

Whitepaper

Author

Eric Leman

Director, Solutions Specialist

Contact Us

Americas
+1.212.553.1653

Europe
+44.20.7772.5454

Asia-Pacific
+852.3551.3077

Japan
+81.3.5408.4100

Regulatory Big Data: Meeting the Costs & Challenges of Granular Reporting

Highlights

- » Regulatory authorities require ever-more granular and standardized data. The Bank Integrated Reporting Dictionary (BIRD) framework, available as a “public good,” sets out clear definitions of data attributes that are required to meet Analytical Credit Datasets (AnaCredit) reporting.
- » Banks are currently ill-equipped to meet these regulatory demands in an efficient and cost-effective way. Ironically, one of the key short-term reasons is the need to meet a plethora of regulatory deadlines, though the underlying problems are the heterogeneity of data sources, lack of data availability, and poor data quality.
- » It is particularly challenging for banks to collect the counterparty information demanded by AnaCredit reporting to the European Central Bank (ECB) and multiple national central banks.
- » It was always possible to process big data, but until recently it was too expensive to invest in it for any one project. The difference is that today, the technology for a big data infrastructure is proven and available in flexible configurations with predictable costs. It provides functionality for the storage, management, and analysis of large amounts of structured and unstructured data, quickly, reliably, and flexibly. These solutions can be on premise, but they can also be hosted and managed as software as a service (SaaS) – which means banks will be able to delegate IT related tasks completely to a third party.
- » If banks want to get unstuck from the regulatory granular reporting once and for all, now is the time to invest in big data technology.

CONTENTS

- Highlights01
- Introduction03
- The regulators' perspective03
- The banking IT perspective03
- The business must take the lead04
- AnaCredit and the data collection challenge04
- Counterparty information might not be available in-house05
- Using technology to solving AnaCredit transactional reporting challenges05
- Benefits beyond AnaCredit06

Introduction

Reporting on large amounts of data poses a significant challenge for many financial institutions as they seek to meet regulatory demands. A survey of 35 risk and compliance professionals attending a recent Moody's Analytics webinar revealed that a third (66%) regarded poor data quality as a serious barrier to the adoption of regulatory big data projects. More than half (51.4%) named data availability, while the same number cited IT infrastructure issues. These three challenges rated considerably higher in terms of difficulty than challenges relating to the regulation itself. For example, national reporting requirements were cited by only 23% of participants and timeline to comply was cited by only 34% of participants. 43% of respondents stated that their firms were actively using big data for regulatory purposes, while 40% have plans to invest in big data for regulatory purposes. Only 17% had no such plans. The overwhelming majority (66%) are currently using in-house developed platforms for dealing with regulatory data collection.

Since the financial crisis of 2008, regulators have sought the most robust, most current and most transparent data on which to base their decisions. The survey results confirm our view that many banks rightly regard big data technology implementations as the best approach to satisfy these requirements. But many also see major challenges along the way, both in the near term, to clean, reconcile and standardize data, and in the longer term in the way they are supervised. These issues are once again coming into sharp focus with the AnaCredit deadlines approaching.

AnaCredit is the European System of Central Bank's (ESCB) project to establish data sets containing detailed information on individual bank loans in the euro area. Data sets will be harmonized between national central banks (NCBs) across all member states. Each NCB has its own local requirements in terms of the attributes that must be reported, and the formats in which the reports must be delivered.

The regulators' perspective

Traditionally, national regulators and the ECB have collected aggregated data for monetary policy and financial stability purposes, as well as data by institution for supervisory purposes. One of their greatest challenges lies in the enormous heterogeneity of data between the various institutions, countries, economic sectors, and market segmentations. This heterogeneity eludes meaningful analysis using aggregated data and averages. Regulators are therefore demanding greater

granularity to be more specific, for example by adding dispersion measures to the aggregates and enabling analysts to analyze the underlying causes and distributions of any potential risks that they identify.

The ECB has therefore initiated several projects to provide such detail: granular information on loans (AnaCredit) security holdings statistics (SHS), transactions, the unsecured money market (EBOR), and derivatives (EMIR). This trend implies a requirement to manage vast amounts of data. For example, AnaCredit itself seeks to build a data set covering some 60-70 million loans between legal entities and considers 80-90 attributes for each loan.

Standardization is required to make these massive volumes of data transparent and easy to understand and analyze. The more granular the data becomes, the more important it is to ensure quality and consistency. Therefore, the ECB has initiated the Banks' Integrated Reporting Dictionary (BIRD) project. BIRD offers standardized definitions of which data must be extracted from banks' internal IT systems, and which transformation rules must be applied, to derive the specific final regulatory figures demanded by the authorities. BIRD is a voluntary initiative, available as a "public good," a form of "crowd intelligence" pursued by the ECB, seven national central banks, and approximately 30 commercial banks across Europe. BIRD definitions and rules are well-developed and have already been published for AnaCredit. The first data collections according to BIRD are scheduled to be posted in September 2018. Standards for security holdings statistics (SHS) are also at an advanced stage of development, with the involvement of nearly 30 of the most significant firms in the Euro area and will be extended to around 130 institutions next year. Money market data is at an advanced stage of development for internal analysis, while EBOR standards are at the early stages and to be published in 2020. A data quality project is now underway for EMIR reporting on the derivatives market.

The ECB recognizes the need to invest in big data technologies and the requisite data science skills required to analyze the massive volumes of granular information that these projects generate.

The banking IT perspective

In parallel with addressing regulatory demands, big data offers the potential to solve a wide range of internal challenges faced by banks. To date banks have traveled a long way on their risk management and regulatory compliance journey.

When compared with the detail and rigor applied to internal accounting, banks have evolved from scarcity of data and a lack of rigor at the turn of the last century, to generating vast amounts of information for regulatory and internal purposes today. However, it is our view, shared by many banks, that they currently find themselves in the worst possible situation: the requests from the regulators are considerable, but the ability of banks to respond to them with detailed and standardized reports is varied. The heterogeneity of IT landscapes and data structures not only between banks, but also within banks, is a significant inhibiting factor. The end game for banks is to make IT systems leaner and information more transparent and readily available in data lakes that are accessible to regulators, with all the attributes required for the projects described in the preceding section. This change will bring benefits for the banks themselves in terms of better decision-making. But there is a long way to go.

Getting there involves the acquisition of new big data technologies and skills, and combining these elements with the older disciplines of computer science, statistical analysis, and risk management. This process would be challenging on its own, and will be all the more so because banks are facing short-term regulatory compliance deadlines. Bank IT teams recognize that there is considerable overlap between the data and attributes needed to meet internal and regulatory reporting requirements. There are benefits of merging teams and putting data into a common store as part of a more coordinated information management strategy. For example, senior managers have greater confidence that reports are consistent when they are based on the same data, and there will be less need for reconciliations.

Thus, the benefits of an integrated data-driven strategy are evident. However, bank IT teams still face considerable barriers:

- » Communicating the benefits to senior managers
- » Justifying the cost of new IT investments, even if they are cheaper than traditional database technologies
- » Limited direction from the regulators
- » Data for regulatory reporting is "siloes" in many separate stores
- » Focus on current regulatory milestones ("box ticking") makes it difficult to construct a long-term strategy
- » Lack of appetite for organizational transformation – merging functions and departments does not necessarily mean job losses, but it certainly entails considerable HR planning and new skills must be brought on board

- » Lack of sponsorship: a successful data-driven strategy cannot be driven or owned by IT – leadership must come from the functions that understand the data

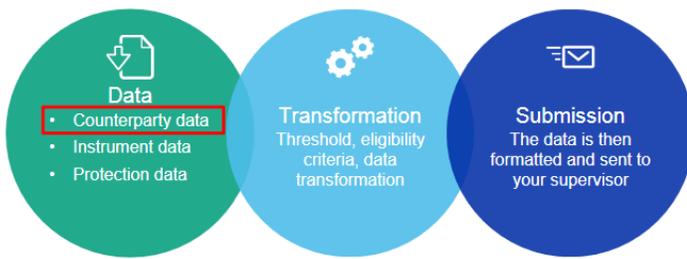
The business must take the lead

The last barrier is the most significant. IT owns the "plumbing" but is in no position to design effective data structures that meet the needs of the business. The relevant heads of the business functions must be mandated to come together and articulate the needs to IT. Further high-level and coordinated guidance from the regulatory community– which is being promoted through BIRD – could provide a greater stimulus for banks to grasp the nettle. That said, the issue for banks is not "regulatory big data" but "big data" period. If banks can succeed in managing data strategically for risk and finance functions, meeting present and future requests from the authorities will be much easier.

AnaCredit and the data collection challenge

Institutions in all Eurozone states are obliged to participate in the AnaCredit project. While member states of the European Union outside the Eurozone can voluntarily participate, Denmark and Sweden have announced their intention to do so. As shown in Figure 1, AnaCredit focuses on the collection of granular credit data and its transformation and submission in the formats required to address the data needs of ESCB members.

AnaCredit does not serve traditional supervisory purposes. Its central purpose is to identify data about the issuers of any credit instrument equal to or exceeding €25,000 extended to a legal entity (for most national discretion). Collecting counterparty data immediately poses a serious practical challenge for banks because, while some of the information might exist for regulatory reporting purposes in different silos, this heterogeneity is the source of inconsistencies at the granular level. Therefore, to standardize and harmonize the data, banks would need to ask individual clients to provide information, which is no easy task. This task is further complicated by the requirement for banks to report for all branches and most subsidiaries, while banks in different jurisdictions face slightly different reporting requirements depending on the local regulator.



“24 data fields requested by ECB for Counterparty data”

Overview > Source ORBIS of Anacredit > Counterparty data > Mapping tables > Automated process

Figure 1: Overview of the AnaCredit collection, transformation, and submission process.

For example, a French holding company would need to report the exposures on its loans in France, together with those of its German and UK branches to the French national central bank, the Banque de France. The German branch would also report its own activities to the German Bundesbank. The German subsidiary then would do the same. Under the terms of AnaCredit, the German Branch is an “Observed Agent” for the French holding company, and a “Reporting Agent” on its own. A similar scenario applies to branches and subsidiaries in other participating countries. In each jurisdiction, the rules are slightly different for the granular risk indicators, like probability of default, off-balance sheet/on-balance sheet amounts, collateral, guarantees allocation, and so on. Moreover, the timelines vary by jurisdiction.

Counterparty information might not be available in-house

Reliable, high quality and consistent data is the essential first building block for AnaCredit, it is required in vast quantities, and it might not be available in-house. The ECB requires information on 24 counterparty attributes for the AnaCredit reporting process. These attributes include company identifier, immediate parent, ultimate parent, name, legal form, institutional sector, economic activity legal status including details of any legal proceedings, enterprise size, number of employees, balance sheet, annual turnover, accounting standard, legal address, and so on. This information is highly granular and standardized, and the ECB gives clear guidance on the standards for its coding and presentation.

Even when this information is available in-house, much of it is in different repositories, and breaking down the internal walls to bring it all together in a single data repository like a data lake would be a huge task. Moreover, AnaCredit reporting will

be done on a monthly schedule – with a five-day deadline for reporting on new clients – and a bank’s client universe is constantly in flux. Some national discretion also require banks to report “on change” as well as providing regular reports. While some information can be static or only change on a predictable (for example, annual) basis, other information (for example, legal address) can change in a way that cannot be predicted. With the high level of M&A activity, ownership data is volatile.

Using technology to solving AnaCredit transactional reporting challenges

A comprehensive AnaCredit solution would need to answer the big data transactional reporting challenges. To do that with confidence, the solution, as a minimum, would need to have:

A modular platform. The solution containing data transformation, data management, calculation, and regulatory reporting blocks would be able to respond to AnaCredit and beyond.

Data lake and Big Data technology. This technology would allow banks to consolidate, catalog, and reference all their structured and unstructured data, and store it in one convenient place for AnaCredit and future use. The system should use a “schema on-read” approach, which means you do not need to format the data in the data lake. Instead, logical mapping would enable the data to be extracted according to the BIRD model on demand and can equally be mapped onto other schemas.

Solutions using Hadoop, Spark, and Yarn and other new, best-in-class technology, provide functionality for the storage, management, and analysis of large amounts of structured and unstructured data, quickly, reliably, and flexibly.

Implementing the solution as software as a service (SaaS) adds the benefit of outsourcing the hosting and managing of such reporting to the service provider. That means banks do not need to worry about recruiting hard-to-find expertise in these new technologies. With SaaS, only the bank’s business users interact with the software.

Benefits beyond AnaCredit

AnaCredit reporting is an obligation, but there are many benefits to be derived from implementing an automated, streamlined solution to comply with AnaCredit (or other granular, transactional reporting requirement) and set the foundation for future reporting:

Comparability

Data can be easily compared between two banks (for example, two subsidiaries within a holding company, or two banks in the same peer group) or two business units (to compare internally).

Reconciliations

Banks are empowered to ensure consistency and reconciliation with Basel III and other regulatory reporting processes (for example, according to FINREP and COREP guidelines) by using common data definitions and calculations.

Internal on-demand reporting

A data model which goes beyond AnaCredit can be adopted by all. It simplifies and accelerates the creation of on-demand reports for risk and finance.

Reuse of data

Data held in the data lake is “data at rest” so it is always consistent and can be reused in advanced risk engines (for example, for simulations and stress tests). These engines can be supplied by Moody’s Analytics or developed in-house.

Auditability through data lineage

One of the major advantages of this big data-driven solution is that it enables data lineage and therefore reinforces auditability and BCBS 239 compliance.

Future cost reduction

The cost of compliance with future regulatory requirements will be substantially reduced as this solution can be used for future reporting frameworks.

CONTACT DETAILS

Visit us at moodyanalytics.com or contact us at a location below.

AMERICAS

+1.212.553.1653

clientservices@moody.com

EMEA

+44.20.7772.5454

clientservices.emea@moody.com

ASIA (EXCLUDING JAPAN)

+852.3551.3077

clientservices.asia@moody.com

JAPAN

+81.3.5408.4100

clientservices.japan@moody.com