Data Quality Automation Frameworks for Regulatory Reporting in Banking

Pavan Kumar Mantha

Pavanmantha777@gmail.com

Abstract:
The Basel III, BCBS 239 and IFRS 9 are some of the strenuous stipulations that have made the reporting of regulations in the banking sector complex. The problem of data quality assurance is a serious consideration in the broken systems especially when systems that can be easily corrupted, slow and inefficient are being utilized; this is the case with manual systems or the obsolete systems. The provided paper describes how data quality automation structures can solve these problems. By performing a structured review of the academic papers, regulatory reporting and practices in the industry, the paper assesses the effectiveness of automation in enhancing accuracy, efficiency and regulatory reporting and cleanliness. Specific attention is paid to the new technologies including artificial intelligence, natural language processing, cloud infrastructures, and the development of end-to-end automated controls, including the reconciliation, variance detection, and schema drift detection. The other positive aspect of the reduced manual labor, the augmented correspondence and readiness to audit as well as the issues pertaining to transparency are also confirmed in the results. The future applications would be explainable AI and blockchain audit trail.

Keywords: Data Quality, Automation, Regulatory Reporting, Banking Compliance, ETL, Data Governance, Risk Management, Data Validation, Basel III, Financial Regulations.

I. INTRODUCTION

Banking industry is typified with very controlled environment where accuracy, openness, and time-sensitivity of information reporting becomes a significant compliance and stability factor in the operation of the banking industry. The Basel III, BCBS 239 have measures such that the financial institutions exercise strict conditions over the information that is being reported so that the information reported is complete, accurate, and verifiable. Avoidance of these needs may lead to massive fines, loss of image and exposure of the organization to the wider financial system. Therefore, a high degree of data quality has been among the strategic issues of banks worldwide.

Despite this significance, in most financial institutions, regulatory reporting is a manual process or a legacy process. These practices are characterized by disaggregated data system, paper based reconciliations and inaccurate validation procedures. These inefficiencies are known to create delays, high cost of operation and subject banks to non-compliance. The greater the volume and complexity of the regulatory requirements, the lower the level of accuracy and agility of the traditional processes can offer regulators. This is particularly acute in large multinational banks where large amounts of heterogeneous data are to be integrated and checked and reported across more than one jurisdiction.

In order to control these problems, a new method of automation of data quality structures has been introduced. Standardized controls, continuous monitoring, and end-to-end validation are supported with automation structures across the regulatory reporting pipelines. These frameworks reduce the human factor and increase the level of data correctness, as they imply the usage of technologies in reconciliation, variance detection, and tracking schema drift.

This paper will address the design, implementation and effects of data quality automation systems of regulatory banking reporting. It will entail exploration of the technological enablers, advantages, issues and opportunities and how the frameworks will improve compliance, lessening the operational pressures, and organizational resilience amidst increased regulatory pressures.

II. LITERATURE REVIEW

Regulatory reporting history of the banking industry indicates that the banking industry has been heavily depending on manual validation and reconciliation systems. Financial institutions were historically using spreadsheet-based reconciliation, manual variance analysis and disjointed audit trails to maintain that the reported data was accurate. Despite the fact that these processes could work at smaller levels of reporting, they were subject to the human factor, inefficiency, and lack of scalability [1]. As the regulatory requirements continued to swell in terms of size and complexity, it was no longer possible to use manual tactics to process the increasing number of transactions, to comply with the increasing number of multi-jurisdictional compliance requirements. Such a dependence on manual validation was the foundation on the current emphasis on automation and systemic quality control.

In academic terms, data governance has been a crucial concept that organizations have attached to the conceptualisation of regulatory reporting. The theory of data quality management concentrates on the data quality aspects of accuracy, completeness, conformity and timeliness. The governance models have also emphasized the need to align data processes to its risk management principles, regulatory requirement and accountability of the governed institution in the banking industry [2]. In the academic research, the opportunities of automation as a method of mitigating systemic risk by restricting the application of manual interventions have also been noticed. In addition, compliance systems were also analyzed as socio-technical systems in which technology, regulation and corporate culture must cooperate to provide efficient supervision.

In the practices of the industry, these scholarly findings have been manifested by the implementation of big data technologies and regulatory technology (RegTech). RegTech solutions are based on automation, artificial intelligence, and advanced analytics to aid real-time monitoring, anomaly detection, and reporting standardization. Cloud computing, which has been aided by scalable architecture and big data platform has enabled the banks to handle huge amount of structured and unstructured information to satisfy compliance requirement. This has also improved efficiency and increased the ability of the banks to be dynamic in responding to new regulatory demands [3]. The increasing use of machine learning-based anomaly detection and automated reconciliation applications can be seen as a sign of increased interest in applying intelligence to data quality models.

On contrary to these developments, a lot of controversy and challenges are still cropping out of the academic and commercial circles. One of the problems is transparency. Automated systems, in particular those driven by a machine learning model, are likely to be black-box systems, i.e. the reasoning by which a data point is being corrected or which anomaly has been found is hard to describe. This unaccountability creates a challenge to regulators who need to have accountability and traceability of reporting procedures. The second ongoing issue is scaling the structures among the global organizations which have diverse data structures [4]. Even though local efficiency can be achieved through automation, the implementation of automation by the multinational organization produces the issue of fragmentation, interoperability, and the possibility of vendor lock-in.

The literature gap is that there has been an acute deficit of incorporation of automated data quality structure that is specifically utilized in regulatory banking reporting. Despite the overall data governance and the RegTechs, not a multitude of frameworks is constructed to align the end-to-end requirements of reconcile and detect variance, schema drift monitoring, and real-time regulatory compliance. This failure proves the need to carry out a research on holistic structures that do not only increase the efficiency, but balance transparency, accountability, and scalability of the regulatory reporting lifecycle.

III. METHODOLOGY

The work has adopted a structured literature review approach in order to compare the academic and industrial knowledge on the automated data quality model of regulatory banking reporting. The reason behind this design was to have a structured, comprehensive and thematically organized of the body of knowledge under consideration [5]. The review involved peer-reviewed journal articles, regulatory reports, and industry reports to present information about the theoretical and practical implementations.

The sources of analysis relied on the available scholarly databases such as IEEE and ACM that provided technical information on automation, data governance and compliance orders. Besides, reports of the major regulatory bodies, such as the Basel Committee on Banking Supervision, the European Central Bank, and the Financial Conduct Authority, were reviewed in order to reflect changing supervisory expectations [6]. Technology provider and consultancy industry white papers were also included to help define the feasible implementations, case studies and new innovations in regulatory technology.

The thematic categorisation was used in the research methodology to categorise the insights into four major domains: drivers of automation adoption, existing frameworks, implementation challenges, and regulatory implications. This thematic form helped create a comparative insight into how various stakeholders conceptualise and operationalise data quality with respect to compliance [7].

Lastly, a comparative framework was used to distinguish between manual, semi-automated and fully automated methods. Manual methods were evaluated based on the human intervention dependency and the probability of error. Semi-automated methods were considered to gain efficiency and biased integration of technology. Fully automated methods were discussed in terms of scalability, accuracy, and compliance with the regulatory requirements. Such an ordered approach allowed creating an evidence-based synthesis of the literature in a balanced way.

IV. RESULTS

A. Drivers of Automation

The growing nature of regulations is the primary factor that makes it necessary to automate the process of implementing banking regulatory reporting. The frameworks that impose strict standards on the banks to provide clear, true and timely information are Basel III, BCBS 239 and IFRS 9. Such regulations extend past the capital sufficiency and computation of risk and will involve information to prove data ownership and auditability along reporting pipelines. These raised expectations can no longer be achieved through the conventional manual validation and reconciliation procedures. In the meantime, financial institutions are struggling to cope with the very high rate of increase in the volumes and complexity of data due to the plurality of products in a portfolio, the demands of international business and multi-jurisdictional regulation [8]. This is even further complicated by the fact that compliance is virtually real-time and the regulators are more demanding constant reporting rather than occasional reporting.

B. Core Value of Automation

The automation frameworks are very admirable since they cement the foundations of data quality of regulatory reporting. Such systems enhance the accuracy, completeness as well as timeliness of the system because it contains standardized routines of validation in all steps of the reporting life cycle. Automation minimizes human interaction hence minimizing chances of human error and variation that are likely to weaken compliance. In addition to this, automation also eliminates the cost-efficiency of the process as it enables the elimination of the redundant procedure of validation and reconciliation and enables the qualified staff to direct their efforts towards other more productive analytical and strategic tasks. Another significant advantage is that audit preparedness since automated records and check points avail clear cut records of data management and efficiency in control to the regulators. The automation will help the institutions to assume the initiative of fulfilling their regulatory expectations by eliminating the inefficiencies in their operations and producing standard outputs.

C. Role of Advanced Technologies

Of essence is the value of the advanced technologies in the creation of the models of data quality automation of the regulatory banking reporting. Machine learning and artificial intelligence widely find their use to spot anomalies, suspicious trends and make predictions irregularities in between complex reporting data [9]. The techniques provide predictive intelligence, which increases proactive data quality control, and also it decreases the amount of hand interventions. It can also be used in complementing practices of compliance, which involves converting non-formalized regulatory documents into actionable and structured regulations, by moving the gap between legal requirements and operation reporting requirements to each other.
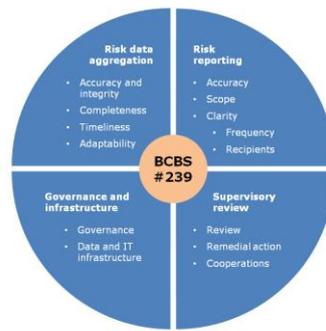
Fig. 1. Regulatory required data quality as decision-making support. Source: [10]

Among the most important innovations, the architecture of the end-to-end automated checks which attempt to deal with the underlying data integrity problems can be cited. Reconciliation is relevant to offer uniformity of different data sources and systems. Variance detection is the process that is applied to determine whether there is any alteration of value either over time or category of reporting that is not foreseen [11]. The schema drift check does not involve any data model transformation, which erases downstream inconsistencies. All these lead to accuracy and reliability of the Basel III, CECL and CCAR reporting.

D. Challenges and Risks

Nevertheless, even with these advantages of automation, there are still certain new problems and threats. Among the most critical ones is the explainability and transparency issue, in particular, with regard to AI-based validation models. Regulators and auditors would require transparency into how anomalies would be detected and rectified, but complex algorithms will probably be considered black-boxes. This inability to be interpreted causes an issue of responsibility and credibility. Also, an issue with the banks is vendor lock-in, the banks may be using proprietary automation solutions which might result in rigidity in terms of long-term contracts which are expensive. Another issue, the uncertainty of the regulation, as the compliance requirements keep on diversifying in this way, therefore, the continuous redesign of the automation systems is needed.

V. DISCUSSION

The results indicate that a trade-off exists between the automation time and the threat of governance in the regulatory reporting. Although the automation systems will render a high level of savings in labor and will enhance the quality of the regulatory filing, it presents new challenges to the administration process. High levels of supervision should be given to very automated systems to increase transparency, explainability and adhering to the changing supervisory requirements. Devoid of such safety measures, such efficiency achieved under automation may be compensated by the fact that there are more regulatory investigations and operational deficits on the list of risks [12].

One of the biggest impacts of the shift towards automated data quality frameworks on the change in reactionary compliance to proactive monitoring is the shift. The old models were based on daily checks and manual balancing following processing of data in-hand had been facilitated.

Fig 2: End to End Automated Data Quality Framework for Regulatory Reporting Banking.

The case with banks and automation has immense implications to an organization. This will cause the organizations to invest in emerging proficiency skills that will involve technical proficiency in automation systems featuring governance oriented positions like data stewardship. The execution of automation structures in a holistic manner between innovation and compliance will be achieved by establishing data quality centres of excellence (CoE). The model of governance should change so that the inter-functional governance of the compliance, technology, and risk management teams could be conducted [13].

Relativism will show that insurance and fintech are developing the same way other industries are. Insurers are unleashing claims validation automation and real-time analytics is being used by fintech firms to detect fraud and comply.

VI. CONCLUSION
As it has been defined in this paper, data quality automation structures are highly helpful to regulatory reporting in the banking sector. Compliance is also enhanced in terms of accuracy of compliance because automation saves time used to report, saves money by minimizing the number of human operations and standardized built-in controls. Nevertheless, these efficiencies are compromised with issues of governance especially in regard to explainability, regulatory flexibility and cyber safety threats. The findings highlight the fact that automation is impossible to happen when there are no powerful overseeing and governing structures.
It is suggested that a hybrid solution to reach a balance between the efficiency and responsibility will be the synthesis of skills of the human factor and machine-assisted validation. Those structures are the assurance that the automation products manageable and controllable and employs technology in achieving the goals of scalability and accuracy. The other directions are the usage of explainable artificial intelligence that increases the rates of transparency and trust towards automated processes.

REFERENCES:
[1] M. Hendayun, E. Yulianto, J. F. Rusdi, A. Setiawan, and B. Ilman, "Extract transform load process in banking reporting system," *MethodsX*, vol. 8, p. 101260, 2021, doi: https://doi.org/10.1016/j.mex.2021.101260.

[2] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy Artificial Intelligence," *Government Information Quarterly*, vol. 37, no. 3, p. 101493, Jul. 2020, doi: https://doi.org/10.1016/j.giq.2020.101493.

[3] I. Anagnostopoulos, "Fintech and regtech: Impact on regulators and banks," *Journal of Economics and Business*, vol. 100, no. 1, pp. 7–25, Nov. 2018, doi: https://doi.org/10.1016/j.jeconbus.2018.07.003.

[4] Basel Committee on Banking Supervision, "Principles for Effective Risk Data Aggregation and Risk Reporting," Bank for International Settlements, 2013.

[5] M. Becker, K. Merz, and R. Buchkremer, "RegTech—the application of modern information technology in regulatory affairs: areas of interest in research and practice," *Intelligent Systems in Accounting, Finance and Management*, vol. 27, no. 4, pp. 161–167, Jun. 2020, doi: https://doi.org/10.1002/isaf.1479.

[6] European Banking Authority, "Supervisory Reporting: Implementing Technical Standards (EU No. 680/2014)," European Banking Authority, 2021.

[7] R. P. Buckley, D. W. Arner, D. A. Zetzsche, and R. H. Weber, "The road to RegTech: the (astonishing) example of the European Union," *Journal of Banking Regulation*, May 2019, doi: https://doi.org/10.1057/s41261-019-00104-1.

[8] P. Kavassalis, H. Stieber, W. Breymann, K. Saxton, and F. J. Gross, "An innovative RegTech approach to financial risk monitoring and supervisory reporting," *The Journal of Risk Finance*, vol. 19, no. 1, pp. 39–55, Jan. 2018, doi: https://doi.org/10.1108/jrf-07-2017-0111.

[9] J. Truby, R. Brown, and A. Dahdal, "Banking on AI: mandating a proactive approach to AI regulation in the financial sector," *Law and Financial Markets Review*, vol. 14, no. 2, pp. 110–120, Apr. 2020, doi: https://doi.org/10.1080/17521440.2020.1760454.

[10] European Central Bank, "ECB Guide to Internal Models," ECB Banking Supervision, 2019.

[11] S. Schelter, J.-H. Böse, et al., "Automatically Tracking Metadata and Data Quality in ML Pipelines," in Proc. 12th ACM Int. Conf. on Management of Digital Ecosystems (MEDES), 2020.

[12] T. Butler and L. O'Brien, "Understanding RegTech for Digital Regulatory Compliance," *Disrupting Finance*, pp. 85–102, Dec. 2018, doi: https://doi.org/10.1007/978-3-030-02330-0_6.

[13] SAS Institute, "Data Quality for Analytics and AI in Financial Services," SAS White Paper, 2020.