



Approach Paper on Automated Data Flow From Banks to Reserve Bank of India

November 2010

Table of Contents

Document Overview	3
Chapter 1 – Guiding principles for defining the approach	5
Chapter 2 – Assessment Framework	6
Chapter 3 – Common End State	8
Chapter 4 – Benefits of Automation	13
Chapter 5 – Approach for Banks	15
Chapter 6 – Proposed Roadmap	22
Chapter 7 – Summary	24
Annexure A – Assessment framework	25
Model for Banks (Excel worksheet)	29
Annexure B - Task Details	39
Annexure C - Returns Classification (Excel worksheet)	59
Appendix I – Other end state components	60
Definitions	66
Appendix II – Global case study	68
References	73

List of Abbreviations

Acronym	Description
BSR	Basic Statistical Returns
CBS	Core Banking Solution
CD	Compact Disc
CWM	Common Warehouse Meta model
CDBMS	Centralised Data Base Management System
DBMS	Database Management Systems
DBS	Department of Banking Supervision
DSS	Decision Support System
e-mail	electronic mail
ETL	Extract Transform Load
FII	Foreign Institutional Investor
FTP	File Transfer Protocol
FVCI	Foreign Venture Capital Investors
GL	General Ledger
HO	Head Office
HTTP	Hypertext Transfer Protocol
IT	Information Technology
JDBC	Java Database Connectivity
LoB	Line of Business
LoC	Letter of Credit
LoU	Letter of Undertaking
MIS	Management Information System
MoC	Memorandum of Change
ODBS	Open Database Connectivity
ODS	Operational Data Store
OMG	Object Management Group
ORFS	Online Returns Filing System
OSMOS	Off-site Surveillance and Monitoring System
PDF	Portable Document Format
PL/SQL	Procedural Language/Structured Query Language
RBI	Reserve Bank of India
RRB	Regional Rural Bank
SARFAESI	Securitization and Reconstruction of Financial Assets and Enforcement of Security Interests
SCB	Scheduled Commercial Bank
SDMX	Statistical Data and Metadata eXchange
SOA	Service Oriented Architecture
SSI	Small Scale Industries
STP	Straight Through Processing
UCB	Urban Cooperative Bank
USD	United States Dollar
XBRL	eXtensible Business Reporting Language

Document Overview

The document contains the following chapters:

Chapter 1 - Guiding Principles for defining the approach: This chapter deals with the guiding principles and of the Approach for Automated Data Flow.

Chapter 2 - Assessment Framework: This chapter provides the detailed methodology (based on a score sheet) for self assessment by banks to determine their current state along Technology and Process dimensions as People dimension being an abstract and dynamic dimension, it is left to the individual banks to calibrate their strategies suitably. Based on the scores of self-assessment, the banks would be classified into clusters.

Chapter 3 - Common End-State: This chapter describes the end state, i.e. the stage to be reached by the banks to achieve complete automation of data flow. The end state has been defined across Process and Technology dimensions.

Chapter 4 - Benefits of Automation: This chapter elaborates the benefits of automation for the banks. It also defines the challenges and critical success factors for this project.

Chapter 5 - Approach for Banks: This chapter provides the approach to be followed by the scheduled commercial banks to achieve the end state. The approach has four key steps. Each step defines a number of activities and tasks that need to be accomplished to complete the step. The Chapter is divided into two sections. The first section deals with the standard approach which has to be followed by all banks and the second section deals with the specific variations from the standard approach and ways to deal with it.

Chapter 6- Proposed Roadmap: This chapter provides a framework for grouping the returns submitted to the Reserve Bank. Depending on the group of returns and bank's cluster (as defined in chapter on Assessment Framework), maximum estimated timelines have been indicated for completing the automation process.

Chapter 7 – Summary: This chapter provides the conclusion for the entire approach paper.

Annexure A – Assessment Framework – Scoring Matrix and Sample Assessment: This annexure provides the scoring matrix to be used by the banks for the purpose of

self assessment. The annexure also provides an illustrative assessment for a representative bank.

Annexure B – Task details for the Standard Approach and an illustrative example of approach adopted by bank: This annexure describes the task details required to complete an activity as defined in the chapter on Approach for banks. It also illustrates the adoption of the approach by a sample bank.

Annexure B1 – Task Details - “Prepare Data Acquisition layer”

Annexure B2 – Task Details - “Prepare Data Integration & Storage layer”

Annexure B3 – Task Details - “Prepare Data Conversion Layer”

Annexure B4 – Task Details - “Prepare Data Submission Layer”

Annexure B5 – Illustrative Approach for a bank

Annexure C – Returns Classification: This annexure provides an Excel based classification of representative returns.

Appendix I - Other End State Components, Technology Considerations and **Definitions**: This appendix describes the component-wise technology considerations for the End-state architecture. It also provides definitions of terms used in the document.

Appendix II – Global Case Study: This appendix describes the process followed by a global regulator for automating their data submission process.

Chapter 1

Guiding Principles for defining the Approach

The guiding principles define the basis on which the entire approach has been framed. These are applicable to all banks regardless of their overall maturity in Process and Technology dimensions.

1.1 Guiding Principles: The guiding principles used in this Approach are detailed below:

Guiding Principles	Rationale
Independent of the current state technology	• The approach defined in this document needs to be independent of the current state technology.
Need for common end state	• As the suggested approach is common across all banks, the end state is referred to as the common end state.
Need for flexible and inclusive approach	• The stated approach is inclusive of the needs of Indian banking industry particularly since the banks function at various levels of maturity on Process and Technology fronts. • The approach (not the end state) must be flexible to allow customization/modification as per a specific bank's requirements.
Need to protect the investment made in technology and infrastructure	• The banks have invested heavily in building their technology infrastructure. The approach seeks to provide a mechanism to leverage on the existing infrastructure and technology within the bank.
Classify return in logical groupings	• Enables the bank to classify the returns into smaller groups and allow for prioritization of resources and requirements. • Allows the banks to conduct the automation of data flow in phases.
Need for extendibility of approach to UCBs and RRBs	• The approach needs to be flexible, scalable and generic so that the approach can be extended to UCBs and RRBs.

Conclusion

The guiding principles ensure that the approach to achieve the common end state as mentioned above is independent of the current technologies being used by the banks and can be used by all the banks irrespective of their current level of automation. It seeks to leverage on the huge investment made by banks in technology infrastructure.

Chapter 2

Assessment Framework

The assessment framework measures the preparedness of the banks to achieve automation of the process of submission of returns. This framework covers the Technology and Process dimensions. This chapter explains the process to be followed by the banks to carry out a self-assessment of their current state.

2.1 Banks' Current State Assessment Framework

The end-to-end process adopted by the banks to submit data to Reserve Bank can be broken into the following four logical categories.



Figure 1: Information Value chain for Reserve Bank return submission

Banks have to assess themselves on dual dimensions i.e. Process and Technology across each of the above four logical categories which will lead to the assigned scores. Based on the overall score, the bank's preparedness and maturity to adopt a fully automated process will be assessed.

2.2 Methodology Adopted for Self Assessment

2.2.1 The overall methodology for assessment of current state is defined below:

- a) Self Assessment by banks based on defined parameters: Banks are required to assess themselves on Technology and Process dimensions. The parameters for Technology and Process dimensions and the corresponding assessment principles are given in **Annexure A**.
- b) Calculation of standardized score for each parameter: Each of the parameters is scored on a scale of 0-1. Due to this no single parameter can skew the results when the scores are aggregated.
- c) Calculation of the category score: The category scores are calculated by using the average of standardized scores of all applicable parameter(s) assessed for the category.
- d) Calculation of overall maturity score for each dimension: The overall maturity score is calculated using the weighted average of the individual category scores. Weights for assessing the overall maturity score, are calculated across four categories of assessment – namely data capture, data integration & storage, data conversion and data submission.
- e) Classification of bank as per the Technology and Process maturity score bands: The overall Technology/Process maturity score is used to place the bank in one of the clusters on the two dimensional axis viz. Process maturity vs. Technology maturity as illustrated below:

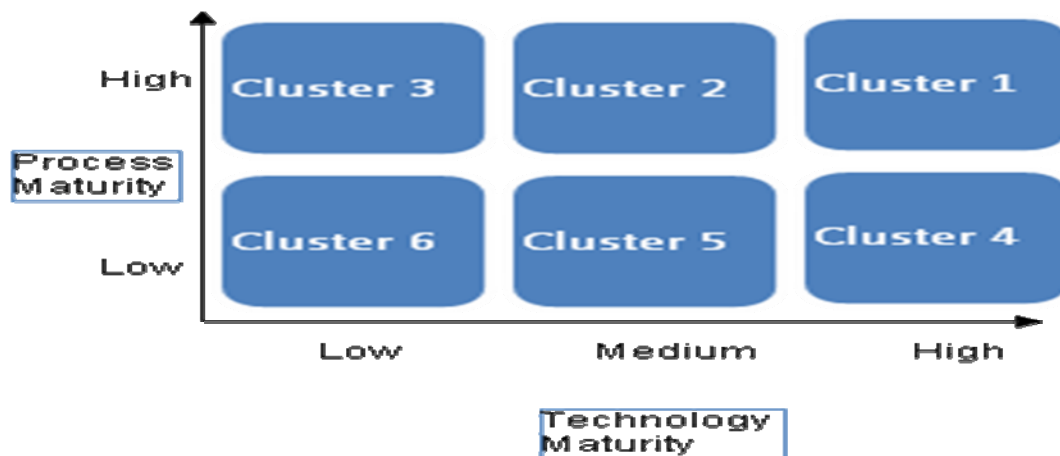


Figure 2: Technology and Process maturity based bank clustering

2.2.2 As an illustration, a sample assessment of a bank has been provided under **Annexure A**. A spread sheet with pre-defined fields to enable banks to carry out the self assessment exercise has been provided in paragraph 2 of **Annexure A**. The classification of the bank will define the implementation roadmap for reaching the end state.

Conclusion

Based on the level of Process and Technology maturity, each bank is placed in one of the six clusters illustrated above. The timeframe for automation for each group of returns will depend on the cluster in which the bank is placed. This in turn will determine the timelines for the banks to achieve the common end state for all returns.

Chapter 3

Common End State

The common end state is the state of complete automation for submission of the returns by the banks to RBI without any manual intervention. To achieve the objective of automated data flow and ensure uniformity in the returns submission process there is need for a common end state which the banks may reach. The common end state defined in this chapter is broken down into four distinct logical layers i.e. Data Acquisition, Data Integration & Storage, Data Conversion and Data Submission. The end state covers the dimensions of Process and Technology.

3.1 Data Architecture

Data Architecture refers to the design of the structure under which the data from the individual source systems in the bank would flow into a Centralized Data Repository. This repository would be used for the purpose of preparation of returns.

3.1.1 The conceptual end state architecture representing data acquisition, integration, conversion and submission is represented in Figure 3 given below. Under this architecture, the Data Integration & Storage layer would integrate and cleanse the source data. Subsequently, the Data Conversion layer would transform this data into the prescribed formats. The transformed data would then be submitted to Reserve Bank by the Data Submission layer.

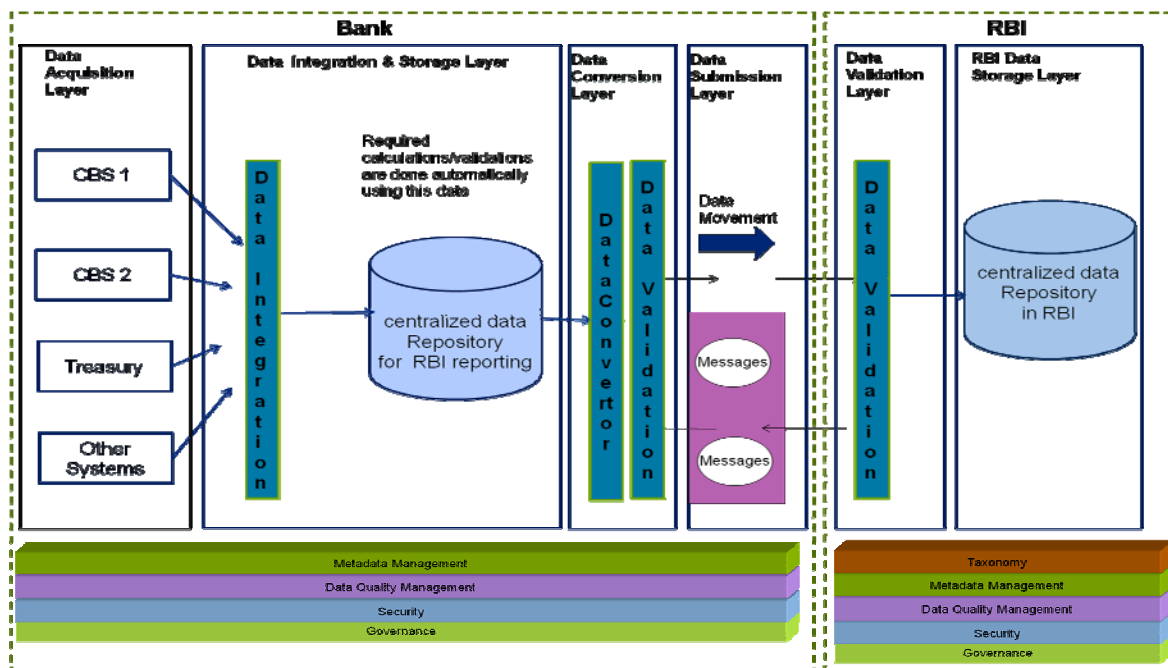


Figure 3: Conceptual Architecture for End State for banks to automate data submission

The details of each layer are enumerated below:

- (a) **Data Acquisition Layer:** The Data Acquisition layer captures data from various source systems e.g. - Core Banking Solution, Treasury Application, Trade Finance Application, etc. For data maintained in physical form, it is expected that this data would be migrated to appropriate IT system(s) as mentioned in the chapter on approach for banks.
- (b) **Data Integration & Storage Layer:** The Data Integration & Storage layer extracts and integrates the data from source systems with maximum granularity required for Reserve Bank returns and ensures its flow to the Centralized Data Repository (CDR). Banks having a Data Warehouse may consider using it as the CDR after ensuring that all data elements required to prepare the returns are available in the Data Warehouse. To ensure desired granularity, the banks may either modify source systems or define appropriate business rules in the Data Integration & Storage layer.
- (c) **Data Conversion/Validation Layer:** This layer converts the data stored in the CDR to the prescribed formats using pre-defined business rules. The data conversion structure could be in the form of a simple spreadsheet to an advanced XBRL instance file. The Data Conversion layer will also perform validations on the data to ensure accuracy of the returns. Some common validations like basic data checks, format and consistency validations, abnormal data variation analysis, reconciliation checks, exception report, etc. would be required to be done in this layer.
- (d) **Data Submission Layer:** The Data Submission layer is a single transmission channel which ensures secure file upload mechanism in an STP mode with the reporting platforms like ORFS. In all other instances, the required returns may be forwarded from the bank's repository in the prescribed format. The returns submission process may use automated system driven triggers or schedulers, which will automatically generate and submit the returns. When the returns generation process is triggered, the system may check if all the data required to generate this return has been loaded to the central repository and is available for generating the return. It may start preparing the return only after all the required data is available. The Data Submission layer will acknowledge error messages received from Reserve Bank for correction and further processing.

3.2 Process Elements

The key characteristics of the processes need to be defined for each of logical layers:

Logical Layer	Key Characteristics to be captured as part of the Process Documentation
Data Acquisition	Process may document: • Data fields in the return • Computations required

Logical Layer	Key Characteristics to be captured as part of the Process Documentation
	• Data owners, roles and responsibilities • Source systems for the different data fields • System access privileges • Frequency of data capture from the source systems Process may be reviewed periodically and/ or when there is a change in the sources systems or introduction of new reports, change in reporting format etc. Any change in the process may be made after a thorough assessment of the impact of the change on the end-to-end submission process Any change in the process may be documented and approved by an authorized committee which has representations from various departments
Data Integration & Storage	Process may document: • Data classification in the reports • Business rules which synchronize the data • Business rules to cleanse and classify the data • Data quality and validation checks to ensure the correctness of the data • Fields to be captured during audit trail • Length of the time the data may be stored in the CDR • Method of backup and archival after the elapsed time • System details where the archived data is stored and the process of retrieval from the archived data store Process may be reviewed periodically (frequency depending on the change in the sources systems, introduction of new reports, change in reporting format etc.)
Data Conversion	Process may clearly document • Reserve Bank requirements (codes, format, classification) for each return along with the interpretation of the fields • Procedures for performing data validation to ensure correctness of format and content of the return. The data validation rules include checks

Logical Layer	Key Characteristics to be captured as part of the Process Documentation
	for basic data validations and format validations as well as validations to ensure consistency, validity of the data. • Defined business rules to transform the data as per Reserve Bank requirements • Fields to be captured during the audit trail • Procedure for performing audits listing the roles and responsibilities, frequency of the audit Process may be reviewed periodically and/or whenever there is a change in the reporting format
Data Submission	Process may clearly document • Submission schedule for each return • Return submission process (ORFS or others) • Procedure to check compliance to Reserve Bank requirements (in terms of the format, syntax, timeliness of submission etc.) listing the roles for checking the compliance • Assigned responsibility to check for the delivery acknowledgement • Corrective procedure in case of failed data validation • Procedure to investigate into discrepancy, if any

3.3 Returns Governance Group (RGG) Organization Structure

3.3.1 The governance of the returns submission process at the banks is mostly distributed across the various business departments with little or no centralization. Presently, governance for returns submission, if any, is limited to the following kind of roles:

- (a) Central team, if any, acting as “Facilitator” - Routing queries to respective departments, monitoring report content, verification of returns, adherence to timelines, interpretation guidelines, clarifications from Reserve Bank on queries.
- (b) Central team, if any, acting as “Watchdog” – Conducting periodic internal compliance audits, maintaining a repository of who files what returns, and provides updates on regulatory changes to internal users.

3.3.2 In order to strengthen the return submission process in an automated manner, it is suggested that banks may consider having a Returns Governance Group (RGG) which has representation from areas related to compliance, business and technology entrusted with the following roles and responsibilities:

- (a) RGG may be the owner of all the layers indicated in the end state from the process perspective. The role of the RGG may be that of a “Vigilante and Custodian”.
- (b) Setting up the entire automation process in collaboration with the bank’s internal Technology and Process teams. This is to ensure timely and consistent submission of returns to Reserve Bank. Though the data within the repository could be owned by the individual departments, the RGG could be the custodian of this data.
- (c) Ensuring that the metadata is as per the Reserve Bank definitions.
- (d) Management of change request for any new requirement by Reserve Bank and handling ad-hoc queries.

3.3.3 In the context of the automated submission system being adopted by the banks, the RGG, will have the following responsibilities for each layer of the return submission end state:

Layer	Responsibility
Data Acquisition	- Defining fault limits for errors - Defining threshold limit for deadlines/ extensions on timelines - Regulatory/ Guideline Changes as and when directed by the Reserve Bank - Introducing select repetitive ad-hoc queries as standard queries in the system
Data Conversion	- Validating the syntax of the reports - Checking for compliance to Reserve Bank requirements
Data Submission	- Enforcing data transmission timelines - Delivery acknowledgement - Investigation of anomalies

Conclusion

Each bank may be at a different level of automation, but as defined in this chapter, the end state will be common for all the banks. To achieve this, the banks may require a transformation across dimensions of Process and Technology. The key benefits derived by the bank in implementing the end state are defined in the next chapter.

Chapter 4

Benefits of automation

By adopting an automated process for submission of returns, the banks would be able to submit accurate and timely data without any manual intervention. This process would also enable the banks to improve upon their MIS and Decision Support Systems (DSS), etc.

4.1 Key Benefits of automated data flow for banks: The automation of data flow will benefit the banks in terms of the following:

- Improved Timelines
- Enhanced Data Quality
- Improved Efficiency of Processes
- Reduced Costs
- Use of the CDR for MIS Purposes

4.2 A comparison of the issues with the current scenario and benefits of automation is given in the table below:

Parameter	Issues with Current State	Benefits of Automation
Timelines	Preparing the data for submission to Reserve Bank is a challenge owing to the extent of manual intervention required for the entire process.	• The end-to-end cycle time will be reduced as data will now be available in a CDR • The entire process of submitting the data will be driven by automated processes based on event triggers, thereby requiring minimal manual intervention. This would ensure that the returns are submitted to Reserve Bank with minimal delays.
Quality	Data Quality and Validation is not driven by defined processes. Hence the onus of validating and ensuring the data is entered correctly is dependent on the individuals preparing the returns. The data is not integrated across the different source systems. This results in data duplication, mismatches in similar data, etc. Metadata is not harmonized across different applications.	• Minimal manual intervention reduces errors and ensures better data quality • A well defined process for ensuring data quality ensures that the source data is cleansed, standardized and validated so that data submitted to Reserve Bank would be accurate and reliable. • Data validation checks in the data integration & storage layer as well as in the data conversion layer ensures that the data being submitted is correct and consistent. • The data would be harmonized by ensuring standard business definitions across different applications.

Parameter	Issues with Current State	Benefits of Automation
Efficiency	Being a manual process, efficiency is person dependant.	Automation will eliminate dependence on individuals and improve efficiency.
Cost	Returns preparation is a non-core activity for the bank which does not generate any revenue. Manual processes require more manpower.	- Automation will reduce the manpower required for returns generation. So, more manpower will be available for revenue generating activities. - Automation will also improve decision making process and MIS within the bank leading better business strategies and improved bottom lines.
Utilisation of CDR for MIS purposes	Use of spread sheet or CBS for generating MIS MIS is generated by each source application individually.	- The CDR created can be used for internal MIS purposes also. This would free up the core operational systems for transactional purposes. - The CDR would have integrated data from across different source systems which will allow banks to have a consolidated view.

Conclusion

With the realization of the end state objective of automated data flow, the banks may be benefited in terms of enhanced data quality, timeliness, and reduced costs.

Chapter 5

Approach for banks

This Chapter defines the approach to be followed by banks to reach the common end state. The approach, based on the Guiding Principles defined in Chapter 1, is divided into two sections:

(a) Standard Approach to be followed by all the banks - The Standard approach assumes that the bank has little or no automation for submission of returns due to which it is possible to provide an end-to-end roadmap for achieving total automation of data flow. The standard approach comprises of four key steps which is divided into activities and further divided into tasks & sub-tasks, which are only indicative and not exhaustive.

(b) Variations (Customisation required due to variations from the Standard Approach) - Variations refer to situations where certain level of automation is already in place at the banks-end and would therefore require customisation of the existing systems to achieve total automation. In other words such customisation would mean carrying out changes and modifications to the existing architecture with respect to the technology and processes dimensions with a view to reach the common end state.

5.1 The Standard Approach Overview

5.1.1 The high level schema for the proposed approach is represented in Figure 4. The approach comprises of four logical steps. Each of the steps is further subdivided into activities. The activities are further broken down into the tasks required to complete the activity.

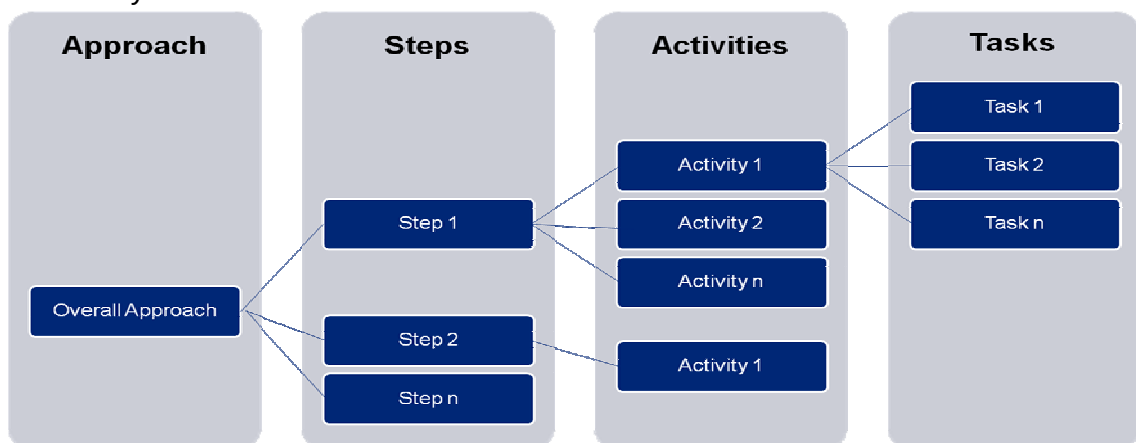


Figure 4: Approach Design

The four steps involved in the standard approach for automation of data flow would involve building of the following layers as illustrated below:

- Data Acquisition layer
- Data Integration & Storage layer

- Data Conversion layer
- Data Submission layer

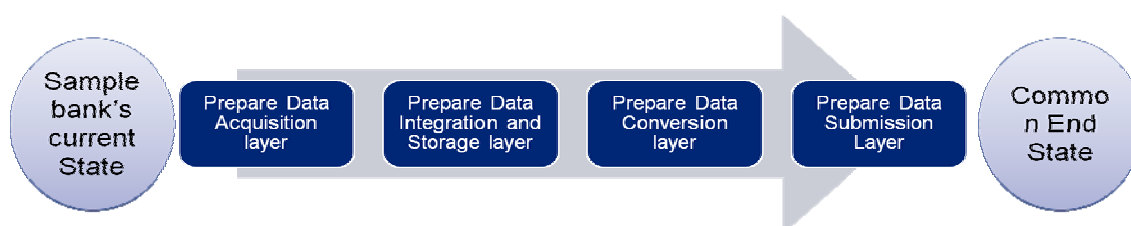


Figure 5: Steps to automate data submission process

5.1.2 Each of these steps is discussed below along with activities required to achieve the objective underlying each step. The figure below lists a few illustrative activities involved in the four steps. Further the activities are classified into tasks and sub-tasks as given in Annexure B.

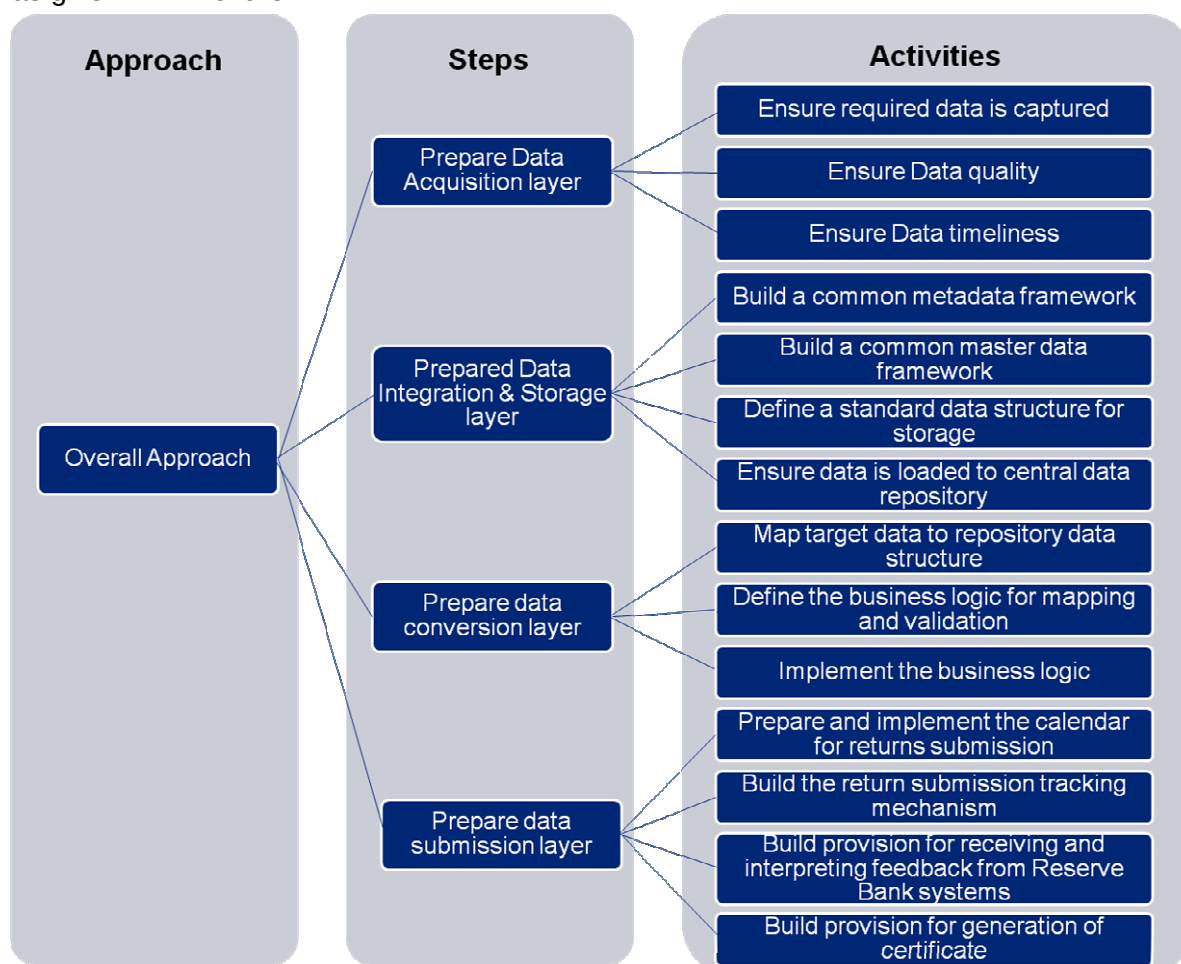


Figure 6: Illustrative steps and activities to be followed in the Standard Approach

5.1.2.1 Building Data Acquisition Layer: This step focuses on building the source data. This will ensure availability and quality of the data in the source systems of the banks. The activities identified under this step are detailed below.

- a) **Ensure Data Capture:** The data components required to be submitted in the returns are mapped to the data elements in the source systems of the acquisition layer. The components which cannot be mapped need to be identified so that the source system can be enhanced accordingly. This activity can be further broken down into tasks as discussed in Annexure B.
- b) **Ensure Data Quality:** Data quality is ensured by improving the data entry processes and by introducing post entry data validation checks. This activity can be further broken down into tasks as discussed in Annexure B.
- c) **Ensure Data Timeliness:** The availability of data elements may be matched with the frequency prescribed by Reserve Bank for the relative returns. This activity can be further broken down into tasks as discussed in Annexure B.

5.1.2.2 Building Data Integration & Storage Layer: The source data from the Data Acquisition layer may be extracted, integrated and stored in the CDR. This repository would be the single source of information for all returns. This step outlines the activities and tasks to be performed to build this layer. The activities identified under this step are detailed below:

- a) **Build a common master data framework -** Master data consists of information about customer, products, branches etc. Such information is used across multiple business process and is therefore configured in multiple source systems of the bank. As part of this activity, the inconsistencies observed in capturing this master data are studied to ensure building of a common master data. This activity can be further broken down into tasks as discussed in Annexure B.
- b) **Build a common metadata framework:** Metadata is information extracted from the data elements from the functional and technical perspective. The tasks required to build the common metadata framework are detailed in Annexure B.
- c) **Define a standard data structure for storage:** The CDR stores bank-wide data required for preparation and submission of returns. The flow of information from Source systems to the CDR will require a well defined data structure. The data structure may be based on the master data and metadata framework built in first two activities. This activity can be further broken into the tasks as given in Annexure B.
- d) **Ensure data is stored in Centralized Data Repository:** Data from the source systems would be stored in the Centralized Data Repository. The stored data would be integrated on the basis of well defined business rules to ensure accuracy. This would involve initial as well as incremental uploading of data on a periodic basis. This activity can be broken into tasks as given in Annexure B.

5.1.2.3 Build Data Conversion Layer: The Data Conversion Layer transforms the data into the prescribed format. This layer is an interface between data storage layer and submission layer. The activities identified under this step are detailed below:

- a) Mapping of prescribed data with repository data -To prepare the required return, it is essential to ensure mapping of prescribed data with the corresponding structures in the CDR. This activity can be broken into tasks as given in Annexure B.
- b) Design the business logic for mapping and validation: As the data available in the CDR will not be in the same structure and format as prescribed, business logic will need to be defined for this conversion. Since data conversion layer is the final layer where the data transformations take place before submission, the business logic will also include data validation rules. The data validation rules will include basic data checks, format validations, consistency validations, abnormal data variations and reconciliation checks. This activity can be broken into tasks as given in Annexure B.
- c) Availability of infrastructure: Successful implementation of business logic would depend upon availability of suitable infrastructure. This activity can be broken into tasks as given in Annexure B.

5.1.2.4 Build Data Submission Layer: The systems and codes in this layer would have the capability to authorise the data flow into the Reserve Bank's systems. This layer will also have the ability to receive messages and information from the Reserve Bank systems. The Data Submission layer would certify that the data submitted to Reserve Bank is reliable and generated using automated system. The activities identified under this step are detailed below:

- a) Adhering to the calendar of returns-Based on the frequency of submission of returns, this activity would generate a detailed calendar with reminders and follow-ups. This activity can be broken into tasks as given in Annexure B.
- b) Tracking mechanism for return submission-This activity provides for a mechanism to track the submission of returns and generate status reports. This activity can be broken into tasks as given in Annexure B.
- c) Provision for feedback from Reserve Bank systems-Under this activity, a mechanism would be developed to receive feedback to act as a trigger for further action. This activity can be broken into tasks as given in Annexure B.
- d) Provision for generation of certificate -Under this activity, the banks may be able to certify that the return submission is fully automated with no manual intervention. This activity can be broken into tasks as given in Annexure B.

5.2. Variations to the Approach

The standard approach would be customized by banks on the basis of the current state of technology. These customization cases are called variations of which. A few illustrative examples have been discussed in Figure 7 below:

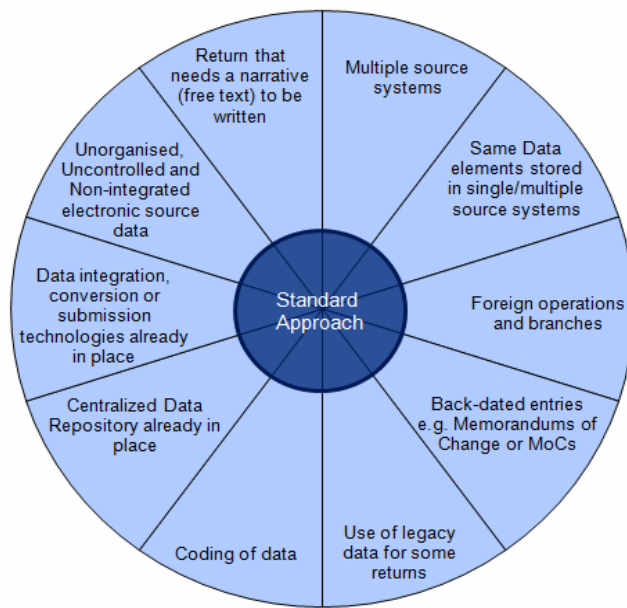


Figure 7: Variation to Standard Approach

The details of the variations are as follows:

5.2.1 Multiple source systems: For data residing in multiple source systems, there would be a need for integration of data. In such cases, customization of the standard approach would be needed at the time of developing a CDR.

- a) A detailed mapping of both master and transactional data, wherever applicable, may be carried out across all the source systems. This mapping may be at the same level of granularity. The common elements across both these systems must be mapped to understand the consolidation process.
- b) The metadata and master data framework would need to be designed keeping in mind the sharing of source system data across multiple systems.
- c) The integration of data across the source systems would continue to take place through the Data Integration & Storage layer and the integrated data would be available in the CDR.

5.2.2 Same Data elements stored in single/multiple source systems: Generally, each data element may be stored in a single and definitive source system. However, there would be instances of same data elements being stored in multiple source systems. It may also be possible that within the same system there may be multiple modules and data flows from one module to another. In such cases, there is a need to identify the appropriate source system from where the data element has to be taken for further processing. The alternatives to be followed for such a scenario could be:

- Use the original source of data i.e. the source where data is generated for the first time.
- Use the latest source of data
- Use the data from the most significant system
- Use the most appropriate source of data

For master data, there may be instances where different attributes for the master data will be sourced from different systems. Therefore the source system with enriched data, highest data quality with latest updates may be used. There may be instances of duplicate source of data within the same system or module. In such cases, data from the most appropriate source may be taken.

5.2.3 Foreign operations and branches: The standard approach assumes that the bank is operating only in India and hence has a common set of business and operational guidelines under which it operates. However, banks with foreign operations will have to follow different set of guidelines. The local nuances and business environment may warrant different business practices.

In such cases, design of the CDR as well as the Data Integration & Storage layer will have additional considerations and corresponding activities and tasks. The key considerations and corresponding activities for this variation are discussed below:

- Location of data: In some cases the source data for foreign operations may not be available in India. In such cases, a periodic extract of data must be made available for the CDR.
- Local regulations and norms: The regulatory requirements of each foreign country are different. Therefore, there may be a need to maintain multiple accounting books. The data from these local accounting books may be integrated using the Data Integration & Storage layer and the common data structure in the CDR.
- Different time zones: Different time zones at different locations lead to different closing times. This leads to either non-updation or delay in updation of data.

5.2.4 Back-dated entries e.g. Memoranda of Change or MoCs: MoCs are given affect by making changes in the financial statements. . Back dated entries are required to be made in the source systems to incorporate these changes, which is a challenge in the context of automated data flow. This can be handled by designing an appropriate module which will integrate the MoCs and generate a revised GL for further processing.

5.2.5 Use of legacy data for some returns: In some cases, legacy data may be required for submission of returns for which it may be necessary to store the same in the CDR. This data is static and is typically stored in tapes or maintained in physical form in registers. In such a case, the CDR may be enhanced for capturing such legacy data.

5.2.6 Coding of data: Coding and mapping of master and transactional data would be required for some of the returns to synchronise it with the codes prescribed by Reserve Bank e.g. BSR coding, sector coding for Priority sector returns, etc. This can be done either by updating the source systems or building the same outside the source systems.

5.2.7 Centralized Data Repository: If a bank already has a centralized data repository in place, then the same may be leveraged for achieving automated data flow to Reserve Bank. In cases where data is partially available in the CDR, the bank may

enhance the integration layer for capturing the relevant data required for submission to Reserve Bank.

5.2.8 Data integration, conversion/submission technologies: Banks may choose to leverage on the investment already made in technologies while building the various layers as prescribed in the end state architecture, rather than procuring new technology.

5.2.9 Unorganised, Uncontrolled and Non-integrated electronic source data: It may be possible that a bank may be using desktop-based or standalone technology tools e.g. spread sheet, etc. for storage of certain source data. While this data may be in electronic form, it is dispersed across the bank. Therefore, there is no control on addition, deletion or updation of this data and no audit trails are available to track this data. As this data is not organised and reliable, it cannot be used for the automated data flow to Reserve Bank. In such cases, the data may be brought under a controlled mechanism or be included in the source system.

5.2.10 Return that needs a narrative (free text) to be written: Some of the returns may require a narrative to be written by the bank before submission. Such narratives are not captured in any source system and are required to be written during the preparation of the returns. In such cases a separate system can be developed where the user can enter all the narratives required for the returns. These narratives will be stored in the central repository. The Data conversion layer while generating the returns will include these narratives in the appropriate returns. Storing the narratives in the central repository will ensure that the complete return can be regenerated automatically as and when required.

Conclusion

Banks may carry out a self-assessment, based on the Technology & Process dimensions and place itself appropriately in one of the prescribed clusters. The banks can then follow the standard approach by applying variations, if any, for achieving the end state.

Chapter 6

Proposed Roadmap

6.1 Due to varied levels of computerisation in banks as also due to a large number of returns to be submitted by them, it is necessary to plan and execute the steps for achieving automated data flow between banks and Reserve Bank in a time-bound manner. To achieve this objective, it is important for banks to carry out a thorough assessment of their current state of computerisation, through a prescribed methodology. This would make it possible for banks to place themselves in different clusters, as has been illustrated in Annexure A. Simultaneously, the returns required to be submitted by banks to Reserve Bank, would be categorised into five broad groups.

6.2 Based on a quick study made on eight banks, maximum time-lines for achieving automation have been estimated for different clusters of banks for each group of returns. They are indicated in Table below.

Sl. No.	Return-Group	Estimated maximum timelines (in months)					
		Bank Cluster 1	Bank Cluster 2	Bank Cluster 3	Bank Cluster 4	Bank Cluster 5	Bank Cluster 6
1	Simple	4-6	7-9	7-9	5-7	8-10	10-12
2	Medium	7-9	10-12	13-15	9-11	12-14	15-17
3	Complex I	10-12	13-15	16-18	12-14	15-17	18-21
4	Complex II	13-15	16-18	19-21	15-17	18-20	21-23
5	Others	NA	NA	NA	NA	NA	NA

Table 1: Timelines for achieving the end-state of automated data flow from banks

6.3 To begin with, a common-end state would be achieved when a bank is able to ensure automated data flow to Reserve Bank for four out of the five groups of return (Simple, Medium, Complex I and Complex II). It may not be possible for the Banks to achieve complete automation for the returns falling in the fifth group viz. "Others" as data required for these returns may not be captured in IT Systems. However, it is envisaged that in due course, this group would also be covered under the automated dataflow.

6.4 Incidentally, as a collateral benefit of automation, the banks would be able to channelize their resources efficiently and minimize the risks associated with manual submission of data.

6.5 The proposed Roadmap comprises of the following-:

6.5.1 Returns Classification Framework

As mentioned above, the returns currently submitted by the banks to Reserve Bank, have been classified into five broad groups on the basis of the feedback provided by eight sample banks by adopting a Moderate approach. Each return was assigned to the group which was chosen by majority of banks (refer to the Annexure C for the detailed Returns Classification on 171 representative set of returns). Each of these groups of return can be considered for automation individually without having an adverse impact on the other returns.

6.5.2 Implementation Strategy

Due consideration has been given to the different clusters of banks and the different categories of returns to ensure that the bank has flexibility in implementing the approach based on their current state without compromising on the overall vision of automating the entire process of submission of returns to Reserve Bank. The suggested timelines have been given in ranges to factor in differences across banks within a cluster.

6.5.3 Key Challenges

The banks might face following challenges on account of converting to a fully automated state of submission:

- (a) The data required to prepare the returns might be captured in more than one source application. Hence all the source systems must be integrated prior to submission of the returns through complete automation.
- (b) Currently MoCs are given effect outside the core applications; hence this process will need to be integrated into this solution.
- (c) The data transformations and business rules are not well documented; hence prior to undertaking the automation project, it might be necessary for banks to prepare a detailed process documentation to capture this information.
- (d) Legacy data will need to be corrected for mapping to the Reserve Bank published codes and definitions.
- (e) Data granularity will need to be standardized across the source systems.

Conclusion

The roadmap enables the bank to take a phase wise approach to achieve the end state. The bank can decide to break the above defined return groups into smaller groups. These sub-groups can then be implemented either in a sequential manner or in parallel depending on the ability of the bank to deploy the requisite resources. However, care must be taken to ensure that the overall time lines are adhered to.

Chapter 7

Summary

7.1 An automated data flow from the IT systems of banks to Reserve Bank with no manual intervention will significantly enhance the quality and timeliness of data. This will benefit the Indian banking sector by enabling more informed policy decisions and better regulation. Over and above the regulatory benefits, there will be many other benefits in disguise to the internal operations, efficiency and management of banks. This can be understood by taking a closer look at the different layers of the end-state given in chapter on common end state.

7.2 Firstly, the data acquisition layer will ensure higher data quality and higher coverage of source data by the systems in electronic form. This will eliminate any inefficiency in the operational processes of the bank as most of the processes can be automated using systems. Secondly, major benefits will be derived from the harmonized metadata in the bank at the data integration & storage layer. This will allow for smooth communication and lesser data mismatches within the bank which often take place due to difference in definitions across departments. Thirdly, the centralized data repository will have a wealth of data that can be easily leveraged by the bank for internal reporting purposes. This not only saves major investment towards a separate reporting infrastructure but also provides management of the banks with recent data of high quality for decision support. Last but not the least, the automated flow of data from banks will have significant benefits in costs and efficiency of the returns submission process itself. The automated flow of data will ensure a smooth process with minimal delays in submission of data that is of much higher quality.

7.3 The realization of the above benefits will accrue to the bank when the automation of data flow is accomplished and the final automated end-state has been reached. The end-state has a layer-by-layer flow of data from the place where it is generated to the central repository and finally to Reserve Bank. To achieve the end-state the banks need to use the constituents of the paper in an integrated manner as given below:

- a) Banks need to carry out a current state assessment across the various technology and process parameters defined in chapter on Assessment Framework. This will help the bank identify the cluster that it falls in.
- b) Banks need to study the proposed roadmap. Banks need to take one return group at a time and study the approach provided in chapter on approach for banks.
- c) While studying the approach, banks need to customize the approach based on the return group, variations applicable to the bank and also on the business model and technology landscape of the bank.
- d) A further fine tuning of the approach can be done by carrying out the assessment separately for different return groups and not for the bank as a whole.

Annexure A – Assessment Framework – Scoring Matrix and Sample Assessment

Assessment Framework – Parameters, scoring matrix and guidelines

The parameters and guidelines for conducting self assessment on technology and process dimensions are as defined below:

1(a) Parameters and guidelines for Technology dimension:

Category	Parameter	Guiding Principle for Assessment		
		Low (Score assigned – 0)	Medium(Score assigned – 1)	High (Score assigned – 2)
Data Acquisition	% of data captured in IT system vis-a-vis data available in individual files and or physical forms.	0-75% of all data captured in IT system	75-90% of all data captured in IT system	90-100% of all data captured in IT system
Data Acquisition	Level of data quality checks/ assessment at data source	No data quality checks available in transactional system(s)	Data quality checks available for some fields in transactional system(s) done as a one-of exercise	Data quality checks available for all fields in transactional system(s) and done on a regular basis
Data Integration & Storage	% of source systems having common data integrated across them	0-50% of all systems have data integrated with each other	50-75% of all systems have data integrated with each other	75-100% of all systems have data integrated with each other
Data Integration & Storage	Near Real time vs. batch integration of data	Data from the different systems is integrated using periodic batch file transfers?	Data from the different systems is integrated on near real time basis?	NA
Data Integration & Storage	Availability of centralized data repository	There is no central integrated data repository available for regulatory reporting purpose	There is a central integrated data repository available for regulatory reporting purpose	NA

Category	Parameter	Guiding Principle for Assessment		
		Low (Score assigned – 0)	Medium(Score assigned – 1)	High (Score assigned – 2)
Data Integration & Storage	Granularity of data in the repository	No repository available	Master data is at the maximum granular level required for RBI returns while transactional data is stored in summarized/ aggregated forms	All data captured in centralized data repository is at maximum granularity level required for RBI return, as in the transactional system
Data Integration & Storage	Metadata availability (if there are multiple source systems)	Metadata is not harmonized across applications	Metadata is harmonized across applications	NA
Data Integration & Storage	Audit trail from integrated data to originating data source	Audit trails are not maintained	Audit trails are maintained within individual application with no end to end integration to check lineage of data	An end to end audit trail is maintained from originating system to the final data submission to Reserve Bank
Data Conversion	Availability of automated extraction tools	Manual data extraction using MIS reports	Scripts are used for data extraction for the returns	Automated extraction and querying tools are used for data extraction
Data Conversion	Availability of Return preparation tools	Manual data preparation using Excel reports And templates	Scripts are used for preparing the data in the Reserve Bank formats	Automated reporting tools are used for preparing the returns
Data Conversion	System handled transformation, classification, mapping as per Reserve Bank requirements	Transformations and /or classification of data required for preparing the returns are done using manual entry	Transformations and /or classification of data required for preparing the returns are done using excel based	Transformations and /or classification of data required for preparing the returns are done using automated tools

Category	Parameter	Guiding Principle for Assessment		
		Low (Score assigned – 0)	Medium(Score assigned – 1)	High (Score assigned – 2)
		templates		
Data Conversion	Availability of automated analysis and validation tools	Analysis and validation of the data for reporting is done manually	Analysis and validation of the data for reporting is done using a checklist	Analysis and validation of the data for reporting is done using an automated tool
Data Submission	Data Entry mechanism for Data Submission	The submission to Reserve Bank for ORFS returns is done through manual keying of data	The submission to Reserve Bank for ORFS returns is done using an XML file upload	NA
Data Submission	Validation of Data Entered	Data is validated by Maker only	Data is validated using a Maker-Checker	Data is uploaded, hence no validation required

1 (b) Parameters and guidelines for Process dimension:

Category	Parameter	Guiding Principle for Assessment		
		Low (Score assigned – 0)	Medium (Score assigned – 1)	High (Score assigned – 2)
Data Acquisition	Documented processes for data capture, listing the right granularity of data, source systems, and departments involved	Ad-hoc processes with little or no process definitions	Well defined policies and procedures being followed	NA
Data Acquisition	Periodic review/ approval of data capture process	Ad-hoc review process	Defined review process	NA
Data Integration & Storage	Documented processes for data integration, listing data owners, mode of data transfer (email, hard copy, automated), guidelines for data integration & collation	Ad-hoc processes with no clear guidelines regarding data ownership and integration procedures	Standard procedures followed with well-defined data ownership and guidelines	NA
Data Integration & Storage	Periodic review of data integration process	Ad-hoc review process	Defined review process	NA
Data Integration & Storage	Documented guidelines or checklist for maintaining data quality	No guidelines or checklist followed	Maker checker arrangement in place to ensure data quality	Checklists and guidelines used for data quality monitoring
Data Conversion	Documented processes for data extraction and preparation , listing the mapping required (to BSR), format of the data	Ad-hoc processes with little or no process definitions	Well defined policies and procedures being followed	NA
Data Conversion	Documented Audit process (with defined roles, responsibilities, timeframe etc.)	Ad-hoc audit process with unclear roles and responsibilities	Timely Audit process in place with clearly outlined responsibilities and timeframe	NA
Data Submission	Documented processes for report submission, listing the in-charge for submission and mode	Individual responsibilities with little or no structured	Structured processes with well-defined	NA

Category	Parameter	Guiding Principle for Assessment		
		Low (Score assigned – 0)	Medium(Score assigned – 1)	High (Score assigned – 2)
	of submission (ORFS, XBRL, hard copy, email)	processes	responsibilities	
Data Submission	Periodic review of data submission process	Ad-hoc review process	Defined review process	NA
Data Submission	Tracker Process for tracking status of submissions (e.g. % of reports submitted on time, pending submissions)	Little or no tracking in place	Tracking procedure to ensure timely submissions	NA
Data Submission	Documented process in case of error (root cause analysis, audit trail for racing back)	Issues handled on a need case basis	Well defined procedures and responsibilities outlined for issue resolution	NA

1 (c) Self Assessment using the Excel Model

The banks will need to conduct a self assessment using an excel model. The bank must evaluate themselves on their Technology and Process Maturity based on the parameters defined above and assigns a rating of High, Medium or Low based on their evaluation. The response must be populated along with the rationale in the “Technology Maturity” and “Process Maturity” worksheets. The overall maturity score will be automatically calculated using the embedded logic within the worksheet.

[Click here to view the Excel Model for banks.](#)

The Technology maturity score is on a scale of 0-1 and hence 3 bands of 0-0.33 (Low), 0.33-0.67 (Medium), 0.67-1 (High) are defined. Similarly the Process maturity score is on a scale of 0-1 and hence 2 bands of 0-0.5 (Low), 0.5-1 (High) are defined.

2. Sample Assessment

A Sample Assessment for one of the bank is illustrated below as reference:

2(a) Technology Dimension:

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
Data Acquisition	% of data captured in IT system vis-a-vis data available in files/forms	Is all the data captured in IT system vis-a-vis data available in files/forms	Medium	About ~85 % data is captured in automated IT system	1	0.50	0.50	0.44
Data Acquisition	Level of data Quality checks/assessment at data source	Are the data quality checks done at the source system level (e.g. - in CBS) for verifying data quality (a sample list of checks could be data format checks, cross field validations, mandatory data checks etc)	Medium	CBS and other transaction systems have some field level checks to validate the data type and other rules	1	0.50		
Data Integration & Storage	% of source systems with integrated data	What percentage of source systems are integrated in terms of the data to make sure the data across systems is consistent	Low	There is a partial integration of data in terms of customer-ids only	0	0.00	0.25	0.28
Data Integration & Storage	Near Real time vs. batch integration	Is the integration of data done in real	Low	The integration is done in batch	0	0.00		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
		time or as periodic batch update		mode				
Data Integration & Storage	Availability of centralized data repository	Is there a centralized data repository containing harmonized data from various operational systems	Low	There is a MIS repository which only captures CBS data	0	0.00		
Data Integration & Storage	Granularity of data model for the data repository	What is the granularity of data model for data repository	High	The MIS repository contains the most granular data from CBS	2	1		
Data Integration & Storage	Metadata availability (if there are multiple source systems)	Is there a metadata repository available to ensure the consistent data definitions are maintained across multiple source systems	Low	There is no harmonized metadata available	0	0.00		
Data Integration & Storage	Audit trail from integrated data to originating data source	Is there an audit trail when consolidating/integrating data to trace back the originating system	Medium	The audit trail is captured in the source systems only	1	0.5		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
Data Conversion	Availability of automated querying tools	Is there some kind of automated querying tool(s) used for extracting data from IT systems	Medium	Scripts are used to extract data from source systems	1	0.50	0.25	0.17
Data Conversion	Availability of Report generation tools	Is there some report generation tool that is available to provide reports	Low	Excel is predominantly used for preparing returns	0	0.00		
Data Conversion	System handled transformation, classification, mapping as per Reserve Bank requirements	Is the transformation/classification of data to suit the Reserve Bank requirements handled outside the system using some solution like MS Excel etc	Medium	The transformation/codification on returns data is handled using Excel	1	0.50		
Data Conversion	Availability of automated analysis and validation tools	Is there an automated tool for defining and executing business rules for validation of data in place	Low	Analysis and validation is manual	0	0.00		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
Data Submission	Manual keying in the data or submission as batch file	Is the data to be submitted keyed in manually field by field or is it a batch upload process	Low	Data is keyed in manually in the form based interface provided by Reserve Bank submission systems	0	0.00	0.25	0.11
Data Submission	Validation of Data Entered	What is the validation mechanism used to verify correctness of data in the submission system	Medium	Verification of manually entered data is through a maker-checker mechanism	1	0.50		
IT maturity score							0.36	
IT maturity Level							Medium	

2(b) Process Dimension:

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
Data Acquisition	Documented processes for data capture, listing the right granularity of data, source systems, and	Is there a documented processes for data capture, listing the source systems, or departments	Medium	A documented process is used to define the list of data sources and the owner department	1	1.00	0.50	0.44

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
	departments involved	involved		ents				
Data Acquisition	Periodic review/ approval of data capture process	Does a centralized committee or department exist to review/ approve data capture process changes	Low	The review of the source system and the data acquisition process is ad-hoc	0	0.00		
Data Integration & Storage	Documented processes for data collation, listing data owners, mode of data transfer (email, hard copy, automated), guidelines for data integration & collation	Is there a documented processes for data collation, listing data owners, mode of data transfer (email, hard copy, automated), guidelines for data integration & collation	Medium	The data owners and the format for sharing data is well defined	1	1.00	0.50	0.28
Data Integration & Storage	Periodic review of data collation process	Is there a centralized committee or department to review/ approve data	Low	The review of the source system and the data acquisition process	0	0.00		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
		collation process changes if there is a change in the data ownership or systems		is ad-hoc				
Data Integration & Storage	Documented guidelines or checklist for maintaining data quality	Does documented guidelines or checklist for maintaining data quality exist	Medium	Maker checker process used as the mechanism for review of data quality	1	0.50		
Data Conversion	Documented processes for data extraction and preparation, listing the mapping required (to BSR), format of the data	Is there a well defined policies and procedure for data extraction	Low	The data conversion process is ad-hoc without defined procedure to do data transformations	0	0.00	0.50	0.17
Data Conversion	Documented Audit process (with defined roles, responsibilities, timeframe etc.)	Is there a documented processes for data extraction and preparation, listing the mapping required (to BSR),	Medium	The data conversion process is subject to periodic review and audit at well	1	1.00		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
		format of the data		defined intervals				
Data Submission	Documented processes for report submission, listing the in-charge for submission and mode of submission (ORFS, XBRL, hard copy, email)	Does a centralized committee or department exist to review/approve data preparation process changes, if there is a change in data format required by Reserve Bank	Medium	The compliance department acts as a centralized department to oversee the submission	1	1.00	0.75	0.11
Data Submission	Periodic review of data submission process	Is there a documented Audit process (with defined roles, responsibilities, timeframe etc?)	Medium	There is periodic review of the submission process with well defines roles for all involved personnel	1	1.00		
Data Submission	Tracker Process for tracking status of submissions (e.g. % of reports	Is there documented processes for report submission, listing the	Medium	The tracker process is defined to ensure the return	1	1.00		

Category	Parameter	Definition	Response	Rationale	Assessment Score	Standardized Score	Category Average	Category weights
	submitted on time, pending submissions)	in-charge for submission and mode of submission (ORFS, XBRL, hard copy, email)		submission timelines are known and any delay is tracked				
Data Submission	Documented process in case of error (root cause analysis, audit trail for racing back)	Is there a centralized committee or department to review/approve data submission process changes	Low	Any issues identified in submission are handled on a case to case basis	0	0.00		
Process maturity score							0.53	
Process maturity Level							High	

Thus as per the maturity scores, on the process v/s technology maturity grid, the bank is classified as follows:

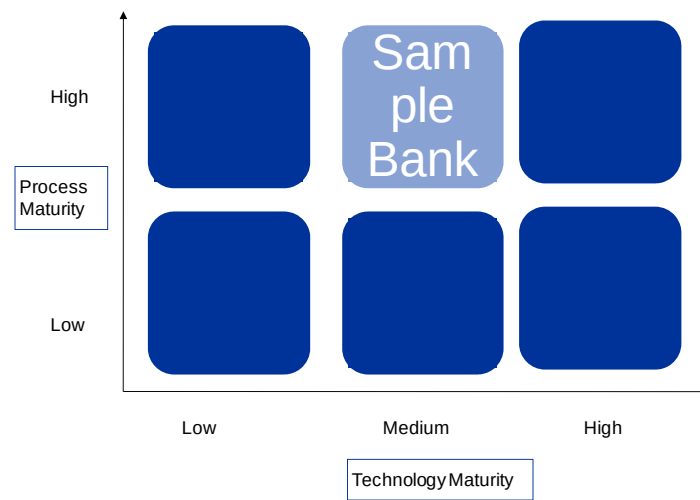


Figure 8: Current maturity based classification of the sample bank based bank

Annexure B

Task details for standard approach and illustrative approach for a bank

Task Details for the standard approach

This chapter provides the task and sub-task (if applicable) level details of the approach provided in Approach for banks. The high level schema for the proposed approach is represented below in Figure 12. The entire approach is constructed in a hierarchical manner with overall approach comprised of four steps. Each of the steps is further subdivided into activities with each of the activity defined as multiple micro level tasks and each task further broken down to sub-tasks, wherever applicable.

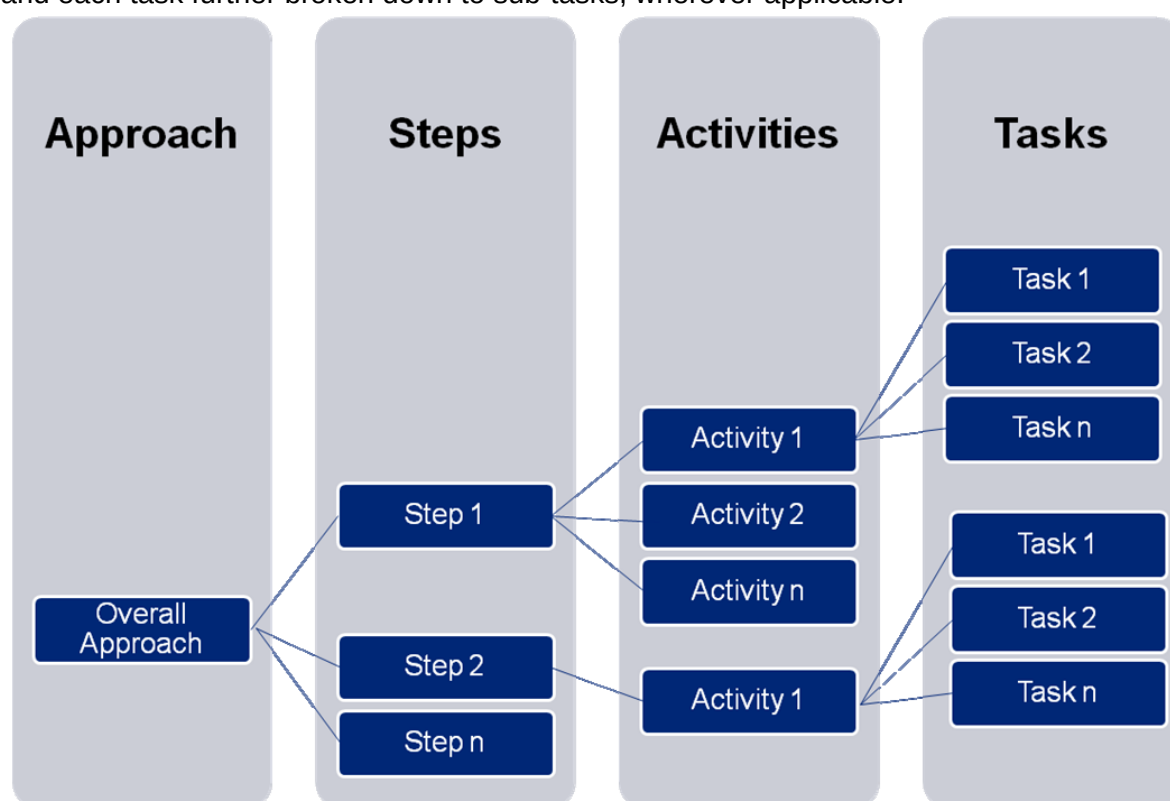


Figure 9: Approach Design

Each of these steps is discussed above in chapter on **approach for banks** along with the constituent activities. The detailed level tasks and sub-tasks are given in the four Annexures– B1, B2, B3 and B4. On completion of the tasks and sub-tasks, an illustrative approach for a bank to reach the end-state has been provided in Annexure B5.

Annexure B1

Task Details - “Prepare Data Acquisition layer”

1. Prepare Data Acquisition layer: The task level details for this step are defined under the activities listed below:

1(a) Ensure required data is captured: This activity can be further broken down into 6 major tasks:

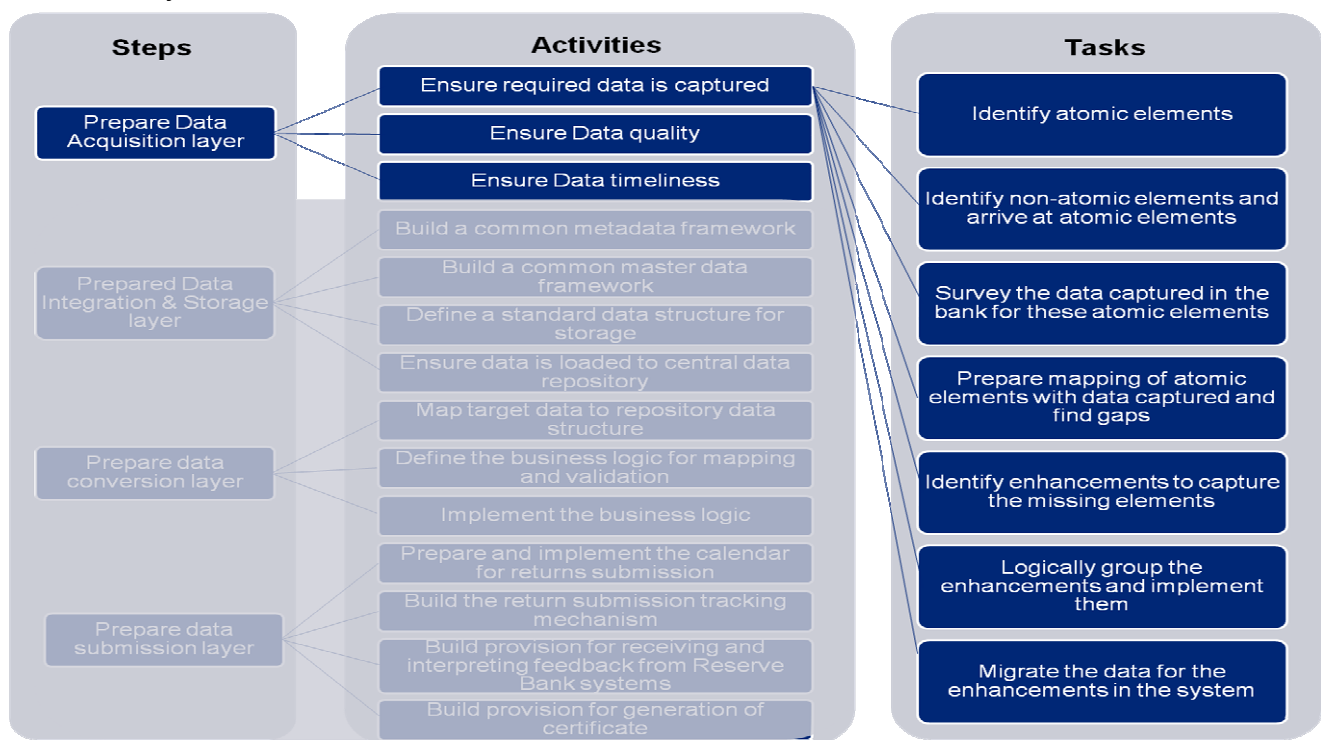


Figure 10: Prepare Data Acquisition – Ensure Required Data is captured

- Identify the atomic elements required for preparing the returns
- Identify the non-atomic elements required for preparing the returns. For these elements, de-cipher the calculation logic to arrive at the atomic elements required.
- Using the atomic data elements list prepared through task (i) and (ii) above and carry out a survey across the bank of the data being captured. This survey may cover both the data being captured through operational systems or as physical documents.
- Prepare a mapping between the data required and the data captured to identify the gaps. The gaps will be classified into one of the following types:
 - Data required but not captured by the bank
 - Data required but captured only in physical form
 - Data required but captured in electronic form that is not controlled (e.g. excel sheets or other desktop tools)
 - Data required and captured in source systems

- Identify the enhancements required to capture the missing elements against each identified gap. Filling a gap may be done by:
 - Implementing a new system(s)
 - Modifying an existing system
 - Adding new modules in current system(s)
- The identified enhancements will be logically grouped based on similar characteristics for the purpose of implementation. Depending on the system enhancement required, there might also be a need to modify and/or introduce new data capture processes.
- The historical data is migrated into the enhanced system. The quantum of data migrated must ensure that all requirements as per the Reserve Bank return formats are satisfied. For reasons of consistency, comprehensiveness or analysis, the bank may choose to update the data in full.

1(b) Ensure Data Quality: The tasks defined below need to be performed to ensure data quality of the source data:

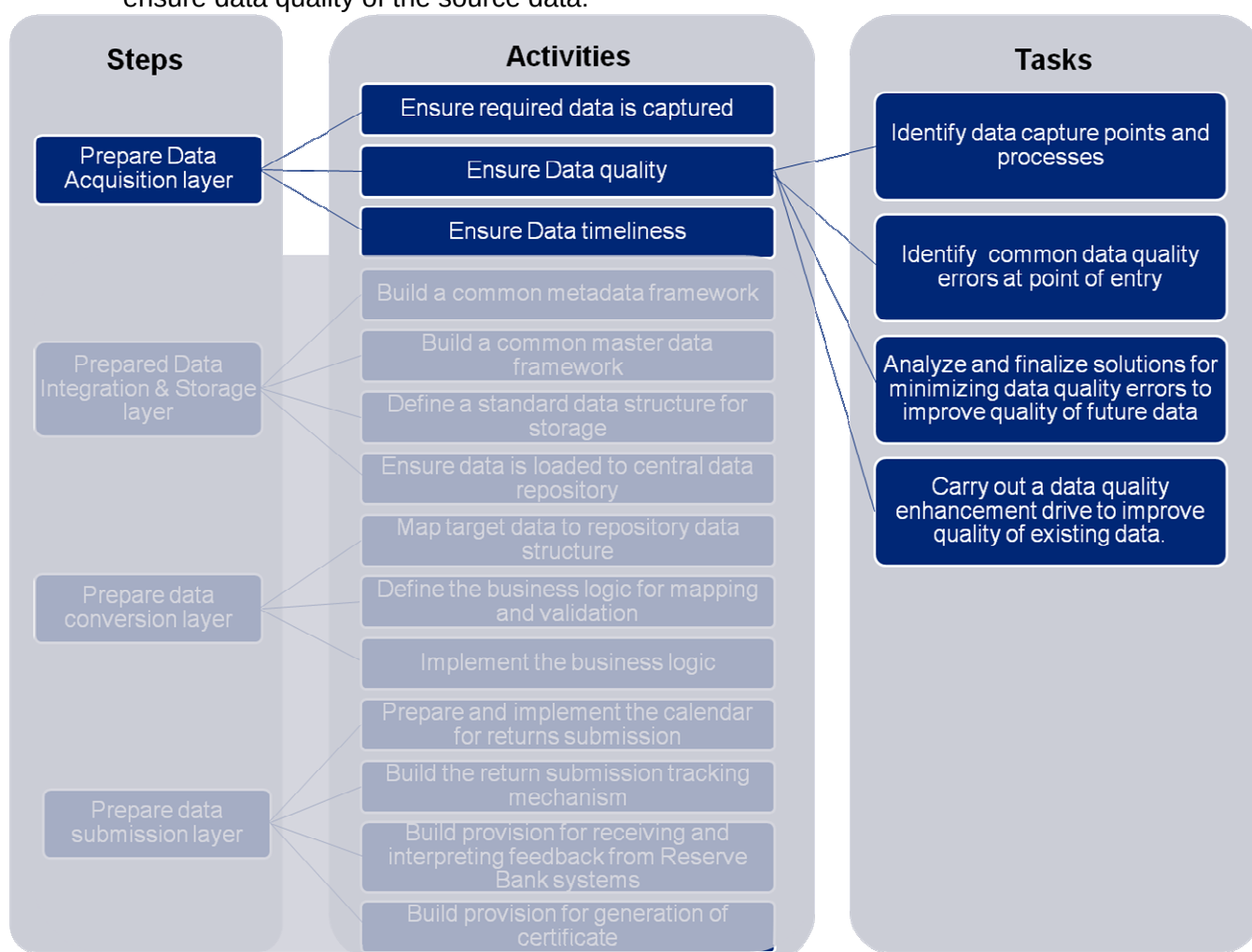


Figure 11: Prepare Data Acquisition – Ensure Data Quality

- Identify the points of capture of the data and the processes at each of these points of capture where the data quality gets impacted

- Identify the most common types of errors being committed at the data capture location e.g. always taking the first item in drop down, always providing a standard value for a mandatory item as the same is not available for entry, etc.
- Analyze and finalize solutions for addressing each of the data quality issues identified above. This may include mechanisms both in terms of systems and processes like (this is a sample list and not exhaustive)
 - Maker-Checker to be introduced
 - Enhancing systems to make required fields mandatory
 - Introducing data validation checks, e.g. range checks, no mismatch of gender and salutation, etc.
 - Introducing data type based business rules
 - Introducing standardized master data based data entry (using drop downs)
 - Introducing cross element business rules to ensure mismatch of data is minimized, e.g. no mismatch may be allowed of customer gender and customer salutation.
- A separate data quality enhancement drive may be undertaken, to augment the quality of the historical data. This may include (this list is a sample and not exhaustive)
 - Data standardization using standard databases for master data like Pin code, states etc.
 - Data cleansing using data quality tools and by defining standard business rules.
 - Data enrichment using similar data across the database
 - Re-entry of customer data based on promotional campaigns to receive updated data from customers

1 (c) Ensure Data Timeliness: The tasks defined below need to be performed to ensure data timeliness:

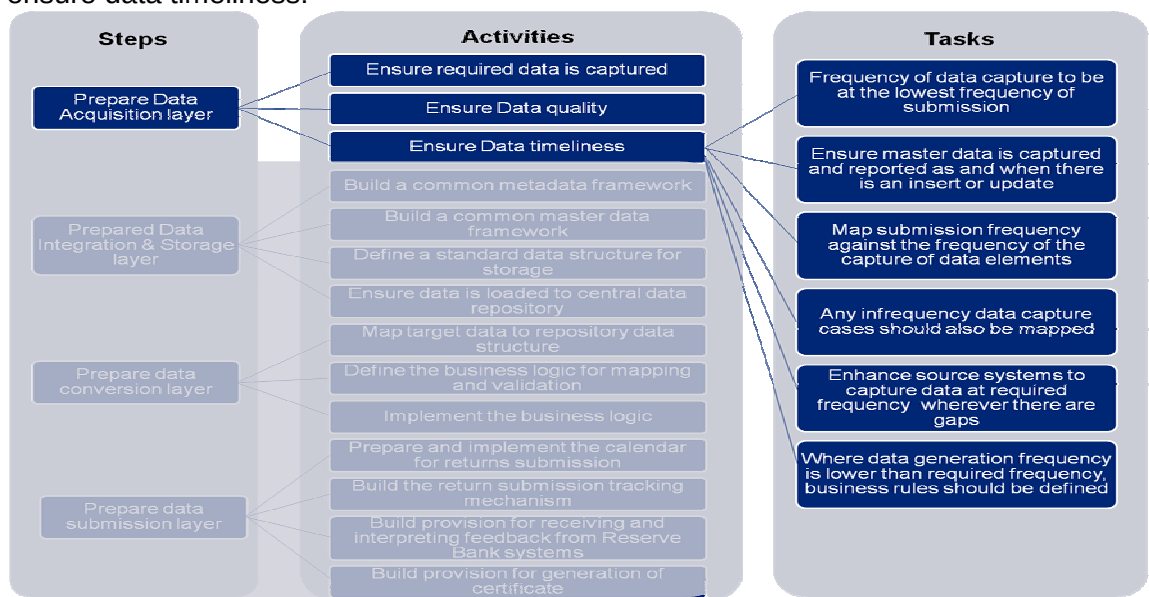


Figure 12: Prepare Data Acquisition – Ensure Data Timeliness

- The frequency of capture for transactional data must be at least at the lowest frequency of submission.
- Master data must be captured and reported as and when the record is inserted and or updated in the system.
- The submission frequency of the elements listed above is mapped to the source data elements capture frequency. This is to ensure that all data captured in the source is at least captured as often as is required by the Reserve Bank mandated submission frequency.
- Any exceptions that require infrequent data capture must also be recorded and mapped, e.g. for the month-end submission of a weekly return, a weekly capture may not be sufficient.
- In case any data is not captured at the desired frequency, the source needs to be enhanced to capture this information at the desired frequency.
- Exceptions to this might be found where the data generation frequency is lower than the frequency requested by Reserve Bank. In such case business rules must be defined to pro-rate the data. E.g. Expense data is required to be submitted fortnightly, but the employee expenses data is available only on a monthly basis.

Annexure B2 – Task Details – “Prepare Data Integration & Storage layer”

1. Prepare Data Integration & Storage layer

The task level details for this step are defined under the activities listed below:

1(a) Build a common master data framework - The tasks defined below need to be performed to build the common master data framework:

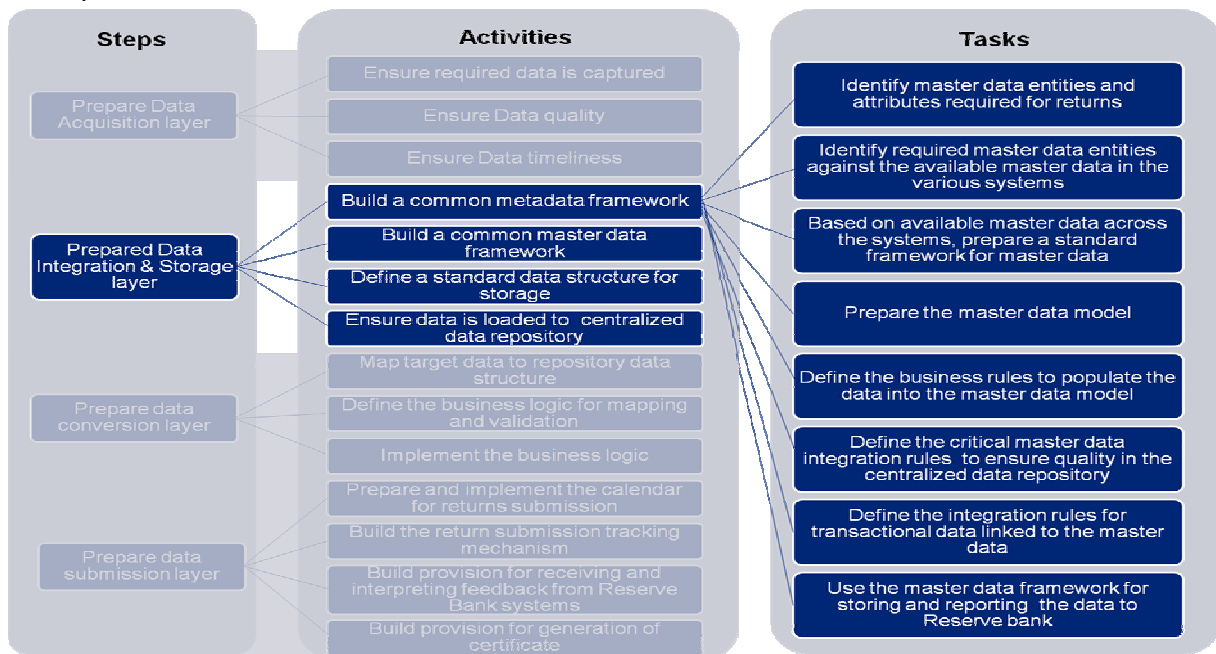


Figure 13: Prepare Data Integration& Storage Layer – Build a common Metadata Framework

- Identify the master data entities and their attributes that are required for the Reserve Bank returns. This would need a scan of the returns and listing the master data entities required for each return. These entities are then consolidated along with a listing of the distinct attributes required by each of the returns.
- Map these identified entities required for Reserve Bank returns across the master data available in the different systems of the bank. This would also identify the duplication of master data across the systems in the bank.
- Based on the identified master data entities available in the source systems, prepare a standard framework for master data storage. The structure must identify all the overlapping attributes across the systems.
- Once the entities and the attributes have been defined the master data model can be prepared along with the mapping to the source systems.
- To populate the data into this model, the rules for integrating common attributes across each entity need to be defined E.g. If the address of a customer is in 2 systems then based on which one is more recent and reliable, the primary source for the address can be defined.

- The master data integration rules will be critical in ensuring the quality of the data stored in Centralized Data Repository.
- The integration rules defined above need to be extended to define the integration rules for the transactional data linked to the master data. The transactional data will now need to link to the new integrated master data.
- The common master data framework will be used by the Centralized Data Repository as the base for storing and reporting the data for Reserve Bank returns.

1(b) Build a common metadata framework: The tasks defined below need to be performed to build the common metadata framework:

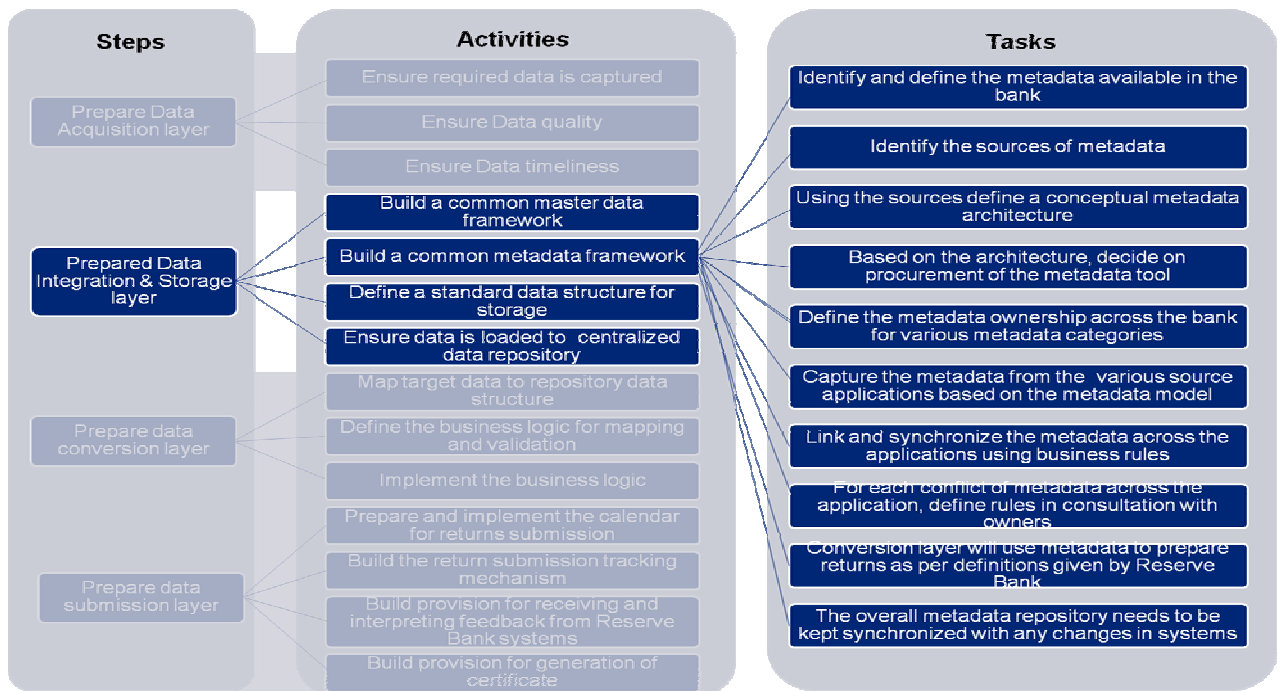


Figure 14: Prepare Data Integration& Storage Layer – Build a common Master Date Framework

- Identify and define the types of meta-data available in the bank.
- Identify the source of meta-data e.g. data dictionaries, policy manuals, tacit information, catalogues, Reserve Bank circulars and guidelines etc.
- Using the above defined sources, define a conceptual architecture for the metadata environment based on data redundancy, scalability, etc.
- Based on the conceptual architecture defined, a decision on procurement of a metadata tool will need to be taken.
- The metadata ownership may be clearly defined across the bank for various metadata categories. The owners need to ensure that the metadata is complete, current and correct.
- Capture the metadata from the individual source applications based on the metadata model for the individual source applications.
- The captured metadata is linked across the applications using pre-defined rules. The rules to be applied for synchronization of metadata also need to be defined.

- In case there is a conflict of metadata between two applications, the definitions of both the systems need to be studied. A rule will need to be defined in consultation with both system owners, as to how to resolve this conflict. The rule might require re-naming one of the source elements to ensure uniqueness of the data elements across all systems.
- The conversion layer will use the Metadata to prepare the returns for Reserve Bank. This is to ensure that the data being submitted is as per the definitions mandated by Reserve Bank.
- The overall metadata repository needs to be kept synchronized with any changes in the source or target systems.

1 (c) Define a standard data structure for storage: The tasks defined below need to be performed to define the standard data structure for storage:

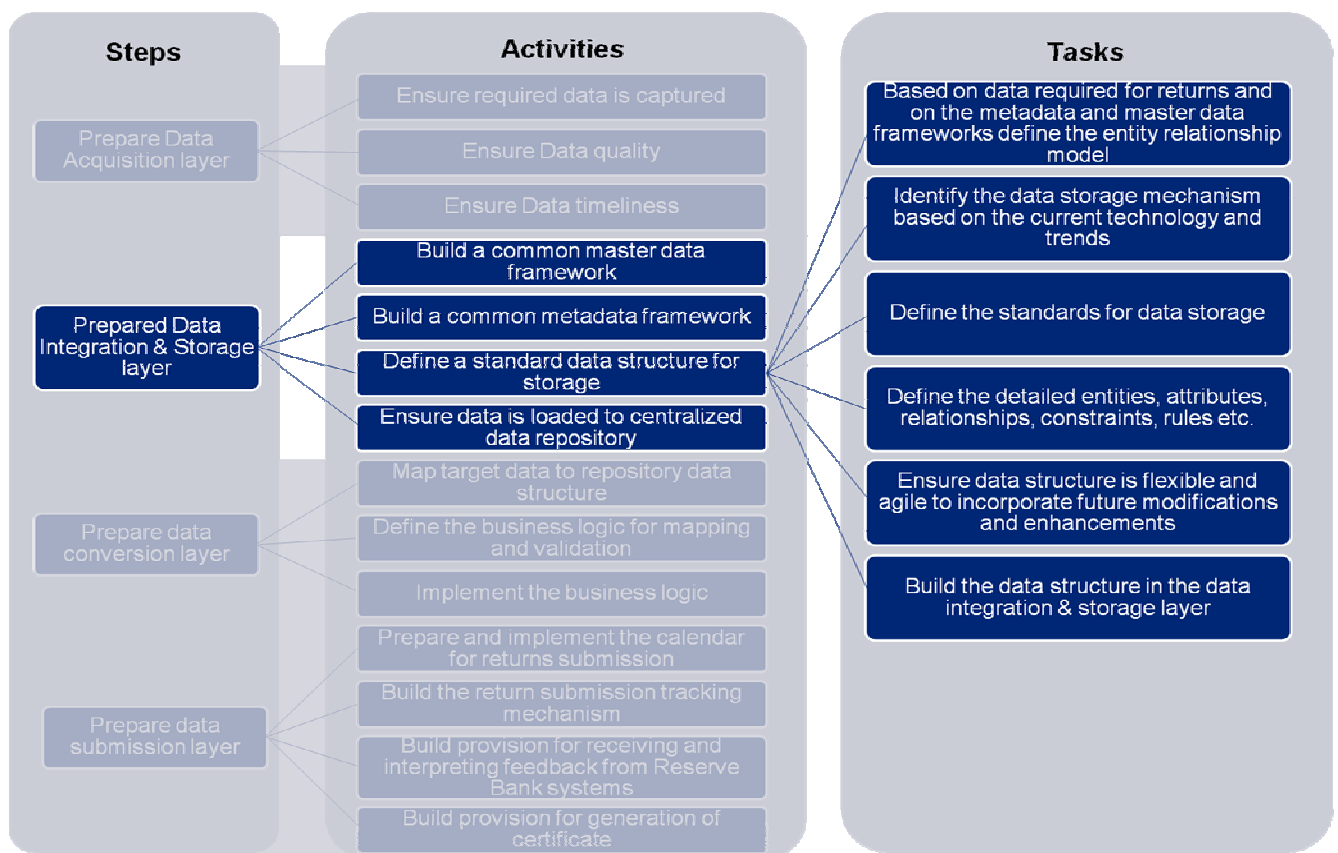


Figure 15: Prepare Data Integration& Storage Layer – Define a standard Data Structure

- Based on the data required for Reserve Bank returns and the metadata and master data frameworks defined above, define the entity relationship model for the centralized data repository.
- Identify the data storage mechanism that will be used based on the current technology and trends.
- Define the standards for the data storage including areas like naming conventions, data types (if applicable) etc.
- Define the detailed entities, attributes, relationships, constraints, rules, etc.

- The data structure may be flexible and agile to incorporate future modifications and enhancements.
- Build the data structure in the data integration & storage layer and configure all the relevant definitions and attributes. This might need procurement of the relevant hardware and software where the data will be stored.

1(d) Ensure data is loaded to centralized data repository: This has two main aspects (A) First time loading of data and (B) Incremental loading of data on a periodic basis. The tasks defined below need to be performed for the first time loading of data:

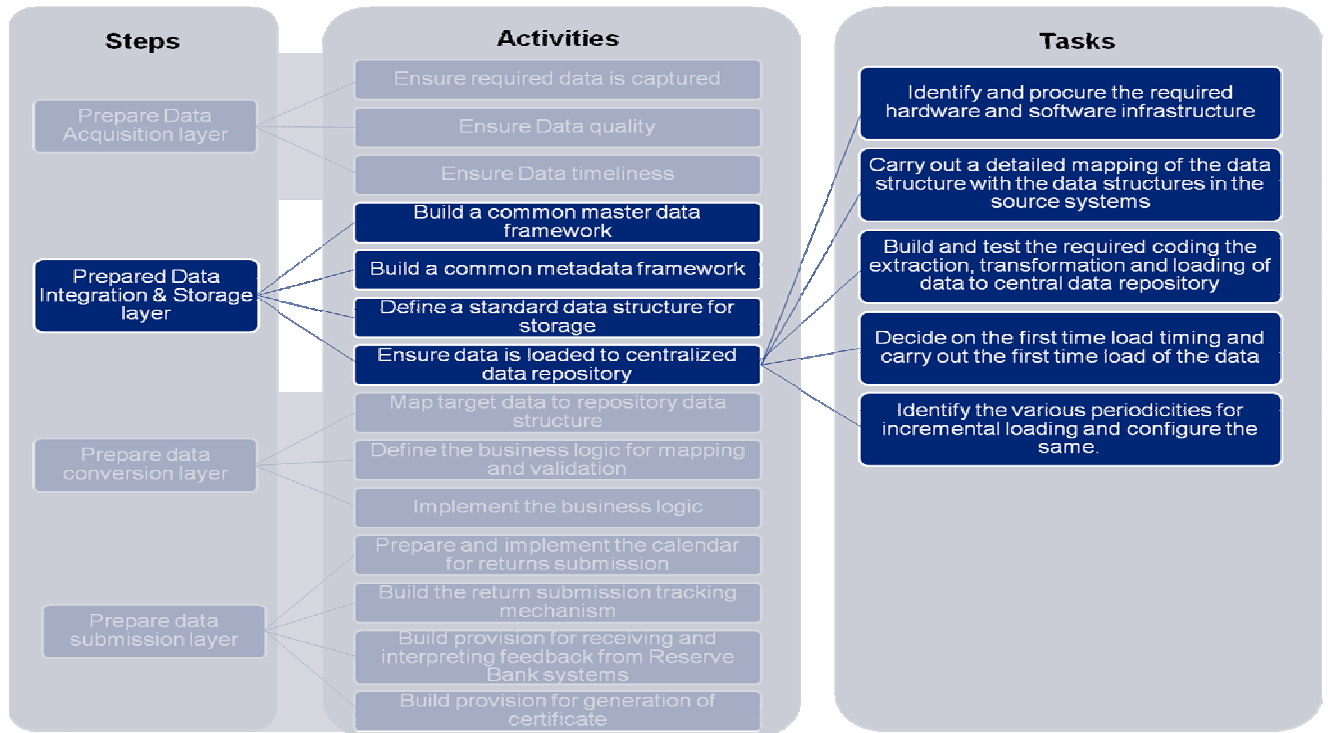


Figure 16: Prepare Data Integration& Storage Layer – Ensure data is loaded to Centralized Data Repository

- Identify and procure the required hardware and software infrastructure. For data loading this will include procurement of software tools for data extraction, transformation and loading as well as hardware for the same.
- Carry out a detailed source mapping of the data structure in the repository with the source systems data storage structures at the entity and attribute level. The mapping will include details of the business rules and logic to be applied to the source system data for arriving at the data structure in the centralized data repository. The business logic may be designed such that the data from source systems gets integrated seamlessly with any existing data in the centralized data repository (this may take place if the approach is implemented in an incremental fashion)
- Build and test the required coding for the extraction, transformation and loading of data from source systems to centralized data repository. The transformations will be based on the rules and logic defined above. Ensure that appropriate routines for error handling while loading are also built.

- Decide on the timing for doing the first time loading and accordingly carry out the first load of data into the centralized data repository.
- The tasks defined below need to be performed for the incremental loading of data on a periodic basis:
 - Identify the periodicity of loading required for the incremental loading. The periodicity may be different for different sets of data.
 - Design the business logic and rules that need to be built for the various sets of data that need to be loaded incrementally at different periodicity.
 - Build and test the required coding based on the design above. Ensure that appropriate routines for error handling during loading are built.
 - Prepare the operational plan for running the incremental loading code at the right timing for ensure that there is no data that is either missed or loaded twice.
 - Implement the operational plan on the centralized data repository.
 - Prepare and implement the data retention and archival plan for long term operations of the centralized data repository.

Annexure B3 – Task Details -“Prepare Data Conversion Layer”

1. Prepare Data Conversion Layer

The task level details for this step are defined under the activities listed below

1(a) Map target data to repository data structure- The tasks defined below need to be performed to build the mapping of target data to source data:

- Map each atomic data element in each return required to the corresponding data element within the repository.
- The non-atomic data elements need to be further broken down into their distinct atomic elements and the conversion logic applied to these elements. Map each of these atomic elements to the corresponding data elements in the repository.
- Prepare a detailed mapping document that maps the returns to the data elements in the repository

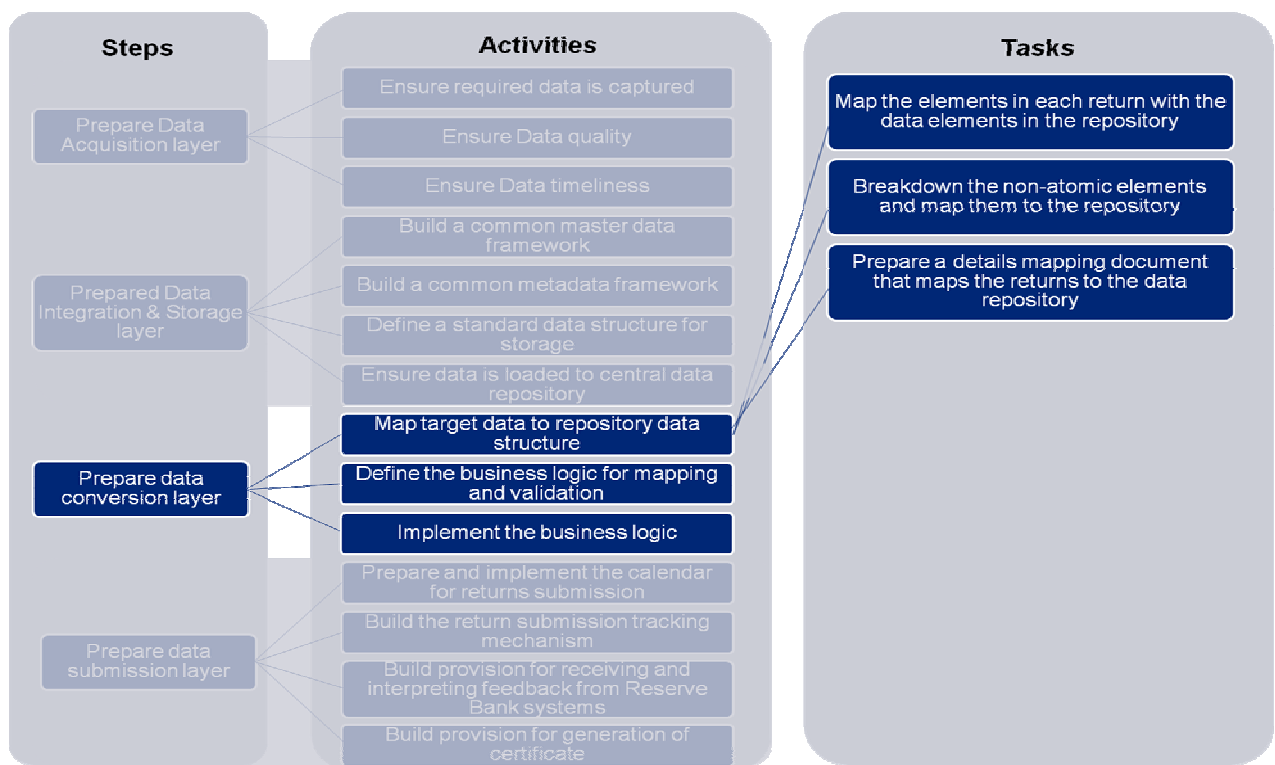


Figure 17: Prepare Data Conversion Layer–Map target data to repository data structure

1(b) Design the business logic for mapping and validation: The tasks defined below need to be performed to design the business logic for conversion:

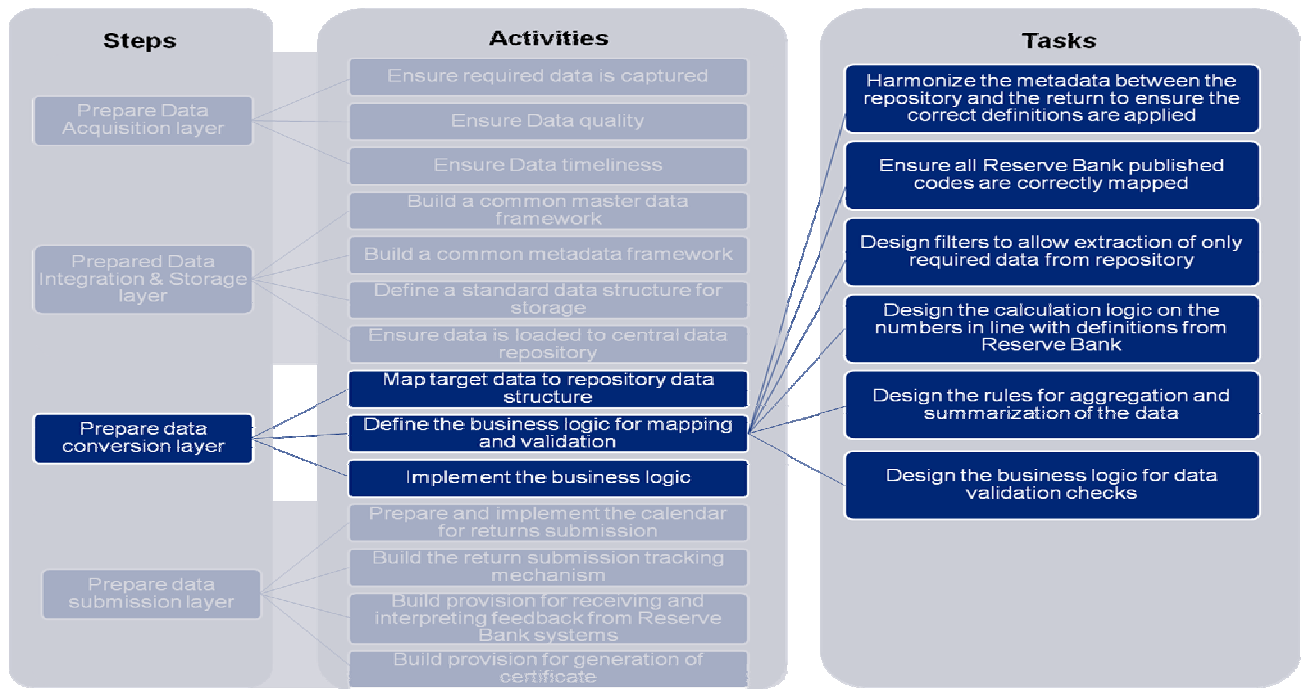


Figure 18: Prepare Data Conversion Layer –Define the Business logic for mapping and Validations

- Harmonize the Metadata between the repository and the return to ensure the correct definitions are applied whilst applying the rules.
- Ensure that all Reserve Bank published codes are correctly mapped and define appropriate filters as required by the return, e.g. filter by gender to prepare gender specific returns.
- Design Filters to allow extraction of only the required data from the repository.
- Design the calculation logic on the numbers in line with the definitions provided by Reserve Bank.
- Design rules for aggregation and summarization based on the return requirements.
- The business logic for data validation checks including the basic data checks, format validations, consistency validations, abnormal data variations and reconciliation checks.

1(c) Implement the business logic: The tasks defined below need to be performed to implement the business logic:

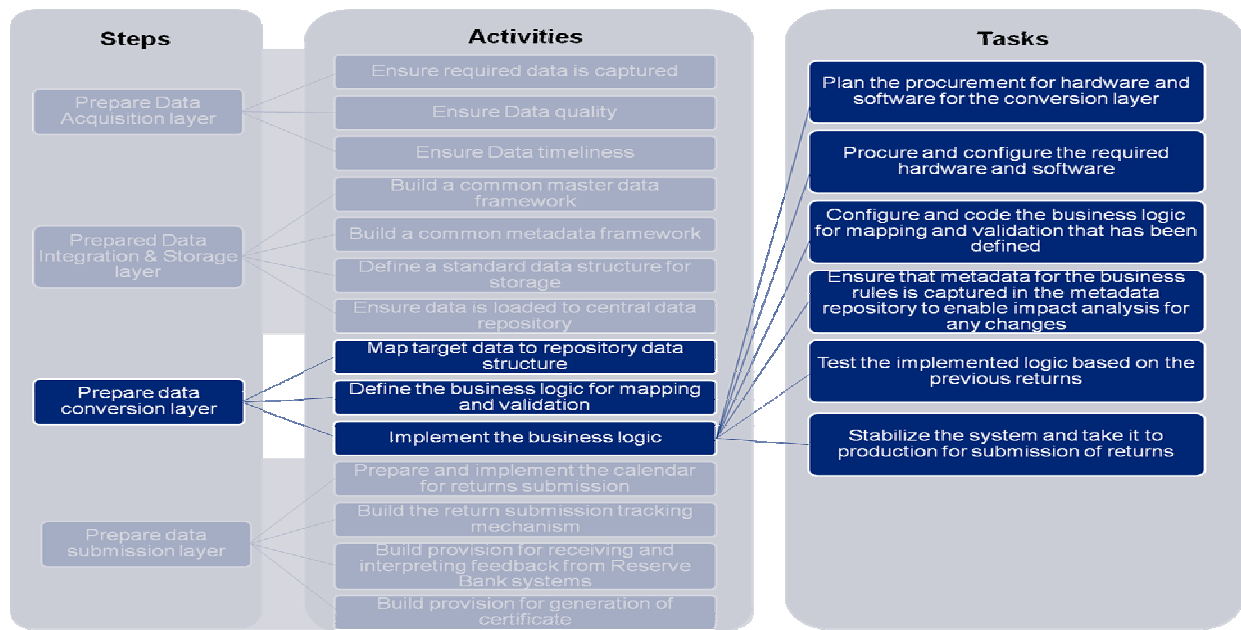


Figure 19: Prepare Data Conversion Layer – Implement the Business Logic

- Plan the procurement (if any) for the hardware and software required for implementing the data conversion layer.
- Based on the plan the identified hardware and software may be procured and configured for use.
- The defined business rules and logic may be configured and coded into the tools.
- The metadata for these rules may be captured in the centralized metadata repository. This will enable impact analysis for any subsequent changes in the logic.
- The logic may be tested using previous data and returns to ensure it is as per Reserve Bank's requirements.
- Once stabilized, the data conversion layer is ready.

Annexure B4- Task Details - “Prepare Data Submission Layer”

1. Prepare Data Submission layer

The task level details for this step are defined under the activities below.

1 (a) Prepare and Implement the calendar of returns submission - The tasks defined to prepare and implement the calendar of returns are:

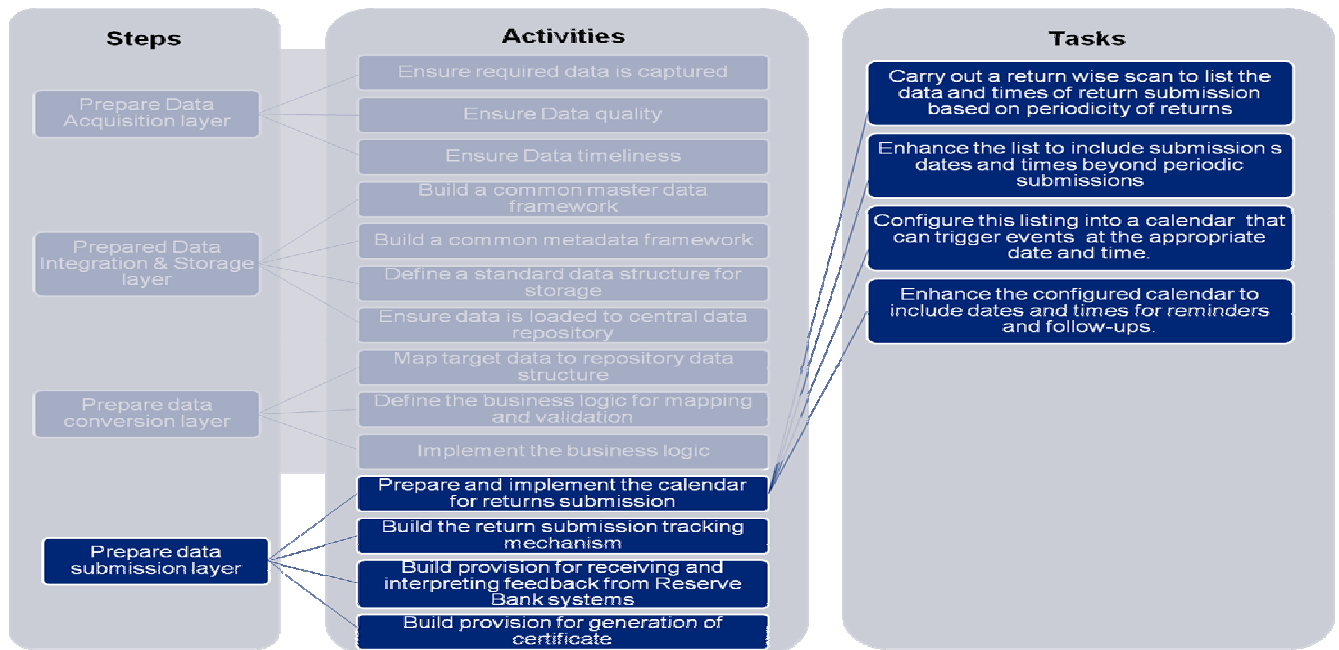


Figure 20: Prepare Data Submission Layer – Prepare and Implement the Calendar for Returns Submission

- Carry out a return wise scan and list out the dates with times of submission for each return based on the frequency of submission of the return
- Enhance the list of submission dates by adding the extra submission dates beyond the periodic submission e.g. month-end, year-end etc.
- Configure the listing of dates with times into the data submission layer. The calendar may trigger events for preparation of each return followed by submission.
- Enhance the returns submission calendar to include dates and timelines for sending reminders and follow-ups to all stakeholders.

1(b) Build the return submission tracking mechanism. The tasks defined for building the tracking mechanism are:

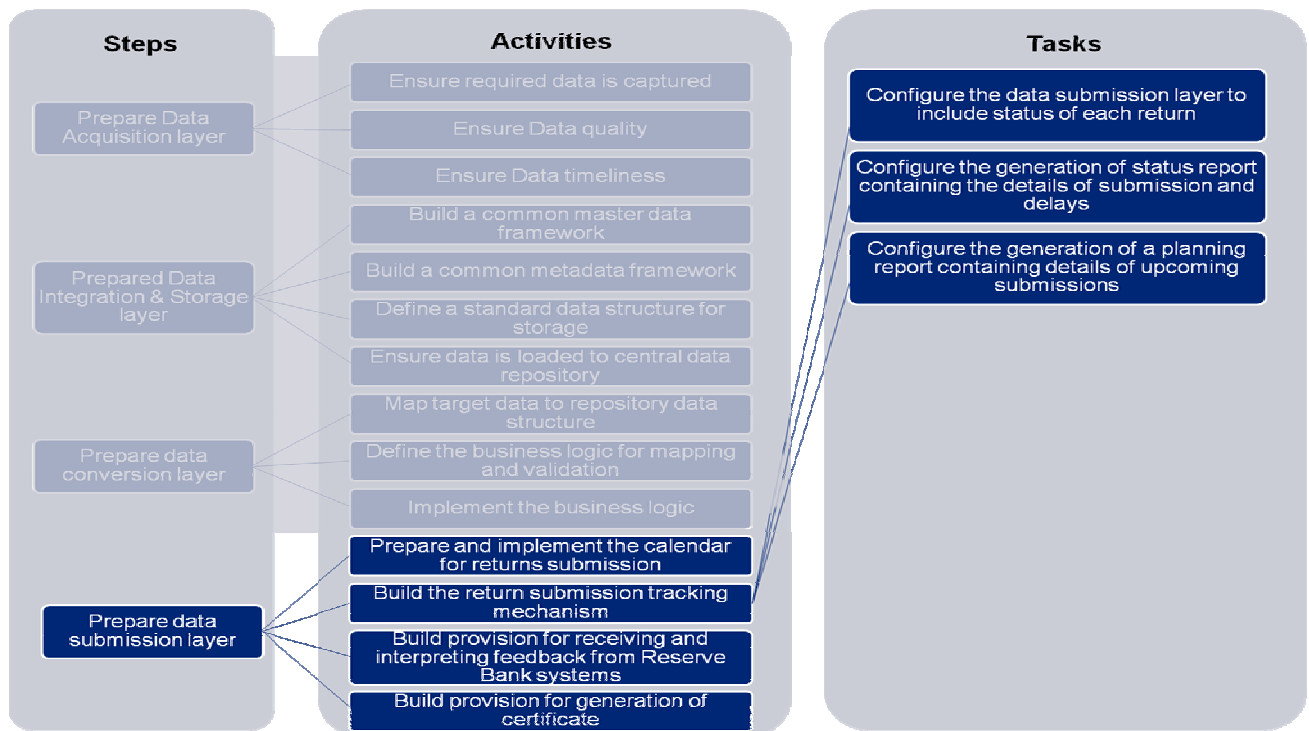


Figure 21 : Prepare Data Submission Layer – Build the return submission tracking mechanism

- Configure the data submission layer to include the status of each return submission. In case of multiple attempts, appropriate details may be updated in the status to understand the history of submission.
- Configure generation of status report from data submission layer that contains the details of submissions made and delays, if any.
- Configure generation of planning report from data submission layer that contains details of the submissions that are expected over the next week or month along with past history of the same submissions.

1(c) Build provision for receiving and interpreting feedback from Reserve Bank systems. The tasks defined to build provision for receiving and interpreting feedback from Reserve Bank are:

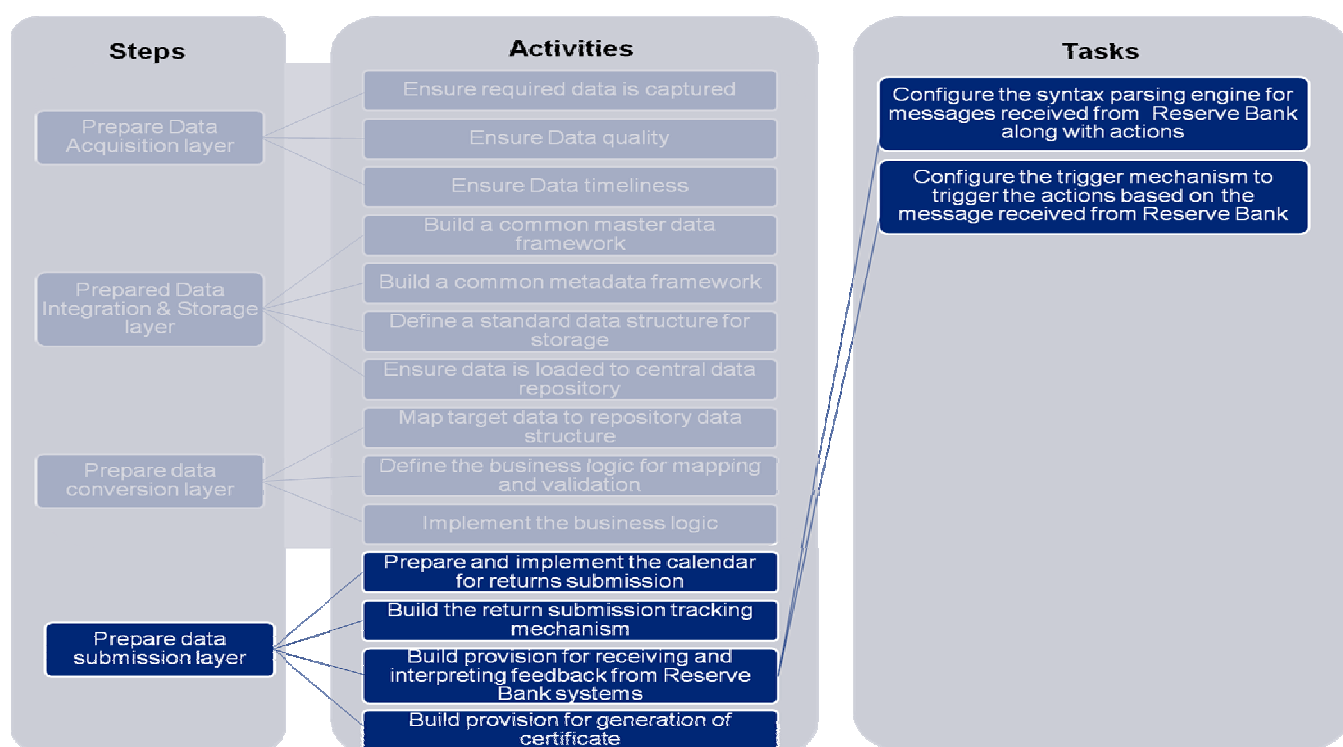


Figure 22: Prepare Data Submission Layer – Build provision to receive feedback from Reserve Bank

- Configure the syntax parsing engine for any messages received from Reserve Bank. Reserve bank messages can be acknowledgements, validation error etc. The engine may also generate the actions to be taken based on the message.
- Configure the trigger mechanism to trigger the actions generated from the parsing engine. The triggers may either be generation of communication like emails, SMS or preparation of returns. Once the relevant action is taken, the re-submission of the return, as applicable, may also be tracked by the data submission layer.

1(d) Build provision for generation of certificate: The tasks defined to build provision for generation of certificate are:

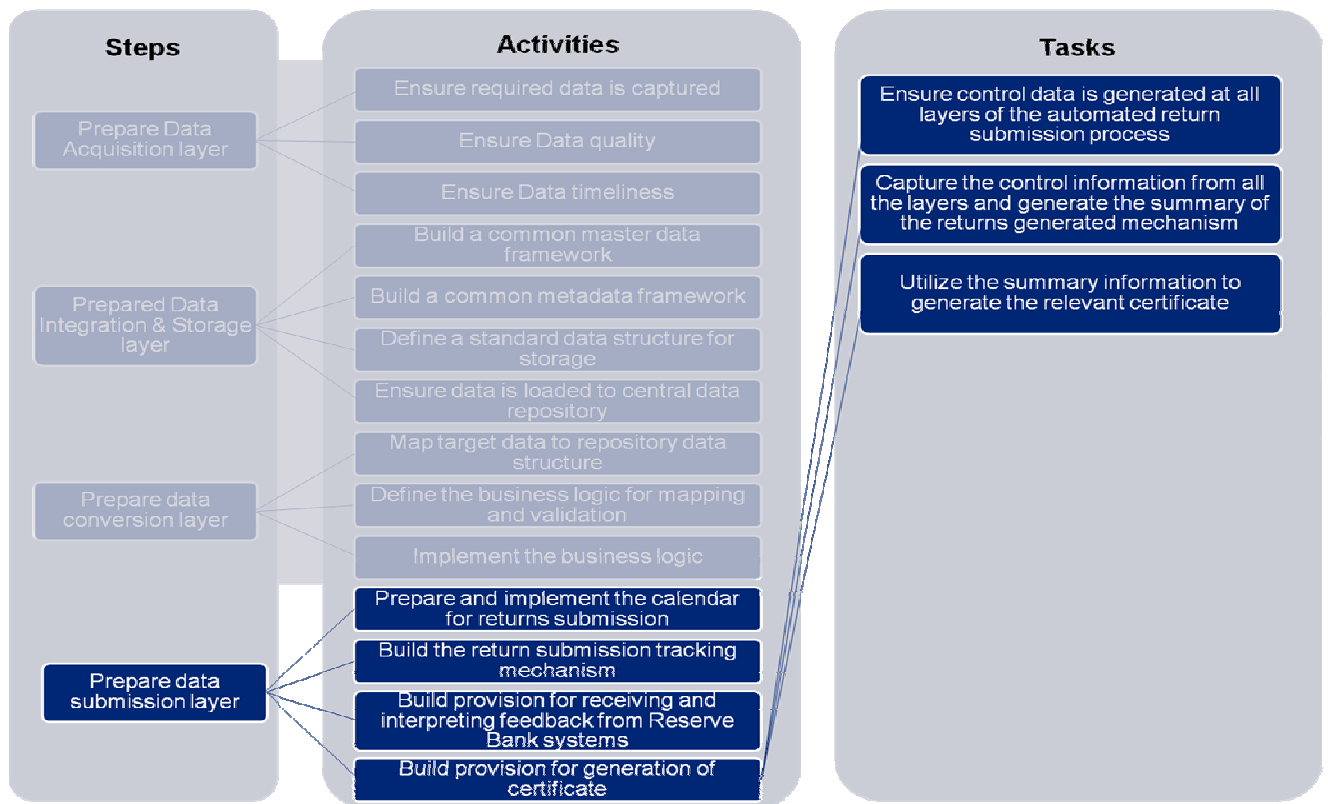


Figure 83: Prepare Data Submission Layer – Build provision for generation of certificate

- The step by step process followed for preparation of the return at all the layers may generate control output that gets can be accesses by the data submission layer. This output will have the relevant information of the success of the generation of return in automated fashion.
- The data submission layer will access this control information at each layer for all the returns and generate the summary information of the returns generated in automated fashion.
- This summary information can in turn be utilized to generate a certificate of the returns that are automated and the overall percentage of the returns that are automated.

Annexure B5 – Illustrative Approach for a bank

1. Introduction to the current state of the bank

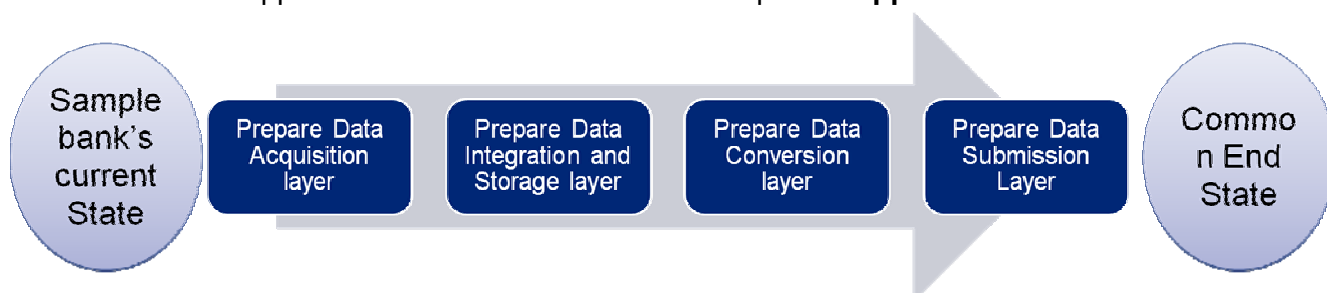
The bank is a scheduled commercial bank with pan India presence in terms of branches, ATMs, regional and zonal offices. All the Bank's branches are live on the Core Banking Solution. The bank has foreign operations and foreign branches in addition to their domestic operations. Apart from core banking solution, the bank also has other technology based applications for treasury (front and back office). The bank is currently in the process of implementing an Enterprise Data Warehouse (EDW). CBS data is also loaded into a legacy credit application.

Though the data is available centrally in the core banking solution, the data extract from the CBS is sent to the Branches for auditing and preparation in the desired format, prior to submission. Most of the departments within the Bank manually key in the data into the Reserve Bank applications for submitting the returns.

2. Assumptions

- The bank has performed their Self Assessment and based on the derived score have found themselves to fall under cluster 5, i.e. Medium on Technology and High on Process. Based on the defined score they have defined the timelines required for the implementation. The same have been shared with Reserve Bank.
- The bank is starting with the first group of returns, i.e. returns generated using the source applications and it has decided to implement all the returns within the group together.
- The bank has undertaken a data cleansing drive on their historical transactional data
- Some data that is required for Reserve Bank reporting exists in manual registers.
- In case there are duplicate elements being stored in the source systems, the bank has made a policy decision to consider the most current data as their system of record

3. **Approach:** The approach for the sample bank to traverse from its current state to the common end state of complete automation of returns is defined as per the standard approach and variations covered in chapter on **approach for banks**.



The key steps and the activities to be undertaken by the bank to become automated are as defined below:

Step 1 : Prepare Data Acquisition Layer: The data elements that are not captured or stored in any source system will be brought under a controllable defined IT system which then would be added into the data acquisition layer. The key activities that the banks need to do are as follows:

- (a) The bank needs to conduct a mapping exercise between Reserve Bank reporting requirements and the data being captured across the different source systems available in the bank.
- (b) In case the data is available in a physical format the same needs to be converted into a controlled system format.
- (c) In case the data is not available in any source system, the source system will need to be enhanced to incorporate this new element which will then be uploaded into the Centralized Data Repository which is a part of the Data Integration & storage layer. A separate business process re-engineering exercise might also be required to capture this additional data.
- (d) The source systems under the data acquisition layer will be reviewed to ensure that enough checks are in place at the time of data capture. This is to ensure that all the data required for reporting is captured with adequate quality.

Step 2 : Prepare Data Integration & Storage layer: This step will ensure that the Centralized Data Repository serves as the single source for preparing all Reserve Bank returns. The key activities for performing this step are as follows:

- (a) Conduct a data profiling exercise on the transactional data and identify priority cleansing areas. The data will be cleansed prior to loading into the repository. Multiple iterations might be required to completely cleanse the historical data.
- (b) A one-time exercise must be conducted to synchronize the data between the credit application and the CBS to ensure consistency of legacy data. All changes applied to the credit application must be replicated in the CBS using a bulk upload facility. Henceforth the credit application will be synchronized with the core banking system on an on-going basis at a specified frequency. The bank must make a policy decision to not allow any changes to the data being uploaded into the credit application by the branch personnel. Data for the Centralized Data Repository will be sourced from the core banking solution.
- (c) Harmonize the metadata definitions to aid the common understanding of definitions and data across the bank. The definitions need to be collected from various sources like internal system documents, Reserve Bank circulars, tacit knowledge, policy documents, bank's manuals etc. A central metadata repository will be used to store this data. All applications would have access to this repository.
- (d) Define the business rules to integrate the data across the different systems. Wherever, data resides in more than one system, the most recently updated record will be considered for loading into the repository, subject to the metadata being same.
- (e) All MoCs will be identified and an application created to store these MoCs. This application will also be considered as a source system under the data acquisition layer.
- (f) Identify the master data entities and their attributes to define a master data model. Integrate the attributes of master data entities across different source systems and link the transactional data to the new integrated master data.

- (g) Define a flexible and agile data model for EDW by considering the Reserve Bank's regulatory requirement and metadata and master data frameworks.
- (h) Define the storage mechanism of centralized data repository in accordance with the data model.
- (i) The timestamps for worldwide operations may be marked as per IST while loading the data so as to ensure that common cut-off is taken for all data irrespective of the time zone.
- (j) Define appropriate operational parameters to make the repository functional by identifying initial data loading schedule, incremental data loading schedule, data retention, archival and data security guidelines.

Step 3: Prepare data conversion layer

- (a) Conduct mapping of data required by the Reserve Bank to the Centralized Data Repository and define the logic and formats required to prepare the Reserve Bank returns.
- (b) Implement the data conversion layer using a tool or customized scripts.
- (c) Generate the previous period returns submitted by the bank using the new system. This is to test whether the logic has been correctly defined.

Step 4: Prepare data submission layer

- (a) Do a test submission of the prepared returns to Reserve Bank's returns receiving system.
- (b) Implement a tracking system that allows defining the submission calendar and can track the return submission status.
- (c) Define event triggers based on possible response codes to be received from Reserve Bank for tracking purposes.
- (d) Conduct a process and system audit at the end of the implementation to ensure that the data being reported to Reserve Bank is as per the Reserve Bank guidelines for automation.
- (e) Cut-over to the new automated submission process for submitting the identified set of returns to Reserve Bank.
- (f) The bank will also need to provide a certificate to Reserve Bank declaring that the return group was generated using automation with no manual intervention.
- (g) For the remaining returns, the bank continues with the earlier process of submitting the data to Reserve Bank.

After completion of this project the bank undertakes the second group of returns. The activities required for this group will be the same as defined above for the first group.

Annexure C – Representative Returns Classification

[Click here to view the Returns Classification Excel](#)

Appendix I – Other End State components, Technology Considerations and Definitions

1. Data Quality Management

Data Quality Management may involve following activities:

- (a) **Data Profiling:** is a systematic exercise to gather actionable and measurable information about the quality of data. Typical Data profiling statistics include specific statistical parameters such as - Number of nulls, Outliers, Number of data items violating the data-type, Number of distinct values stored, Distribution patterns for the data, etc. Information gathered from data profiling determines the overall health of the data and indicates the data elements which require immediate attention.
- (b) **Data Cleansing:** process may be enabled for detecting and correcting erroneous data and data anomalies prior to loading data in the repository. Data cleansing may take place in real-time using automated tools or in batch as part of a periodic data cleansing initiative. Data Cleansing is done by applying pre-defined business rules and patterns to the data. Standard dictionaries such as Name matching, Address Matching, Area – Pin code mapping, etc. can also be used.
- (c) **Data Monitoring:** is the automated and/or manual processes used to continuously evaluate the condition of the bank's data. A rigorous data monitoring procedure may be enabled at banks to handle the monitoring of data quality. Based on the data monitoring reports, corrective actions will be taken to cleanse the data.

2. Data Privacy and Security

Data Privacy and security is an essential part of the bank's end state vision so as to prevent any unauthorized access, modification or destruction of sensitive data. The security needs to be ensured across the solution at various levels:

- (a) **Network Security:** This is designed to act as a checkpoint between the internal network of the bank and the Reserve Bank's network(s). The data must be transmitted using a secure, encrypted format to ensure there is no interception of data.
- (b) **Application security:** This provides the ability to the administrator to access the Centralized Data Repository on the banks' side and limit the access to the authorized users only. There may be appropriate log of activity to ensure that details like user-id, time of log in, data elements viewed etc. are captured.
- (c) **Database security:** This forms a security layer around the Centralized Data Repository so as to control and monitor access, update rights and for deletion of data held within the database.
- (d) **Workstation Security:** Workstation security is done through the usage of solutions like anti-virus protection, screen saver passwords and policies around installation of personal software and internet access.

3. Data Governance

The data organizational structure and setup of governance processes that manage and control the return submission process may be clearly defined and documented as part of the overall data governance policy within the bank. This helps in communicating the standard procedures to all stakeholders effectively. The key constituents of a data governance model that the bank needs to put in place are:

- (a) Clearly defined roles and responsibilities across all the departments involved in the returns submission process.
- (b) Clear accountability may be put in place by defining data ownerships and process ownerships across the departments and at each layer in the process.
- (c) Key metrics e.g. - number of instances of data validation failures at Data Validation Layer may be defined, recorded and tracked.
- (d) Clear decision-making guidelines and protocols, escalation process for decision resolution procedures may be put in place.
- (e) Open sharing of governance information i.e. governance process models and definitions may be available to anyone in the overall governance organization who needs them.

The data governance group for the Returns submission process at the bank's end must consist of members from the IT function and the compliance department and each of the main business groups owning the data within the bank.

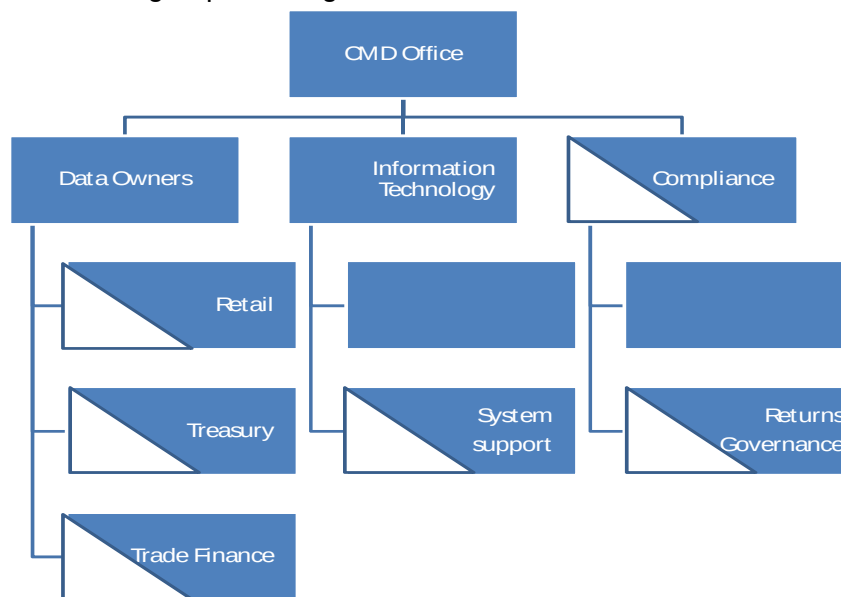


Figure 24: Model data governance structure for banks

From the Reserve Bank perspective the governance model may define the following:

- (a) As far as possible, there may be a single point of contact, clearly identified by the bank for communication with Reserve Bank for returns related issues. In case of multiple points of contact, the segregation of responsibilities for each point of contact may be clearly defined and may be in-line with the principles defined in this section. Banks also need to provide justification for having multiple points of contact and why single contact will not suffice.
- (b) There may be a well defined escalation matrix for each point of contact defined by the bank. The escalation matrix may lead up to the ultimate authority at the banks' end with no exceptions.
- (c) The data governance organization defines the basis on which the ownerships of data and information will be segregated across the bank.

While there can be numerous models for the same, the three typical models are – Process Based, Subject Areas Based and Region Based.

(a) Process based model defines data ownership for each layer in the process separately – whereby the ownership for same data at data acquisition layer and data submission layer will reside with different owners.

(b) Business Function based model defines data ownership based on the data subject area e.g. foreign exchange data, Priority sector data etc.

(c) Region Based model defines data ownership based on the regions e.g. North Zone data, South Zone data etc.

Given the nature of regulatory information and the process it is recommended that the data ownership may be primarily based on the business function. Based on the size, while some banks may need to also have a layer of regional ownership, the eventual model may be subject area based.

4. Audit Trail

Detailed audit trails must be captured to cover the entire submission process. The audit trail must provide the ability to re-create the return if required in a controlled manner and also offer itself to detailed audit functions. The key element to be captured by the audit trail must include:

- (a) Source of the Data
- (b) Transformations including and format changes applied
- (c) User / System doing the transformations
- (d) Date for the transformations

The audit trail can either be maintained using a monitoring tool or a standard operating process documentation which defines all the details required for audit purposes.

5. Change Management

There is a need to define a change management process to manage and maintain the automated data flow architecture. While change management is a continuous process, some possible change management scenarios that could occur for the Automated Data Flow process are as mentioned below:

5.1 Modification in existing return owing to:

- Reserve Bank changes the calculation/ format/ codes for the return, without requiring the addition of any new data element: This change in the return generation can be achieved by the bank by modifying the business rules in the data conversion layer.
- Introduction of a new data element in a return which was previously submitted as part of some other return: The data conversion layer needs to be modified to process the data element in the new return and perform any transformation and validation associated with the data element.
- Introduction of a new data element in a return which previously was not required to be submitted to Reserve Bank. On mapping the element though it is found to exist in the source systems: The data model of the Centralized Data Repository will have to be modified to accommodate the new data element(s). Henceforth this data will have to be sourced from the appropriate source systems and processed using data integration, conversion and submission layers.

- Introduction of a new data element, which was previously not required to be submitted to Reserve Bank. On mapping the element is found to be not available within any source system: The system change will include enhancement to capture the data in appropriate source system(s). Thereafter it is required to define the data flow from source system to centralized data repository with any transformational and validation rules as applicable.

5.2 Modification in the source system:

- If any new source system is added then the data from this source system may also be loaded in the Centralized Data Repository. In case some new functionality has been introduced, the impact on the data repository must be assessed.
- Request for Ad-hoc information by Reserve Bank: In case of ad-hoc requirements, if the data is readily available in the repository, the data conversion layer can be used to create the ad-hoc data directly. If the data is readily available in the repository, the data conversion layer can be used to create the ad-hoc data directly. If the data is not found in the repository, the additional data element must be sourced using the steps defined earlier. The new return format will be created using the data conversion layer.
- Introduction of new return by Reserve Bank: The data elements of the return must be mapped to the data already available within the Centralized Data Repository. If the data is readily available in the repository, the data conversion layer can be used to create the new return directly. In case the data is not found in the repository, the additional component must be sourced using the steps defined earlier. The new return format will be created using the data conversion layer.

6. Exception Handling

An acknowledgement will be generated by Reserve Bank once the data is successfully received and validated by Reserve Bank. The acknowledgement can be in the form of an e-mail sent to the respective banks or as per a mutually agreed information exchange format.

If there is an error in either the validation or receipt of data by Reserve Bank, an error message will be sent to the concerned bank. The message shall indicate the type of error and the error details. The corrective action will be driven through systems and processes within the bank, to find out the cause of the exception and take appropriate remedial action.

7. Data Retention and Archival

The data archival policy may be defined in line with regulatory guidelines as well as the internal policy of the bank, e.g. if the bank has its internal policy to retain the treasury data for five years while Reserve Bank requires a seven year data then the retention period may be at least seven years.

8. Metadata Management

(a) Metadata is the means for inventorying and managing the business and technical data assets. Metadata plays a critical role in understanding the data architecture for the overall system and the data structures for individual data elements. It also enables

effective administration, change control, and distribution of information about the information technology environment.

(b) The metadata across all the individual applications must be collated using a centralized metadata repository. The metadata to be captured across the different applications must be standardized. This is to ensure the same unique characteristics are captured and checked for consistency across the applications.

(c) The metadata (data definitions) may be synchronized across various source systems and also with the Reserve Bank definitions. This is important to ensure that all transformations have been correctly applied to the data as per the Reserve Bank definitions.

9. Master Data Management

Master data is the consistent and uniform set of identifiers and extended attributes that describe the core entities of the bank. Some examples of core entities are parties (customers, employees, associate banks, etc), places (branch locations, zonal offices, regions or geographies of operations, etc.) and objects (accounts, assets, policies, products or services). Master data management essentially reduces data redundancy and inconsistencies while enhancing business efficiencies. The essential components of master data management are as follows:

(a) Defining Master Data: Using the data available in the different applications a unique master record is generated. This record can consist of elements from different systems, without any duplication of information.

(b) Creating Master Data: Once the elements forming the unique record for the particular data entity are identified the source mapping document is prepared. Wherever the master data is available in more than one source application, the element which is most accurate and recent must be considered for the master record formation.

(c) The storage, archival and removal of Master data may be in compliance with the guidelines provided in the Data Retention and Archival Policy.

10. Technology Considerations for the End State

10.1 Data Acquisition Layer – Comprises of all the source systems which capture the data within the bank such as the Core Banking Solution, Treasury Application, Trade Finance Application and any other systems used by the bank to capture transactional, master, and /or reference data.

10.2 Data Integration & Storage Layer-

10.2.1 Data Integration - The data integration can be done either by using an ETL (Extract-Transact-Load) tool or custom developed scripts or a middleware application.

- **ETL Tool** - An ETL tool is used to describe processes and tools for extracting data from multiple sources in different formats, applying business rules (transforming the data), and loading the data into a target system, in this case the Centralized Data Repository. This activity is usually done in batch mode with large quantities of data or using trickle feed mechanism for near real-time data.
- **Custom Developed Scripts** - Custom developed scripts can be built using technologies like PL/SQL, Perl etc. wherein the scripts are developed and

optimized as per the specific needs of the bank. These scripts perform the task of extraction, transformation and loading of data from source system(s) to target databases as per the defined rules. However, the usage of custom developed scripts requires maintenance of the scripts supported by a well defined change management process. This is to ensure that the scripts are as per the defined business requirements.

- **Middleware Application-** Middleware tool is a software component that connects different application software that may be working on different platforms. Examples of Middleware include Enterprise Application Integration software and messaging and queuing software. These tools can be used to integrate different applications and technologies to allow seamless data interchange. These tools are also used for transforming the data formats as per pre-defined industry formats such as SWIFT, RTGS, NEFT, etc.

10.2.2 Data Storage - Data storage is a defined repository for the data. The data can be stored using an Operational Data Store, Data Mart or a Data Warehouse.

- **Operational Data Store** - Data stored in the operational data store (ODS) is a current, volatile and updatable repository of transactional data. The data stored in the ODS is refreshed as frequently as the transactional system and typically stores current data. The ODS is generally used for tactical decision making.
- **Data Warehouse-** the Data Warehouse is defined as a subject-oriented, integrated, time variant and non volatile collection of data used in strategic decision making. Unlike an ODS the data warehouse stores historical and current data. A data warehouse is an environment of primarily read-only data storage in which data is transformed, integrated, and summarized for effective decision support across the whole enterprise.
- **Data Mart** - A Data Mart is a subset of an organizational data warehouse, usually focussed on a specific subject area. It usually is used as a downstream analytical application on top of the data warehouse.

10.3 Data Conversion Layer - The data conversion process to convert the bank's data in the format specified by Reserve Bank can be achieved using standard reporting formats like XBRL, XML, Excel based files or flat files. This layer can be developed either using data conversion tools available in the market or customized scripts. These scripts must be developed to ensure appropriate application of the business logic required for generating the returns, perform appropriate validations and convert the data in the specified format. There are several applications available in the market which would allow conversion of data from the bank's internal format the XBRL or XML formats. The conversion solutions can be implemented using either client-server approach or as a series of services using service oriented architecture (SOA). Owing to the varied nature of the return formats, it might be found that no single solution satisfies all requirements. In such a case a mix of custom developed scripts and readily available tools might be required to develop this layer.

10.4 Data Submission Layer - The data submission from the banks to Reserve Bank can be achieved using web based submission systems such as a secure web portal where the data can be uploaded directly or by using a batch loading (FTP) process, template based web page entries loading (HTTP), and custom developed

Web Services. The submission process can be triggered automatically using a scheduler. Data submission can be designed to support either “push” or “pull” mechanisms. If the data submission is initiated by the bank then it is considered to be a “Push”. In a “Pull” mechanism the submission process is initiated by Reserve Bank.

10.5 Data Validation Layer - A business rules engine can be used to do the data validation on the Reserve Bank side. Also, a custom built script can be developed to do the validation instead of using the business rules engine.

(a) Security - Generally available, database management systems allow for defining role based access to the data. Along with access control, additional security measures in the form of data encryption, identity management can be used to ensure there is no unauthorized access to the data. Data security may also be maintained whilst transmitting the data to Reserve Bank.

(b) Metadata – Metadata describes other data. It provides information about a certain data element's content. For e.g. the metadata for Account balance can be range of valid values, data type, currency used etc.

(c) Metadata Management - Metadata can be maintained using either the Open Information Model (OIM) or the Common Warehouse Meta model (CWM) depending on the overall architecture prevalent in the bank. The standard that has gained industry acceptance and has been implemented by numerous vendors is the Common Warehouse Meta model. The metadata management tool can be sourced as a separate package or can be part of the integrated suite of data integration and warehouse solutions. The metadata repository may be stored in a physical location or may be a virtual database, in which metadata is federated from separate sources

(d) Data Quality Management - Data Quality Management can be done either using a packaged tool or using custom developed scripts. Though custom developed scripts are cheaper and more flexible than a packaged tool. However, scripts are not very often preferred as they take longer to develop and are not easy to maintain on an on-going basis. The standard tools available in the market have pre-defined dictionaries which allow automated tool based profiling, cleansing and standardisation with minimal manual intervention.

(e) Reporting Tools - Reporting tools can be used by the bank to prepare the returns in the format desired by Reserve Bank. These tools provide the flexibility to do both standard reporting, i.e. where formats are pre-defined as well ad-hoc reporting, i.e. where format is defined by the user on the fly. In addition to allowing the user to create the report for internal consumption, the tools also provide for the ability to create external reports in XBRL or SDMX formats. The tools also provide facilities to export the reports in various formats such as: PDF, Microsoft Office Word format (.doc, .docx), Microsoft Office Excel format (.xls, .xlsx), Web page based reporting, etc.

11. Definitions

(a) Atomicity – Atomicity is the measure of granularity. Data which cannot be further broken down into smaller components is known as atomic data.

(b) Backdated entries in accounts – These are entries in accounts that are made effective from an earlier date. These changes are generally required after an audit is

done. For e.g. if an audit done on August 26th recommends a change in one of the entries for July 25th, such a change will be termed as a backdated entry.

(c) Centralized Data Repository – Refers to a repository of enterprise-wide data as required for preparing the returns to be submitted to Reserve Bank.

(d) Complete Automation – Is defined as no manual interventions required from the point of capture of data within the core system to the submission of the return to Reserve Bank.

(e) Data Granularity – Is defined as the level of detail of the data. Higher the granularity, deeper the level of detail. For e.g. a repository that stores daily account balances is more granular than a repository that stores monthly account balances.

(f) IT System / Core Operational System - The system refers to a technology based solution that supports collection, storage, modification and retrieval of transactional data within the Bank.

(g) People Dimension – This refers to the human resources aspect of the returns generation process which includes availability of resources, the roles and responsibilities of people involved, their skills and training.

(h) Process Dimension – This refers to the procedures and steps followed in the returns generation process. This will also cover the validation and checking done on the returns generated and review mechanism followed before submitting the return. It also includes the mechanism for submitting returns to Reserve Bank and receiving feedback or answering queries from Reserve Bank.

(i) Technology Dimension – This refers to all the IT systems used or required for the entire returns generation and submission process. It includes the systems required for collecting data, processing it and submitting the return.

Appendix II – Global Case Study

Banking - FDIC Call Report, CEBS

1. Many regulators across the world share some common challenges in their reporting functions owing to the nature of their requirements. Some of the most common ones are as defined below:

- Securely obtaining data that can be entered automatically and seamlessly into systems
- No re-keying, reformatting and / or other "translation" required to be done on the data.
- Reducing costs through automation of routine tasks.
- Quickly and automatically identifying errors and problems with filings.
- Validating, analyzing and comparing data quickly, efficiently and reliably.
- Shifting focus and effort of the concerned filers on analysis and decision-making rather than just data manipulation.
- Promoting efficiencies and cost savings throughout the regulatory filing process.

Regulators in the banking sector in the United States of America recognized these challenges and undertook a modernization project to overcome them. Members of the Federal Financial Institutions Examination Council (FFIEC), the Federal Deposit Insurance Corporation (FDIC), the Federal Reserve System (FRS), and the Office of the Comptroller of the Currency (OCC) sought to resolve these challenges through a large-scale deployment of XBRL solutions in its quarterly bank Call Report process. In addition, through the modernization project, the FFIEC also sought to improve its in-house business processes.

2. Legacy Data Collection Process –

A private sector collection and processing vendor acted as the central collection agent for the FFIEC. After receipt of the data from the agent, the FFIEC Call Agencies processed the data. The FRS transmitted all incoming data received from the agent to the FDIC. The FDIC and FRS then performed analysis on the received data and independently validated the data series for which each was responsible. The validation process consisted of checking the incoming data for "validity errors," including mathematical and logical errors, and "quality errors." Checking for quality errors included tests against historically reported values and other relational tests. FFIEC Call Agency staff corrected exceptions by manually contacting the respondents. They entered corrections and/or explanations into the FDIC's Call System and the FRS's STAR System. In some cases, the respondents were required to amend and resubmit their Call Report data.

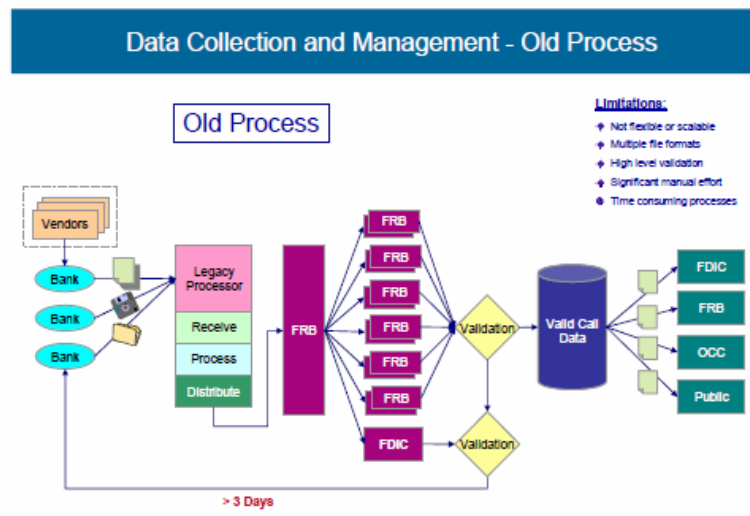


Figure 27

The FDIC was responsible for validating data of approximately 7,000 financial institutions, and used a centralized process at its Washington, DC headquarters. Historically, the agencies exchanged data continuously to ensure that each had the most recent data that had been validated by the responsible agency. Each agency maintained a complete set of all Call Report data regardless of the agency responsible for the individual reporting institution.

In addition to reporting current data quarterly, institutions were also required to amend any previous Call Report data submitted within the past five years as per the requirement. Amendments submitted electronically were collected by means of the process described above. Often the institution contacted the agency, and the agency manually entered only the changes to the data. The validation and processing of Call Report amendments were similar to those for original submissions. But, in this case an agency analyst reviewed all amendments before replacing a financial institution's previously submitted report. Amendments transmitted by the institutions using Call Report preparation software always contained a full set of reported data for that institution. Once the data was collected from all the respondents and validated by the agencies, the data was made available to outside agencies and to the public.

3. Technology Used in Automation Project

The Call Agencies relied on the "Old Process" for decades, introducing enhancements in piecemeal fashion. The Call Modernization project sought to reinvent and modernize the entire process in order to make it more useful for the regulatory community and its stakeholders. At the same time, it aimed to provide a relatively neutral transparent change to financial institutions. Early in the project, Call Report preparation software vendors were invited to participate in a roundtable discussion of reporting requirements and practices with an eye towards finding ways to improve it. Based on the findings of those discussions, the FFIEC identified areas to target for improvement in undertaking an inter-agency effort to modernize and improve the legacy process.

It was decided that the FFIEC may continue to provide data collection requirements that include item definitions, validation standards, and other technical data processing standards

for the banking institutions and the industry. The banking institutions would continue to utilize software provided by vendors or use their own software to compile the required data. The updated software would provide automated error checking and quality assessment checks based on the FFIEC's editing requirements. The editing requirements would have to be met before the respondent could transmit the data. Thus, all the data submitted would have to pass all validity requirements, or provide an explanation for exceptions. The regulatory agencies believed that quality checks built into the vendor software may play a key role in enhancing the quality and timeliness of the data. Placing the emphasis on validating the Call Report data prior to submission was deemed more efficient than dealing with data anomalies after submission.

The FFIEC was interested in exploring the use of a central data repository as the "system of record" for Call Report data. The data would be sent using a secure transmission network. Potentially, a central data repository would be shared among the regulatory agencies, and possibly with the respondents, as the authentic source of information. Once the central data repository received data, a verification of receipt would be sent to the respondent confirming the receipt. If a discrepancy was discovered in the data, online corrections would be made in the Centralized Data Repository directly by the respondent or by the regulatory agencies during their review.

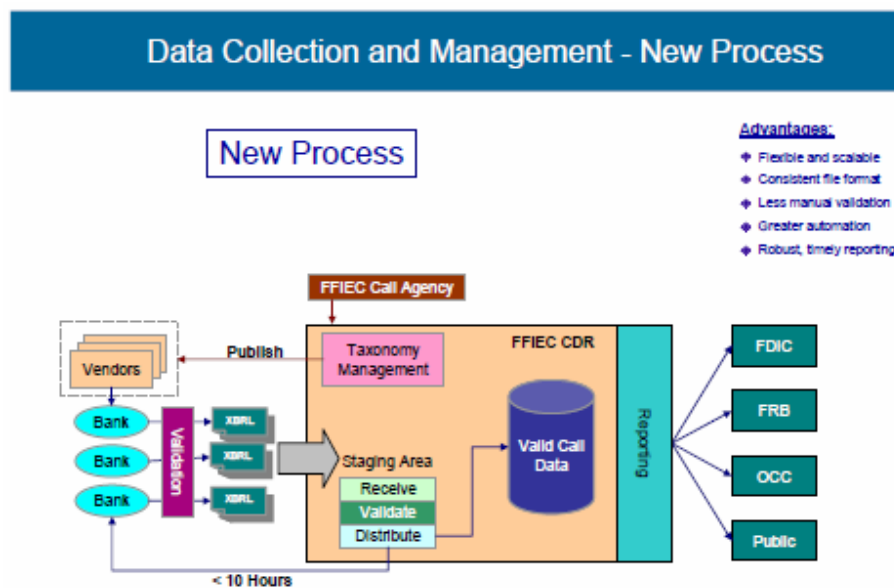


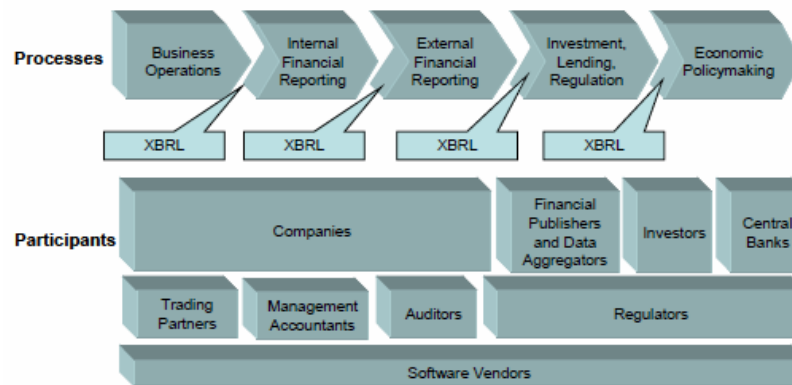
Figure 28:

The new system, known as the Central Data Repository (CDR), is the first in the U.S. to employ XBRL on a large scale and represents the largest use of the standard worldwide. The CDR uses XBRL to improve the transparency and accuracy of the financial reporting process by adding descriptive "tags" to each data element. The overall result has been that high-quality data collected from the approximately 8,200 U.S. banks required to file Call Reports is available faster, and the collection and validation process is more efficient.

4. End state

The FFIEC targeted five specific areas for improvement.

Vendor Software - The FFIEC provided Call Report software vendors with an XBRL, version 2.1 taxonomy.



Secure Transmission - A high level of security was needed in all phases of the data submission. Security had to encompass the entire process, from entry point to delivery point. The transmission process had to be automatic, with little or no input from the filing institution.

Verification of Receipt - A verification or notification mechanism was required to enable automatic reply to the institutions when the transmission of the data had been completed. In addition, institutions needed to be able to verify receipt of their transmission by logging into the CDR system.

Online Corrections - Respondents had to be notified if corrections were needed to the transmitted data. The institutions would have access to their data in the central data repository system. The online correction capability needed to be available in a real-time mode.

Central Data Repository - A Centralized Data Repository, "system of record," that banks, vendors and the agencies could use to exchange data needed to be created.

5. Benefits

Improvements to the data collection process have reaped immediate benefits in terms of timeliness and quality of data for the banking agencies. The CDR utilizes XBRL to enable banks to identify and correct errors before they submit their data to the federal banking agencies. Consequently, initial third quarter 2005 data submissions were of a high quality received days sooner than in previous quarters, when most data validation occurred only after the initial submission to the agencies.

Top line results of Call Report Modernization Project Using XBRL

Value Realized	Results Under New Process with CDR	Old Results Under Legacy Process
1. CLEANER DATA		
Requirements regarding data accuracy are better documented and more easily met.	95% of banks original filings met CDR requirements – logical business relationships must be true e.g. reported credit card income on the income statement may have a corresponding asset on the balance sheet, and banks were able to provide written explanations for any situations that exceed FFIEC tolerances.	66% clean when received – banks did not have the capability to provide notes when submitting data.
2. Data Accuracy		
Data adds up – 100% of mathematical relationships sum, no follow up required.	100% of data received met mathematical requirements – total accuracy and reliability.	30% of banks original filings did not meet requirements – not fully accurate.
3. Faster Data Inflow		
Requirements regarding data accuracy are better documented and more easily met.	CDR began receiving data at 4pm on October 1, 2005— less than one day after the calendar quarter end.	Data received weeks after the calendar quarter – not as timely.
4. Increased Productivity		
Staff can take higher case loads and are more efficient – agencies save money.	550 to 600 banks per analyst – an increase of 10-33%.	450 to 500 banks per analyst – less productive.
5. Faster Data Access		
Agencies receive data sooner and have the capability to publish it almost immediately; public can use data sooner and make better-informed decisions sooner.	As fast as within one day after receipt.	Within several days after receipt.
6. Seamless Throughput		
FFIEC Agencies and Call Report Software Vendors consume the same taxonomies, test changes prior to implementation, and	XBRL taxonomies provide the ability to make changes within minutes/hours, depending on number of changes.	Within days/weeks, depending on number of changes.

ultimately bankers are using the same requirements as the agencies created through XBRL taxonomies.		
---	--	--

References

<http://www.xbrl.org/us/us/FFIEC%20White%20Paper%2002Feb2006.pdf>

http://www.xbrl.org/CaseStudies/Spain_XBRL_06.pdf

<http://www.fujitsu.com/global/casestudies/INTSTG-bank-of-spain.html>

http://www.xbrl.es/downloads/libros/White_Paper.pdf

Satty, T. L. (1980). The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York: McGraw-Hill.

-----End of Document-----