, paper-based collections were on tabular format, that were designed to be filled in and consumed by humans. I have been around long enough to be able to show you an example that I worked on as a young statistician. The data collection template for ESA 95, looking very nice on paper and, later, even in Excel format. In the early days of the digital transition, data collection still used the same tabular formats. It was still based around the needs of human consumption. I am speaking about tabular formats where the title and the row and column headers used non-standardised labels.

As they were designed to be consumed by humans, these tabular format data collections contained a limited number of aggregated data points which focused on the immediate needs of the person who designed the template. Consequently, the data collections were only able to address the questions for which they were designed, meaning they had limited analytical capabilities. If any stakeholder subsequently identified something that required further analysis or a new crisis emerged, a new data collection would need to be defined and pushed through the European legislative process over a period of years and, finally, added to the reporting burden on institutions, in many cases introducing new redundancies and overlaps.

I am sure many of you will remember those times very well, and many of you can even think of ongoing data collections that still follow this pattern.

From template-driven to data-driven

In the process of digitalisation, it soon became clear that it was much more powerful to move to data-driven data collections. Instead of collecting data using tabular layouts, let us design datasets and let us identify all the relevant attributes. These datasets could still produce the tables that humans consume but would also significantly improve the analytical capabilities.

Data points were defined through a combination of well-defined attributes instead of the non-standardised labels used for template titles, rows and columns in old-school tables. Categorising the data by dimensions and attributes also facilitated an understanding and standardisation of data, as it was possible to identify any data point using a combination of attributes.

Data-driven collections should ease the burden on reporting agents, as collections will be better structured, easier to standardise, less heterogeneous and, initially, fewer in number, as it will not be necessary to frequently come back to reporting agents. Moreover, as these data-driven collections facilitate understanding among all stakeholders, they will also contribute to improving data quality.

Statistical authorities also started to use the Statistical Data and Metadata eXchange (SDMX) standard for defining and exchanging datasets. In the field of supervisory reporting, the first ideas around data point models were also developed at this time and were subsequently implemented by several authorities.

That was in the early 2000s, but the data collections were still mainly focused on aggregated indicators.

The financial crisis – towards granular data collections

Soon after, however, it was made evident by the 2007-08 financial crisis that more granular data were needed to better identify financial risks. While aggregated data collections are designed to answer specific analytical questions, granular data collections provide a more flexible framework in which data can be used to answer questions that pop up after the data collection was implemented.

Just think about the type of questions that appear after unexpected events such as the invasion of Ukraine. How will it affect the European Economy? Which banks/companies are most affected? What if the European Union imposes sanctions? Where are the Russian assets? Many of these questions cannot be answered with existing aggregated data collections. It is simply impossible to design a data collection that includes all the aggregated indicators needed to answer all the unexpected questions that could be asked in the future. The answer was to go micro. Thanks to the well-designed granular data collections implemented during and after the financial crisis, regulators were better prepared to get answers from the data to these unexpected questions. Granular data collections are more flexible, increase analytical capabilities exponentially and reduce the need for ad hoc data collections. You can imagine them as tiny Lego bricks which we can combine to get new answers without having to define new collections.

At the same time, the evolution of computers and networks opened up possibilities that were not even thinkable years before. Processing volumes and speeds increased, while processing costs fell. New technology made it possible to execute large queries, processing millions of data points and getting responses in seconds.

Furthermore, granular data collections initially require less effort from reporting agents. Yes, the volumes are much bigger, but the data are more homogeneous, closer to the operating systems of reporting agents and require fewer transformations.

Thanks to these developments, authorities have been collecting more granular data with the objective of being better equipped to respond swiftly to the next crisis, whatever it may be. Examples of such data collections at the ECB are AnaCredit, which collects data on individual loans, and Securities Holdings Statistics.

The present – where are we now?

Now it is 2023 and in Europe we have a mix of all the data collection types I have described. Over the last 25 years, multiple data collections have been defined by various authorities at European and national level. Some of these have a template mindset and collect aggregated indicators, while others collect granular data.

Current data collections were largely established in silos. Although a certain level of harmonisation was aimed at and several countries have moved down the path towards standardising and integrating data collections, from a global and certainly a European perspective the solution is still suboptimal. Data dictionaries and data models are not standardised across domains or even across countries within the same domain. This makes the work of combining and integrating data from different datasets challenging, limiting the analytical capabilities. In addition, it requires significant effort from reporting agents due to the need to report to different authorities in different countries using different codification schemes and different technical formats.

The future – integrated and real time

Can we imagine a future in which data are integrated and available to all authorities, minimising the burden on reporting agents? What do we need to make this future real?

The answer is collaboration. European authorities need to jointly:

- standardise the way in which data are described by defining reference data and common dictionaries;
- define a common business workflow, putting an end to doing everything 20 times.

This will also enable process automation and machine-to-machine communication.

We at the ECB, together with other authorities, are convinced this is the direction to go and we are working on different initiatives to achieve this desirable future.

I would like to share some information in particular about two ongoing initiatives:

1) the Integrated Reporting Framework (IReF);
2) the Banks' Integrated Reporting Dictionary (BIRD).

IReF

Our first strategic initiative is the Integrated Reporting Framework (IReF), which will integrate the Eurosystem's statistical requirements for banks into a single standardised reporting

framework directly applicable across the euro area with the aim of achieving maximum harmonisation of statistical reporting. A common data model and common data dictionary will be applied, as well as a common data collection framework. All requirements will be part of one legal act without overlaps.

IReF will thus address some of the integration challenges previously mentioned, helping banks with their data reporting obligations by standardising them, while also reducing redundancies. In addition, having one reporting framework independent of the country in which a reporting agent is located will benefit banks operating in multiple countries. In short, it will reduce the reporting burden.

IReF will also facilitate the work of authorities in processing the data as they can rely on common transparent processes and data sharing, facilitating cross-border comparison, improving data quality and reducing time to market.

Finally, analysts and policymakers will also benefit as IReF will facilitate the combination of data coming from different datasets in a seamless and flexible way. They will be able to get answers to new questions more easily and more quickly, requiring fewer ad hoc inquiries and reducing the time needed to gain insights and take decisions.

By now you will have probably concluded that IReF is a key project for achieving the objective of collecting integrated statistical data. Its principles are define once, regulate once, report once. IReF is a very ambitious initiative to which the Eurosystem is strongly committed and will see the light of day in 2027.

BIRD

The Banks' Integrated Reporting Dictionary (BIRD) is aimed at the collaborative development of a harmonised data dictionary and a harmonised data model that specifies how data can be extracted from banks' internal IT systems to generate the reports required by authorities. BIRD also provides transformation rules which can be applied to the data extracted from banks' internal IT systems to produce these regulatory reports.

In essence BIRD describes how the "define once" and "report once" principles can be implemented in the banking industry. on the banking industry site.

It is important to highlight that BIRD is a collaboration between authorities and banks in which the ECB, the European Banking Federation, national central banks and several participating commercial banks are working closely together to develop and maintain BIRD. It is a "public good" and its adoption is voluntary.

BIRD is, however, broader in its scope than IReF, paving the way to even wider integration of statistical, prudential and resolution reporting.

In the context of this wider integration, let me share a few thoughts on how European authorities will collaborate in a new Joint Bank Reporting Committee.

Collaboration among European authorities – towards a Joint Bank Reporting Committee

Establishing IReF and BIRD will not be the end of our mission, since statistical integration is one very significant and important first step in a bigger journey aimed at integrating all regulatory reporting by banks, including not only statistical but also prudential and resolution reporting obligations.

In this context, we have been collaborating at the European level through the Informal Coordination Group on Integrated Reporting by Banks which has also been preparing the ground for a new Joint Bank Reporting Committee (JBRC) that will foster collaboration among relevant European institutions and authorities. The Committee will involve European and national authorities and other stakeholders (the banking industry and the wider public). It will provide support and advice on the development and implementation of an integrated reporting system aimed at increasing efficiencies in the reporting process and reducing reporting costs.

The JBRC will provide non-binding advice and assist in translating existing and new user needs into integrated reporting requirements. It will also foster the development of a common regulatory data dictionary and a common data model for prudential, resolution and statistical reporting. And it will advise on ways to enhance coordination and data sharing among authorities to avoid overlapping and duplication of data requests.

In a nutshell, the objective of the JBRC is to facilitate coordination among authorities and pave the way for a single semantically and syntactically integrated data dictionary in the area of banks' regulatory reporting with the objectives I already mentioned of reducing the reporting burden, enhancing analytical possibilities and bringing new insights to policymakers.

In the context of European collaboration, I would also like to mention the ongoing collaboration between the European Banking Authority (EBA), the ECB and the European Insurance and Occupational Pensions Authority (EIOPA) aimed at establishing a single methodology for modelling reporting requirements. A group of experts from European and national authorities analysed the Data Point Model (DPM) and concluded that the new version of the DPM (the DPM refit) is a valid candidate and could be used to describe the IReF requirements. Pending the defining of DPM governance arrangements to ensure that the maintenance and evolution of the DPM will cover the needs of the EBA, the ECB and EIOPA, we are working towards using the DPM to describe the regulatory requirements of the three authorities.

Timeliness – real time

I don't want to finish today without touching on another aspect of the future of digital reporting – timeliness. Making the most of our data and taking timely decisions implies the timely availability of data. Whatever value can be extracted from our datasets will become worthless if it is not identified and used at the appropriate moment.

Very recently we have read in the news about Silicon Valley Bank (SVB). According to reports, Twitter fuelled the run on SVB in real time. Just one tweet seen by 2.4 million people and retweeted 3,500 times caused panic and the fastest withdrawal of deposits ever seen. According to SVB's former Chief Executive Officer, USD 42 billion in deposits were withdrawn

from the bank in ten hours or “roughly USD 1 million every second”. People read news and tweets in real time, people can withdraw deposits in real time, people can access their bank statements in real time using their smartphones, but regulators get quarterly or monthly reports from banks after weeks of delay. Can regulators take timely decisions under these circumstances or are they just running behind the curve?

Can you imagine the benefits of having data from banks in real time? Supervisors would be able to implement alert systems that could flag any bank’s health problems immediately with the possibility of reacting quickly and addressing the problem at an early stage or even of running anticipatory analyses.

Real-time integrated data collection can only be done if it is based on mutually agreed definitions and requires no further intervention with consequent delays. It is therefore important to eliminate silos and reach a consensus based on achieving a high-performing and stable financial system and a level playing field among institutions.

I hope you can all now see that the roadmap I outlined before is a pre-condition for leveraging and taking advantage of the benefits that real-time data could bring. Integrated data collection, IReF and the JBRC are prerequisites for exploiting the full potential of real-time data.

It is also important to mention that we may need to change some paradigms in the area of data collection. For real-time data collection we could complement the push approach, where reporting agents push data to the regulator, with a pull approach, where regulators can also pull data from reporting agents when needed. We could move from regulatory reports to an alerts-based system based on events or application programming interfaces (APIs) acting as sensors that allow regulators to monitor the health of our financial system on a continuous basis. We could move from regulatory reports at a fixed frequency to continuous streams of data flowing from reporting agents to regulators.

The technology is already here, and the only question in my view is how long it will take to change the mindset of regulators and reporting agents and implement this vision.

At the same time, I don’t want to be naïve and it is important to remember that we should not expect all data to be collected in real time. The basis of many regulatory indicators is the accounting process, which by definition requires a number of judgement calls on classifications, valuations, etc. This requires time and preparation and will always involve some lag, while other indicators could be collected in real time and at much higher frequency, such as, for example, liquidity indicators.

In the future we will probably see a combination of all these options – regular reports, real-time data streams, APIs, sensors, alerts, etc. – addressing all the different needs and constraints. Anyhow, I am sure that, if we want it to be effective, it should be based on standardised, harmonised, integrated data at the European level and collaboration between European authorities. The projects and activities that I have mentioned today – IReF, BIRD and JBRC – are all heading in this direction and preparing the ground for the hopefully not so distant future.