From Data Management to Data Governance: Experience of the Reserve Bank of India

- Archana Dilip and Debasis Nandi*

Abstract

The Reserve Bank of India is an integral part of the official statistical system in the country. It compiles and disseminates macro-financial statistics in areas related to banking and financial systems, currency, payments system, monetary policy surveys, foreign exchange management, external accounts, private corporate sector, and public finance including public debt management. The Reserve Bank's system of data and information management has evolved over the years guided by its statutory role, promotor of robust financial architecture, policy and monitoring needs, and emerging requirements of the growing economy and financial system. It disseminates considerable information as a 'public good' and also constantly strives for refinement in relevance, quality and timeliness of data to support data-driven policies and monitoring.

Currently, activities like building an information management system based on data warehousing approach for data receipt, processing, production, storage, retrieval and dissemination; standardisation of reporting of financial data; developing a statistical system for maintaining data quality are largely centralised. This paper reviews the existing information management structure in the Reserve Bank with respect to the three critical elements of data governance, *viz.*, people, processes and technology. It also discusses the information governance through an inter-departmental return governance group, which aims to rationalise the process of rationalisation regarding data elements, periodicity, new prescription/s and system/s in a concerted manner. The Reserve Bank has deployed tools and technology for managing each stage of the data lifecycle, namely data collection, compilation, dissemination and access. Given the diversity of reporting entities with different levels of technology, use of multiple channels for data collection and their handholding is necessary. The investment in technology for information management and handholding of reporting entities, when necessary, has ensured continuous flow of information to the public even during the period of lockdown induced by the COVID-19 pandemic.

The present structure has consolidated the learnings from the past in terms of standardisation of data collection methods; rationalisation of returns/statements; and reduction in the time required to deploy a system for new data/modification of existing data due to the centralisation of overall information management. The scope for implementation of global best practices has ensured increased

* Archana Dilip (adilip@rbi.org.in) is Assistant Adviser and Debasis Nandi (dnandi@rbi.org.in) is Director at the Department of Statistics and Information Management, Reserve Bank of India. The views expressed in this paper are those of the authors and are not necessarily shared by the Reserve Bank of India.

standardisation across the organisation. The envisaged changes for implementing a data governance framework at the organisation level include: (a) implementing an enterprise-level data model, (b) standardisation of return introduction/modification, (c) maintenance of metadata and definitions, (d) implementing element-based reporting, (e) provision of standardised supplementary information, (f) conduct of periodic reviews, and (g) creation of principles and policies related to data lifecycle. Finally, the paper discusses the guidelines provided for implementing data governance at the regulated entities and recommends increased engagement of the regulated entities and other financial regulators; development of a data quality framework with appropriate metrics to evaluate the quality of data; and implementation of data privacy decisions and protection for efficient data governance.

From Data Management to Data Governance: Experience of the Reserve Bank of India

The Reserve Bank of India is an integral part of the official statistical system in the country. It compiles and disseminates macro-financial statistics in diverse areas with focus on banking and financial systems, currency, payments system, monetary policy surveys, foreign exchange management, external accounts, private corporate sector, and public finance including public debt management. The Reserve Bank's system of data and information management has evolved over the years guided by its statutory role, promotor of robust financial architecture, policy and monitoring needs, and emerging requirements of the growing economy and financial system. It disseminates considerable information as a 'public good'. The Reserve Bank also constantly strives for refinement in relevance, quality and timeliness of data to support data-driven policies and monitoring.

Data are collected as part of statutory, regulatory, supervisory, policy and research requirements that is used for compilation of certain key banking statistics at aggregate level. The Reserve Bank collects data through prescribed formats in the form of returns. Various data owner departments at the Reserve Bank prescribe returns to obtain data from regulated entities such as banks, co-operative banks, non-banking finance companies and financial institutions. The data collected and disseminated by the Reserve Bank may be classified based on (i) granularity and frequency, (ii) collection/generation mode, (iii) purpose/sector, (iv) confidentiality/nature. The granularity of data may be transactional-level, unit-level, account-level, aggregated data or distributed data aggregates.

With the efforts related to collecting and disseminating reliable statistics starting as early as 1945[†], followed by the computerisation of data management over the years and subsequently setting up the Centralised Database Management System (CDBMS) in 2002 — the corporate data warehouse that was operational for internal users, the Reserve Bank has a rich experience in effective data management. With the internet deployment of the CDBMS in 2004, dissemination of statistics got centralised and the data warehouse was christened as Database on Indian Economy (DBIE) - the portal for data dissemination of the Reserve Bank. During the same period, efforts for collection of data through online platform were initiated in response to the need for bringing in uniformity, standardisation, and ease of maintenance.

The data requirements are ever-changing and an increase in the sources of data has been witnessed in the recent times. As the decision making at the organisation has become more data-driven, there is an increased demand for quality data that is reliable. Currently, activities like building an information management system based on data warehousing approach for data receipt, processing, production, storage, retrieval and dissemination; standardisation of reporting of financial data;

[†] Some of the surveys that provided valuable information to the Bank and paved for data-driven decision making include: Annual surveys of the ownership of demand deposits with scheduled banks as well as the investments of scheduled banks in Government securities of 1945, a census of India's foreign assets and liabilities with end-June 1948 as the reference period and the All-India Rural Credit Survey, with 1951-52 as the reference period.

developing a statistical system for maintaining data quality are largely centralised. Over the years, the footprint of this department has expanded to every part of the organisation (Patra, 2022). Further, the existing data warehouse of the Reserve Bank is getting overhauled with the implementation of the Centralised Information Management System (CIMS) to cope up with the scalability required for handling the rising requirements of dealing with macro and micro data and providing advanced analytical solutions. CIMS is envisaged to provide a comprehensive solution for overall information management and streamlining the data flow to the Reserve Bank (RBI, 2022a).

The Data Governance Institute defines data governance as "… the system of decision rights and accountabilities which describe who can take what actions with what information, and when, under what circumstances, using what methods." Data governance frameworks consist of data office, risk management, and change management (Smallwood, 2014). Further, data governance is the process used by an organisation to manage consistency, quality, security, usability, as well as availability of the organisation's data, and added that it was a critical topic that should be addressed with urgency (Putro *et al.*, 2016). As stated by Fernandes *et al*. (2012), a data governance framework should consist of the basic precepts of process, policies, and people, specifically data stewards.

Literature highlights the three critical elements of an effective data governance framework as people, processes and technology. The paper is structured based on these three elements — Section II introduces the people who are responsible for data and information management at the Reserve Bank, followed by the existing processes in place detailed in Section III. Section IV briefs about the technology that facilitates the people involved and the processes with respect to each stage of the data lifecycle. The envisaged changes with respect to data governance are introduced in Section IV. Guidelines provided by the Reserve Bank for implementing data governance at the regulated entities along with other recommendations for the RBI are discussed in Section V. Section VI provides the concluding remarks.

II. People

With the computerisation of information management, the existing data compilation systems give priority to departmental information systems, which are the primary feeders of the management information system (RBI, 2014). The existence of these separate departmental information systems has necessitated the need for dedicated personnel at the individual departments (who are involved in prescribing returns, ensuring timely submission of quality data and compilation of required statistics) and the Department of Statistics and Information Management (DSIM), which manages the data warehouse to which all data flows.

The need for a data governance structure was recognised much earlier at the Reserve Bank and it was suggested that an inter-departmental group be formed so that the process of rationalisation regarding data elements, periodicity, need for provisional returns can be carried out in a concerted

manner (RBI, 2015). An interim structure was formed in end-2015 to manage the reporting by regulated entities in the form of returns. Presently, there is an inter-departmental group, Return Governance Group (RGG) which is responsible for vetting the introduction of new returns by any department, modification of existing returns, approving definition of new data items across the complete universe of returns/data requirements of RBI, etc. The RGG is headed by the Executive Director in charge of DSIM. The RGG has members from several returns prescribing departments of the Reserve Bank.

Recognising the important aspects of data management, different committees/groups comprising senior officers have been formed for reviewing the existing practices and recommending best practices to be followed. The details of the important committees/ group related to data management that were formed in the last decade are provided in Table 1.

Table 1: Details of Important Committee/Groups Related to Data Management

S.No.	Name of the Committee/Group	Important Objective	Chairperson/Important Members	Year
1	Committee on Data and Information Management in the Reserve Bank of India	To address the concerns about data gaps and to place the information management on a more technologically mature footing	Executive Director (Chairperson)	2014
2	Committee on Data Standardization	To bring about synergy and uniformity of efforts being undertaken in data reporting and data standardisation	Senior officers from departments concerned with data management, external member from regulated bank.	2015
3	Inter-departmental Task-force for Harmonization of Banking Statistics	To provide "harmonised" definitions of major balance sheet/ profit and loss/ off-balance sheet items covered in the banking/ regulatory returns received across RBI departments.	Senior officers from departments prescribing returns	2014-17
4	Regulations Review Authority- Committee for Rationalisation of Returns/Statements	To review the returns and streamline as well as rationalise the reporting mechanism	Executive Director in charge of DSIM	2022

Source: Reserve Bank of India

III. Processes

The existing practices followed with respect to each stage of the data lifecycle are detailed in the following sub-sections.

III..1 Data Collection

The main vehicle for collecting information from regulated entities is the 'returns' submitted by the regulated entities in the prescribed format and periodicity. Data is collected through various channels

involving XBRL standard[‡]; collected from various sources; or compiled at the concerned department at the Reserve Bank. Given the diversity of reporting entities with different levels of technology, use of multiple channels for data collection and their handholding is necessary. The consolidated list of returns prescribed the Reserve Bank for different reporting entities has been made available on the RBI website since 2018. The list includes information on format, frequency of reporting, targeted entities, relevant circulars and the concerned department.

The data definitions and compilation instructions are provided to the regulated entities by the concerned department in the form of circular or instructions. In view of frequent divergence observed in the data reported by banks for the same item across various returns prescribed by RBI and with an objective to ensure quality and uniformity of the inputs from the banks, the Reserve Bank had constituted an inter-departmental Task-force in December 2014, to provide harmonised definitions of major balance sheet/ profit and loss/ off-balance sheet items covered in the banking/ regulatory returns received across RBI departments. Subsequently, harmonised definitions for around 189 data elements that are required to be reported in multiple returns have been provided on the RBI website.

Important issues faced earlier included (i) duplication of data elements across returns, (ii) inconsistency in reported figures across returns, and (iii) increasing burden on banks for collecting, processing and submitting such returns (RBI, 2015). However, the Returns Governance Group has made several contributions in improving the quality of returns, streamlining of data elements and definitions, as well as, in rationalising several returns (RBI, 2022a).

III.2 Data Compilation

The work related to data compilation is mainly undertaken by the assigned Data Makers and Data Checkers of the concerned unit. Data compilation at the respective units requires the definition of roles of Data Maker and Data Checker. The first level of compilation is done by the assigned Data Makers based on applicable data manuals. The correctness of the data is then checked by the respective Data Checker by applying the validation checks. The validated data is then transmitted to the data warehouse for dissemination.

III.3 Data Dissemination

The Reserve Bank of India, a signatory of the Special Data Dissemination Standards (SDDS) of the IMF, undertakes the responsibility of disseminating information under certain data categories, such as, analytical accounts of the banking sector, analytical accounts of the central bank, balance of payments, international reserves and exchange rates. The RBI also provides a National Summary Data Page (NSDP) on the website to provide quick access to a single comprehensive source of economic and

[‡] XBRL (eXtensible Business Reporting Language), is a XML-based non-proprietary open standard that is used to prepare, exchange and publish business information.

financial data consistent with the data categories and components described in the subscriber's metadata. Currently, information is disseminated through three important channels—-(i) statistical publications like Statistical Tables Relating to Banks in India, Handbook of Statistics on Indian Economy, Current Statistics Section of RBI Bulletin, Weekly Statistical Supplement, (ii) data/press releases on the website of the Reserve Bank, (iii) DBIE. The DBIE has become an important data communication channel of the Reserve Bank (RBI, 2014).

III.4 Data Access

Data access rights are provided by the concerned data owner department. If a new user require access to a data, the permission is sought from the data owner. The data owner then informs the data warehouse support to issue the required rights. The user lists are being maintained and the same are subjected to periodic review.

IV. Technology

The Reserve Bank's investment in technology for information management has ensured continuous flow of information to the public even during the period of lockdown induced by the Covid-19 pandemic (Das, 2022). Presently, around 250 returns are being collected from various financial entities through various modes including eXtensible Business Reporting Language (XBRL)-based filing system, file uploads, emails and Electronic Data Submission Portal (EDSP). Returns have been developed under the XBRL system in a phased manner starting from 2008. A study focussing on estimating the effectiveness of the adoption of the XBRL standard in the Indian banking sector found that overall efficiency of the reporting system has increased to a great extent and has aided in research and analysis with considerable reduction in errors in reporting (Kumar *et al*., 2019). The RBI's EDSP that collects data in the form of delimited text has become an important portal for account level and branch level data related to credit and debit.

Although standardisation with respect to data collection has been achieved to a great extent, there are around 69 returns that are being collected in physical copies or through email (RBI, 2022a). Some data collecting departments have their custom-made legacy systems for data reporting. Currently, there are separate systems to collect data pertaining to overseas investment, export/import data, audit regulation, cyber security, risk-based supervision, *etc*. Data is also collected directly from other statistical bodies and RBI's internal systems such as Real Time Gross Settlement, Core Banking System (eKuber).

All periodic publications have been integrated with the DBIE, the data warehouse of the Reserve Bank. Apart from the important publications, data is disseminated in the form of time-series reports on statistics related to seven sectors *viz*., real sector, corporate sector, financial sector, financial markets, external sector, public finance, and socio-economic indicators. This is a unique feature of the dissemination portal enabling users to access the relevant data through publications or statistics as per

convenience and understanding of data. DBIE has provision for data visualisation, reports, dashboards and analytics. Data collected from around 138 returns by RBI from its about 10,000 regulated entities are moved the data warehouse for storage and processing.

The Central Repository of Information on Large Credits, popularly known as CRILC system, has been developed and implemented under XBRL and DBIE system. Apart from added features of using Personal Authentication Number (PAN) as the unique identifier and facilities to generate automatic alerts and to perform processing of large volume of data, CRILC may be viewed as the first system in the Reserve Bank where information submitted by banks are also shared with the banks for timely intervention. The Central Fraud Registry (CFR) has been developed along the same lines for ensuring that the people / entities involved in fraudulent activities are known to the banks and are kept out of the system. Similarly, the Liberalised Remittance Scheme (LRS) database built leveraging the technology of XBRL and DBIE captures the transaction-wise and purpose-wise remittances made by the individuals.

V. Data Governance Envisaged at RBI

The implementation of the RGG has aided in rationalisation of returns which in turn helps in reduction of reporting burden. Standardisation of data collection has been achieved to a large extent across the organisation. Further, centralisation of overall information management has helped in cutting down the time required to implement and deploy a system for new data/modification of existing data. The experience gained from implementing earlier systems and deployment of dedicated personnel have facilitated this reduction in time. This is evident from the expansion of the EDSP, within a short period of its implementation, to cover diverse returns like payment frauds, natural calamity returns, unit level data for inflation expectations survey of households. However, there is further scope for implementation of best practices to ensure adoption of global standards and increased standardisation across the organisation.

The different aspects of data governance and the process of bringing in more standardisation in data management across the organisation, thereby enhancing data quality and reducing the reporting burden of regulated entities have been explored by many expert committees. The recommendations provided by the expert committees, the implementation of which will make way for enterprise-level data governance have been consolidated below. The implementations of these recommendations are at different stages at the Reserve Bank.

i. Implementing an enterprise-level data model: The existing data compilation mechanism gives priority to information systems at the departmental-level and therefore there is a need to implement an enterprise-wide data model (RBI, 2014).

ii. Standardisation of return introduction/modification: While introducing a new return, it may be ascertained that the same is not available in the existing universe of reporting. Approval of RGG

may be obtained mandatorily for return introduction/modification. Any ad hoc return introduction for a period of more than six months may be subject to rigorous scrutiny and approval by the RGG (RBI, 2022a).

iii. Maintenance of metadata and definitions: Data Structure Definitions (DSDs) with details of variables, definitions, statistical attributes and other time series attributes may be prepared and maintained as a part of the metadata directory. This will facilitate dissemination of statistics using the Statistical Data and Metadata Exchange (SDMX) standard (RBI, 2014). The metadata may be harmonised across different systems and with the RBI definitions for regulatory reporting (RBI, 2015). Further, RRA 2.0 recommended that a single updated master file containing definition of data items may be made available on the RBI website. Return prescribing departments may provide definitions of all major data items in the master file at the time of prescribing a return (RBI, 2022a).

iv. Implementing Element-based reporting: It was recommended by RRA 2.0 that in the future, element-based reporting may be explored as the main vehicle for data submission (RBI, 2022a). In this context, it may be noted that the Reserve Bank is in the process of implementing the new data governance framework through flexible element-based repository (EBR) with the facility to convert from return-based repository (RBR) by carrying out forward and reverse engineering to ensure completeness (RBI, 2022b).

v. Provision of standardised supplementary information: Data validation rules, guidance notes and frequently addressed questions and solutions may be shared with the reporting entities and be made available in a consolidated manner for easy accessibility (RBI, 2022a). This will help banks and other reporting entities to further improve the quality of data reported to the RBI. Standardisation of regulatory instructions with uniform formats may provide better understanding to the regulated entities.

vi. Conduct of periodic reviews: A periodic data quality assessment process may be implemented across different domains and feedback on quality issues may be provided to data owners (RBI, 2015). Further, the coverage of statistical publications and returns with respect to relevance, content, continuity and rationalisation may be subject to periodic review (RBI, 2015; RBI, 2022a).

vii. Creation of principles and policies related to data lifecycle: Various policies related to information management and governance *viz.*, data management policy, access control policy, data dissemination policies may be formulated. The dissemination policy may be based on the principles of transparency, comprehensiveness, relevance and timeliness (RBI, 2015).

VI. Recommendations for Data Governance

VI.1 Overview of RBI's Recommendations for Data Governance at Regulated Entities

The Basel Committee on Bank Supervision (BCBS) issued BCBS 239 in 2013 which covers the principles for effective risk data aggregation and risk reporting. The Principle 1 of BCBS 239 recognises the importance of having a strong data governance framework, risk data architecture and an IT infrastructure as the prerequisites in ensuring compliance of other principles. Data governance strategy addresses not only the present needs but future data and information requirements (Diokno, 2019). The RBI Committee on Standardization of 2015 made some recommendations regarding the key components that may be incorporated in the data governance architecture of banks.

The Committee recommended that policies may be formulated to clearly define data governance organisational structure, data ownership, roles and responsibilities, processes and procedures, ensure standardisation, data quality management, data dissemination and use of appropriate metrics. Data governance may be monitored by the Audit Committee of a bank. An executive level Data Governance Committee may also be formed to oversee the exciting information management. Data ownership may be clearly stated for all key data management platforms. The committee recommended the conduct of internal audits and periodic reviews of data governance processes. The 'Principles for effective risk data aggregate and risk reporting' may be followed by domestic systemically important banks. The Reserve Bank may also facilitate data governance at banks and validate critical banks' applications (RBI, 2015).

VI.2 Other Recommendations

After reviewing the existing literature, organisational framework, the benefits reaped and the envisaged changes, the following are recommended:

i. There may be increased engagement of the regulated entities and other financial regulators. This will help in seamless data compilation and to address the issues arising from the data-centric competitions and misuses in the digital era.

ii. The development of a data quality framework with appropriate metrics to evaluate the quality of data periodically at different levels may be explored. The principles of data quality framework may be aligned with that of the values and vision of the organisation. This will allow taking prompt actions in addressing data issues.

iii. With respect to consent-based sharing of personal data, Niti Ayog has released the Data Empowerment and Protection Architecture (DEPA). DEPA will empower individuals by operationalising a regulatory, institutional and technology design for secure data sharing (Niti Ayog, 2020). Accordingly, the Reserve Bank of India issued Master Directions in creating consent managers in the financial sector to be known as Account Aggregators. The

implementation of necessary directions on data privacy and protection may be explored based on evolving conditions.

VII. Conclusion

The Reserve Bank collects and disseminates data as a part of its statutory functions as well as in its role as a promoter of a robust financial system for India to support economic growth through efficient financial intermediation. The paper discussed the Reserve Bank's journey of data management by reviewing the practices followed with respect to the three important precepts of data governance, namely, people, processes and technology. The paper identified standardisation of data collection methods; rationalisation of returns; and reduction in the time required to implement and deploy a system for new data/modification of existing data due to the centralisation of overall information management as the key advantages of the existing structure. The envisaged changes with respect to data governance include implementing an enterprise-level data model, standardisation of return introduction/modification, maintenance of metadata and definitions, implementing element-based reporting, provision of standardised supplementary information, conduct of periodic reviews and creation of principles and policies related to data lifecycle. Increased engagement of the regulated entities and other financial regulators; development of a data quality framework with appropriate metrics to evaluate the quality of data at different levels; and implementation of necessary directions on data privacy and protection have been recommended for efficient data governance.

References

Das, S (2022). Inaugural Address at the Annual Statistics Conference of the Department of Statistics and Information Management, Guwahati. June.

Diokno, B (2019). Speech at the data governance seminar, Manila, June.

Fernandes, L. M., O'Connor, M., & Weaver, V. (2012). Big data, bigger outcomes. *Journal of AHIMA*, *83*(10), 38-43.

Kumar, P., Kumar, S. S., & Dilip, A. (2019). Effectivness of the Adoption of the XBRL Standard in the Indian Banking Sector. Journal of Central Banking Theory and Practice, 8(1), 39-52.

NITI Aayog (2020). Data empowerment and protection architecture, August.

Patra, M. D. (2022). Speech on Statistics and Information Management in the RBI –Where We Are; Where We Want to Be. Annual Statistics Conference of the Department of Statistics and Information Management at Guwahati. June.

Putro, B. L., Surendro, K., & Herbert. (2016, February). Leadership and culture of data governance for the achievement of higher education goals (Case study: Indonesia University of Education). In *AIP Conference Proceedings* (Vol. 1708, No. 1, p. 050002). AIP Publishing LLC.

Reserve Bank of India (2014). Report of the Committee on Data and Information Management in the Reserve Bank of India.

Reserve Bank of India (2015). Report of Committee on Data Standardisation.

Reserve Bank of India (2022a). Report of the Regulations Review Authority 2.0.

Reserve Bank of India (2022b). Annual Report 2021-22.

Smallwood, R. F. (2014). Defining the Differences Between Information Governance, IT Governance, & Data Governance. *Aiim*, 1-3.