# The Magic of Data Lineage

*Making regulatory data the best available business information*

*by Giovanni Butera*and Peter Tierney***

In the financial services sector, the key driver of Data Lineage — the crucial ability to know where data came from, track it through various systems and learn how it changes as it moves through systems — has been the Regulatory Reporting environment. With the number and diversity of sources, the astounding number of data items and the complexity of systems, Data Lineage analysis has always been a daunting challenge.

But there is also magic to this challenge.

Making regulatory data available to the business eliminates the need for business units to collect and review data that regulatory and compliance functions have already gathered and reviewed for quality. The regulatory challenges of tracking data lineage have actually turned regulatory data into a wealth of quality information that can drive growth and create business value.

*Dr. Giovanni Butera is acting CEO and Managing Director, Head of Data Management and Innovation at Nixora Group. M: +61 410 860 036 E: giovanni.butera@nixoragroup.com.au

**Peter Tierney is General Manager-Asia at AxiomSL, M: +65 8127 0944, E: ptierney@axiomsl.com

**AxiomSL** combines deep industry expertise with an intelligent data management platform to deliver regulatory and risk reporting, liquidity, capital and credit, operations, trade and transactions, and tax analytics. Our global footprint spans 70 regulators across 50 jurisdictions, surveilling more than 4,000 regulatory filings. We currently serve national, regional and global financial institutions with more than $39 trillion in total assets. For more information, visit www.axiomsl.com or connect with us on LinkedIn.

**Nixora Group** is the implementation partner of Axiom SL in Australia and New Zealand. Nixora also partners with Oracle, IBM, ASG and Moody's Analytics for the implementation of their solutions in financial institutions and large organizations.
If you would like to know how to take full advantage of Nixora proposition, call us today for a consultation and for a proof of concept. Email us at contact@nixoragroup.com.au, connect with us on LinkedIn, follow us on Facebook or visit our web site nixoragroup.com.au.

## Introduction - An excursus on regulatory compliance

Until a few years ago, most Financial Institutions (FIs) took a predominantly manual approach to Regulatory Reporting (RR), usually via internal spreadsheets and macros. These tools provided the comfort of leveraging on established procedures across the organisation that enabled business users to visualise, organise and report the information.

The activities of RR were focused on end results: as long as the reports were completed and issued on time, executives paid little attention to the actual process of compiling all of the data. The business challenge was to improve efficiency of a process which, by its own nature, was cumbersome and complex. The activities of collection, manipulation and aggregation of data elements for various regulatory reports were typically spread across the organisation in Excel files, CSV extracts and Access databases (the so-called end-user-computing, or EUC tools), none of which appeared in formal IT architectural maps.

Such an approach was inherently limited in terms of monitoring and addressing changes, particularly when it came to extensibility and scale. For example, when a financial institution operates in several jurisdictions and is obliged to comply with multiple regulations. It was also evident that regulations would continue to evolve and the need for agility put immense strain on tactical, stop-gap solutions. Moreover, practices of freely changing or adjusting data, formulae and formatting heightened the risk of human errors and inconsistencies, making it very difficult to establish a clear audit trail.

The "Principles for Risk Data Aggregation & Reporting" issued by the Basel Committee in 2013 (BCBS 239) opened the door to substantial change of the regulatory approach. It contains 14 principles covering four areas: governance and architecture, risk data aggregation, risk reporting, and supervisory review. Taken together, the principles urge banks to clean up the fragmented data standards across their business lines and legal entities, which hitherto had prevented them from achieving a comprehensive view of their risk exposures.

The BCBS 239 was a seminal regulation[1] and signalled an important change in the approach adopted by regulators. Regulatory Reporting was historically viewed by FIs as just a matter of putting numbers in a box and submitting the reports on time. Today, executives from different departments and levels of responsibility are called to attest to the accuracy of figures, and must be able to demonstrate that the processes implemented to access data and produce reports are robust.

[1] Regulators in Singapore, Australia and across the globe continue to step up their pressure on banks to achieve more transparency and consistency. During the past two years, global regimes such as the Basel III's LCR (Liquidity Coverage Ratio), NSFR (net stable funding ratio), IFRS 9 (the 9th International Financial Reporting Standard) and CRS (the Common Reporting Standard), including updated reporting requirements for Large Exposures, IRRBB (Interest Rate Risk in the Banking Book) and SA-CCR (Standardized Approach for Counterparty Credit Risk) were rolled out, raising the bar for reporting by financial institutions.

## Changing regulatory environment

### Financial Institutions' Perspective

Financial institutions responded to the challenge of enhancing the quality of RR data in two stages.

In the first instance they focused on enhancing their ability to locate and audit the activity taking place in and across multiple files. In many cases the scope of EUC tools extended beyond its original purpose, as these tools were essentially impractical for managing the increasing complexity of the new regulatory reporting rules.

Inevitably, as the business challenge moved from timely completion and delivery of reports to demonstrating that numbers are correct and IT systems are capable of producing reports quickly and reliably, a second instance emerged. FIs then gravitated towards the creation of data hubs (or data lakes) with the inspiring idea of generating regulatory reports directly from central data sources (i.e. the "single source of truth").

As new RR requirements swept across the sector, there was an increasing need for FIs to integrate data from a multitude of systems, produce much more granular information at higher frequencies while also providing full traceability through to the underlying source data – all this to achieve compliance. Many FIs took this wave of change as an opportunity to step back and evaluate alternative business models[2].

### System Providers' Perspective

IT system providers and software vendors also saw an opportunity to develop specific regulatory reporting solutions to enable FIs to meet the new challenges. Regulators and FIs around the globe have shown explicit support for this model because of the key advantages of simplifying, automating and streamlining the overall RR activity[3].

Although the offering of RR solutions and services differ substantially from one provider to the next, the range of RR services managed through an external provider can cover all aspects of the RR operations, from technical aspects (Application Service Provider - ASP) to business operations (Business Process Outsourcing - BPO) of the RR activities, within both the Run-The-Bank framework as well as for amendments in regulatory reporting within the Change-The-Bank framework.

## Tracking Data Lineage - Data Lineage in a RR environment

Nowadays, Financial Regulators are not just looking at collecting data from regulatory reports, they want to understand the models underlying the reports and whether the data used in the models is both accurate and consistent. What becomes crucial therefore is the ability to accurately track and understand what happens to data as it passes through diverse processes from one end of the organisation to the other, from origin to reporting.

### A Single Version of Truth

The BCBS 239 *in primis* made FIs realize that they need a "single version of the truth" in order to (1) effectively prove to regulators how they arrive at their numbers, and (2) allow for reconciliation between reports.

Ideally, the desired outcome for FIs is to be able to click on any element of a report that a regulator has queried and see the raw data as well as the transformations and calculations that were used to create that element. The sheer number and diversity of data sources, data items and the complexity of applications make data lineage analysis a daunting challenge[4]. But without data lineage an organisation has little or no visibility on the consistency between reports, processes and data definitions.

Most data appearing within a regulatory report was generated for the purpose of running the FI, not reporting to the regulator. This means that the journey of a piece of data from a front office system to the regulatory return might involve hundreds of steps, logical aggregations, transformations and calculations across multiple systems. Organisations can start losing track very quickly as to why and how a piece of data appeared in a regulatory return, and what happened to it in the journey.

A major side effect of not being able to track data lineage is the inability for an organisation to quickly act and respond to critical situations. For example: when supervisors ask for clarification on the origins of specific data elements, reasons for changes to data from a previous reporting period, or clarifications on the consistency of such changes in relation to data elements in other reports. In many instances the outcome of poor data governance is not just stress and panic, more often it is the loss in competitive advantage, hefty compliance fines and reputational impact.

The crucial questions are summarised by the 7 Ws:

- "**W**here did the data originate?",
- "**W**hen was it captured?",
- "**W**ho is using the data?",
- "**W**hy is it meaningful?",
- "**W**hat are the data points it has passed through?",
- "In what **W**ay has the data changed as it passed through those points?", and, importantly,
- "Ho**W** good is the quality of the reported data?"

To answer these questions, one must be able to trace up and down the data stream, from one data point to another without gaps, and be able to properly trace any errors back to their source and ensure that reported data is sufficiently accurate and complete.

## The Benefits of Data Lineage

The overall benefits of data lineage from a regulatory and compliance perspective are straightforward. It simplifies vast amounts of complex information and provides visibility of the data's flow through various data points from source to destination. This provides FIs with:

- **The ability to actively monitor regulatory changes and to provide updates within their systems** thereby allowing organisations to meet ever-changing regulatory requirements on an ongoing basis across multiple jurisdictions

- **A workflow management approach** that monitors and determines the process from start to finish, from automating loading of data, execution of calculation processes and preparation of the reports, with logs of approved or declined changes, including report level validation rules and electronic delivery (XBRL, XML, etc.)

- **A full audit trail**, including the ability to drill-up and down the underlying data to provide proof of quality, and dashboards for an instant view of a firm's regulatory reporting status

***"Even though governance and regulatory compliance are the major drivers for tracking data lineage, that does not mean that the complex exercise of tracking RR data is useful merely for ticking off the compliance box."***

Because regulators are interested in gaining information about certain aspects of the business and want to make sure such information is reliable and trustworthy, RR data has become the most reliable business information available for making relevant operational, tactical and strategic decisions. Data lineage allows greater understanding of the relationships between data items across reports, individual lines of business and the overall enterprise, which means the common argument for the dichotomy between regulatory and business data no longer stands.

Indeed, the most sophisticated FIs have organised their regulatory reporting systems to allow further processing of all of its data, and to look for ways to effectively generate value from such data. While the crucial function of a RR system is to certify the quality and lineage of all data, it should ideally be able to group and extract such information to support strategic business decision making.

## Regulatory data and business value

Generally speaking, the primary advantage of making RR data available to the business helps eliminate the need for duplication through the data collection process. Data that risk, regulatory and compliance functions have already gathered should be shared across the organisation and re-used as required since this data has been reviewed for quality and comes from a "single source of truth", with full traceability and auditability across the entire data river.

Recent overhauls of RR Standards in Singapore and Australia have been consistent with international initiatives to enhance the breadth and quality of statistical and economic data gathered from the financial services sector. The MAS (Monetary Authority of Singapore) 610 and the Australian Prudential Regulation Authority's EFS (Economic and Financial Statistics) overhauls introduced several new reporting requirements, along with a myriad of requirements introduced by the Risk and Compliance (R&C) regulations. These represent the bulk of a financial institutions' reporting information[2]. FIs in these jurisdictions needed to broaden extensively the range of information provided in areas such as lending, deposits, fees, counterparty information, and derivatives, residency, sector and industry classification.

Specifically, the information required by supervisors for FIs to comply with their RR standards range from balance sheet data (including foreign currency, securitisation, derivatives), with a focus on financing and funding, (including interest rate/cost of funds, loan application/origination, deposit/security characteristics), to securities financing (such as collateral). But also required is information such as residency status, loan purpose, loan application, loan serviced with foreign income, first home owner grant, cost of funds, offset accounts, and excess repayments.[3]

Such regulations have also increased data granularity. For example, income/expense and asset/liabilities data are more disaggregated and specific, and while counterparty' details are required on most forms (e.g., residency, related party), including the breakdown of counterparty by economic sector, industry classification has been extended to include greater details on the financial sectors and business size classification.

Most of the RR information can be used to generate business value when analysed in a specific business context by using advanced data science capabilities (AI&ML) in conjunction with specific additional internal or external information. In fact, by integrating RR data with forward-looking data analytics and machine learning algorithms, the more forward-thinking FIs are able to realize internal operational efficiencies, business advantage and cost savings on top of satisfying regulatory reporting needs. These FI are also able to eliminate the duplication of business and data management activities, thereby freeing up resources to work on revenue/value generating activities, including opportunities for cross selling/upselling, creation of new products, defining target strategies, enhancing stress testing and macroeconomic modelling for analysis and forecasting.

[2] Key components of these enhancements include making sure that Basel III requirements are appropriate for their markets and optimizing FIs core balance sheets reporting.

[3] In most cases, regulations specify the accuracy thresholds that information provided must achieve to comply with the standards. FIs must ensure that the scope of an internal audit includes a review of the policies, processes and controls put in place by management for compliance with the reporting standards.

Nonetheless, RR obligations are only a modest part of the overall set of the regulatory data FIs have to provide to supervisory authorities. The largest set of regulatory reporting data requirements are for Risk and Compliance (R&C) — that is, all risk reporting data including Pillar III disclosure and all disclosure within the Basel II framework (credit, market and operational risk), AML regulation, Compliant logs obligations, Financial Crime and Fraud regulation.

For example, Pillar III reporting obligations, including detailed quarterly series of information about the risk of the bank portfolio, grouped by client segment, geographic area, and products. This information combined with economic and other factors may reveal valuable patterns and interactions that are useful from a risk portfolio management perspective. This data—in addition to transactional data—is also readily available and can help target marketing campaigns based on improved understanding of customer risk patterns and segment behaviour, enhancing cross and up selling.

As in the case of Anti Money Laundry (AML) data analysis, the more progressive FIs have enhanced its Know Your Customer (KYC) program and, as part of this effort, created a dedicated data science team to analyse accounts and identify suspicious transactional patterns. They were able to demonstrate to the regulators that the accounts were compliant with AML regulations. Rather than stop there, these institutions took advantage of insights from their KYC analysis to generate additional value for the business: they used insights on customer travel patterns and foreign transactions to make targeted travel insurance and specific credit card offers.[4]

The value that comes from tracking data lineage and sharing RR and R&C data with business units is enormous. Its boundaries seem defined only by limitations in the collaboration between various business units and R&C departments, and in the capabilities of wise and experienced business analysts.

[4] The Extra Mile: risk Regulatory and Compliance data drive business value, PWC, 2015.

## Tune in to the magic of data lineage

FIs all over the world have converged or are converging towards the integration of data from different critical systems into a centralised data lake or hub to address the so called “single source of truth” requirement for regulatory data

Understandably, leading financial institutions have been exploring strategic and business uses of risk, regulatory and compliance data they already own to drive operational efficiencies, lower costs and increase revenues.

The ability to gain valuable insight from RR data/R&C data and integrating this information with business operations cannot realistically be implemented overnight. However, the organisations that have tracked data lineage and shared RR data and analytical results across the organisation are realizing a number of benefits. Innovative FIs see regulatory compliance as the trigger for:

- Innovation of the traditional approach to data management, and an inspiration for implementing best business practices;
- Customising products based on enhanced knowledge of each customer risk profile and risk appetite;
- Identifying new customer segments and potential new products through a better understanding of customer risk patterns and behaviour, enhancing cross and up selling.

The availability of a data tracking tool for an integrated information model that supports full forwards and backward navigation for each individual data element in any report has helped cutting-edge FIs to change their reporting practices. This has had a major impact on how these institutions manage their balance sheet, credit risk, customer relationship and other practices.

More importantly, it has impacted on how FIs use regulatory data while complying with a vastly expanded set of new and evolving reporting requirements.

The magic of data lineage is evident in how leading financial institutions have capitalised on the information assets they already collect and analyse to successfully leverage risk, regulatory and compliance information to drive profitable growth.

Many of these leading institutions faced challenges as they looked to repurpose risk, regulatory, and compliance data, but considering the enormous investments and efforts already made, re-using this data to create business value is where we see the competitive edge in the drive for efficiency.

## Case Studies – the use of regulatory data for business purposes

While a lot of attention has been focused on how to meet the challenges of new regulatory reporting requirements, banks are becoming aware of the potential of adding value to their business lines via data collected and destined for regulatory reporting. Regulatory data need not be limited to regulatory reporting, so let's take a look at a few cases where benefits can be had.

### Case 1 – Financial statement data and economic outlook

Most regulatory data — as required in Australia by APRA's Economic Financial Statistics (EFS) collection or in Singapore by the MAS 610 Notice — contains detailed information regarding both the Balance Sheets and Profit & Loss Statements of banks. This information has a higher degree of granularity than in the past, including data on deposits divided by geography, source, currency and industry sector. The same goes with data on loans but with additional information such as whether a loan is interbank or to a specific industry; moreover, such loans are grouped according to industry segmentation, which means this data can reveal the industry where most loans are concentrated. Given that loan-to-value ratio is a requirement of regulatory reporting, a bank will be providing a more refined assessment of risk on those loans. The point here is that all of this information is collected and provided monthly, while in some cases a bank must systemically provide this data to regulators every quarter.

So, let's say a bank has five years' worth of monthly data, which means there are 60 data points allowing the bank to see a trend over those five years. The bank's strategic planning department can then compare this trend to trends in macro-economic variables (GDP, Inflation, CPI, etc) and find links between the economic indicators and how parts of the loan portfolio has moved over that five-year period. Such a comparison can give banks an incisive glimpse into the future. By looking at the economic outlook for the year to come, the bank can forecast how the portfolio will move and can thereby make strategic decisions for creating business value and putting advantage on the Balance Sheet.

### Case 2 – Loan to Value ratio forecast

Connecting and comparing loan-to-value ratio on mortgages to external data about the economic outlook can provide another insight on how risky parts of the bank's portfolio might become. Convention says that the bigger the ratio, the riskier is the mortgage or the bank's exposure to that mortgage (the worst-case scenario being that the amount of the loan is greater than the value of the house). But there are more factors that impact on the loan-to-value ratio than house prices, and there could be a stronger relationship between the loan-to-value ratio and the other economic variables (current and forecast) that better describe the overall exposure of the bank in its loan-to-value ratio. Two examples could be the occupancy ratio and the "breakeven occupancy ratio".

### Case 3 – Large Exposure reporting requirements

New regulations require a bank to report not only on the large exposures in its portfolio but on the counterparties of these exposures. That is, banks must monitor groups of counterparties with specific economic relationships or dependencies, and must report on whether these relationships are reliable. The aim is to mitigate the risk of the default of one counterparty adversely impacting the bank's customer or causing the default of other counterparties.

Large exposures are therefore very complex when it comes to regulatory reporting. While a bank will have information on a particular counterparty, it will not necessarily have information for all of the counterparties within the group. And because the information requested on all of the counterparties may not be readily available, this means the bank must invest a great amount of time and resources to comply. There are no two ways around it: because the bank must get this job done for regulatory compliance, then it's best for the bank to start with a mindset that allows it to take advantage of the effort and expense from a commercial perspective. Having an overall picture of the relationships between the counterparties, the bank can identify which of the counterparties within this same sector could be its potential customers.

Moreover, considering that the bank is now aware of the line of business between its customer and the counterparties — information that is generally not available — the bank has a clearer basis or motive to invest in a particular industry. If the economic outlook for the next year is looking good for mining companies, for example, then the bank would want to have heavy exposure in that sector or vice versa. Regulatory data could be the key to making confident strategic decisions on how to balance the portfolio or unbalance it.

### Case 4 – Pillar III Risk Reporting

Pillar III requires all banks to publish risk assessments of their portfolio divided into groups of customers at a very granular level. For every group of customers, the bank needs to report the degree of risk for customers, counterparties, sector or industry, or class of borrowers depending on the size of the portfolio. A risk assessment for any of those components of the portfolio is important information for a bank's marketing and credit departments. Although these departments already have business strategies in place based on customer information already collected — such as type of job, income, postcode, even a Facebook profile — it would make sense for these departments to consider an additional dimension of risk according to the different segments of the portfolio. Pillar III data can be very relevant for targeted campaigns, as, for example, when a bank markets a service to two groups of customers with similar characteristics but with different risk profiles. With Pillar III data the bank can better determine a preference of one group over another, or how to better target the least risky counterparties within a specific sector.

### Case 5 – AML and cross selling

When a bank initiates Anti-Money Laundering (AML) controls on its portfolios, it encounters hundreds of thousands of exposures — for example, every customer who has spent or sent money overseas will likely be analysed for AML. Yet the red flags that show up, the points where there really might be AML activity, often represent only ten to fifteen violations. The bank however has been analysing positions or transactions that run into the hundreds of thousands, and would obviously be privy to valuable information. The bank can easily identify which of its customers travel overseas often and why (it could be for business, holiday, medical, family reasons), and with which overseas companies their customers have an established relationship. One way in which AML data collected for regulatory reporting can be put to business purposes is for the bank to develop specific services that it can offer to a particular set of customers — in this case, frequent overseas travellers. One such service is a credit card that can be used in any country without exchange currency costs. Another could be to provide a letter-of-credit offering to customers who have interactions with countries the bank would like to invest in or have exposure in.

## References


A-Team, How to Get Data Lineage Right for Regulatory Compliance and Business Value, Sep. 2018

APRA, Modernizing Economic and Financial Statistics, Sep. 2018

Cluster Seven, Regulatory Reports and Data Lineage, Nov. 2016

Deloitte, Regulatory reporting: a business value contribution, New York, 2015

Deloitte, MAS Notice 610/1003 Regulatory Reporting, Executive Briefing, May 2015.

Ernst & Young, Regulatory Reporting: Implementing the proposed MAS Notice 610, 2017

Federal Financial Analytics, The Regulatory Price-Tag: Cost Implications of Post-Crisis Regulatory Reform, July 2014.

IDC, Worldwide IT Spending 2013-2018, Risk IT Spending Guide, June 2014.

Loughlin Marty, Transcending Regulatory Compliance: Creating Business Value from Data Governance, June 2016.

Medici, Regulatory Reporting: Don't Play Catchup! Mar. 2017

McKinsey – IFF survey, Living with BCBS 239, By Holger Harreis, Tony Ho, Jorge Machado, Peter Merrath, Kayvaun Rowshankish, and Asin Tavakoli, May 2017

McKinsey & Company, The future of bank risk management, by Philipp Härle, András Havas, and Hamid Samandari.

Monetary Authority of Singapore, Notice 610 of submission of statistics and returns, May 2018

Nixora Group, Economic and Financial Statistics: have you spotted the opportunity, 2018

PWC, The extra mile: Risk, regulatory, and compliance data drive business value, May 2015


Collins Square, Level 23, Tower Five
727 Collins Street - Melbourne Vic 3008
Ph: +61 3 9101 8038 Email: Contact@nixoragroup.com.au