

# **Report of Committee on Data Standardization**



**Department of Information Technology, Central Office  
Reserve Bank of India**

**February, 2015**

---

# **Report of Committee on Data Standardization**



**Department of Information Technology, Central Office  
Reserve Bank of India**

**February, 2015**

---



**Report of the Committee on  
Data Standardization**

(Dr. A S Ramasastry)  
Member

(Dr. A R Joshi)  
Member

(Shri G P Borah)  
Member

( Shri Kajal Ghosh)  
Member

( Shri A Narayanan)  
Member

( Shri N Raghupathi )  
Member

(Shri Sanjay Singhi)  
Member

( Shri Alope Chatterjee )  
Member

(Shri Devesh Lal)  
Member

(Smt. V. Radha)  
Member

  
(Shri P Parthasarathi)  
Convenor

## Table of Contents

| <b>Sr.No.</b> | <b>Content</b>   | <b>Page</b> |
|---------------|--|-------------|
|               | Abbreviations  |             |
|               | Acknowledgement  |             |
|               | Executive Summary  | i-xi        |
| 1             | Introduction   | 1-2         |
| 2             | Chapter I:Data Standardization - Setting the Context - Perspectives and overview on standardization in general | 3-6         |
| 3             | Chapter II:Issues relating to data gaps, data quality and data standardization from global perspective         | 7-12        |
| 4             | Chapter III: Data standards in banking/financial sector  | 13-20       |
| 5             | Chapter IV:Data management and data quality issues in banks  | 21-24       |
| 6             | Chapter V:Data governance architecture in banks  | 25-30       |
| 7             | Chapter VI: Data standardization in regulatory reporting – Commercial banks, NBFCs and UCBs                    | 31-38       |
| 8             | Chapter VII:ADF Project – Implementation by banks  | 39-42       |
| 9             | Chapter VIII :XBRL project of RBI – status and way forward   | 43-48       |
| 10            | Chapter IX: Automating data flow between banks and RBI   | 49-54       |
| 11            | Chapter X:Data Quality Assurance Process at RBI  | 55-56       |
| 12            | Chapter XI:Perspective on System-wide improvements   | 57-66       |
| 13            | Chapter XII: Trends and new developments in future   | 67-71       |
| 14            | Annexures  | 72-114      |
| 15            | References   | 115-116     |

---

## Abbreviations

|       |  |
|-------|--|
| ACB   | Audit Committee of Board                                   |
| ADF   | Automated Data Flow  |
| BI    | Business Intelligence                                      |
| BIS   | Bank for International Settlements                         |
| BSR   | Basic Statistical Returns                                  |
| CBS   | Core Banking Solution                                      |
| CDS   | Credit Default Swap  |
| CIO   | Chief Information Officer                                  |
| CTO   | Chief Technology Officer                                   |
| DBIE  | Database of Indian Economy                                 |
| DBS   | Department of Banking Supervision                          |
| DGI   | Data Gap Initiative  |
| DIT   | Department of Information Technology                       |
| DNBS  | Department of Non-Banking Supervision                      |
| DQF   | Data Quality Framework                                     |
| DSB   | Department of Supervision- Banking                         |
| DSIM  | Department of Statistics and Information Management        |
| DTCC  | Depository Trust and Clearing Corporation, US              |
| DW    | Data Warehouse   |
| ECB   | European Central Bank                                      |
| EDW   | Enterprise Data Warehouse                                  |
| ETL   | Extraction, Transformation and Loading                     |
| FIBIM | Financial Instrument Business Information Model            |
| FIBO  | Financial Industry Business Ontology                       |
| FIX   | Financial Information Exchange Protocol                    |
| FpML  | Financial Product Markup Language                          |
| FSB   | Financial Stability Board                                  |
| Gol   | Government of India  |
| GSBPM | Generic Statistical Business Process Model                 |
| IDRBT | Institute for Development & Research in Banking Technology |
| IMF   | International Monetary Fund                                |
| IMM   | Information Management Meta model                          |
| ISO   | International Organization for Standardization             |
| IT    | Information Technology                                     |
| KDM   | Knowledge Discovery Meta model                             |
| LEI   | Legal Entity Identifier                                    |
| MDDL  | Market Data Definition Language                            |
| MISMO | Mortgage Industry Standards Maintenance Organization       |
| MIS   | Management Information System                              |
| NBFCs | Non-Banking Financial Companies                            |

|             |  |
|-------------|--|
| NBFCs-D     | NBFCs - Deposit taking                                 |
| NBFCs-ND    | NBFCs - Non-deposit taking                             |
| NBFCs-ND-SI | NBFCs- Non-Deposit taking - Systemically Important     |
| NIBM        | National Institute of Bank Management                  |
| ODM         | Ontology Definition Meta model                         |
| OECD        | Organization for Economic Co-operation and Development |
| OLAP        | On-line Analytical Processing                          |
| OMG         | Object Management Group                                |
| ORFS        | Online Returns Filing System                           |
| OWL         | Web Ontology Language                                  |
| RBI         | Reserve Bank of India                                  |
| RBS         | Risk-Based Supervision                                 |
| RDBMS       | Relational Database Management System                  |
| RDF         | Resource Description Framework                         |
| RTGS        | Real Time Gross Settlement                             |
| SCB         | Scheduled Commercial Bank                              |
| SDMX        | Statistical Data and Metadata Exchange                 |
| SFTP        | Secure File Transfer Protocol                          |
| SSM         | Senior Supervisory Manager                             |
| StCB        | State Co-operative Bank                                |
| UBD         | Urban Banks Department                                 |
| UML         | Unified Modeling language                              |
| UCB         | Urban Co-operative Bank                                |
| UNECE       | United Nations Economic Commission for Europe          |
| XBRL        | eXtensible Business Reporting Language                 |
| XSD         | XML Schema Definition                                  |
| W3C         | World Wide Web Consortium                              |

## **Acknowledgements**

The Committee acknowledges with gratitude the role of Deputy Governor Shri.H.R.Khan in supporting and guiding the Committee. The Committee gratefully acknowledges the detailed guidance and encouragement by the Executive Director Shri.G.Padmanabhan and for laying down the detailed framework for the work of the Committee. The Committee also deeply acknowledges the valuable guidance of previous CGM-in-charge Shri.A.S.Ramasastri, (currently Director, IDRBT). The Committee also gratefully acknowledges the detailed guidance and key inputs received for enhancing the report from Shri.S.Ganesh Kumar, present CGM-in-charge, DIT,CO. The Convenor acknowledges the cooperation extended by the members of the Committee in completing the task entrusted to it.

The Group wishes to thankfully acknowledge the Secretarial support provided by the Policy Division of DIT, Central Office consisting of Ms.Nikhila Koduri, GM(since transferred), Shri. Devesh Lal, GM and Ms. Jolly Roy, AGM for conduct of meetings of the Committee, preparing materials for discussion and compiling draft materials.

The Committee gratefully acknowledges the contribution by Shri.N.Suganandh, DGM, DIT, CO for his detailed research assistance and dedicated efforts in compiling the report.

Useful inputs from Shri.Purnendu Kumar, Asst. Adviser, DSIM on chapter on XBRL project is also acknowledged.

---





## Executive Summary

Reserve Bank has always emphasized the importance of both quality and timeliness of data to enable transforming the data into information that is useful for decision making purposes. To achieve this, uniform data standards are of vital importance. RBI issued the Approach paper on ADF in 2010 and laid down a detailed road map for its implementation by the banks. Banks initiated measures to implement the ADF. However, the progress in implementation varies across banks.

Simultaneously, work is also underway to develop the XBRL schema for returns which enables standardization and rationalization of various returns with internationally accepted best practices of electronic transmission of data apart from work on Harmonization of Banking Statistics. This Committee has been set up to bring about synergy and uniformity of efforts being undertaken in the area of data reporting and data standardization.

In the context of terms of reference to the Committee, the examination of any given stream of thought on data standardization would necessitate addressing a canvas of issues relevant to our context so as to fully address its different dimensions. Accordingly, the following aspects were examined and covered by the Committee:

- (i) Discussion of basic conceptual perspectives and various data/information standards in financial sector
- (ii) Data quality, data gap issues – international developments
- (iii) Issues in data management and data quality in banks
- (iv) Need for Data governance framework in banks
- (v) Data standardization and reporting – commercial banks, NBFCs, UCBs
- (vi) ADF Project – Status, Issues and way forward
- (vii) XBRL project – status and way forward
- (viii) Automating data flow from banks to RBI
- (ix) Data Assurance process at RBI
- (x) Data standardization - System-wide Perspective
- (xi) New developments and future perspectives

Summary of recommendations are furnished below:

### **(I) Data standards**

- 1) Regulatory data reporting standards - The data exchange standards need to be based on open standards and allow for standardization of data elements and minimizing data duplication/redundancy. RBI had already embarked on XBRL(which is based on XML platform) as the standard platform for a set of

regulatory returns which may be continued for rest of returns or data elements. Thus, XBRL may be the reporting standard for regulatory reporting from banks to RBI.

- 2) The well-known Statistical Data and Metadata eXchange (SDMX) is recommended as a standard for exchanging statistical data.
- 3) In regard to standardisation of coding structures, in accordance with international standards, the ISO 4217 currency codes and ISO 3166 country codes can be used.
- 4) System of National Accounts (SNA) 2008 can be used for classification of institutional categories.
- 5) International Standard Industrial Classification (ISIC)/ National Industrial Classification (NIC) codes for economic activity being financed by a loan may be incorporated.
- 6) The ISO 20022 standard is recommended to be the messaging standards for the critical payment systems. RTGS in India already uses the ISO 20022 message formats.
- 7) In order to acquire single view of transactions in respect of a customer, banks are required to allot unique customer id. Given that no single identifier can represent all categories of customers of banks, the differentiation may need to be made by mapping with the identifiers presently available. Recently, Clearing Corporation of India Limited (CCIL) has been selected to act as a Local Operating Unit in India, for issuing globally compatible unique identity codes (named as legal entity identifier or LEI) to entities which are parties to a financial transaction in India. Given the LEI initiative, efforts to facilitate LEI for legal entities involved in financial transactions across financial system needs to be expedited to maximise coverage over the medium term.
- 8) Given the complexity of some of the corporate entities with numerous subsidiaries including step down subsidiaries, there is a need for usage of LEI or similar methodology to link the complex hierarchy of any corporate to facilitate ease of identification of total credit exposure of corporate groups. While it is reported that LEI application of CCIL has provision for the same, the utility may need to effectively leveraged to map the corporate group hierarchy.
- 9) While presently LEI architecture caters to legal entities involved in financial transactions, ultimately LEI or similar system needs to be made broad-based to incorporate other categories of customers like partnership firms and individuals.
- 10) For conduct of electronic transactions and reporting purposes in financial markets, well known international standards like ISO based standards can be considered where possible.
- 11) In order to take up data element/return standardisation through standardising or harmonising definitions, efforts of earlier working groups (Committee on rationalisation of returns and Committee on harmonisation of banking statistics) can be consolidated by setting up an inter-departmental project group within RBI

which can work in a project mode so as to ensure comprehensive and effective implementation of standardisation and consistency of data element definitions across complete universe of returns/data requirements of RBI.

## **II. Key components of Data governance architecture in banks and related aspects**

- (i) Committee recommends that key components of data governance architecture in banks may incorporate following aspects:
- (ii) Formulation of Data governance or information management policy with emphasis on various aspects like data governance organisational structure, data ownership, definition of roles and responsibilities, implementation of data governance processes and procedures at individual functions/departments, development of a common understanding of data, data quality management, data dissemination policy and management of data governance through metrics and measurements.
- (iii) Overall oversight of data governance may be with the Audit Committee of Board (ACB) of a bank or a specific Committee nominated by the Board.
- (iv) Formation of executive level Data Governance Committee or entrusting responsibility to existing information management committee if already existing. Data governance responsibilities and accountabilities should be clear, measured and managed to ensure sustained benefit to the bank.
- (v) Data Ownership related aspects to be considered include overall responsibility for data, assigning ownership of key data elements to data controllers or stewards, assigning data element quality within business areas, implementing data quality monitoring and controls, providing data quality update to management/data governance committee and providing data quality feedback to business data owners.
- (vi) 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. Data ownership needs to be primarily based on the business function.
- (vii) Platforms and data warehouse/s need to employ common taxonomies/data definitions/meta data. 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 need to be linked across the applications using pre-defined rules. The rules to be applied for synchronization of metadata also need to be defined.

- (viii) The metadata (data definitions) may be synchronized across various source systems and also with the RBI definitions for regulatory reporting.
- (ix) To help drive data governance success, measurements and metrics may be put in place which define and structure data quality expectations across the bank for which various data governance metrics and measures would be required. Data needs to be monitored, measured, and reported across various stages: data acquisition, data integration, data presentation and data models, dictionaries, reference data and metadata repositories.
- (x) Internal audit function to provide for periodic reviews of data governance processes and functions and report on the issues to ACB.
- (xi) Detecting and correcting the faulty data manually is very tedious and time consuming. It is in this context the validation methods based on statistics, machine learning and pattern recognition gain importance. Many DBMS, DWDM product vendors now offer Data Profiling, Data Quality, Master Data Management services. Banks can take advantage of all these tools and techniques to keep their data clean.
- (xii) Various focussed data quality assessment and improvement projects need to be undertaken.
- (xiii) Banks can also endeavour to establish a centralised analytics team as a centre of excellence in pattern recognition technology and artificial intelligence (AI) to provide cutting edge analysis and database tools or information management tools to support business decisions.
- (xiv) Apart from providing enhanced focus during AFI/RBS, data governance mechanisms in banks may also be examined intensively through focussed thematic reviews by DBS of RBI. Based on outcome of thematic reviews, detailed guidance may be issued to banks to address issues identified during review.
- (xv) Banks, in particular domestic SIBs, may also be advised to keep in context BIS document “Principles for effective risk data aggregation and risk reporting” as part of their information management process.
- (xvi) Guidance on Best practices on data governance and information management can be formulated by IDRBT.
- (xvii) RBI may facilitate creation of Data Governance Forum under the aegis of IBA or learning institutions like CAFRAL or NIBM with other stakeholders like IDRBT, RBI, IBA, banking industry technology consortiums and banks, to assist in development of common taxonomies/data definitions/meta data for banking system.
- (xviii) Bank Technology Consortiums under the aegis of IDRBT and other stakeholders like banks can validate critical banking applications like CBS and provide guidance on expected minimum key information requirements/validation rules and address to the extent possible different customizations across banks. While specifying key regulations, RBI may

also endeavour to specify any key system related validation parameters and details of data quality dimensions expected from concerned regulated entities.

- (xix) Committee also prepared an illustrative list of data aspects pertaining to credit function that would need to be addressed by commercial banks to facilitate data standardization, data comparability and data reliability across banks.

### **III. Data standardization in regulatory reporting–Commercial banks, UCBs, NBFCs**

- 1) XBRL platform may be gradually expanded across the full set of regulatory returns.
- 2) Robust internal governance structure needs to be set up in regulatory entities with clear responsibilities and accountabilities to ensure correct, complete, automated and timely submission of regulatory/supervisory returns.
- 3) Regulatory reporting - Commercial Banks:
  - a. Adoption of uniform codes among different returns of RBI will reduce inconsistency among returns. For eg. DSIM of RBI collects industry classification of credit as per NIC codes while other departments use different classification. Each bank has adopted its own approach to map NIC codes. Analysis of mappings of some banks showed apparent divergences. Hence, as part of data standardisation efforts, data collected by other returns may also be brought in alignment with the usage of common or standardised codes incorporated in BSR.
  - b. The BSR codes need to be updated based on latest NIC 2008 classification. The BSR codes may be reviewed periodically and updated. Further, it should be possible to establish one-to-one mapping of sector/ industry codes in various other regulatory returns from the same.
  - c. The nature of returns are generally dimensional in nature, consisting of various components like measures, concepts , elements, attributes, dimensions and distributions. A suitable data model may be generated to facilitate element-based, simplified and standardised data collection process by RBI under a generic model structure that is suitable for both primary and secondary data.
  - d. There is a need to ultimately move over to “data” centric approach from the current “form” centric approach. Under a data centric approach methodology, any data point must be expressed by its “primary element” and all additional dimensions necessary to their identification. As “form” centric approach is oriented to the visualization of the data in certain format, it may be used for reviewing purpose only. Thus, from a

medium term perspective, moving from return based approach to data element based approach needs to be considered

- e. The values of various attributes and dimensions should be standardised to enable the collation of data from different domains.
  - f. Suitable data sets with varied nature like hierarchical, distributional or dimensional can be created to facilitate submission of data in summarised or granular form as the case may be from the central repository of the banks.
  - g. A good feedback mechanism from banks to RBI and vice versa can help maintain uniqueness in data definitions. As and when multiple data definitions from XBRL taxonomy given by RBI map to same data element in any bank, they need to be flagged to RBI.
  - h. Phased implementation of various standardised data definitions can be commenced based on elements which were already standardised.
- 4) NBFCs and UCBs:
- a. Rationalisation of returns needs to be attempted for NBFCs and UCBs. An exercise carried out earlier indicated significant duplication in the information provided through various returns. The Committee recommends that these returns may be rationalised by identifying major data elements and removing duplicate data elements.
  - b. The Committee recommends an online data collection mechanism for larger NBFCs and Tier II UCBs.
  - c. In due course, after rationalisation exercise, data element based return submission through XBRL may also be initiated.
  - d. Suitable data model and robust meta data system may be developed.

#### **IV. ADF implementation by banks**

- 1) **Use of ADF for Internal MIS** - The RBI Approach Paper highlighted usage of ADF platform for generating internal MIS as one of the key benefits of ADF. In this regard, banks may explore using the platform for generating internal MIS and other uses. Indicatively some aspects include :
  - NPA Management Automation Module
  - Automation of SLBC Returns
- 2) Detailed survey can be carried out by RBI to ascertain the status of ADF implementation by banks. Feedback may also be obtained from DBS regarding any issues relating to ADF implementation obtained during AFI/RBS examination process. Any manual intervention from source systems to ADF central repository needs to be ascertained. Independent assurance on the ADF central repository mechanism in individual banks may also be verified. This would enable assessment of the quality and comprehensiveness of ADF

implementation by individual banks. Any specific issues may be taken up with concerned banks for remediation.

- 3) Banks may take steps to enable the ADF Platform to cater to the Risk Based Supervision(RBS) data requirements by suitably mapping the RBS data point requirements. Thus, the ADF structure should be made use of and aligned to the RBS set-up so that synergies can be built-in, data quality and consistency can be enabled and the overall system can be made more efficient.
- 4) Existing ADF platform needs to be leveraged by prescribing the necessary granular data fields to be captured by banks to achieve consistency and uniformity in regulatory reporting.
- 5) Banks may also port the necessary details required by RBI as indicated in Guidelines on “Framework for Revitalizing Distressed Assets in the Economy - Guidelines on Joint Lenders' Forum (JLF) and Corrective Action Plan (CAP)” in February, 2014 under ADF central repository platform.
- 6) Depending on the requirement of RBI regarding granularity of data, ADF system needs to be suitably updated to provide for the requisite granular data fields at the central repository level. The ADF system of the banks should be designed flexibly to accommodate any anticipated changes in the format of return, i.e., addition and deletion of data elements.

## **V.XBRL Project of RBI**

- 1) Similar forms can be taken together within/ across the departments of RBI and thus common reporting elements can be arrived at. Rationalisation /Consolidation of returns before taking up the returns pertaining to a department must be done. The rationalisation / consolidation of returns may be examined and reviewed on a periodic basis.
- 2) For granular account level data and transactional multi-dimensional data, RBI may develop and provide specific details of RDBMS/text file structures along with standardised code lists and basic validation rules so that banks can run the validation logics to ascertain that the datasets are submission-ready. In this connection, XBRL based data element submission may also be explored.
- 3) It is expected that banks would generate the instance document from the Centralised Data Repositories (CDR) and submit the same to RBI without manual intervention. The banks should validate the generated instance documents based on the XBRL taxonomy and validation rules before sending them to the Reserve Bank. Thus, the present approach of spreadsheet(Excel) based submission of XBRL returns needs to be given up ultimately.
- 4) An Inter-Departmental Data Governance Group (DGG) for the RBI as a whole may be formed, so that the process of rationalization regarding data elements, periodicity, need for provisional returns can be carried out in a concerted

manner. All future returns to be prescribed by any department may be routed through the DGG, to avoid duplication.

- 5) As part of its data governance activities, the DGG may also pro-actively identify any data gaps in the evolving milieu and prepare plan of action to address the gap.
- 6) The XBRL taxonomy must include data definitions so as to completely leverage the utility offered by XBRL.
- 7) The XBRL taxonomy should be designed flexibly so as to take care of the anticipated changes in the format of return, i.e., addition and deletion of data elements.
- 8) The XBRL based submission by financial companies to MCA should be shared across the regulators as required.
- 9) Since new tools/software are developed for leveraging XBRL, there needs to be process of continuous monitoring of new developments so as to examine their utility and possible value addition.
- 10) Ultimately, the logical location for storage of XBRL data is a Data Warehouse. Therefore the existing Data Ware House needs to be revamped with Next Generation Data Ware House capabilities.

## **VI. Recommendations on Automating data flow from banks to RBI**

- 1) Using secure network connections between the RBI server and the bank's ADF server, the contents of the dataset can be either pulled through ETL mode or pushed through SFTP mode and loaded onto the RBI server automatically as per the periodicity without any manual intervention. Pushing of data by banks could enable easier management of the process at RBI end. An acknowledgement or the result of the loading process can be automatically communicated to the bank's ADF team for action, if necessary.
- 2) The validation schemes may also be expressed in XBRL/XML form so that the systems at banks automatically understand the requirement, accordingly process their data and return the data to RBI, without any manual intervention. This would enable a fully automated data flow from banks to RBI even with dynamic and changing validation criteria.
- 3) While the traditional RDBMS infrastructure in place in RBI may be used for storage and retrieval of aggregated and finalized data, Big-data solutions may also be considered for micro and transactional datasets given their high volume, velocity and multi-dimensional nature. Big Data solutions also help enhance analytical capability in the new data paradigm particularly in the area of banking supervision.
- 4) The enterprise-wide data warehouse (EDW) of RBI should be made the single repository for regulatory/supervisory data pertaining to all regulated entities of RBI with appropriate access rights. Any unstructured components pertaining to RBS data may be maintained in EDW using new tools available for such items.



- 5) As a key support for risk based supervision for commercial banks, internal RBI MIS solution needs to seamlessly generate two important sets of collated information: (i) Risk Profile of banks (risk-related data – mostly new data elements), and (ii) Bank Profile (mostly financial data – DSB Returns and additional granular data) based on data supplied by banks.
- 6) Once the system stabilises, the periodicity of data can be reviewed so as to obtain any particular set of data at shorter intervals or even up to near real time.

## **VII. Recommendations on data quality assurance process**

- 1) Exclusive data quality assurance function can be created under the information management unit of RBI.
- 2) A data quality assurance framework may be formulated by RBI detailing the key data quality dimensions and systematic processes to be followed. The various key dimensions include relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence. The framework may also be periodically reviewed.
- 3) Various validation checks like sequence check, range check, limit check, existence check, duplicate check, completeness check, logical relationship check, plausibility checks, outlier checks are among the key checks which need to be considered and documented for various datasets with assistance from domain specialists.
- 4) Usage of common systems for data collection, storage and compilation would help provide environment for robust implementation of systematic data quality assurance procedures.
- 5) Deployment of professional data quality tools as part of the data warehouse infrastructure could also provide for comprehensive assessment of data quality dimensions.
- 6) Whenever data are received and compiled, quality assessment reports that summarize the results of various quality checks may also be generated internally.

## **VIII. System-wide Improvements:**

- 1) Given that standards are considered a classic public good, with costs borne by a few and benefits accruing over time for many entities, active involvement of regulators and Government is now internationally acknowledged as key towards solving the collective action problems created by these disincentives. Inter-regulatory forums could help facilitate improvements in data/information management standards across the financial sector to benefit all stakeholders and furthering collaboration with international stakeholders.
- 2) A separate standing unit Financial Data Standards and Research Group may be considered with involvement of various stakeholders like RBI, IBA, banks,

ICAI, IDRBT, SEBI, MCA, NIBM, CAFRAL etc for looking at the financial data elements/standards and to try to bring them into holistic data models apart from mapping with applicable international standards.

- 3) Regulators like RBI, SEBI, MCA are in the process of undertaking various XBRL projects. Given the benefits offered by XBRL and its usage across the globe by regulatory bodies, all the regulators may explore possibilities of commonalities in taxonomy and data elements to the extent possible. Protocols and formats may be formulated for sharing of the data among themselves.
- 4) In regard to OTC derivatives, one of the issues being debated is data portability and aggregation among the trade repositories spanning countries and jurisdictions. Hence, it is important to be cognizant of the needs of uniformity of standards across the globe and the need for our repository framework to have sufficient flexibility to conform to international standards and best practices as they evolve depending upon their relevance in the Indian context.
- 5) Ultimately, from a banking system perspective full benefit would arise by enabling transactional and accounting systems in banks to directly tag and output data in formats like XBRL to maximize efficiency and benefit. Thus, there is need for integration of standard formats like XBRL in internal applications/accounting systems of banks. The present scope of XBRL data definitions have to be further extended to cover in depth data definitions covering almost all data elements that are required to carry banking business.
- 6) In respect of knowledge sharing and research, various measures recommended include (i) Research by IDRBT regarding ways and means of leveraging new data technological platforms like XBRL for enhancing overall efficiencies of banking system (ii) conducting of pilot for enhancing leveraging of technologies like XBRL for internal uses by banks.
- 7) Standard Business Reporting, which involves leveraging technologies like XBRL by Government for larger benefits beyond the field of regulatory reporting, is being implemented in various countries like Australia and Netherlands. The same may be explored in India by Government of India in a phased manner.
- 8) As the leveraging of machine readable tagged data reporting increases, the audit and assurance paradigm also need to get re-engineered to carry out an electronic audit and electronic stamp of certification using digital signatures.
- 9) Committee recognizes that coordinated efforts are being carried out by various organizations which have developed standards like FIX, FpML, XBRL, ISD etc for laying the groundwork for defining a common underlying financial model based on ISO 20022 standard. Costs of migration and inter-operability would be key factors going forward.

- 10) As had been indicated in Chapter II, comparability of financial data across countries is a key challenge faced globally. Increasing adoption of IFRS across countries is a positive development. While there is large number of convergence in capital standards via Basel I and Basel II, there are variations in details and level of implementations across countries. While the G-20 Data Gap initiative is a work in progress, there is also need for international stakeholders to analyse and examine how technologies like XBRL can help facilitate ease of comparability of data as also to identify differences between countries in respect of financial/regulatory measures and reporting rules in an automated manner.
- 11) Financial instrument reference database could be explored with focus on key components relating to ontology, identifiers and metadata and valuation and analytical tools, akin to such initiatives in US.
- 12) A single Data Point Model or methodology at international level can be explored for the elaboration and documentation of XBRL taxonomies
- 13) Govt has plans to establish Financial Data Management Centre (FDMC) as a repository of all financial regulatory data. Large investments already made by the individual regulators also needs to be factored in.
- 14) There is also need to incorporate training and education on the new technologies like XBRL by various academic bodies as also training/learning institutions so as to help in capacity building and to improve the availability of trained resources.

## **IX. Future trend and developments**

- 1) Committee recommends that research/assessment of new developments in technology and financial data/technology standards need to be made a formal and integral part of the information system governance of banks and the regulator.
- 2) Banking technology research institute IDRBT may carry out research on new technologies/development and serve as a think tank in this regard.
- 3) Banks may explore Big Data solutions for leveraging various benefits of the new paradigm concerned with volume and velocity of data.
- 4) Any financial technical data standards needs to be of the nature of open standards, inter-operable and scalable in nature. Due impact assessment and pilot run would also be necessary before implementing on larger scale.

## **X. Implementation of key recommendations**

- 1) The suggested timeframe for implementing the recommendations of the Committee is indicated at Annex VII.



# Introduction

## Committee on Data Standardization

### THE GENESIS

The IT Vision 2011-17 document of the Reserve Bank had emphasized the importance of both quality and timeliness of data for its processing into useful information for MIS and decision making purposes. To achieve this, uniform data reporting standards are of vital importance. Banks are in various stages of implementation of the Automated Data Flow (ADF), a project initiated by the Reserve Bank to ensure smooth and timely flow of quality data from the banks to the Reserve Bank. Simultaneously, work is also underway to develop the XBRL schema for regulatory returns which enables standardization and rationalization of various returns with internationally accepted best practices of electronic transmission of data as also on Harmonization of Banking Statistics. It is in this context that RBI had approved the formation of a Committee for Data Standardization which inter-alia will bring about synergy and uniformity in the efforts being undertaken in the areas of data reporting and data standardization.

DIT CO had initiated the Automated Data Flow (ADF) project in August 2010 and as enumerated in ADF Approach Paper, it was envisaged that under the project data from various source systems of banks would flow to a centralised MIS server within the bank which in turn would be used to furnish regulatory reports and returns to RBI in an automated manner without any manual intervention. The banks were at varied stages of implementing the project. A comprehensive Status Note was put up to top management. Based on the directions of top management, various distinct efforts were initiated for addressing issues around quality of regulatory data reported by the banks to RBI.

The main brief of DIT Committee on Data Standardisation is, *inter alia*, to bring about a synergy and uniformity into data reporting and data standardisation process. The committee has representatives from various departments of the Bank such as DSIM, DIT, DBS, and a large Regional Office in addition to experts from IT firm Infosys, chief of returns Governance Group of three scheduled commercial banks (SBI, ICICI, and HSBC), IDRBT with Shri P. Parthasarathi, CGM, DIT as convener of the committee.

The members of the Committee are given below:

| Department/Office/Institute |                                      | Name                      |
|-----------------------------|--------------------------------------|---------------------------|
| <b>From RBI</b>             |                                      |                           |
| 1                           | DIT (Convenor)                       | Shri P Parthasarathi, CGM |
| 2                           | DSIM                                 | Dr. A R Joshi, Adviser    |
| 3                           | Chennai RO( Presently In Mumbai ,RO) | Shri G P Borah, CGM       |
| 4                           | DBS,CO                               | Shri Alope Chatterjee, GM |
| 5                           | DIT                                  | Shri Devesh Lal, GM*      |
| <b>Other Organizations</b>  |                                      |                           |
| 6                           | IDBRT                                | Dr. A S Ramasastrri       |
| 7                           | IDRBT                                | Dr. V Radha               |
| 8                           | SBI                                  | Shri Kajal Ghose          |
| 9                           | HSBC                                 | Shri A Narayanan          |
| 10                          | ICICI                                | Shri Sanjay Singhvi       |
| 11                          | Infosys                              | Shri C N Raghupathi       |

\*Shri.Devesh Lal replaced Ms.Nikhila Koduri, GM subsequent to her transfer.

The terms of reference of the committee are as follows:

- (a) to study the Quality Assurance Procedures and function relating to data in few other central banks and analyse it's possible emulation at the RBI.
- (b) to study the existing reporting system and its usage in various departments in the Bank.
- (c) to peruse the data standards implemented by XBRL and data warehouse and
- (d) to recommend on uniform data standards for banking system
- (e) any other related issues

The approach of the Committee was to form sub-groups for focussed study of subject areas under remit of the Committee. The secretarial assistance for the committee was provided by DIT, CO.

## Chapter I

### Data Standardization -Setting the Context - Perspectives and overview on standardization

#### **1.1 Introduction**

A fundamental issue for enabling effective decision making is obtaining consistent, reliable and robust data. In the field of banking too, data is a key component and affects or impacts the information and knowledge gleaned from the same. The major developments in finance and the speed with which it has occurred, have been built upon standardized methods to exchange data. This chapter sets the context and leads to further unraveling of the various dimensions of data standardization and related issues in subsequent chapters of the report. This chapter indicates in general the importance of standardization and then delves into the world of IT standards by giving an overview of standardization in the IT sector, brief history of IT standards and importance of data standards.

#### **1.2 Concept of Standards**

The term “standard” has various definitions. While various dictionaries offer multiple definitions , a commonly offered definition is one offered by the International Organization for Standardization (ISO). The website of ISO defines a standard as a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose.

#### **1.3 Benefits of standardization**

As per ISO, standards help bring technological, economic and societal benefits. They help to harmonize technical specifications of products and services making industry more efficient and breaking down barriers to international trade. ISO has documented several research and case studies espousing the various empirical benefits of standards.

The value of standards to business and indeed the economy at large comes from the effective and efficient adoption, uptake of a standard across its target population. Thus, the effective adoption and diffusion of a standard is vital to the standardization process for economic benefits to be realized .

#### **1.4 Classification of standards**

There could be either horizontal or vertical standards. Every industry has standards that are specific to the business domains they deal with. Finance, Manufacturing, Medical etc. are examples of industry sectors that maintain their own standards. Thus, each of these areas creates vertical standards. These vertical standards often overlap in areas of applicability, particularly in respect of data representation and information technology, resulting in standards that have a similar purpose but are different in terms of implementation and the vocabulary or terminology used to specify the standard.

Information technology standards are generally applicable across multiple industry sectors. Because of their general applicability, these standards are often referred to as horizontal standards. Given the rapid developments in information technology arena, many of the key horizontal technology standards are falling within the broad category of Information and Communication Technology (ICT). In recent years there has been a push to implement industry-specific vertical standards using existing horizontal technology standards, for example XML-based vertical standards such as XBRL. In this case, the horizontal syntax standard XML is used to create specific vertical applications for financial/accounting data.

#### **1.5 Information Technology standards**

Individuals, businesses and governments throughout the world use Information Technology (IT) extensively. In order to facilitate this extensive use of IT, systems need to be interconnected and work across applications, organizations and geographic locations. This has resulted in a dramatic jump in network connections, a proliferation of computing devices and varied uses of IT based applications.

These interactions highlight the critical need for a comprehensive and consistent set of standards within the IT sector. Standards activities in the IT sector were said to have begun in the 1960s. Early standardization efforts were for certain programming languages and protocols for moving information around. Further, the early standards were reported to have been mainly focused on the syntax rather than describing content in terms of the nature of the information being standardized .

Subsequently, computer readable forms of syntactic specifications have emerged like XML (eXtensible Markup Language). The interoperability and data exchange across different vendors, platforms, applications and software is dependent upon standardized interfaces, protocols, services and formats. Therefore, standards in



information technology can help the portability and compatibility of systems and hence enable ease of exchanging information between systems.

## **1.6 Data standards**

The data standards help facilitate the exchange of data between systems, aggregation of data from multiple source systems, comparison of data among unrelated systems and automation of processes for storing, reporting, and processing data. Data standards define the format, content, and syntax of data, providing a common language that enables precise identification of entities and instruments, the relationships among them, and the data to describe them. Data standards also can help enhance data quality by supporting consisting metadata.

Standardized Definitions used across the public and private sectors improve the value of data for analysis. When key terms are not clearly defined, financial analysts are unable to accurately interpret and compare data, resulting in a lack of confidence in the results. Standards are particularly needed when data include a common term that can be understood in various ways. Standardized formats help analysts aggregate and compare data, and automate processes for storing, reporting, and processing data. It is important to consider how data may be used and to apply a standard format, even to routine information.<sup>1</sup>

## **1.7 Data standardization in Indian banking system**

The foresaid discussion provides a general overview of the concept of standards, its different dimensions and regarding data standards.

In the context of terms of reference to the Committee, Committee opines that the examination of any given stream of thought on data standardization would necessitate addressing canvas of issues so as to fully address its different dimensions. Accordingly, the following aspects were examined and covered by the Committee:

- (i) Discussion of basic conceptual perspectives and various current data/information standards in financial sector
- (ii) International developments relating to data quality, data gap and data standardisation issues
- (iii) Issues in data management and data quality in banks
- (iv) Need for Data governance framework in banks
- (v) Data standardization and reporting – commercial banks, NBFCs, UCBs
- (vi) ADF Project – Status and Issues
- (vii) XBRL project – status and way forward
- (viii) Automating flow of regulatory data from banks to RBI

---

<sup>1</sup>Office of Financial Research(OFR), USA, Annual report (2014)

- (ix) Data quality assurance process at RBI
- (x) Data standardization - System-wide Perspectives
- (xi) New developments and future perspectives

The subsequent chapters elucidate the assessment of the Committee on these dimensions and the recommendations thereon.

## **Chapter II–Data gaps, data quality and data standardization– International developments**

### **2.1 Introduction**

Financial crisis had revealed various issues pertaining to data quality. In the aftermath of the same, various data quality issues were identified and specific initiatives to address the same are being carried out. This chapter details the various issues pertaining to data quality, data gaps and data standardization arising out of financial crisis and various steps currently being taken in this regard. The experience of the financial crisis led to a call by the Group of Twenty (G-20) Finance Ministers and Central Bank Governors for the IMF and the Financial Stability Forum (FSF), the predecessor of the FSB, “to explore gaps and provide appropriate proposals for strengthening data collection.” As indicated by BIS, “the emphasis in the 1990s and early 2000s among the international community for comparability, consistency and quality of data within and across countries remains relevant.”

### **2.2 Context**

The integration of economies and markets, as evidenced by the financial crisis spreading worldwide, highlights the critical importance of relevant statistics that are timely and internally consistent as well as comparable across countries. The international community has made a great deal of progress in recent years in developing a methodologically consistent economic and financial statistics system covering traditional datasets, and in developing and implementing data transparency initiatives. While within macroeconomic (real sector, external sector, monetary and financial, and government finance) statistics, the System of National Accounts (SNA) is considered the main organizing framework, for macro-prudential statistics, an analogous structure/framework is not yet in place, but there is on-going progress in developing a consensus among data users on key concepts and indicators, including in relation to the SNA.<sup>2</sup>

While it is generally accepted that the financial crisis was not the result of a lack of proper economic and financial statistics, it exposed a significant lack of information as well as data gaps on key financial sector vulnerabilities relevant for financial stability analysis. Some of these gaps affected the dynamics of the crisis, as markets and policy makers were caught unprepared by events in areas poorly covered by existing information sources, such as those arising from exposures taken through complex instruments and off-balance sheet entities, and from the cross-border

---

<sup>2</sup>FSB, IMF: The Financial Crisis and Information Gaps (2009)

linkages of financial institutions. Broadly, there is a need to address information gaps in three main areas that are inter-related - the build-up of risk in the financial sector, cross border financial linkages and vulnerability of domestic economies to shocks.<sup>3</sup>

Further, for efforts to improve data coverage and address gaps to be effective and efficient, requires action and cooperation from individual institutions, supervisors, industry groups, central banks, statistical agencies, and international institutions. Existing reporting frameworks should be used where possible. While data gaps may be an inevitable consequence of the ongoing development of markets and institutions, these gaps are highlighted, and significant costs incurred, when a lack of timely, accurate information hinders the ability of policy makers and market participants to develop effective policy responses.

Consequently, staff of the IMF and the FSB Secretariat, in consultation with official users of economic and financial data in G-20 economies and key international organizations, identified 20 recommendations that need to be addressed. Some of these include:

- The need to strengthen the data essential for effectively capturing and monitoring the build-up of risk in the financial sector. This calls for the enhancement of data availability, both in identifying the build-up of risk in the banking sector and in improving coverage in those segments of the financial sector where the reporting of data is not well established, such as the nonbank financial corporations.
- The need to improve the data on international financial network connections. This calls for enhanced information on the financial linkages of global systemically important financial institutions (G-SIFIs), as well as the strengthening of data gathering initiatives on cross-border banking flows, investment positions, and exposures, in particular to identify activities of nonbank financial institutions.
- The need to strengthen the data needed to monitor the vulnerability of domestic economies to shocks.
- The need to promote the effective communication of official statistics to enhance awareness of the available data for policy purposes.

These recommendations were endorsed by the G20 finance ministers and central bank governors at their meeting in Scotland in November 2009.

---

<sup>3</sup> Ibid (2009)

Figure 1  
Overview of the 20 Recommendations

|   | Conceptual/statistical framework needs development  | Conceptual/statistical frameworks exist and ongoing collection needs enhancement   |
|---|---|--|
| <b>Build-up of risk in the financial sector</b>       | #3 (Tail risk in the financial system and variations in distributions of, and concentrations in, activity)<br>#4 (Aggregate leverage and maturity mismatches)<br>#6 (Structured products) | #2 (Financial Soundness Indicators (FSIs))<br>#5 (Credit default swaps)<br>#7 (Securities data)  |
| <b>Cross-border financial linkages</b>                | #8 and #9 (Global network connections and systemically important global financial institutions)<br>#13 and #14 (Financial and nonfinancial corporations' cross-border exposures)          | #10 and #11 (International Banking Statistics (IBS) and the Coordinated Portfolio Investment Survey (CPIS))<br>#12 (International Investment Position (IIP)) |
| <b>Vulnerability of domestic economies to shocks</b>  | #16 (Distributional information)  | #15 (Sectoral accounts)<br>#17 (Government finance statistics)<br>#18 (public sector debt)<br>#19 (Real estate prices)                                       |
| <b>Improving communication of official statistics</b> |   | #20 (Principal global indicators)  |

Source: FSB

### 2.3 How should the data be organized and reported?

Standardization of data reporting allows efficient aggregation of information for effective monitoring and analysis. It was therefore “important to promote the use of common reporting systems across countries, institutions, markets, and investors to enhance efficiency and transparency. Standardized reporting allows the assemblage of industry-wide data on counterparty credit risk or common exposures, thus making it possible for stakeholders to construct basic measures of common risks across firms and countries”.<sup>4</sup>

### 2.4 Who should have access to the data?

It is also acknowledged that the enhanced data collection by regulatory and/or supervisory agencies must be accompanied by a process for making data available to key stakeholders as well as the public at large. This is consistent with the “public good” nature of data, while safeguarding the confidentiality concerns of both the

<sup>4</sup> Enhancing Information on Financial Stability, Background paper prepared by Delheid Burgi-Schmelz, Alfredo Leone, Robert Heath and Andrew Kitili for Irving Fischer Committee Conference, 2010

home and host regulators and supervisors. Differences in accounting standards across countries were expected to be addressed through the legislative framework.

## **2.5 Legal Entity Identifier**

Introducing a single global system for uniquely identifying parties to financial transactions is expected to offer many benefits. There is widespread agreement among the global regulatory community and financial industry participants on the merits of establishing such a legal entity identifier (LEI) system. The system would provide a valuable 'building block' to contribute to and facilitate many financial stability objectives, including: improved risk management in regulated entities; better assessment of micro and macro prudential risks; facilitation of orderly resolution; addressing financial fraud; and enabling higher quality and accuracy of financial data overall. But despite numerous past attempts, the financial industry has not been successful in establishing a common global entity identifier and it is reported to be lagging behind many other industries in agreeing and introducing a common global approach to entity identification.

The financial crisis has provided a renewed spur to the development of a global LEI system. International regulators have recognized the importance of the LEI as a key component of necessary improvements in financial data systems. The value of strong co-operation between private sector stakeholders and the global regulatory community is widely accepted in this context.

The lack of a common, accurate and sufficiently comprehensive identification system for parties to financial transactions raises many problems. A single firm may be identified by different names or codes which an automated system may interpret as references to different firms.

The ultimate aim is to put in place a system that could deliver unique identifiers to all legal entities participating in financial markets across the globe. Each entity would be registered and assigned a unique code that would be associated with a set of reference data (e.g. basic elements such as name and address, or more complex data such as corporate hierarchical relationships). Potential users, both regulators and industry, would be granted free and open access to the LEI and to shared reference information for any entity across the globe and could build this into their internal automated systems. A high quality LEI would thus offer substantial benefits to financial firms and market participants that currently spend large amounts of money on reconciling and validating counterparty information, as well as offering major gains to risk managers and the regulatory community in relation to the identification, aggregation and pooling of risk information.<sup>5</sup>

---

<sup>5</sup>FSB(2012) - A Global Legal Entity Identifier for Financial Markets

## **2.6 BIS – Principles for effective risk data aggregation and risk reporting**

The financial crisis revealed that many banks, including global systemically important banks (G-SIBs), were unable to aggregate risk exposures and identify concentrations fully, quickly and accurately. This meant that banks' ability to take risk decisions in a timely fashion was seriously impaired with wide-ranging consequences for the banks themselves and for the stability of the financial system as a whole. The Basel Committee issued “Principles for effective risk data aggregation” in 2014 which is expected to strengthen banks' risk data aggregation capabilities and internal risk reporting practices. Implementation of the principles will strengthen risk management at banks - in particular, G-SIBs - thereby enhancing their ability to cope with stress and crisis situations.

The principles pertain to various aspects like governance, data architecture and IT infrastructure, accuracy and integrity, completeness, timeliness, adaptability, accuracy, comprehensiveness, clarity and usefulness, frequency and distribution.

The Basel Committee and the Financial Stability Board (FSB) expect banks identified as global systemically important banks (G-SIBs) to comply with the Principles by 1 January 2016. In addition, the Basel Committee strongly suggests that national supervisors also apply the Principles to banks identified as domestic systemically important banks (D-SIBs) three years after their designate on as such by their national supervisors.

A summary of principles and detailed requirements is indicated at Annex I.

## **2.7 Conclusion**

Recent years have seen significant progress in the availability and comparability of economic and financial data. However, the present crisis has thrown up new challenges that call for going beyond traditional statistical production approaches to obtain a set of timely and higher-frequency economic and financial indicators, and for enhanced cooperation among international agencies in addressing data needs. Organizational issues also need to be tackled, especially in developing common and standardized datasets on exposures of G-SIFIs. **Thus, data standardization, data quality issues and data gaps have become key focus in the aftermath of financial crisis.**





## **Chapter III- Data standards in banking/financial sector**

### **3.1 Introduction:**

This chapter is concerned with various aspects relating to current data and information standards in the financial sector used globally. Standards make modern commerce possible. Data standards allow the exchange of data between systems; aggregation of data from multiple sources; comparison of data among unrelated systems; and automation of processes for storing, reporting, and processing data. This chapter provides context for usage of standards in finance and discusses on various types of standards and Committee's recommendations are provided.

### **3.2 Context for standards in finance**

The exponential growth in computing power and the resulting proliferation of competing protocols and standards, led to a significant increase in the level of complexity required to assemble, maintain, and evolve business information systems. In 1987, John Zachman created the Zachman Framework, the first attempt at organizing the complete set of information required to manage and maintain systems planning for large organizations. The Technical Architecture Framework for Information Management (TAFIM) was first published in 1991, with an emphasis on non-proprietary, open systems architectures and subsequently emerged the Open Group Architecture Framework (TOGAF) system approach, which has significantly influenced government and defense-related Enterprise Architecture (EA) approaches.

### **3.3 Enterprise architecture framework**

MIT Centre for Information Systems Research defined EA as the specific aspects of a business that are under examination: EA is the organizing logic for business processes and IT infrastructure reflecting the integration and standardization requirements of the company's operating model. The operating model is the desired state of business process integration and business process standardization for delivering goods and services to customers. The Enterprise Architecture body of knowledge defines EA as a practice, which "analyzes areas of common activity within or between organizations, where information and other resources are exchanged to guide future states from an integrated viewpoint of strategy, business and technology.

The Zachman framework helps in terms of defining interoperability and standards for data. It provides for a structured way of viewing an enterprise. It consists of a two dimensional classification matrix based on the intersection of six communication questions (What, Where, When, Why, Who and How) with five levels of reification,

intended to transforming the most abstract ideas into more concrete ideas at the operations level.

### **3.4 Industry consortiums**

One of the fundamental requirements for information and communication technology is interconnection and interoperability. The need to develop standard interfaces and communication protocols behind which multiple companies could compete in terms of service offering spawned an entire industry of standards organizations, referred to as consortiums. There are estimated to be roughly 500 industry standards consortiums. Consortiums are often specific purpose with a specific time horizon, while some continue on and have quite a long time horizon.

### **3.5 The Internet standards organizations**

The collection of standards referred to as the Internet standards are managed by a group of related organizations like Internet Engineering Task Force (IETF), Internet Society (ISOC), Internet Architecture Board (IAB), Internet Corporation for Assigned Numbers and Names (ICANN).

### **3.6 International Standards - International Organization for Standardization (ISO)**

The International Organization for Standardization (ISO), located in Geneva, Switzerland is the premier standards body responsible for the development and management of international standards. The standards work is performed within ISO's Technical Committees, their subcommittees and working groups. The listing of various ISO standards for banking/finance is indicated at Annex II.

### **3.7 Horizontal technology standard consortiums**

Three organizations were formed to provide horizontal information technology solutions in response to technical innovations. While the World Wide Web Consortium(W3C) has mainly stayed to its core mission of supporting XML horizontal technologies, both the Organization for the Advancement of Structured Information Standards(OASIS) and the Object Management Group (OMG) have evolved since their origin.

World Wide Web Consortium (W3C) - The W3C was created to support open standards for the World Wide Web in 1994 by Tim Berners-Lee, the inventor of the World Wide Web. The major contributions from the W3C include: HTML (HyperText Markup Language); The Extensible Markup Language (XML) in 1996; the Semantic Web in 2001; and Web Services in 2002. Many of the financial data-interchange formats used widely like ISO 20022, XBRL, FpML, FIXML are based upon XML.

The Object Management Group (OMG)-The Object Management Group (OMG) was created in 1989 as part of the open systems movement. The mission of the OMG has evolved along with the industry. While initially it was responsible for creating a heterogeneous distributed object standard, some of OMG's latest initiatives are focused on the financial services industry and semantics. The Financial Domain Task Force (FDTF) is a partnership between the OMG and the EDM Council.

### **3.8 Standards for Identification**

There is also considerable variations around business practices by firms providing both standard and proprietary identifiers. A few standard financial identifiers are indicated below:

Currency codes- The Codes for Representation of Currencies and Funds (ISO 4217) defines the three character currency codes that are used to identify currencies.

Country codes- The Codes for the Representation of Names of Countries and their Subdivisions(ISO 3166) standard provides standard codes for countries.

Market identifier codes- The Codes for Exchanges and Market Identification (MIC) (ISO 10383) standard provides standard codes for markets and venues.

Identification of financial instruments -The ISIN is a standard beneath ISO TC68.The standard is administered by the Association of National Numbering Agencies. The ISIN was created in response to the globalization of markets where firms traded and held positions in securities across countries. The ISIN provides a country prefix in preceding the existing national identifiers.

Legal entities and parties- A new Legal Entity Identifier (LEI) standard has been approved as an ISO standard.

### **3.9 Standards for code lists**

Between the static structure and dynamic data are data items that change at a slower rate. These data items are at the same time data, but also have a structural role.

Genericode- Genericode is an XML based standard that provides a structure for managing code lists.

SDMX- Statistical Data and Metadata eXchange (SDMX) is a standard for exchanging statistical data. While broadly suited for most types of statistics, it was initially developed by central bankers and has been largely used for economic statistics. The original sponsoring institutions were the Bank for International Settlements, the European Central Bank, Eurostat, the International Monetary Fund (IMF), the Organization for Economic Co-operation and Development(OECD), the United Nations Statistics Division, and the World Bank. With the release of v2.1 in 2011, the information model of SDMX can be viewed as becoming a horizontal technology for managing statistical time series of any type.

XBRL-The use of the XBRL terminology for managing code lists has also been explored as an alternative standard.

### **3.10 Financial technology standards**

Data technology standards support how information is collected, aggregated and transported. An example of a standard is the extensible Business Reporting Language (XBRL), which defines methods to allow machine readable data on financial statements that can then be transported, exchanged, and stored in a consistent manner. Standards exist both for the basic data element level and as lists and messages to represent financial instruments and financial transactions. Standards also exist for the syntaxes or physical representation of the elements.

Within the financial services industry, there are multiple messaging standards being used, and internationally the Standards Coordination Group has come up with an approach that leverages and includes these standards into a broader framework without reinventing and creating redundant messages that increase implementation costs and create uncertainty or confusion for the industry.

The investment roadmap issued in September, 2010 indicates commitment of each concerned organization (FIX, FpML, SWIFT, XBRL, ISITC and FISD) to the ISO 20022 business model by laying the groundwork for defining a common underlying financial model and ensuring some level of interoperability by producing a consistent direction for utilization of messaging standards and communicating that direction clearly to the industry. The respective business processes included in the roadmap are or will be incorporated within the ISO 20022 business model and the model allows for ISO 20022 XML based messages to be created to support the business processes, while at the same time provides in certain circumstances for existing domain specific syntaxes and protocols to be maintained in order to protect the investments of market participants. The organizations have reported that they are committed to meeting on a consistent basis to ensure the roadmap continues to accurately depict the standards environment.

ISO 20022 - ISO 20022 is a methodology used by the financial industry to create message standards for a few functions. Its business modelling approach allows users and developers to represent financial business processes and underlying transactions in a formal but syntax-independent notation. As the scope of this standard covers the global financial services industry, this allows coverage across various business areas (securities, payments, foreign exchange, for example), and across asset classes. The ISO 20022 method starts with the modeling of a particular business process along with details of relationships and interactions between the actors, and the information that they must share to execute the process also are identified. The output is subsequently organized into a formal business model using

UML (Unified Modeling language). The formal business model of the process and the business information needed to support this particular process are organized into “business components” and placed into a repository. ISO 20022 describes a Metadata Repository containing descriptions of messages and business processes, and a maintenance process for the Repository Content. The Repository contains a large amount of financial services metadata that has been shared and being standardized across the industry.

The Financial Information Exchange Protocol (FIX) -The FIX Protocol was created beginning in 1992. It is reported to be a standard of the securities front office. Many instructions relating to interest, trade instructions, executions etc., can be sent using the FIX protocol. FIX supports equities, fixed income, options, futures, and FX. In 2003, FIXML was optimized to greatly reduce message size to meet the requirements for listed derivatives clearing. FIXML is reported to be widely used for reporting of derivatives positions and trades in the USA.

The eXtensible Business Reporting Language (XBRL) - XBRL, eXtensible Business Reporting Language, is an XML based data technology standard that makes it possible to “tag” business information to make it machine readable. As the business information is made machine readable through the tagging process, this information does not need to be manually entered again and can be transmitted and processed by computers and software applications. The use of XBRL has expanded into the financial transaction processing area also in recent years.

The Financial Product Markup Language (FpML) -The Financial Products Markup Language was created in response to the increased use and rapid innovation in over-the-counter financial derivatives. It uses the XML syntax and was specifically developed to describe the often complicated contracts that form the base of financial derivative products. It is widely used between broker-dealers and other securities industry players to exchange information on Swaps, CDOs, etc.

ISO 20022 and XBRL- In June 2009, SWIFT, DTCC and XBRL US commenced an initiative to develop a corporate actions taxonomy using XBRL. Each of the elements in the corporate actions taxonomy corresponds to a message element in the ISO 20022 Corporate Actions Notification message. This initiative attempts to bring together a XBRL standard with the standard used by financial intermediaries to announce and process corporate actions events.

ISO 8583 – It is used for almost all credit and debit card transactions, including ATMs. This type of messages are exchanged daily between issuing and acquiring banks.

### **3.11 Reference data standards**

The 2008 financial crisis brought to fore the fact that this is one of the most neglected areas in the financial services industry. In respect of investment area, reference data define the financial instruments that are used in the financial markets.

1. Market Data Definition Language (MDDL)- MDDL was created to provide for a comprehensive reference data model as an XML based interchange format and data dictionary for financial instruments, corporate events, and market related, economic, and industrial indicators, which started in 2001 by the Software and Information Industry Association's Financial Information Services Division. The direct adoption of MDDL is very limited though it is reported that MDDL's value as a reference model continues.

2. OpenMDDDB - FIX Protocol Ltd. and FISD jointly developed a relational database model for reference data derived from MDDL in 2009. It also provides support for maintenance and distribution of reference data using FIX messages. The EDM Council, FISD, and FIX had entered into an agreement for EDM Council to facilitate and support evolution of the Open MDDDB by the community.

### **3.12 Standards for representing business model**

#### **1. ISO 20022**

As a result of the integration of ISO 19312 into ISO 20022, the financial messaging standard was expanded to be a business model of reference data for the financial markets in addition to the messages.

#### **2. Financial Information Business Ontology (FIBO)**

The EDM Council and the OMG created a joint working group, the Financial Data Task Force, "to accelerate the development of 'sustainable, data standards and model driven' approach to regulatory compliance. "The initiative focused on semantics instead of traditional modeling techniques.

### **3.13 Recommendations on data standards**

- 1) Regulatory data reporting standards - The data exchange standards need to be based on open standards and allow for standardization of data elements and minimizing data duplication/redundancy. RBI had already embarked on XBRL(which is based on XML platform) as the standard platform for a set of regulatory returns which may be continued for rest of returns or data elements. Thus, XBRL may be the reporting standard for all regulatory reporting of structured data.
- 2) The ISO 20022 standard is recommended to be the messaging standards for the critical payment systems. RTGS in India already uses the ISO 20022 message formats.
- 3) The well known Statistical Data and Metadata eXchange (SDMX) is recommended as a standard for exchanging statistical data.

- 4) In regard to standardisation of coding structures, in accordance with international standards, the ISO 4217 currency codes and ISO 3166 country codes can be used.
- 5) System of National Accounts (SNA) 2008 can be used for classification of institutional categories and International Standard Industrial Classification (ISIC)/ National Industrial Classification (NIC) codes for economic activity being financed by a loan may be incorporated.
- 6) In order to acquire single view of transactions in respect of a customer, unique customer id is allotted by individual banks. Given that no single identifier can represent all categories of customers of banks, the differentiation may need to be made by mapping with the identifiers presently available. Recently, Clearing Corporation of India Limited (CCIL) has been selected to act as a Local Operating Unit in India, for issuing globally compatible unique identity codes (named as legal entity identifier or LEI) to entities which are parties to a financial transaction in India. Given the LEI initiative, efforts to facilitate LEI for legal entities involved in financial transactions across financial system needs to be expedited to maximise coverage over the medium term.
- 7) Given the complexity of some corporate entity with numerous subsidiaries including step down subsidiaries, there is a need for usage of LEI or similar methodology to link the complex hierarchy of any corporate may be considered to facilitate ease of identification of total credit exposure of corporate groups. While it is reported that LEI application of CCIL has provision for the same, the utility may need to effectively leveraged to map the corporate group hierarchy.
- 8) While presently LEI caters to legal entities involved in financial transactions, ultimately LEI or similar system needs to be made broad-based to incorporate other categories of customers like partnership firms and individuals.
- 9) For conduct of electronic transactions and reporting purposes in financial markets, well known international standards like ISO based standards can be considered where possible.
- 10) In order to take up data element/return standardisation through standardising or harmonising definitions, efforts of earlier working groups (Committee on rationalisation of returns and Committee on harmonisation of banking statistics) can be consolidated by setting up an inter-departmental project group within RBI which can work in a project mode so as to ensure comprehensive and effective implementation of standardisation and consistency of data element definitions across complete universe of returns/data requirements of RBI.





## **Chapter IV – Issues in data management and data quality in banks**

### **4.1 Introduction**

With the advent of technology in banking, huge volume of data is produced and stored digitally. The transformation of banking in the form of anywhere/virtual banking has also resulted in increased information availability which necessitates banks to implement robust information management processes to facilitate effective Decision Support system. The ability of organizations to capture, manage, preserve and deliver the right information at the right time to the required personnel is one of the key success factors of Information Management. This chapter highlights various issues relating to data quality and challenges faced by banks in regard to data management and data quality.

### **4.1 Key issues relating to information management**

KPMG India's Information Enabled Banking Survey (IEB) was conducted amongst select 10 private sector banks in South India during 2013. IEB was aimed to provide an overview of Information Management landscape across small and medium private sector banks.

The various key issues highlighted in the survey included:

#### **Data Quality**

Only 33% of the respondents use automation in their monthly report generation process and 44% have minimal manual intervention so as to reduce Data Quality issues. Banks tend to collect information across multiple locations and multiple formats, thus potentially creating non standardized data. 67% of the respondents claimed to have clean and standardized data across systems.

#### **Standardization**

Standardization across business eliminates data duplication, data redundancy and cost associated with resolving these issues. All surveyed banks have standardized reporting formats at the local branch, region, zone and at the corporate level. Banks faced challenges in standardizing the report generation methodology despite having a standardized reporting form and clearly defined output.

Banks use different technologies and databases to capture and store information. One of the key issues they may need to focus on is to address 'multiple of truth' scenarios to availability of 'single version of truth'. Key observation was that standardization is needed at Extract, Transform and Load (ETL) process which would result in process efficiency, reduced manual intervention and reduced costs.

### 4.3 Data standards -Common issues in data and MIS in commercial banks

Generally, various data quality and MIS related issues observed by RBI include:

- 1) Incomplete information and issues relating to fields in the core systems
  - (i) Identifiers Missing or Invalid
  - (ii) Incomplete or incorrect format of contact details like address, PIN code, contact numbers etc
  - (iii) Date of Birth Missing
  - (iv) Consumer Name Invalid/incomplete
  - (v) Ownership Indicator Invalid
  - (vi) Date of Birth Invalid
  - (vii) Ownership Indicator Missing
  - (viii) Invalid status of accounts
  - (ix) Non updation of accounts
  - (x) Non closure of accounts despite amount overdue and current balance being zero
  - (xi) Incorrect entry of values/measures in respect of deposit or loan accounts
- 2) Issues with configuration of business products in the system
- 3) Incorrect activity or sector codes
- 4) No fields for capturing certain key fields which are either maintained manually or entered in ad-hoc or generic fields in the system
- 5) Data captured in electronic form that is not controlled (e.g. excel sheets or other desktop tools)
- 6) Though CBS was available, manual compilation of data from the branches
- 7) STP or near real time interface unavailable between various business systems and accounting systems
- 8) Incomplete master data and reference data
- 9) Deficiencies in mapping of data for assessing risks like liquidity risks

All these factors could potentially impact business aspects like capital management and capital ratios, asset quality monitoring, funds and liquidity management ultimately impacting effective risk management. All these aspects indicate need for enhanced processes and procedures for data and information management in the bank and need for robust and standardized metadata.

Committee identified various data related aspects relating in specific to credit area that would need to be addressed to facilitate standardization across banking system. These are detailed at Annex III.

#### 4.4 Key Drivers and challenges for the banking system

The various new key drivers and challenges in the new milieu in banking include the following:

- (i) The regulatory environment in which banks in India are functioning is undergoing a paradigm shift. Apart from the basic approaches for handling major risk categories, Basel II further entails progressive advancement to sophisticated but complex risk measurement and management approaches to credit, market and operational risks depending on the size, sophistication and complexity of the respective banks. Some of the banks have applied to Reserve Bank of India for moving to Advanced Approaches of calculating Pillar I capital.
- (ii) In addition, Pillar 2 and Pillar 3 of Basel II emphasize the need for developing better risk management techniques in monitoring and managing risks not adequately covered or quantifiable under Pillar 1 and increased disclosure requirements. The banks are required to carry out Internal Capital Adequacy Assessment Process which comprises a bank's procedures and measures designed to ensure appropriate identification and measurement of all risks to which it is exposed, an appropriate level of internal capital in relation to the bank's risk profile and an application and further enhancement of risk management systems in the bank.
- (iii) Basel III Capital Regulations has commenced in India from April 1, 2013 and would be fully implemented as on March 31, 2019. There are various direct and related components of the Basel III framework like increasing quality and quantity of capital, enhancing liquidity risk management framework, leverage ratio, incentives for banks to clear standardised OTC derivatives contracts through qualified central counterparties, regulatory prescription for Domestic Systemically Important Banks and Countercyclical Capital buffer (CCCB) framework.
- (iv) The growing emphasis on fair treatment to customers calls for moving over from "Caveat Emptor" (Let the Buyer beware) to the principle "Caveat Venditor" (Let the seller beware) and focus on comprehensive consumer protection framework in financial sector in India.
- (v) Globally heightened regulatory requirements in respect of KYC / AML practices to prevent banks from being used, intentionally or unintentionally, by criminal elements for money laundering or terrorist financing activities.
- (vi) Extensive leverage of technology for internal processes and external delivery of services to customers requiring robust IT governance and Information security governance framework and processes in banks.
- (vii) In the background of growing volume of non - performing assets and restructured assets causing concern for the financial as well as the real sector in India, a framework for revitalizing distressed assets in the economy has been implemented with effect from April 1, 2014. The Framework lays down guidelines

for early recognition of financial distress, information sharing among lenders and co-ordinated steps for prompt resolution and fair recovery for lenders.

(viii) Impending developments in regulatory policies and economic environment are likely to result in banks facing a far more competitive environment in the coming years. As banks' customers – both businesses and individuals - become global, banks will also need to keep pace with the customer demands and develop global ambitions. The challenge for banks will be to develop new products and delivery channels that meet the evolving needs and expectations of its customers.

**Thus, there is a need for effective information management practices and robust MIS. This calls for a robust data governance framework in banks.**

## **Chapter V – Data governance architecture in banks**

### **5.1 Introduction**

The issues highlighted in previous chapter call for a robust data governance or information management architecture in banks. Specific focus need to be accorded to data quality through various mission mode strategies and as an ongoing exercise. This chapter delineates the recommendations of the committee on key components of robust data governance architecture in banks and key practices to address data quality issues.

### **5.2 Data governance**

Bob Seiner states in his book Non-Invasive Data Governance, that Data Governance is the formal execution and enforcement of authority over the management of data and data related assets. In this case, data governance refers to administering or formalizing, discipline(behavior) around management of data.

According to the Data Governance Institute, a well-defined and complete data governance solution is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, and using what methods.

There are other views or definitions of data governance. Thus, data governance broadly refers to the policies, standards, guidelines, business rules, organizational structures in respect of data related processes performed in an organization.

### **5.3 SSG Risk Appetite Framework and IT Infrastructure**

In the aftermath of financial crisis, the Senior Supervisors Group had indicated following key pre-requisites for implementing comprehensive risk data infrastructure:

#### **5.3.1 The Importance of IT Governance in Strategic Planning and Decision Making**

- (i) Strategic planning processes need to include an assessment of risk data requirements and system gaps.
- (ii) Firms with leading, highly developed IT infrastructures bring together senior IT governance functions, business line units, and IT personnel to formulate strategy.

- (iii) Firms successful in aligning IT strategies with the needs of business line managers and risk management functions have strong project management offices (PMOs) to ensure that timelines and deliverables are met.
- (iv) Firms with effective IT project implementation appoint a data administrator and a data owner with responsibility and accountability for data accuracy, integrity, and availability.
- (v) Firms with high-performing IT infrastructures ensure that the Board committees institute internal audit programs, as appropriate, to provide for periodic reviews of data maintenance processes and functions.

### **5.3.2 Automating Risk Data Aggregation Capabilities**

- (i) Supervisors observed that while many firms have devoted significant resources to infrastructure, very few can quickly aggregate risk data without a substantial amount of manual intervention.
- (ii) Firms with leading practices have very limited reliance on manual intervention and manual data manipulation.
- (iii) Supervisors have observed that an inability to aggregate risk data in an accurate, timely, or comprehensive manner can undermine the overall value of internal risk reporting.
- (iv) Consolidated platforms and data warehouses that employ common taxonomies permit rapid and relatively seamless data transfer, greatly facilitating a firm-wide view of risk.
- (v) Leading firms implement data aggregation processes covering all relevant transactional and accounting systems and data repositories to maintain comprehensive coverage of MIS reporting.
- (vi) Leading firms' MIS practices also include periodic reconciliation between risk and financial data.

### **5.4 Key components of data governance architecture – Committee Recommendations**

Committee recommends that key components of data governance architecture in banks may incorporate focus on the following:

- (i) Formulation of Data governance or information management policy with emphasis on various aspects like data governance organisational structure, data ownership, definition of roles and responsibilities, implementation of data governance processes and procedures at individual functions/departments, development of a common understanding of data, data quality management, data dissemination policy and management of data governance through metrics and measurements.

- (ii) Overall oversight of data governance may be with the Audit Committee of Board of a bank or a specific Committee nominated by the Board.
- (iii) Formation of executive level Data Governance Committee or entrusting responsibility to existing information management committee if already existing. Data governance responsibilities and accountabilities should be clear, measured and managed to ensure sustained benefit to the bank.
- (iv) Data Ownership related aspects to be considered include holding overall responsibility for data, assigning ownership of key data elements to data controllers or stewards, assigning data element quality within business areas, implementing data quality monitoring and controls, providing data quality update to management/data governance committee and providing data quality feedback to business data owners.
- (v) 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. Ideally, data ownership needs to be primarily based on the business function.
- (vi) Platforms and data warehouse/s need to employ common taxonomies/data definitions/meta data. 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 need to be linked across the applications using pre-defined rules. The rules to be applied for synchronization of metadata also need to be defined.
- (vii) The metadata (data definitions) may be synchronized across various source systems and also with the RBI definitions for regulatory reporting.
- (viii) To help drive data governance success, measurements and metrics may be put in place which define and structure data quality expectations across the bank for which various data governance metrics and measures would be required. Data needs to be monitored, measured, and reported across various stages: data acquisition, data integration, data presentation and data models, dictionaries, reference data and metadata repositories.
- (ix) Internal audit function to provide for periodic reviews of data governance processes and functions and report on the issues to ACB.
- (x) Detecting and correcting the faulty data manually is very tedious and time consuming. It is in this context the validation methods based on statistics, machine learning and pattern recognition gain importance. Many DBMS, DWDM product vendors now offer Data Profiling, Data Quality, Master Data Management services. Banks can take advantage of all these tools and techniques to keep their data clean.

- (xi) Various focussed data quality assessment and improvement projects need to be undertaken through various methodologies covering data profiling, data cleansing and data monitoring.
- 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. Implement solutions that address the root causes of the data quality problems. Monitor and verify the improvements that were implemented. Maintain improved results by standardizing, documenting, and continuously monitoring successful improvements
- (xii) Banks can also endeavour to establish a centralised analytics team as a centre of excellence in pattern recognition technology and artificial intelligence (AI) to provide cutting edge analysis and database tools or information management tools to support business decisions.

## **5.5 Other related recommendations on data governance**

- 1) Apart from providing enhanced focus during AFI/RBS, data governance mechanisms in banks may also be examined intensively through focussed thematic reviews by DBS of RBI. Based on outcome of thematic reviews, detailed guidance may be issued to banks to address issues identified during review.
- 2) Banks, in particular domestic SIBs, may also be advised to keep in context BIS document "Principles for effective risk data aggregation and risk reporting" as part of their information management process.
- 3) While specifying key regulations, RBI may also endeavour to specify any key system related validation parameters and details of data quality dimensions expected from concerned regulated entities.



- 4) Guidance on Best practices on data governance and information management can be formulated by IDRBT.
- 5) RBI may facilitate creation of Data Governance Forum under the aegis of IBA or learning institutions like CAFRAL or NIBM with other stakeholders like IDRBT, RBI, IBA, banking industry technology consortiums, banks, to assist in development of common taxonomies/data definitions/meta data for banking system.
- 6) Bank Technology Consortiums under the aegis of IDRBT and other stakeholders like banks can validate critical banking applications like CBS and provide guidance on expected minimum key information requirements/validation rules and address to the extent possible different customizations across banks.
- 7) Committee also identified certain key data aspects relating to credit function that would need to be addressed by banks to facilitate standardization and data comparability across banks.



## Chapter VI - Data standardization in regulatory reporting – Commercial banks, NBFCs and UCBs

### **6.1 Introduction**

This chapter delineates the efforts of earlier initiatives, recommendations of various Committees and provides major recommendations in enabling standardization of data pertaining to commercial banks, NBFCs and UCBs submitted to RBI.

### **6.2 Commercial Banks**

The Reserve Bank of India (RBI) collects large volume of data from banks for its key functions like monetary policy formulation, supervision, regulation and promoting research. Around 220 returns submitted by Scheduled Commercial Banks (SCBs) excluding Regional Rural Banks, the returns submitted to the Department of Banking Supervision (DBS), Department of Banking Operations and Development (DBOD), Department of Statistics & Information Management (DSIM) and Monetary Policy Department (MPD) cover most of the data elements reported by the banks. This pool of banking data spans over various dimensions and granularities with different structures, formats, naming conventions, levels of aggregation and frequencies.

However, at times, these data lack internal consistency which could potentially impact their utility as policy input. Data mismatches can occur due to reasons like different reporting periods, definitional issues, non-uniform / inadequate classifications and coding structures, lack of unified instructions to banks, and also methodology of compilation in banks. Further, similarities in data elements in multiple returns gave rise to problem of inconsistency. Therefore, need was felt to develop harmonised and integrated system of reporting banking data to the RBI by re-examining the entire gamut of the definitions, classification and coding structure of data.

The data/information required for supervision can be classified into two groups based on its source, viz. (i) submitted by banks, (ii) generated/compiled by the supervisor. Further, the form of these data can be classified as either structured (e.g. numeric/ textual) or unstructured (e.g., documents/files). Examples of various data/information types used for supervision are given below:

| Table 1: Type of Information Used for Banking Supervision |                         |   |   |
|---|-------------------------|---|---|
| Data Type   | Submitted by Bank       | Generated/Compiled by Supervisor  |   |
| Structured  | Numerical/<br>Financial | 1) DSB Returns (XBRL)<br>2) Fraud Returns<br>3) FID Returns<br>4) RBS Risk data (Data Collector application)<br>5) Financial Conglomerate Return (Excel)<br>6) Ad hoc data (Data Collector application) | 7) Standard Annexes as part of onsite inspection<br>8) Assessment of key financials/capital including validation/re-assessment of RBS risk data furnished by bank<br>9) Scores for aggregations of various risks as part of IRISc model<br>10) Thematic/Sector/Industry/other bank-wide studies |
|   | Textual                 | 11) RBS Control gap information (Data Collector application)<br>12) RBS Compliance information (Data Collector application)   | 13) Comments/additional information on Control gap and Compliance by SSM<br>14) Comments by Quality Assurance Division  |
| Unstructured  |                         | 15) Annual Reports<br>16) Policy Documents<br>17) Board Minutes<br>18) Reports of External Auditors   | 19) Working documents for supervisory assessment<br>20) Supervisory Reports<br>21) BFS Reports<br>22) Communications to banks   |

Source: Report of the Committee on Data and Information Management in the Reserve Bank of India (2014), RBI

(1) Rationalisation of returns - Presently, data collection in RBI is done on a decentralized basis. Various departments in RBI have prescribed fixed-format returns for specific purposes. While each of the above returns has some distinct features, there are some common data elements among them. Several attempts were made earlier to rationalize return submission. In 1999, DSIM undertook an exercise in which out of the 286 returns in existence, 76 returns were proposed to be discontinued. As a follow up, 39 returns could finally be discontinued. In August 2008, as part of the implementation of XBRL (**eXtensible Business Reporting Language**) based data reporting, returns

were further rationalised and the number of returns to be submitted by Scheduled Commercial Banks (excluding RRBs) was brought down to 223. In order to streamline the process, initiative has been taken in the recent past to store data received through XBRL platform in a centralized database which can be accessed by multiple users.

- (2) Internal Group on harmonisation of banking statistics had after examining the contents of 84 returns on banking data received by the member departments, based on the commonality in data items, 19 returns were identified for harmonization exercise assigned to the Group. To know the extent of data mismatch, the Group studied data from six such returns viz., Form A, Form VIII, Form X, ALE, BSA, and Annual Accounts of banks as on March 31, 2013 of select banks. Based on the same, the Group gave its observations with respect to items which require harmonization in respect of certain asset and liability items. It also provided common definitions for key information blocks.

### **6.3 NBFCs**

Unlike commercial banks, deposits form a very small component of the overall liability of NBFCs as they predominantly rely on institutional sources including bank borrowings and capital/ money markets for their funding requirements. Risk to financial stability from the sector emanates from these inter-linkages between NBFCs and other financial intermediaries and their funding dependencies. Accordingly, the regulatory guidelines are tuned towards discouraging a higher degree of leverage and having adequate capital buffers so as to ensure that any stress on their balance sheets is absorbed rather than transmitted to the financial system.

The total number of NBFCs as on March 31, 2014 are 12,029 of which deposit taking NBFCs are 241 and non-deposit taking NBFCs with asset size of ₹100 crore and above are 465, non-deposit taking NBFCs with asset size between ₹50 crore and ₹100 crore are 314 and those with asset size less than ₹50 crore are 11009. NBFCs-ND with assets of ₹1 billion and above had been classified as Systemically Important Non-Deposit accepting NBFCs (NBFCs-ND-SI) since April 1, 2007 and prudential regulations such as capital adequacy requirements and exposure norms along with reporting requirements were made applicable to them. From the standpoint of financial stability, this segment of NBFCs assumes importance given that it holds linkages with the rest of the financial system.

NBFCs are required to submit various returns to RBI with respect to their deposit acceptance, prudential norms compliance, ALM etc. Detailed instructions regarding

submission of returns by NBFCs have been issued through various company circulars. A list of such returns to be submitted by NBFCs-D, NBFCs-ND-SI and others is as under:

A. Returns to be submitted by deposit taking NBFCs

1. NBS-1 Quarterly Returns on deposits in First Schedule
2. NBS-2 Quarterly return on Prudential Norms is required to be submitted by NBFC accepting public deposits
3. NBS-3 Quarterly return on Liquid Assets by deposit taking NBFC
4. NBS-4 Annual return of critical parameters by a rejected company holding public deposits
5. NBS-6 Monthly return on exposure to capital market by deposit taking NBFC with total assets of ₹100 crore and above
6. Half-yearly ALM return by NBFC holding public deposits of more than ₹20 crore or asset size of more than ₹100 crore
7. Audited Balance sheet and Auditor's Report by NBFC accepting public deposits
8. Branch Info Return

B. Returns to be submitted by NBFCs-ND-SI

9. NBS-7A Quarterly statement of capital funds, risk weighted assets, risk asset ratio etc., for NBFC-ND-SI.
10. Monthly Return on Important Financial Parameters of NBFCs-ND-SI
11. ALM returns:
  - (i) Statement of short term dynamic liquidity in format ALM [NBS-ALM1] - Monthly,
  - (ii) Statement of structural liquidity in format ALM [NBS-ALM2] Half yearly,
  - (iii) Statement of Interest Rate Sensitivity in format ALM -[NBS-ALM3], Half yearly.
12. Branch Info return

C. Quarterly return on important financial parameters of non deposit taking NBFCs having assets of more than ₹50 crore and above but less than ₹ 100 crore

13. Basic information like name of the company, address, NOF, profit / loss during the last three years has to be submitted quarterly by non-deposit taking NBFCs with asset size between ₹50 crore and ₹100 crore.

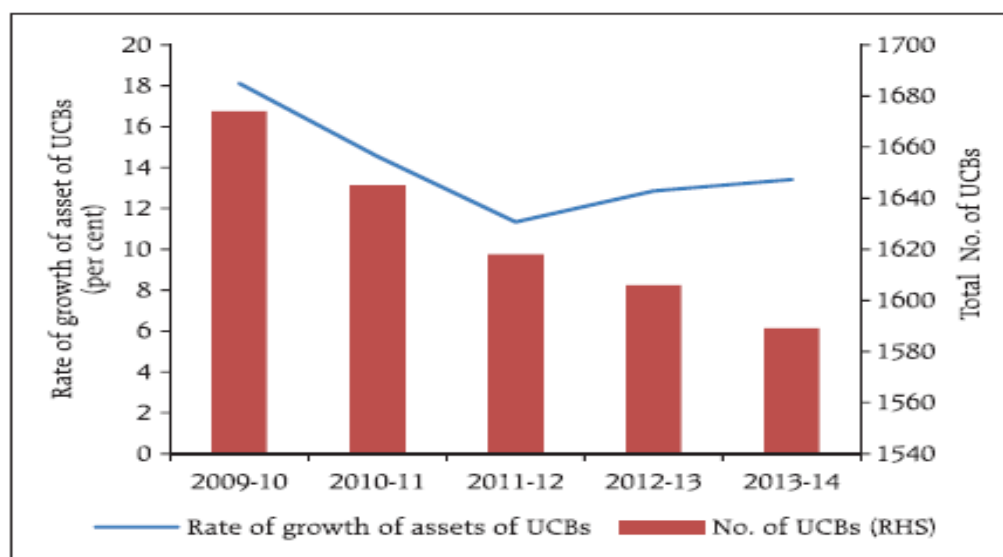
## D. Other Returns

14. With regard to overseas investment a Quarterly Return is to be submitted by all NBFs to the Regional Office of DNBS and also Department of Statistics and Information Management (DSIM)

### 6.4 Urban co-operative banks

2.21 The regulation and supervision of Urban Co-operative Banks (UCBs) is vested with the Reserve Bank in respect of their banking activities. UCBs are primarily classified as scheduled or non-scheduled. Following consolidation, the number of UCBs came down marginally to 1,589 in 2013-14 from over 1,600 a year ago.

2.22 The scheduled UCBs filed around 40 returns. In the UCB sector, about 80 per cent of the assets are held by Tier II UCBs, accounting for only about one-fourth of the total number of UCBs.



Source: RBI supervisory returns.

### 6.5 Recommendations of the Committee:

- 1) XBRL platform may be gradually expanded across the full set of regulatory returns.
- 2) Robust internal governance structure needs to be set up in regulatory entities with clear responsibilities and accountabilities to ensure correct, complete and timely submission of regulatory/supervisory returns.
- 3) Regulatory reporting - Commercial Banks:

- a. Adoption of uniform codes among different returns of RBI will reduce inconsistency among returns. For eg. DSIM of RBI collects industry classification of credit as per NIC codes while other departments use different classification. Each bank has adopted its own approach to map NIC codes. Analysis of mappings of some banks showed apparent divergences. Hence, as part of data standardisation, data collected by other returns may also be brought in alignment with the usage of common or standardised codes used in BSR
  - b. The BSR codes need to be updated based on latest NIC 2008 classification. The BSR codes may be reviewed periodically and updated. Further, should be possible to establish one-to-one mapping of sector/ industry codes in various other regulatory returns from the same.
  - c. The nature of returns are generally dimensional in nature, consisting of various components like measures, concepts , elements, attributes, dimensions and distributions. A suitable data model may be generated to facilitate element-based, simplified and standardised process data collection process by RBI under a generic model structure that is suitable for both primary and secondary data.
  - d. There is a need to ultimately move over to “data” centric approach from the current “form” centric approach. Under a data centric approach methodology, any data point must be expressed by its “primary element” and all additional dimensions necessary to their identification. As “form” centric approach is oriented to the visualization of the data in certain format, it may be used for reviewing .
  - e. The values of various attributes and dimensions should be standardised to enable the collation of data from different domains.
  - f. Suitable data sets with varied nature like hierarchical, distributional or dimensional can be created to facilitate submission of data in summarised or granular form as the case may be from the central repository of the banks.
    - a. A good feedback mechanism from banks to RBI and vice versa can help maintain uniqueness in data definitions. As and when multiple data definitions from XBRL taxonomy given by RBI map to same data element in any bank, they need to be flagged to RBI.
    - b. Phased implementation of various standardised data definitions can be commenced based on elements which were already standardised.
- 4) NBFCs/UCBs:
- a. Rationalisation of returns needs to be attempted for NBFCs and UCBs. An exercise carried out earlier indicated significant duplication in the information provided through various returns. The Committee recommends that these returns may be rationalised by identifying



major data elements and removing duplicate data elements. (A sub-group constituted by the earlier Mohanty Committee had identified a number of elements that should form the basis for rationalisation.)

- b. The Committee recommends an online data collection mechanism for larger NBFCs and Tier II UCBs.
- c. In due course, after rationalisation exercise, data element based return submission may also be initiated.
- d. Suitable data model and robust meta data system may be developed.



## **Chapter VII – ADF Project – Implementation by banks**

### **7.1 Introduction**

The Reserve Bank of India had issued an Approach Paper in 2010 on Automated Data Flow (ADF) that outlines the Guiding Principles, Assessment Framework by Banks, Common-end-State, Benefits of Automation, Approach to be adopted by Banks and the Proposed Roadmap for Implementation of ADF. This chapter reviews the current status on ADF implementation in commercial banks and provides recommendations in this regard.

The Reserve Bank's Approach Paper on ADF (RBI, 2010) envisaged that banks would prepare a central repository that would contain all the data elements required for reporting to the Reserve Bank. Banks that have followed this vision in developing their central repositories will find it easier to migrate to the element-based data reporting paradigm. Banks that have followed a return-based approach would require some changes to their systems.

### **7.2 ADF- Implementation Strategy & Approach by banks**

#### ***Strategy***

In terms of the Strategy and as per the guidance on the Approach Paper under ADF, the Banks have set-up the Returns Governance Group (RGG) for monitoring the ADF implementation and ensuring regular submission of the regulatory returns through the Centralized Data Repository (CDR). In terms of the approach paper referred to above, the RGG consist members from the IT Department and Compliance Department and some of the key business groups that submit important regulatory returns.

#### ***Approach***

In view of the guidelines set-out in the Approach Paper, the banks have set-up processes to ensure compliance under ADF and the approach generally adopted by the banks is as follows:

- The banks have referred the Approach Paper along with the list of 223 returns shared by RBI and compiled the list of applicable returns.
- Integrated the data residing in multiple source systems and mapped to a single Centralized Data Repository (CDR) Platform
- Ensured usage of most appropriate source of data for implementation of ADF

- The coding of data has been done as per the RBI prescribed guidelines e.g. BSR Codes, PSL Coding etc.
- The data is extracted (ETL) at periodic intervals and the return logic is defined around the extracted data input in the CDR
- The unorganized, uncontrolled and non-integrated electronic source data is being uploaded under ADF through Gap Data Screens with Maker & Checker Control
- Return that needs a narrative free text are stored in CDR and uploaded through GAP Data Screens with Maker & Checker Control
- Definition of process to handle Change Management, Data Governance, Data Archival and Data Retention
- Definition of Roles & Responsibilities of respective departments (under ADF) with regards to the following process elements:
  - Data Acquisition
  - Data Integration & Storage
  - Data Conversion
  - Data Submission
  - Random Checking

The banks are in various stages of implementation of the ADF project.

### **7.3 ADF implementation by banks - Recommendations of the Committee**

- 7) **Use of ADF for Internal MIS** - The RBI Approach Paper highlighted usage of ADF platform for generating internal MIS as one of the key benefits of ADF. In this regard, banks may also explore using the platform for generating internal MIS and other uses. Indicatively some aspects include :
  - NPA Management Automation Module
  - Automation of SLBC Returns through the same platform
- 8) Detailed survey can be carried out by RBI to ascertain the status of ADF implementation by banks. Feedback may also be obtained from DBS regarding any issues relating to ADF implementation obtained during AFI/RBS examination process. Any manual intervention from source systems to ADF central repository needs to be ascertained. Independent assurance on the ADF central repository mechanism in individual banks may also be verified. This would enable assessment of the quality and comprehensiveness of ADF implementation by individual banks. Any specific issues may be taken up with concerned banks for remediation.
- 9) Banks may also evaluate and take steps to enable the ADF Platform to cater to the Risk Based Supervision(RBS) data requirements by suitably mapping the RBS data point requirements. Thus, the ADF structure should be made use of and aligned to the RBS set-up so that synergies can be built-in, data

quality and consistency can be enabled and the overall system can be made more efficient.

- 10) Existing ADF platform needs to be leveraged by prescribing the necessary granular data fields to be captured by banks to achieve consistency and uniformity in regulatory reporting.
- 11) Banks may also port the necessary details required by RBI under Guidelines on “Framework for Revitalizing Distressed Assets in the Economy - Guidelines on Joint Lenders' Forum (JLF) and Corrective Action Plan (CAP)” in February, 2014 under ADF central repository platform.
- 12) Depending on the requirement of RBI regarding granularity of data, ADF system needs to be suitably updated to provide for the requisite granular data fields at the central repository level. The ADF system of the banks should be designed flexibly to accommodate any anticipated changes in the format of return, i.e., addition and deletion of data elements.



## **Chapter VIII– XBRL project of RBI – status and way forward**

### **8.1 Introduction**

XBRL related basic information is provided at Annex V and various worldwide project on XBRL are indicated at Annex VI. In this chapter, the various aspects are covered relating to the XBRL project like the scope and coverage, assessment of schema and data elements, coordination among SEBI, MCA and other agencies, impact of changes in XBRL scheme on systems in RBI and banks, NBFC XBRL schema of MCA and the need for a data governance group in RBI particularly for XBRL implementation.

### **8.2 Status of return submission and XBRL Development in RBI**

eXtensible Business Reporting Language (XBRL) is an international standard for business reporting. Within the Reserve Bank, XBRL has been viewed as a natural evolution of the existing Online Returns Filing System (ORFS). While ORFS does the job of data capturing and transmission of returns from the banks to the Reserve Bank, it incorporates no in-built standardization. The developments in respect of XBRL are as follows:

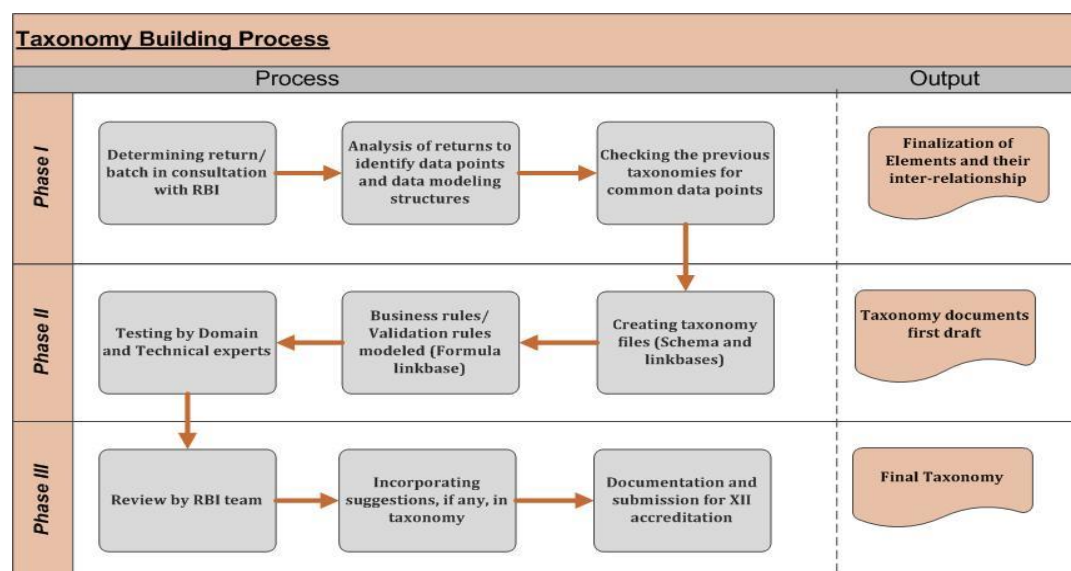
- XBRL enables standardization and rationalization of elements of different returns using internationally recognized best practices in electronic transmission. In the process, XBRL also facilitates rationalization of number of returns to be submitted by the banks, thus reducing the reporting burden on banks. The Reserve Bank **could bring down the number of returns from 291 to 225** (vide RBI press release dated August 14, 2008 and December 17, 2008.)
- The **XBRL phase I** started with the regulatory set of returns, Returns on Capital Adequacy (RCA2) being the first set of returns to be implemented under the XBRL system, followed by the implementation of a statutory return on liquidity viz., Form 'A' (under Sec 42(2) of RBI Act 1934). Gap, Position and cash Balances (GPB) of the commercial banks, a high frequency daily return, was also implemented under XBRL submission. The targeted returns under phase I project also included the implementation of the financial statements of the banks, in addition to the 6 other returns. The banks have been submitting financial statements through XBRL from the financial year 2012-13.
- A High Level Steering Committee (HLSC) headed by Deputy Governor, RBI, is constituted, to oversee the implementation of XBRL project in RBI. The

HLSC had decided that all the reporting will take place in on-line mode through XBRL over the medium term.

- Out of the nearly 280 different reporting forms for the banks and non-bank entities, 58 have been taken up in the Phase II of the XBRL project (from Department of Banking Supervision, Foreign Exchange Department, Urban Banks Department, Department of Statistics and Information Management, and International Debt Management Department). Out of the 58 returns, 50 returns have been deployed for UAT presently.
- In XBRL phase II, data is moved to data warehouse for analysis purpose.
- The taxonomy developed is based on the international standard prescribed by XBRL international, an international body responsible for the implementation of XBRL worldwide
  - XBRL 2.1 and Dimension 1.0 specification
  - Following Financial Reporting Taxonomy Architecture (FRTA) Rules.
- The Ministry of Corporate Affairs has implemented XBRL based submission of financial statements of non-financial companies from the financial year 2010-11 onwards.

### 8.3 Assessment of data elements included in the schema

Following is the flow chart which depicts the taxonomy development process for RBI



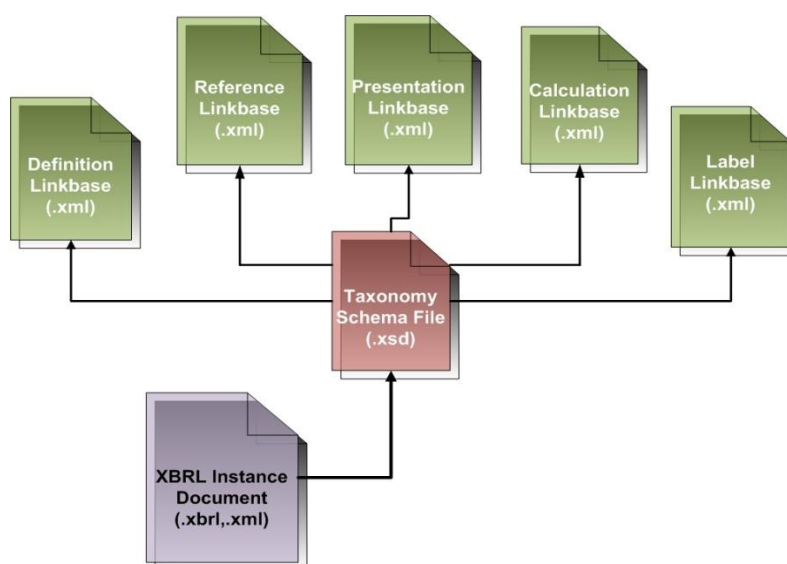
Currently, taxonomy development is done for each return separately. Analysis of reporting elements is more return specific. Though the element identification process is return specific, a single core taxonomy is prepared for all the returns under XBRL. The core taxonomy contains unduplicated list of all the elements across the returns.



The XBRL taxonomy has the facility to incorporate data definitions also. The XBRL taxonomy has reference link base, where presently the reference to the related circular is given. The data definitions can also be incorporated in the reference link base, however, due to various issues including the time involved, the required data definitions have not been provided by RBI. There is scope for further improvements in this regard.

#### 8.4 Data standards implemented by XBRL

The schema file is associated with the Definition, Calculation/Formula, Presentation, Label and Reference linkbases for the defined taxonomy. The diagram below defines the structure of the taxonomy.



**Taxonomy Structure Diagram**

The basic details on XBRL structure is elaborated as part of Annex V.

For Basel II related taxonomy of RBI, the salient features include the following:

- Detailed templates for
  - Credit Risk– on and off BS items, securitization, Market and non-related off BS, failed transactions, counterparty credit risk
  - Market Risk – AFS, HFT, aggregate
  - Operational Risk – BIA template
- Data Model – Primary items and dimensional information
- Examples of Dimensional items include counterparty groups, risk weights, ratings etc.
- Around 400 unique elements including about 25 dimensions

- Granular details elicited when required. However, no attempt to reproduce detailed calculations carried out at bank's end.
- References to RBI capital adequacy guidelines/other relevant RBI circulars for each data element incorporated in the return through reference linkbase.

### **8.5 Data standards implemented by data warehouse/databases**

In a dimensional database, dimensions are required to be logical, clearly identifiable, well defined, mutually exclusive (without any overlapping with other dimensions) and comprehensive (to be able to generate all data requirements). In data dimensional modelling, when data is indexed across various dimensions, it gives rise to consistent data across returns.

Key aspects of an information block include dimensions and numeric data-items called measures. Dimensions relate to characteristics of the account that are of interest and the measures relate to values satisfying one or more of such dimensions. All possible ways, along which dimension can be measured, are named as the list of values associated with the respective dimensions. Measures assume numeric values for the list of values of dimensions.

### **8.6 Coordination among SEBI, MCA and other agencies**

The Ministry of Corporate Affairs had asked the **non-financial companies** (i.e. Manufacturing & Service companies) - with paid-up capital more than ₹5 crore or turnover more than ₹100 crore or listed companies - to file their balance sheet and profit & loss account statements from the financial year 2010-11. However, the entire financial statement data has not been shared with RBI. The taxonomy for financial companies is also ready, but no filing has started so far. SEBI has also taken some initiative for XBRL based filing.

One of the advantages of XBRL based system is the portability across different users. Ultimately, banks need to supply the XBRL data from their single central repository. Further, over the long term for common data elements collected by different regulators, these can be stored in a centralised financial system database which can then be accessible to all the regulators. Hence, there is a need to have clarity of common requirements of data elements for various regulators for which regular co-ordination should be ensured among regulators. In the long term, feasibility of having a centralised financial database, which is accessible to all the regulators may be explored after due analysis of the issues from various dimensions.

### **8.7 Impact of changes in XBRL schema on systems in RBI and banks**

RBI XBRL taxonomy has one core schema in which all the reporting elements are defined. There are separate folders for each return, which contains linkbases capturing the relationship specific to the return. The common schema is internally referred in return specific entry points.

Most of the banks have already created the ADF server. Every time there is a change in the schema, the ADF system in the bank need to be changed. The bank needs to check if they need to change the mapping of data elements generated from their internal systems necessitated by change in taxonomy.

The ADF system of the bank and the XBRL taxonomy should be designed properly to take care of anticipated changes in the format of return, i.e., addition and deletion of data elements. ADF centralized repository system should ideally capture all key business/regulatory data at the most granular level. This not only will serve filing of regulatory returns but also help in meeting future regulatory demands.

### **8.8 NBFC XBRL schema of MCA**

The taxonomy related to the financial statement of financial companies is also ready and filing was initially scheduled to start for the financial year 2012-13, but no filing has started so far. The XBRL based submission by financial companies should be started at the earliest and the relevant data may also should be shared across the regulators.

### **8.9 Need for a Data Governance Group for XBRL implementation in RBI**

At present, various regulatory, operational and policy departments prescribe returns as per their requirement and make changes in the formats / contents as per the evolving needs. The present practice of prescribing return has resulted in

- Duplication of data elements across returns
- Inconsistency in reported figures across returns
- Increasing burden on banks for collecting, processing and submitting such returns

In this connection, there is necessity of an Inter-Departmental Data Governance Group (DGG) for the RBI, so that the process of rationalization regarding data elements, periodicity, need for provisional returns can be carried out in a concerted manner. All future returns to be prescribed by any department may be routed through the DGG, to avoid duplication.

### **8.10 Recommendations of the Committee:**

- 1) Similar forms can be taken together within/ across the departments of RBI and thus common reporting elements can be arrived at. Rationalisation /Consolidation of returns before taking up the returns pertaining to a department must be done. The rationalisation / consolidation of returns may be examined and reviewed on a periodic basis.
- 2) For granular account level data and transactional multi-dimensional data, RBI may develop and provide specific details of RDBMS/text file structures along with standardised code lists and basic validation rules so that banks can run the validation logics to ascertain that the datasets are submission-ready. In this connection, XBRL based data element submission may also be explored.
- 3) In due course, from a medium term perspective, moving from return based approach to data element based approach needs to be considered.
- 4) It is expected that banks would generate the instance document from the Centralised Data Repositories (CDR) and submit the same to RBI without manual intervention. The banks should validate the generated instance documents based on the XBRL taxonomy and validation rules before sending them to the Reserve Bank. Thus, the present approach of spreadsheet(Excel) based submission of returns needs to be given up ultimately.
- 5) An Inter-Departmental Data Governance Group (DGG) for the RBI as a whole may be formed, so that the process of rationalization regarding data elements, periodicity, need for provisional returns can be carried out in a concerted manner. All future returns to be prescribed by any department may be routed through the DGG, to avoid duplication.
- 6) As part of its data governance activities, the DGG may also pro-actively identify any data gaps in the evolving milieu and prepare plan of action to address the gap.
- 7) The XBRL taxonomy must include data definitions so as to completely leverage the utility offered by XBRL.
- 8) The XBRL taxonomy should be designed flexibly so as to take care of the anticipated changes in the format of return, i.e., addition and deletion of data elements.
- 9) The XBRL based submission by financial companies to MCA should be shared across the regulators as required.
- 10) Since new tools/software are developed for leveraging XBRL, there needs to be process of continuous monitoring of new developments so as to examine their utility and possible value addition
- 11) Ultimately, the logical location for storage of XBRL data is a Data Warehouse. Therefore the existing Data Ware House needs to be revamped with Next Generation Data Ware House capabilities. Big Data solutions also need to be explored for enhancing analytical capability in the new data paradigm which would be of particular use in the area of banking supervision.

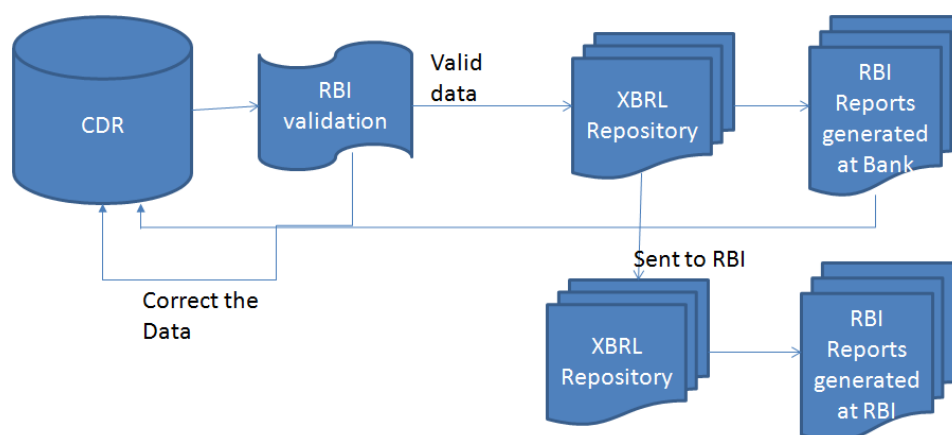
## Chapter IX– Automating Data Flow from banks to RBI’s central repository

### 9.1 Introduction

With the implementation of Phase I of the Automated Data Flow (ADF) project in varied stages by different banks, there is now a need to consider the process of automating data flow from banks to RBI.

### 9.2 Push and Pull mechanisms to and from RBI:

- At present banks generate report at their end, view it and send it once they feel satisfied / informed about what they are sending to RBI.
- Technology allows the data to be sent to RBI in basic elementary data form, from banks to RBI after appropriate mapping is completed. Once the data is received, RBI can generate the reports.
  - This is exactly like the data being transferred from back end source system to CDR (Central Data Repository) using some ETL (Extract, Transform and Load), or data replication that normally happens between data centre and DR centre.
- Alternately, once the mapping is completed, XBRL repository/instance document gets created at bank side, reports get generated, reviewed and once the bank personnel feel fine, then the XBRL repository can be copied / sent to RBI.
- From the XBRL instance data sent from bank, RBI can now generate the reports. Both the reports generated at RBI and bank should be same, even if the vendor solutions and platforms are dissimilar.

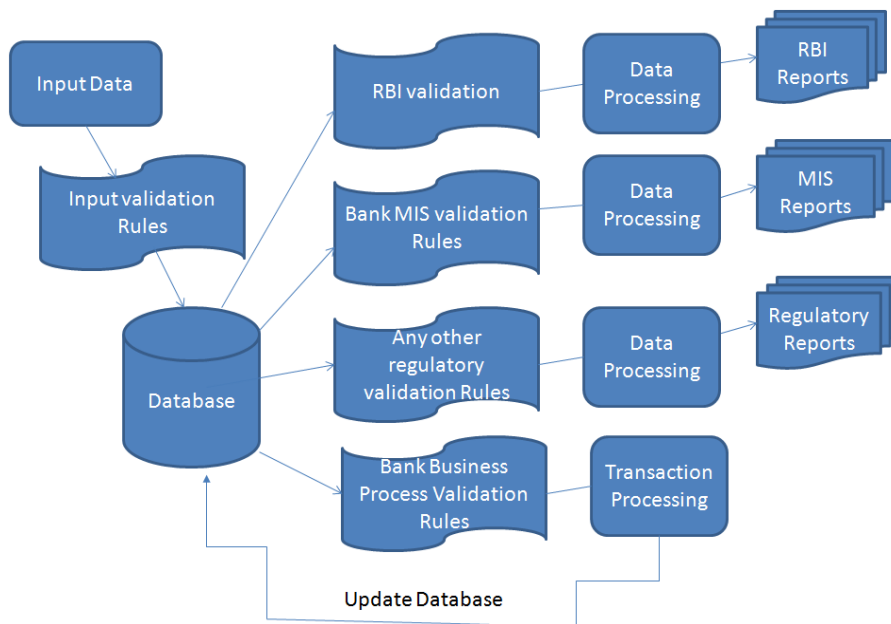


- Alternately, as the correction of data is minimized or there is no need to correct the data, the banks can directly send the validated data in the form of XBRL to the cloud hosted either by RBI. From this cloud, reports can be generated and viewed by both RBI and respective banks.
  - This approach can mitigate duplication in data submission across returns – i.e. same information being submitted across several returns for different department.
  - As the granularity or level of detail of data increases, this approach also relieves the banks from generating the reports, at times with changing formatting requirements.
  - Even in this approach, banks may be allowed to review data to ensure the output is correct prior to RBI generating and finalizing the reports.

### **9.3 Validation Techniques:**

With the ever increasing risk, the demand for precise and accurate data is on the increase. Invalid data is not only is risking the business, but the assessment of overall economy of the country as well.

It is a common practice that any system captures the data, validates it, processes it and stores/communicates the processed data. One of the most expensive and time consuming aspect of data management is data validation. The programming rules of validating data are so far embedded within source systems. The future is that the validation rules have to be decoupled from the task of data processing. By coupling the validation rules with processing, it's very inflexible and time consuming to be in tune with the changing requirements. Not only that, in future a different organization may be setting the validation rules, not just the business owner. Or, we may have to apply different validation rules for different purposes like data capture, transaction processing, regulatory reporting, internal MIS etc. The following figure depicts the scenario.



While capturing the data, input validation rules would be applied and the valid data is saved in database. For any further transaction processing or MIS processing the respective validation rules are applied against the data. The desired output is generated after applying the business process on validated data.

For the above system to be in place, what is needed is a validation rule specification mechanism. There are many such schemes like Excel, DTD (Data Type Definition), XML Schemas, RelaxNG etc. Except DTD, XML and RelaxNG, majority of the schemes are proprietary and platform specific. The support for RelaxNG also is limited.

The aspects in this regard include:

1. XML as common language: RBI's criteria for extracting data from banks may keep changing. Accordingly, new validation schemes for various purposes may need to be defined ex: Male Senior, age > 60; Female Senior Citizen age > 55 etc... Here, no new data element is defined, but the data is filtered or processed as per the criteria.
2. The validation schemes have to be expressed in XML form or in similar compatible form by the system at RBI, so that the systems at banks automatically understand the requirement, accordingly process their data and return the data to RBI, without any manual intervention. This enables a fully automated data flow from banks to RBI even with dynamic and changing validation criteria.
3. The present data capturing methods may not yield 100% valid and quality data into the system, due to a number of reasons like lack of uniform data definitions, legacy systems; volume, and variety of data.

The validation schemes have to be expressed in XML form or in similar compatible form by the system at RBI, so that the systems at banks automatically understand the requirement, accordingly process their data and return the data to RBI, without any manual intervention. This enables a fully automated data flow from banks to RBI even with dynamic and changing validation criteria.

### **(iii) Efficient Mechanisms to transfer large chunks of data**

Certain types of the data required by RBI are very huge and take too much time to send to RBI. Implementing some incremental transfer of the data, similar to backup mechanisms like incremental back up, full back up and differential backups etc can be looked at. If these mechanisms cannot be directly applied to the needs, custom development of such a mechanism may be opted. Alternate mechanisms like WAN bandwidth optimization etc can definitely help in increasing the transfer rates, but still an optimization mechanism at application level can give better control, than an optimization at network level.

## **9.4 Automating data flow between banks and RBI - Recommendations of the Committee**

- 1) Using secure network connections between the RBI server and the bank's ADF server, the contents of the dataset can be either pulled through ETL mode or pushed through SFTP mode and loaded onto the RBI server automatically as per the periodicity without any manual intervention. Pushing of data by banks could enable easier management of the process at RBI end. An acknowledgement or the result of the loading process can be automatically communicated to the bank's ADF team for action, if necessary.
- 2) While the traditional RDBMS infrastructure in place in RBI may be used for storage and retrieval of aggregated and finalized data, Big-data solutions may also be considered for micro and transactional datasets given their high volume, velocity and multi-dimensional nature.
- 3) The validation schemes may also be expressed in XML form or in similar compatible form by the system at RBI, so that the systems at banks automatically understand the requirement, accordingly process their data and return the data to RBI, without any manual intervention. This would enable a fully automated data flow from banks to RBI even with dynamic and changing validation criteria.
- 4) The enterprise-wide data warehouse (EDW) of RBI should be made the single repository for regulatory/supervisory data pertaining to all regulated entities of RBI with appropriate access rights. Any unstructured components pertaining to RBS data may be maintained in EDW using new tools available for such items.
- 5) As a key support for risk based supervision for commercial banks, internal RBI MIS solution needs to seamlessly generate two important sets of collated



information: (i) Risk Profile of banks (risk-related data – mostly new data elements), and (ii) Bank Profile (mostly financial data – DSB Returns and additional granular data) based on data elements supplied by banks.

- 6) Once the system stabilises, the periodicity of data can be reviewed to examine obtain any particular set of data at shorter intervals or even up to near real time.



## **Chapter X –Data Quality Assurance Process at RBI**

### **10.1 Introduction**

Given the increase in coverage as well as granularity of data handled by central banks across the world in the contemporary milieu, the quality assurance processes assume great importance and need to be robust. This chapter reviews practices in few international jurisdictions/entities and considers possible emulation by RBI.

### **10.2 Quality Assurance Procedures in foreign central banks/other entities**

The details of various quality assurance or validation procedures in ECB and Bank of England and entities like OECD and UNECE are indicated at Annex VI.

It is apparent that the key data quality dimensions considered as part of quality assurance framework include relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence. ECB lays emphasis on various checks like completeness checks, internal consistency, consistency across frequency of the same dataset, external consistency, revision studies, plausibility checks and regular quality reporting. The plausibility checks help in identification of reported figures that are significantly different from the usual reporting pattern using statistical quality control techniques.

In a survey of central banks the Denmark National Bank found that central banks have streamlined ‘administrative work’ and ‘outlier evaluations’ and achieved better ‘follow-up communication’ by automating their data validation processes (Drejer, 2012).

As stated by BIS IFC working paper on “Optimizing checking of statistical reports”, the purpose of the data checking process is to ensure that data reported are without substantial errors and that the compiler learns about the story behind important changes in reported figures. In general, the data cleansing process can be divided into two steps: First a data validation process followed by the process of plausibility testing. Data validation rules are built into the reporting system and the data validation process will check whether the data pass or fail the validation rules.

### **10.3 Quality Assurance Process in RBI**

A systematic data quality assurance framework is required to be formulated to provide further enhancement data quality and integrity of data stored in the bank’s data warehouse. International statistical quality frameworks like those developed by UNECE may also be considered.

The framework may detail various aspects relating to key data quality dimensions and the detailed processes to implement the dimensions in a robust manner.

#### **10.4 Recommendations on data quality assurance process**

- 1) Exclusive data quality assurance function can be created under the information management unit of RBI.
- 2) A data quality assurance framework may be formulated by RBI detailing the key data quality dimensions and systematic processes to be followed. The various key dimensions include relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence. The framework may also be periodically reviewed.
- 3) Various validation checks like sequence check, range check, limit check, existence check, duplicate check, completeness check, logical relationship check, plausibility checks, outlier checks are among the key checks which need to be considered and documented for various datasets with assistance from domain specialists.
- 4) Usage of common systems for data collection, storage and compilation would help provide environment for robust implementation of systematic data quality assurance procedures.
- 5) Deployment of professional data quality tools as part of the data warehouse infrastructure could also provide for comprehensive assessment of data quality dimensions.
- 6) Whenever data are received and compiled, quality assessment reports that summarize the results of various quality checks may also be generated internally.

## **Chapter XI – Perspective on System-wide improvements**

### **11.1 Introduction**

There is a need to look into financial system wide perspective and beyond to facilitate efficiency and synergies in reporting to Gol and regulatory agencies, and sharing of data among regulators. A technological standard for data interchange like XBRL can be effective only to the extent organizations use it. It may be mentioned that the ability to automatically pull data that is easily reused and analyzed is dependent upon the data originating in that format. As more entities begin to transmit information according to a specific standard, the value of adopting the standard increases. Thus, key goal is to achieve critical mass of such standards for ensuring the public good of cheap, accurate, timely financial data for everyone.

This chapter outlines some recommendation of the Committee on leveraging of common standards and new technological developments for efficiency and effectiveness from a systemic perspective.

### **11.2 Deliberations and recommendations of the Committee**

#### **(1) XBRL projects by regulators**

Regulators like RBI, SEBI, MCA are in the process of undertaking various XBRL projects. Given the benefits offered by XBRL, all the regulators can explore possibilities of commonalities in taxonomy and data elements and protocols and formats for sharing of the data among themselves.

A separate standing unit Financial Data Standards and Research Group may be considered with involvement of various stakeholders like RBI, IBA, banks, ICAI, IDRBT, SEBI, MCA, NIBM, CAFRAL etc for looking at the financial data elements/standards and to try to bring them into holistic data models apart from mapping with applicable international standards.

Given that standards are considered a classic public good, with costs borne by a few and benefits accruing over time for many entities, involvement of regulators and Government would help solve the collective action problems created by these disincentives. Inter-regulatory forums could help facilitate improvements in data/information management standards across the financial sector to benefit all stakeholders.

## **(2) Integration of standards like XBRL in the core banking system/ accounting software**

XBRL is mostly used to allow the transfer of aggregate performance information, from place to place and organization to organization. But if one wants to drill down to the detail, XBRL Global Ledger or “XBRL GL” provides the capability. Electronic accounting and ERP systems work at a transactional level by storing a range of information about each individual entry in a specialized ledger, which, in turn, is summarized into a general ledger. The data in the general ledger are themselves summarized in order to provide reports. In doing so, especially if the information is moved from its originating system, much or all of the details of each original transaction become unavailable. To address this problem, there is a need for a standard way to capture, archive, transmit and aggregate all of the information contained in the original ledgers, as well as what’s in the general ledger, journal entries etc. There is a need for a standardized way to store all of the operational data and data definitions contained in an accounting system.

Data is at the heart of any business. Some amount of data is required to be shared /communicated to business partners or regulatory authorities and majority of the data is towards conducting their own business efficiently. With these two purposes, two ways of defining financial data elements emerged. (1). Financial Data Service Models like (FSDM from IBM, OFSAA from Oracle, Teradata model, Universal Financial Data Model and the data models by core banking vendors etc and (2) XBRL, which mainly concentrates towards an efficient way of reporting / exchanging the data.

All these days, it has been thought that the amount / level of data elements required by regulatory authorities is much smaller (aggregate) when compared to the data elements captured by the business. But as time progresses, the supervisory bodies are in need of much deeper look into the data. RBI, as part of the XBRL project defined much of the data elements in the form of taxonomies, which are mainly aimed at reporting purposes.

It is time now to see, whether it is feasible to extend the scope of XBRL to cover all the data elements of banking business. This helps in preparing for the future demands in lesser time frame. As such, there is hardly any data generated without any interaction between entities either within the organization or outside. New data is generated or updated always with some interaction among the related entities. And every interaction can be encoded in the form of XML. i.e even the basic transactions like deposits and withdrawal can be in the form of XML messages.

If every bank business element is defined in XML taxonomy, a uniform validation rules for capturing the data, for transaction processing can be possible to be prescribed centrally by an industry body or regulatory authority. While the actual physical carrying of business is left to banks, the rules of carrying the business and enforcing them can easily be applied by regulatory authority using this approach.

This ensures greater transparency in the system, reduces the burden of physical audits by regulatory bodies and acts as an early preventing mechanism for future frauds. If one can achieve universal data definitions across industries, government and society, it would be easy to monitor and protect the assets by embedding smart contracts with good business practices.

The committee feels, the present scope of XBRL data definitions have to be further extended to cover in depth data definitions covering almost all data elements that are required to carry banking business. RBI's XBRL taxonomy and the other standards like IFX, OFX, ISO20022 etc if combined may generate the majority of the data elements etc.

Ultimately, from a banking system perspective the benefit would arise by enabling MIS systems or accounting systems to directly tag and output data in formats like XBRL to maximize efficiency and benefit. Duplication of "Validation layer construction" by each member bank and RBI can be addressed. Thus, there is need for integration of standards like XBRL in internal applications like core banking systems/accounting systems of banks.

### **(3) Proposed plan to set up Financial Data Management Centre**

- 1) The FSLRC has recommended the creation of a Financial Data Management Centre(FDMC) which would be a repository of all financial regulatory data. It is expected to have advanced database management capabilities with electronic data submission, generate a full view of the entire Indian financial system and sharply reduce costs of compliance for financial firms submitting supervisory data to financial agencies. A task force has been setup by Gol.
- 2) Ideally, the primary regulator needs to have the fullest granular data available at its database while FDMC may have the relevant summarized data. Automated connections could be established between the regulatory databases and FDMC. It is critical that processes and protocols are put in place such that the work of regulators do not get impeded in any manner. Large investments already made by the individual regulators may also need to be considered.

### **(4) Knowledge sharing and research:**

Following measures would be helpful in knowledge sharing and research.

- (i) Analysis of new developments in respect of XBRL
- (ii) Conducting of pilot for enhancing leveraging XBRL for internal uses
- (iii) Research and examining ways and means of leveraging new data technological platforms like XBRL for enhancing overall efficiencies of banking system

## **(5) Standard Business Reporting - Leveraging technologies like XBRL by Government for larger benefits**

Essentially, SBR is based on:

- (i) Creating a national financial taxonomy which can be used by business to report financial information to Government. That taxonomy could encompass all financial data from outset or be built up gradually.
- (ii) Using the creation of that taxonomy to drive out unnecessary or duplicated data descriptions.
- (iii) Enabling use of that taxonomy for financial reporting to Government and facilitating straight-through reporting for many types of report direct from accounting and reporting software in use by business and their intermediaries; and
- (iv) Creating supporting mechanisms to make SBR efficient where they do not already exist (a single Government reporting service or portal or gateway etc.)

In fact, SBR has the potential to achieve much more for business and Government. While the initial focus is on financial reporting to Government, the standardization it introduces can be exploited for 'business to business' reporting and for more effective and efficient use of information within Government (including risk assessment which is important to revenue bodies). Commercial banks and their customers might derive significant benefits from the regular provision and analysis of such information. The Dutch project is piloting such a scheme, in conjunction with local banks.

The basic proposition for SBR is the creation of a national financial and business reporting taxonomy that Government and the private sector use to describe data. This can be done leveraging XBRL. It is critical to understand that SBR is not a technology initiative but a policy one which harnesses technology.

## **6) Usage for multiple purposes like credit management by banks**

- Banks do collect lots of information like progress of the business, balance sheet etc from corporate customers, many are on paper at present. This data can be collected in the form of XML/XBRL. The data collected over a period of time can give an in depth look at a corporate customer and can give early warnings to the bank while taking crucial decisions, like loan renewals etc



- Through standards like XBRL, loan and credit management departments can obtain data quickly and reliably via automated reporting, reduce costs in processing data, compare and analyze financial information much more reliably, fully and effectively using automated processes, track financial performance more quickly and efficiently, reach decisions more confidently and provide a quicker response to clients.
- Dutch banking giant ING has reportedly announced that from 1 January 2015 it will start to offer discounts on loan and credit applications for its Small and Medium sized Enterprise customers in the Netherlands. All they have to do is start providing XBRL versions of financial statements through the Dutch SBR platform. SBR in the Netherlands would be moving to a new phase next year, with mandatory filing of XBRL company accounts starting to come into effect. Accounting software and accounting firms are now fully capable of producing XBRL versions of annual accounts along with the mandatory tax filings in XBRL already in place. Hence, retail banks like ING can leverage these new capabilities to allow it to be better informed about its customer's financial profile. The XBRL versions of financial statements will replace the paper and PDF versions of reporting previously used. ING is one of a number of banks in the Netherlands seeking to leverage the work of the SBR program of reporting improvements.<sup>6</sup>

## **(7) Audit and Assurance**

As the leveraging of machine readable tagged data reporting increases, the audit and assurance paradigm also need to get re-engineered to carry out an electronic audit and electronic stamp of certification using digital signatures. Lot of research and developments are happening across the world on the various methodologies in this regard.

## **(8) International comparability of financial information**

As had been indicated in Chapter II, comparability of financial data across countries is a key challenge faced globally. Increasing adoption of IFRS across countries is a positive development. While there is large number of convergence in capital standards via Basel I and Basel II, there are variations in details and level of implementations across countries. While the G-20 Data Gap initiative is a work in progress, there is also need for international stakeholders to analyse and examine how standards like XBRL can help facilitate ease of comparability of data as also to

---

<sup>6</sup> Source: <https://www.xbrl.org/news/cheaper-loans-for-xbrl-filers>

identify differences between countries in respect of financial reporting rules in an automated manner.

Some of the aspects in this regard include:

- (i) Basel II Pillar III disclosures for facilitating market discipline can be aligned through a platform like XBRL for consumption of information in automated manner by the market.
- (ii) Corporate actions are events that impact shareholders or debt holders of a company, such as stock splits, reverse splits, dividends, stock dividends, share buy backs, mergers, exchanges, name changes, etc. The distribution of such information has traditionally occurred via a press release issued by the company or a filing submitted to securities regulators, made available to processors and investors in a format that is not computer-readable. Transforming these messages into format like XBRL eliminates the need to rekey the data, thereby improving timeliness, accuracy and functionality of the information. The XBRL International Standards Board (XSB) in 2014 had requested financial regulators and listed company filers to form a global “Corporate Actions Working Group” to define XBRL taxonomies for standardizing the transmission and consumption of corporate actions events allowing straight through processing through the world’s financial systems. This project will build on the work initiated by the Depository Trust & Clearing Corporation (DTCC), SWIFT and XBRL US in 2009 that resulted in development of a draft taxonomy and business plan. The mission of the XBRL Corporate Actions Working Group will be to define a global standard for corporate actions documents that can be tagged at origination by the issuer or the issuers agent and allow straight through processing of this information to the security holder, tightly integrated with relevant ISO20022 messages.
- (iii) A single Data Point Model or methodology at international level can be explored for the elaboration and documentation of XBRL taxonomies
- (iv) Internationally, various initiatives for categorizing derivatives products for analysis and regulatory action and universal mortgage loan identifier to promote transparency, data aggregation, comparability, and analysis in the mortgage market are happening.
- (v) In the US, OFR has initiated process to create a financial instrument reference database with focus on key components relating to ontology, identifiers and metadata and valuation and analytical tools. Regulators, the financial industry, academics, and the public could potentially use the database to calculate the value of an instrument, compare a group of instruments, or link instruments to other datasets that use the same instrument identification.

## **(9) Increase in skillsets – education and training**

There is also need to incorporate training and education on the new technologies like XBRL by various academic bodies as also training/learning institutions so as to improve the availability of trained resources. For example core skills recommended for a taxonomy architecture role include the following:

- a) Data Modelling – Must understand data modelling methodologies and techniques, relevant taxonomy design options available, potential uses of the data (including internal and external analytics) within the information domain, and be able to bring these together to create a data model that can be used to guide the taxonomy architecture.
- b) Domain Knowledge – Must have a broad theory and technical knowledge of the business domain, current legacy reporting processes, and how reporters will produce instances of the taxonomy and how consumers will consume them.
- c) XBRL Technical Expertise – Must have a solid technical understanding of both XBRL and XML, including how the two technologies differ, so that the taxonomy architecture choices fully leverage the unique features of XBRL. Must also have an understanding of the technical needs of the taxonomy producers/consumers so that appropriate taxonomy design decisions can be made to facilitate correct filings.

### **11.3 Recommendations of the Committee:**

- 1) Given that standards are considered a classic public good, with costs borne by a few and benefits accruing over time for many entities, active involvement of regulators and Government would be key in solving the collective action problems created by these disincentives. Inter-regulatory forums could help facilitate improvements in data/information management standards across the financial sector to benefit all stakeholders and furthering collaboration with international stakeholders.
- 2) A separate standing unit Financial Data Standards and Research Group may be considered with involvement of various stakeholders like RB, IBA, banks, ICAI, IDRBT, SEBI, MCA, NIBM, CAFRAL etc for looking at the financial data elements/standards and to try to bring them into holistic data models apart from mapping with applicable international standards.
- 3) Regulators like RBI, SEBI, MCA are in the process of undertaking various XBRL projects. Given the benefits offered by XBRL and its usage across the globe by regulatory bodies, all the regulators may explore possibilities of commonalities in taxonomy and data elements and protocols and formats for sharing of the data among themselves.

- 4) In regard to OTC derivatives, one of the issues being debated is data portability and aggregation among the trade repositories spanning countries and jurisdictions. Hence, it is important to be cognizant of the needs of uniformity of standards across the globe and the need for our repository framework to have sufficient flexibility to conform to international standards and best practices as they evolve depending upon their relevance in the Indian context.
- 5) Ultimately, from a banking system perspective full benefit would arise by enabling transactional and accounting systems in banks to directly tag and output data in formats like XBRL to maximize efficiency and benefit. Thus, there is need for integration of standard formats like XBRL in internal applications/accounting systems of banks. The present scope of XBRL data definitions have to be further extended to cover in depth data definitions covering almost all data elements that are required to carry banking business.
- 6) In respect of knowledge sharing and research, various measures recommended include (i) Formation of banking sector level forum for data governance possibly under aegis of IBA or IDRBT (ii) Research by IDRBT regarding ways and means of leveraging new data technological platforms like XBRL for enhancing overall efficiencies of banking system (iii) conducting of pilot for enhancing leveraging of technologies like XBRL for internal uses by banks.
- 7) Standard Business Reporting, which involves leveraging technologies like XBRL by Government for larger benefits beyond the field of regulatory reporting, is being implemented in various countries like Australia and Netherlands. The same may be explored in India by Government of India in a phased manner.
- 8) As the leveraging of machine readable tagged data reporting increases, the audit and assurance paradigm also need to get re-engineered to carry out an electronic audit and electronic stamp of certification using digital signatures.
- 9) Committee recognizes that coordinated efforts are being carried out by various organizations which have developed standards like FIX, FpML, XBRL, ISD etc for laying the groundwork for defining a common underlying financial model based on ISO 20022 standard. Costs of migration and inter-operability would be key factors going forward.
- 10) As had been indicated in Chapter II, comparability of financial data across countries is a key challenge faced globally. Increasing adoption of IFRS across countries is a positive development. While there is large number of convergence in capital standards via Basel I and Basel II, there are variations in details and level of implementations across countries. While the G-20 Data Gap initiative is a work in progress, there is also need for international stakeholders to analyse and examine how technologies like XBRL can help

facilitate ease of comparability of data as also to identify differences between countries in respect of financial reporting rules in an automated manner.

- 11) Akin to initiatives in US, financial instrument reference database could be explored with focus on key components relating to ontology, identifiers and metadata and valuation and analytical tools.
- 12) A single Data Point Model or methodology at international level can be explored for the elaboration and documentation of XBRL taxonomies
- 13) Govt has plans to establish Financial Data Management Centre(FDMC) as a repository of all financial regulatory data. Automated connections could be established between the regulatory databases and FDMC. Large investments already made by the individual regulators needs to be considered.
- 14) There is also need to incorporate training and education on the new technologies like XBRL by various academic bodies as also training/learning institutions so as to help in capacity building and to improve the availability of trained resources.



## **Chapter XII – Trends and new developments**

### **12.1 Introduction**

In this chapter, the focus areas include the level of granularity of data, validation techniques based on self learning techniques and pattern recognition at both RBI and banks' ends and other new developments are covered.

### **12.2 Trends and developments**

#### **a. Semantic Web**

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners.

The Semantic Web is a Web of data. There is a lot of data we all use every day, and it's not part of the Web. Now, data is controlled by applications, and each application keeps it to itself. The vision of the Semantic Web is to extend principles of the Web from documents to data. Data should be accessed using the general Web architecture using, e.g., URI-s; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data.

Semantic Web technologies can be used in a variety of application areas- for example: in data integration, whereby data in various locations and various formats can be integrated in one, seamless application; in resource discovery and classification to provide better, domain specific search engine capabilities; in cataloging for describing the content and content relationships available at a particular Web site, page, or digital library.

#### **b. Building blocks of the Semantic Web**

In order to achieve the goals of semantic web, the most important is to be able to define and describe the relations among data (i.e., resources) on the Web. This is not unlike the usage of hyperlinks on the current Web that connect the current page with another one: the hyperlinks defines a relationship between the current page and the target. One major difference is that, on the Semantic Web, such relationships can be established between any two resources, there is no notion of "current" page.

Another major difference is that the relationship (i.e., the link) itself is named, whereas the link used by a human on the (traditional) Web is not and their role is deduced by the human reader. The definition of those relations allow for a better and automatic interchange of data. RDF, which is one of the fundamental building blocks of the Semantic Web, gives a formal definition for that interchange.

The Semantic Web Stack helps illustrate the architecture of the Semantic Web. The various components include XML, RDF, OWL, etc.

#### **d. Semantic Web and Artificial Intelligence**

Some parts of the Semantic Web technologies are based on results of Artificial Intelligence research, like knowledge representation (e.g., for ontologies or rules), model theory (e.g., for the precise semantics of RDF and RDF Schemas), or various types of logics (e.g., for rules). However, it is reported that Artificial Intelligence has a number of research areas (e.g., image recognition) that are completely orthogonal to the Semantic Web . At the same time, development of the Semantic Web brought some new perspectives to the Artificial Intelligence community: the “Web effect”, i.e., the merge of knowledge coming from different sources, usage of URIs, the necessity to reason with incomplete data; etc.

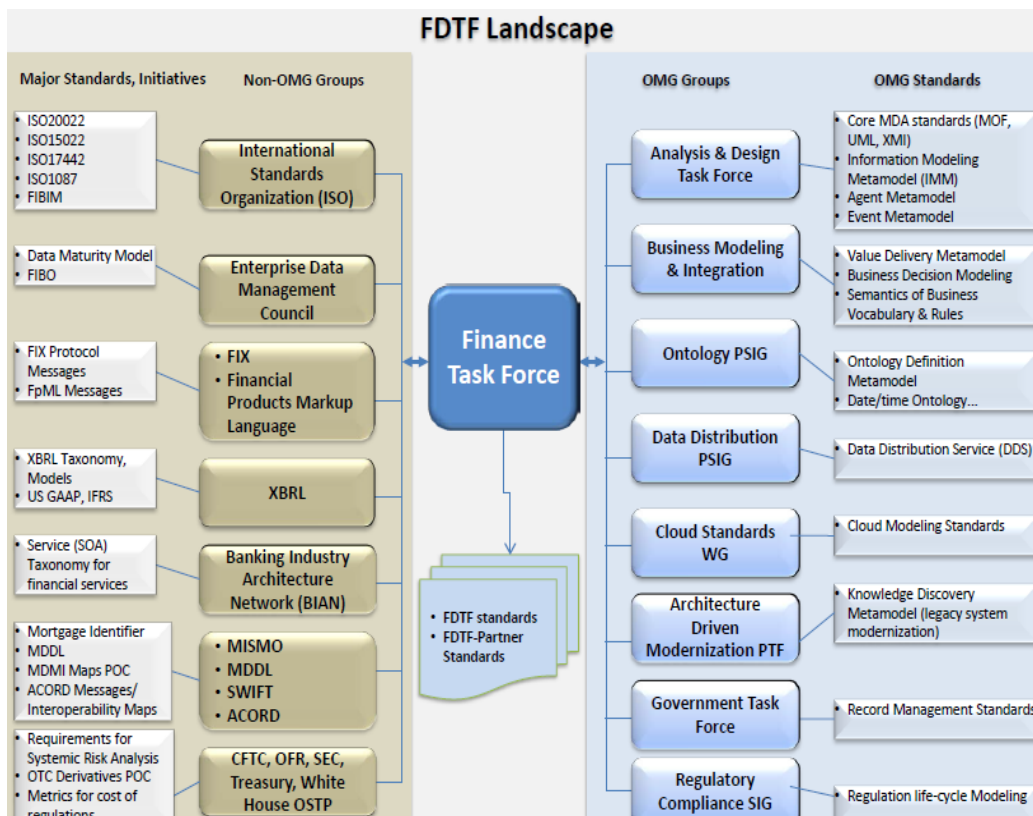
#### **f. Ontologies and Semantic web context**

Ontologies define the concepts and relationships used to describe and represent an area of knowledge. Ontologies are used to classify the terms used in a particular application, characterize possible relationships, and define possible constraints on using those relationships. In practice, ontologies can be very complex with several thousands of terms or very simple (describing one or two concepts only). An example for the role of ontologies or rules on the Semantic Web is to help data integration when, for example, ambiguities may exist on the terms used in the different data sets, or when a bit of extra knowledge may lead to the discovery of new relationships.

#### **h. Future of financial standards**

The following diagram from OMG Finance Task Force (FDTF) indicates the various standards and groups working on various aspects of financial standards.





(Source: FDTF, Object Management Group)

There are numerous technical syntaxes which are being used today to express financial information digitally and there will likely be many others like XBRL (Extensible Business Reporting Language), W3C Government Linked Data, W3C Linked Data, Various forms of RDF and OWL, various forms of XML. One of the most popular technical syntaxes is XBRL. Ultimately, various disparate data sources in the form of XBRL, linked open data, news feeds, market price information etc need to be handled by big data systems or processors to provide useful and actionable knowledge.

From a technology standpoint, it is easy to establish a standard. It is reported that given that the cost for creation of a standard is low, resultantly there is potential to bring about a plethora of competing data standards in the financial services industry. Given such an environment, the future will involve finding new ways for data interoperability between various established standards. Hence, all stakeholders need to become involved with standards-setting organizations today or at least keep abreast of the new developments and provide strategic focus on data issues, at the same time being focussed on new technology trends that can help facilitate progress in the future.

### **12.3 Unification of standards**

Subsequent to ISO 150022, ISO 20022 is the subsequent step towards unification of standards. The task of unification has brought out some semblance of unified messages. However, many challenges need to be addressed yet –redundancy built over a period of time with multiple standards are catering to a single functionality, adoption of unified messages is still going on a slow pace, ability of common standard support changes and new regulations. This would require significant amount of time and efforts. Ultimately, best outcomes can potentially bring about cost reduction, reduction in errors.

### **12.4 Big Data and Banks**

While many large banks leverage their vast data warehouses, Big data is different. It is vast in scope, varied in form and instantaneous in velocity, encompassing data from varied sources like mobile devices, social media applications and website visits as well as information from third-party providers of credit, spending and legal data. It promises to reveal hidden consumer behaviours that may not be immediately apparent Big data potentially allows banks to measure and manage risk at an individual customer level, as well as at a product or portfolio level, and to be much more precise in credit approvals and pricing decisions.

Although banking has always been built heavily reliant on data, today's data paradigm is said to be bigger, faster and more varied, requiring new and different tools. Moreover, big data also holds more promise for mitigating risk and recognising opportunities, especially when novel and diverse data sources are integrated into traditional risk management frameworks.

Thus, there is potential for major role of Big Data solutions for banks ranging from risk management to the development of new products and services, to customer engagement. Big data can assist banks to customise the products to their customers. Big data solutions can help banks enhance information security by making it quicker to flag up blacklisted credit cards, or any cards or accounts with potentially fraudulent activity on them. The ability to use big data during credit scoring or considering loan or credit applications help banks to better identify customers who might not constitute reasonably good credits.

To learn more about the intersection between big data and risk management at banks, the Economist Intelligence Unit (EIU) surveyed 208 risk management and compliance executives at retail banks (29%), commercial banks (43%) and investment banks (28%) in 55 countries on six continents. The results demonstrate that growing numbers of bankers are embracing the analysis and sharing of big data,

but that they still face challenges in applying the results to delivering superior risk management performance—especially around liquidity and credit risk.

The survey asked executives to rate their own institution's performance in controlling and mitigating risk. Those that rated their institution above average were also more likely to use:

- basic big data tools to integrate, manipulate and access structured and unstructured data (35% for the above-average risk managers versus 7% of those rated average or below)
- more advanced big data tools such as predictive analytics and visualisation (33% versus 8%).
- In other words, banks that perform better are more likely to use a variety of different methodologies, including both basic and advanced analytics, to understand and manage their risks. Moreover, they're more likely to bring large amounts of data to bear on risk management problems.

## **12.5 Recommendations of Committee**

- 1) Committee recommends that research/assessment of new developments in technology and financial data/technology standards need to be made a formal and integral part of the information system governance of banks and the regulator.
- 2) Banking technology research institute IDRBT may carry out research on new technologies/development and serve as a think tank in this regard.
- 3) Bank may explore Big Data solutions for leveraging various benefits of the new paradigm concerned with volume and velocity of data.
- 4) Any financial technical data standards needs to be of the nature of open standards, inter-operable and scalable in nature. Due impact assessment and pilot run would also be necessary before implementing on larger scale.

## ANNEXURE

## Annex I

### BIS – Principles and related requirements for effective risk data aggregation and risk reporting

| Principles                         |  | Requirements |  |
|------------------------------------|--|--------------|--|
| Governance/infrastructure          | 1. Governance                              | 1            | Framework established  |
|                                    |  | 2            | Approval of the framework and resources deployed                   |
|                                    |  | 3            | Data quality risks as part of risk framework                       |
|                                    |  | 4            | Policies on data and risk management                               |
|                                    |  | 5            | Full documentation and validation                                  |
|                                    |  | 6            | Independent validation by qualified staff                          |
|                                    |  | 7            | Consideration as part of any new initiatives                       |
|                                    |  | 8            | Assessment of the data aggregation process in case of acquisitions |
|                                    |  | 9            | Unaffected by bank's group structure                               |
|                                    |  | 10           | Aware of technical limitations                                     |
|                                    |  | 11           | IT strategy addresses improvements                                 |
|                                    |  | 12           | Sufficient financial and human resources                           |
|                                    |  | 13           | Board sets reporting requirements                                  |
|                                    |  | 14           | Board's awareness of limitations                                   |
|                                    |  | 15           | Board's awareness of implementation and ongoing compliance         |
|                                    |  | 16           | Overall assessment   |
|                                    |  | 17           | Expected date of full compliance                                   |
| Risk data aggregation capabilities | 2. Data architecture and IT infrastructure | 18           | Business' continuity planning and impact analysis                  |
|                                    |  | 19           | Data taxonomies  |
|                                    |  | 20           | Responsibilities on ownership, quality of data and information     |
|                                    |  | 21           | Role of the business owner   |
|                                    |  | 22           | Adequate controls through the life cycle of data                   |
|                                    |  | 23           | Overall assessment   |
|                                    |  | 24           | Expected date of full compliance                                   |
| Risk data aggregation capabilities | 3. Accuracy and integrity                  | 25           | Controls as to accounting data                                     |
|                                    |  | 26           | Mitigants and controls for manual processes                        |
|                                    |  | 27           | Reconciliation with different sources                              |
|                                    |  | 28           | Access to risk data of the bank's risk personnel                   |
|                                    |  | 29           | Dictionary   |
|                                    |  | 30           | Balance between automated and manual systems                       |
|                                    |  | 31           | Documentation of risk data aggregation processes                   |
|                                    |  | 32           | Measurement and monitoring processes                               |
|                                    |  | 33           | For all material risks   |
|                                    |  | 34           | Process to rectify poor data quality                               |
|                                    |  | 35           | Overall assessment   |
|                                    |  | 36           | Expected date of full compliance                                   |
| Risk data aggregation capabilities | 4. Completeness                            | 37           | Process to identify groups to report risks                         |
|                                    |  | 38           | All material risk data included                                    |
|                                    |  | 39           | Documentation of approaches to aggregate exposures                 |

|                          |                           |  |   |
|--------------------------|---------------------------|--|---|
| Risk reporting practices |                           | 40   | Measurement and monitoring of all material risk data                      |
|                          |                           | 41   | Exceptions properly identified and explained                              |
|                          |                           | 42   | Process to rectify completeness issues                                    |
|                          |                           | 43   | Overall assessment  |
|                          |                           | 44   | Expected date of full compliance  |
|                          | 5. Timeliness             | 45   | Documented timeliness requirements in normal and stress situations        |
|                          |                           | 46   | Capabilities to produce timely information to meet reporting requirements |
|                          |                           | 47   | Capabilities of rapidly producing risk data in stress situations          |
|                          |                           | 48   | Overall assessment  |
|                          |                           | 49   | Expected date of full compliance  |
|                          | 6. Adaptability           | 50   | Ad hoc data requests  |
|                          |                           | 51   | Customisation of data   |
|                          |                           | 52   | Incorporate new internal or external developments                         |
|                          |                           | 53   | Incorporate regulatory changes  |
|                          |                           | 54   | Overall assessment  |
|                          | 7. Accuracy               | 55   | Expected date of full compliance  |
|                          |                           | 56   | Requirements and processes to reconcile reports to risk data              |
|                          |                           | 57   | Automated and manual edit and reasonableness checks                       |
|                          |                           | 58   | Integrated procedure for identifying and reporting data errors            |
|                          |                           | 59   | Accuracy requirements for regular and stress cases                        |
|                          |                           | 60   | Overall assessment  |
|                          | 8. Comprehensiveness      | 61   | Expected date of full compliance  |
|                          |                           | 62   | Reporting in line with business model and risk profile                    |
|                          |                           | 63   | Emerging risks included, in the context of the risk appetite              |
|                          |                           | 64   | Status of measures agreed to deal with specific risks                     |
|                          |                           | 65   | Forecasts and stress tests  |
|                          |                           | 66   | Overall assessment  |
|                          | 9. Clarity and usefulness | 67   | Expected date of full compliance  |
|                          |                           | 68   | Reports tailored to recipients' needs                                     |
|                          |                           | 69   | Balance between data analysis and qualitative explanations                |
|                          |                           | 70   | Differentiated information needs of the board, senior management, etc     |
|                          |                           | 71   | Board determines its own reporting requirements                           |
|                          |                           | 72   | Feedback by the board to senior management                                |
|                          |                           | 73   | Senior management determines its own reporting requirements               |
|                          |                           | 74   | Inventory and classification of risk data items                           |
|                          |                           | 75   | Balance between data and recommendations, conclusions, interpretations    |
|                          |                           | 76   | Periodic confirmation of relevance and completeness                       |
| 77                       |                           | Overall assessment   |   |
| 10. Frequency            | 78                        | Expected date of full compliance   |   |
|                          | 79                        | Requirements for how quickly reports are produced in normal/stress times   |   |
|                          | 80                        | Routine test to produce accurate reports in stress conditions              |   |
|                          | 81                        | Availability of all critical exposure reports shortly in stress situations |   |
|                          | 82                        | Overall assessment   |   |
| 11. Distribution         | 83                        | Expected date of full compliance   |   |
|                          | 84                        | Timely dissemination of reports balanced with appropriate confidentiality  |   |
|                          | 85                        | Periodic confirmation of the timeliness of reports                         |   |
|                          | 86                        | Overall assessment   |   |
|                          | 87                        | Expected date of full compliance   |   |

Source: BIS - Principles for effective risk data aggregation and risk reporting (2013)

## Annex II

### ISO Standards – Banking/finance

#### **ISO Standard**

##### ISO 4217:2008

Codes for the representation of currencies and funds

##### ISO 4217:2008/Cor 1:2008

##### ISO 6166:2013

Securities and related financial instruments -- International securities identification numbering system (ISIN)

##### ISO 8109:1990

Banking and related financial services -- Securities -- Format of Eurobonds

##### ISO 8532:1995

Securities -- Format for transmission of certificate numbers

##### ISO 9019:1995

Securities -- Numbering of certificates

##### ISO 9362:2014

Banking -- Banking telecommunication messages -- Business identifier code (BIC)

##### ISO 10383:2012

Securities and related financial instruments -- Codes for exchanges and market identification (MIC)

##### ISO/TS 10674:2011

Rating services -- Assessment of creditworthiness of non-listed entities

##### ISO 10962:2001

Securities and related financial instruments -- Classification of Financial Instruments (CFI code)

##### ISO 11649:2009

Financial services -- Core banking -- Structured creditor reference to remittance information

##### ISO/TR 13569:2005

Financial services -- Information security guidelines

##### ISO 13616-1:2007

Financial services - International bank account number (IBAN) -- Part 1: Structure of the IBAN

##### ISO 13616-2:2007

Financial services - International bank account number (IBAN) -- Part 2: Role and responsibilities of the Registration Authority

##### ISO/TR 14742:2010

Financial services -- Recommendations on cryptographic algorithms and their use

##### ISO 15022-1:1999

Securities -- Scheme for messages (Data Field Dictionary) -- Part 1: Data field and message design rules and guidelines

ISO 15022-1:1999/Cor 1:1999

.

ISO 15022-2:1999

Securities -- Scheme for messages (Data Field Dictionary) -- Part 2: Maintenance of the Data Field Dictionary and Catalogue of Messages

ISO 15022-2:1999/Cor 1:1999

.

ISO 17442:2012

Financial services -- Legal Entity Identifier (LEI)

ISO 19092:2008

Financial services -- Biometrics -- Security framework

ISO 20022-1:2013

Financial services -- Universal financial industry message scheme -- Part 1: Metamodel

ISO 20022-2:2013

Financial services -- Universal financial industry message scheme -- Part 2: UML profile

ISO 20022-3:2013

Financial services -- Universal financial industry message scheme -- Part 3: Modelling

ISO 20022-4:2013

Financial services -- Universal financial industry message scheme -- Part 4: XML Schema generation

ISO 20022-5:2013

Financial services -- Universal financial industry message scheme -- Part 5: Reverse engineering

ISO 20022-6:2013

Financial services -- Universal financial industry message scheme -- Part 6: Message transport characteristics

ISO 20022-7:2013

Financial services -- Universal financial industry message scheme -- Part 7: Registration

ISO 20022-8:2013

Financial services -- Universal financial industry message scheme -- Part 8: ASN.1 generation

ISO 22222:2005

Personal financial planning -- Requirements for personal financial planners

ISO 22307:2008

Financial services -- Privacy impact assessment

ISO/IEC TR 27015:2012

Information technology -- Security techniques -- Information security management guidelines for financial services



## IT applications in banking

---

### ISO Standard

#### ISO 1004-1:2013

Information processing -- Magnetic ink character recognition -- Part 1: Print specifications for E13B

#### ISO 1004-2:2013

Information processing -- Magnetic ink character recognition -- Part 2: Print specifications for CMC7

#### ISO/IEC 8484:2014

Information technology -- Magnetic stripes on savings books

#### ISO 9144:1991

Securities -- Optical character recognition line -- Position and structure

#### ISO 9564-1:2011

Financial services -- Personal Identification Number (PIN) management and security -- Part 1: Basic principles and requirements for PINs in card-based systems

#### ISO 9564-1:2011/Amd 1

#### ISO 9564-2:2014

Financial services -- Personal Identification Number (PIN) management and security -- Part 2: Approved algorithms for PIN encipherment

#### ISO/TR 9564-4:2004

Banking -- Personal Identification Number (PIN) management and security -- Part 4: Guidelines for PIN handling in open networks

#### ISO/DIS 9564-4

Financial services -- Personal Identification Number (PIN) management and security -- Part 4: Requirements for PIN handling in eCommerce for Payment Transactions

#### ISO/AWI 11568-1

Financial services -- Key management (retail) -- Part 1: Principles

#### ISO 11568-1:2005

Banking -- Key management (retail) -- Part 1: Principles

#### ISO 11568-2:2012

Financial services -- Key management (retail) -- Part 2: Symmetric ciphers, their key management and life cycle

#### ISO/AWI 11568-2

Financial services -- Key management (retail) -- Part 2: Symmetric ciphers, their key management and life cycle

#### ISO 11568-4:2007

Banking -- Key management (retail) -- Part 4: Asymmetric cryptosystems -- Key management and life cycle

#### ISO/AWI 11568-4

Financial services -- Key management (retail) -- Part 4: Asymmetric cryptosystems -- Key management and life cycle

#### ISO 13491-1:2007

Banking -- Secure cryptographic devices (retail) -- Part 1: Concepts, requirements and evaluation methods

#### ISO/DIS 13491-1

Banking -- Secure cryptographic devices (retail) -- Part 1: Concepts, requirements

## ISO Standard

and evaluation methods

### ISO 13491-2:2005

Banking -- Secure cryptographic devices (retail) -- Part 2: Security compliance checklists for devices used in financial transactions

### ISO/DIS 13491-2

Banking -- Secure cryptographic devices (retail) -- Part 2: Security compliance checklists for devices used in financial transactions

### ISO 13492:2007

Financial services -- Key management related data element -- Application and usage of ISO 8583 data elements 53 and 96

### ISO/AWI TR 14742

Financial services -- Recommendations on cryptographic algorithms and their use

### ISO/TR 14742:2010

Financial services -- Recommendations on cryptographic algorithms and their use

### ISO 15782-1:2009

Certificate management for financial services -- Part 1: Public key certificates

### ISO 15782-2:2001

Banking -- Certificate management -- Part 2: Certificate extensions

### ISO 16609:2012

Financial services -- Requirements for message authentication using symmetric techniques

### ISO/AWI 16865

Retail Financial Services Compliance Guideline: PIN Security and Key Management

### ISO/CD 19038

Banking and related financial services -- Triple Data Encryption Algorithm (DEA) modes of operation -- Implementation guidelines

### ISO/TR 19038:2005

Banking and related financial services -- Triple DEA -- Modes of operation -- Implementation guidelines

### ISO 19092:2008

Financial services -- Biometrics -- Security framework

### ISO/CD 20038

Triple DES Modes and Key Wrap

### ISO 21188:2006

Public key infrastructure for financial services -- Practices and policy framework

Source: ISO website

## Annex III

### **Illustrative list of issues relating to credit area that need to be addressed by banks in their systems for enabling data standardization across banking system**

- 1) Ensuring of Unique customer id, correct mapping of individual accounts of a customer under unique id
- 2) Ensuring appropriate and correct security type/security name and security values to be entered in the system for secured advances, status of charge creation, date of last valuation
- 3) Ensuring availability of flag for “restructured” accounts
- 4) Flagging of project loans, incorporating DCCO for project loans, nature of project, location of project, status of project
- 5) Incorporating field for Diminution in fair value for restructured accounts and other details relating to restructuring
- 6) To flag the nature of loan - individual, multiple banking and consortium
- 7) Making use of “group id” concept for automatically calculating group exposures
- 8) Flag for “whether securitized”
- 9) History details for NPA
- 10) Specific provision details to be incorporated invariably where applicable.
- 11) Capturing details of date of sanction of loan and purpose of loan.
- 12) Linking of fund based and non-fund based exposure through common customer id
- 13) Facility for incorporating details of various charges levied other than interest charge
- 14) Flag relating to direct or indirect sector of priority sector and whether interest subvention applicable. Details relating to priority sector – eg. MSE (Micro and Small borrowers), housing loan borrowers in metropolitan centres with population of above 10 lakh, education loans for studies abroad, loans granted to distressed farmers,
- 15) minority communities under weaker section, distressed farmers, minority communities under weaker section, National Rural Livelihood Mission(NRLM) etc
- 16) Incorporating details relating to receipt of stock statements, document expiry, receipt of financial statements.
- 17) Provision for entering details relating to suit filed accounts and details of accounts under SARFAESI.
- 18) Flagging of capital market, commercial real estate exposure in the system
- 19) For agricultural accounts, the long term or short term nature of the agricultural advance based on crop cycle need to be incorporated.
- 20) Capture of date of last renewal of running accounts
- 21) Flagging of written off/technically written off status of accounts

- 22)History of SMA-1, SMA-2 status of accounts
- 23)Need to capture complete guarantor details
- 24)Incorporating internal and external rating details
- 25)Reckoning guarantees invoked and devolved LC accounts as part of principle operating account (for NPA classification)
- 26)Whether advances covered by guarantees from ECGC, CGTMSE etc
- 27)Consistent definitions based on RBI circulars and internal circulars
- 28)Invariably entering unique identifier details in the system
- 29)Making sector, industry and activity codes as mandatory fields in system
- 30)Minimising “Others” category in sector/activity to provide more accurate picture

## **Annex IV**

### **XBRL – Basic concepts, Developments and international case studies**

#### **1.Key aspects of XBRL**

This section provides information foundational to understanding digital financial information. Charles Hoffman, CPA and Raynier van Egmond in their book titled “Digital Financial Reporting” provide a summary of the following ideas, concepts, and terminology relating to digital financial reporting.

- (a) Interactive data** The SEC coined the term “interactive data”. Most business users have used or at least seen a Microsoft Excel pivot table. A pivot table is interactive, or dynamic, in that it can be pivoted to display information in different configurations. Digital financial reports can be made interactive, or dynamic, because of the nature of XBRL.
- (b) Unstructured versus structured information** -Structured which means the information has identifiable structure which can be recognized and utilized by computer software. Structuring information enables computer software applications to leverage that structure and work with the information.
- (c) Differentiating syntax and semantics** - Syntax describes the form of the information and is generally not relevant to a business person. Syntax is important to technical people. Semantics communicates the meaning of the information. Business meaning is key to the digital world. Business users need to work with the meaning of information, not the syntax.
- (d) Interoperability** -Achieving interoperability will result in new cost effective, easy to use, robust, reliable, repeatable, predictable, scalable, secure, auditable, business information exchange across business systems. Some business systems might be internal to an organization, others might be external to an organization.
- (e) Notion of semantic model** -While logical models have their benefits, they still leave something missing: business meaning. A semantic model provides an order of magnitude jump in usability over using a logical model.
- (f) Dimensional Business information is inherently dimensional** – Business information, and particularly financial information, is inherently multidimensional. The multidimensional model is simply a logical model for organizing information. What the multidimensional model does provide is enough agreement to express information so that it can be unambiguously understood by a computer software application, including applications which can render the financial information in a format appropriate for human consumption.

## 2.0 XBRL Framework

The core of XBRL is the XBRL 2.1 Specification. Modules build upon the base XBRL specification, providing additional functionality. The XBRL Specification provides a framework that divides XBRL into two main parts:

**XBRL taxonomies**, which are XML schemas that define the concepts and articulate a controlled vocabulary used by XBRL instances, and XLink linkbases, which provide additional information about those concepts.

**XBRL instances**, which contain the facts being reported, along with contextual information for those facts.

### XBRL Taxonomy Parts

XBRL taxonomies have various physical aspects and express concepts, resources, and relations. These work together to provide the required functionality to express the meaning of business information that is to be exchanged.

### Taxonomy schemas and linkbases

XBRL taxonomies are comprised of two parts:

**Taxonomy schemas** are the XML Schema part of the XBRL taxonomy. Taxonomy schemas contain concept definitions that take the form of XML Schema elements.

**Linkbases** are the XLink part of the XBRL taxonomy and are also XML documents. The term *linkbase* is an abbreviation for link database. Linkbases are physical aspects used to express a logical aspect called networks. Networks are of two types: resource and relation. Resource and relation networks are expressed in the XLink syntax in the form of an *extended link*. Extended links are like containers that hold the data contained within linkbases.

### Discoverable taxonomy sets

A single XBRL taxonomy may be comprised of a set of multiple taxonomy schemas and linkbases. This set is indicated in XBRL as discoverable taxonomy set (DTS). A DTS is governed by various discovery rules, specified by the XBRL Specification, that XBRL processors understand. A DTS can contain any number of taxonomy schemas and/or linkbases and can start from either a taxonomy schema, a linkbase, or an XBRL instance.

### Networks and extended links

Networks are a logical aspect of XBRL expressed physically as a set of linkbases. Linkbases exist within the physical model and are collections of extended links. Extended links work slightly differently in XLink than they do in XBRL. While in XLink, each extended link is physically separated in the case of XBRL, a role attribute is added to an extended link. A network is a collection of all the extended links of a specific type with the same extended link role. An extended link role is akin to a unique identifier expressed as a role attribute of an extended link.

*Resource networks* provide additional information about a concept. The additional information is in the form of an XLink resource. Of the five standard types of linkbases,

label and reference are resource linkbases, and they express resource networks. *Relation networks* express relations between concepts using XLink arcs. Of the five standard types of linkbases, presentation, calculation, and definition are relation linkbases, which they express as relation networks. Relations (expressed as an XLink arc) can have different arc roles to help further categorize relations.

XBRL instances contain the information that is being exchanged. That information is expressed in the form of facts. Each fact is associated with a concept from an XBRL taxonomy, which expresses the concept and either defines it or points to a definition of the concept external to the XBRL taxonomy by using one or more XBRL references. Concepts are associated with an XBRL instance by being part of the DTS.

### **3 Major implementation of XBRL across the world**

#### **i. Banking – US FDIC Call Report**

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.

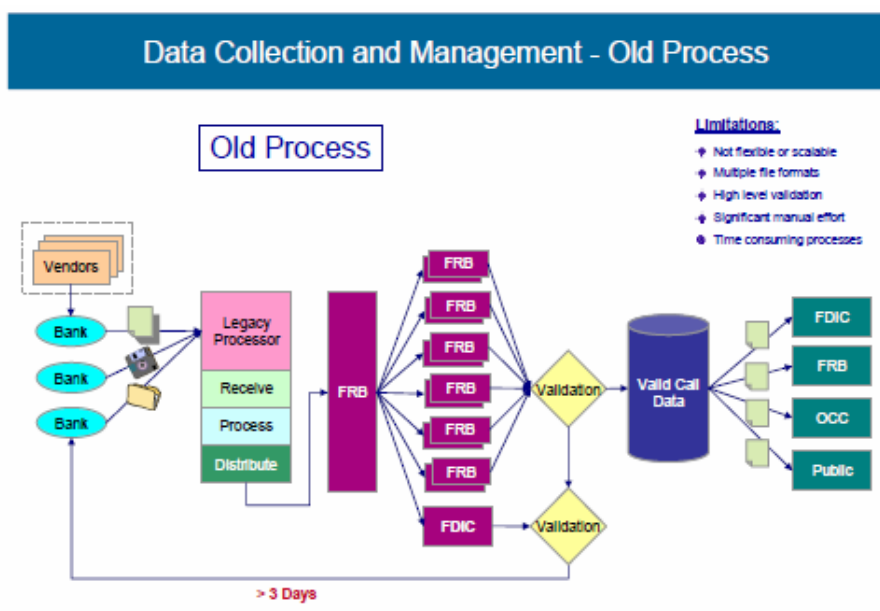
Regulators in the banking sector in the United States of America recognized these challenges and undertook a modernization project to overcome them. Members of the Federal Financial Institutions Examination Council (FFIEC), the Federal Deposit Insurance Corporation (FDIC), the Federal Reserve System (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.



Source: Approach paper on Automated Data Flow, RBI

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

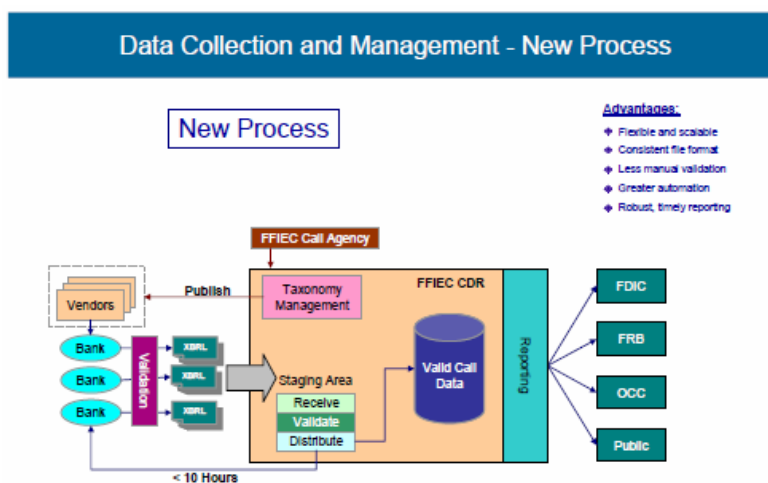


Report preparation software always contained a full set of reported data for that institution. Once the data was collected from all the respondents and validated by the agencies, the data was made available to outside agencies and to the public.

### 3. Technology Used in Automation Project

The Call Modernization project sought to reinvent and modernize the entire process in order to make it more useful for the regulatory community and its stakeholders. 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.

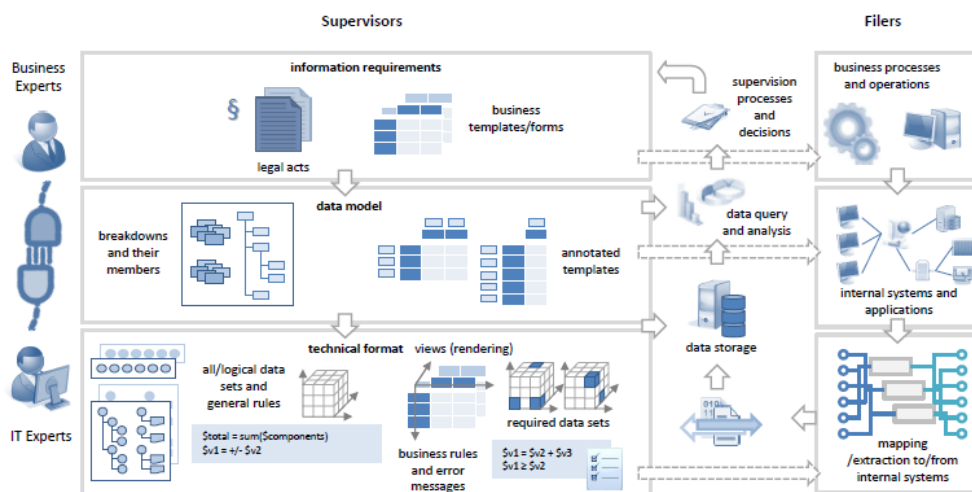


Source: Approach paper on Automated Data Flow, RBI

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.

## ii. European Banking Authority (EBA)

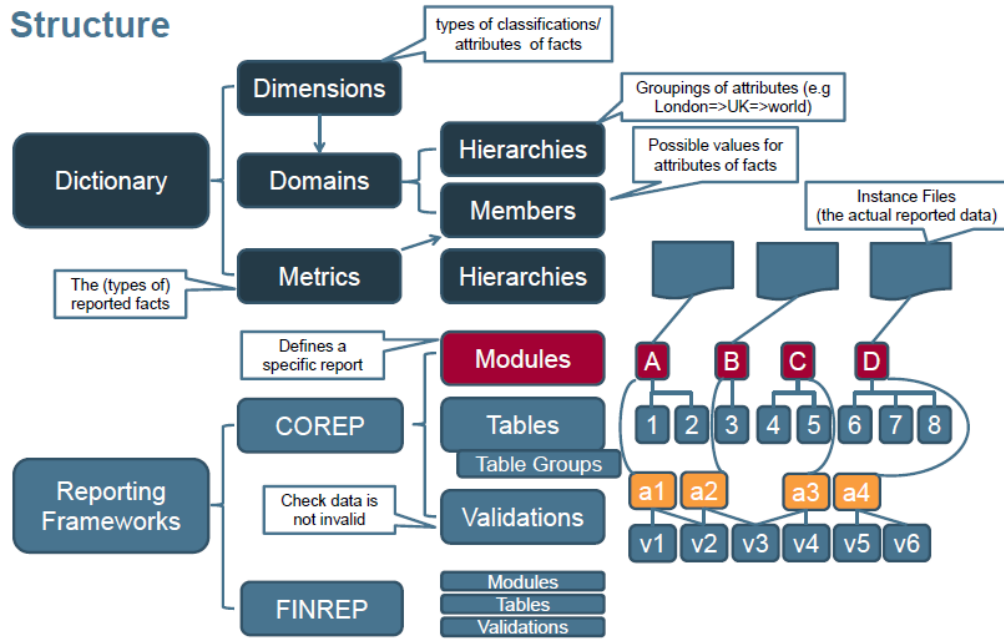
The EBA [earlier called Committee of European Banking Supervisors (CEBS)] is made up of about 27 regulators from the different European Union countries. The members collect solvency and liquidity information for the financial institutions that they monitor. The members are using IFRS for financial reporting by all financial institutions in Europe (rather than the 27 different sets of financial reporting standards used previously) to collect liquidity information. The members use Basel II for financial institution solvency reporting. EBA suggested XBRL as the exchange medium for these standard liquidity and solvency data sets. In addition to each country collecting financial institution information within the country, the members of EBA (the countries) also exchange information among themselves using XBRL. The summary of the process at EBA is indicated in figure below:



Source: EBA

Data Point Model driven Taxonomy is produced by an automated process directly from the DPM Database. For COREP and FINREP reporting, common dictionary is used with same concepts, same dimensions used to categorize things.

Taxonomy structure is indicated below:



Source:EBA

## ANNEX V

### XBRL Projects around the World

| S. No. | Project Name  | Country        | Brief Description  | Status      |
|--------|---|----------------|--|-------------|
| 1      | ACRA XBRL Project   | Singapore      | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |
| 2      | Australian SBR  | Australia      | Use of XBRL for reporting to government  | Implemented |
| 3      | Bank Examination Department of the Bank of Japan XBRL Project | Japan          | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |
| 4      | Bank of Spain (Financial statements)                          | Spain          | Financial statements sent to the Bank of Spain include European reporting frameworks, the Basel II solvency framework (COREP), Financial Reporting (FINREP), and ECB Statistics.         | Implemented |
| 5      | Banque De France  | France         | Use of XBRL by regulators to collect information from those they regulate.   |             |
| 6      | Bombay Stock Exchange XBRL Project                            | India          | Use of XBRL to collect financial information of listed companies.  | Implemented |
| 7      | Borsa Istanbul A.S., PDP XBRL Project                         | Turkey         | Borsa Istanbul operates an electronic disclosure system named as PDP. The project aims to collect financial statements from listed companies and brokerage houses in XBRL format in PDP. | Ongoing     |
| 8      | Borsaitaliana XBRL Project                                    | Italy          | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |
| 9      | Bryant College XBRL Education Initiative                      | United States  |  | Development |
| 10     | Cayman Islands Monetary Authority Project                     | Cayman Islands | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |

| S. No. | Project Name   | Country               | Brief Description  | Status      |
|--------|--|-----------------------|--|-------------|
| 11     | Centrae des Bilans (Belgique)  | Belgium               | Use of XBRL by regulators to collect information from those they regulate.                         | Implemented |
| 12     | Central Bank of Argentina Project  | Argentina             | Use of XBRL by regulators to collect information from those they regulate.                         | Development |
| 13     | CentralenRegistar  | Republic of Macedonia | Use of XBRL by regulators to collect information from those they regulate                          | Development |
| 14     | China Accounting Standards General Purpose Taxonomy Develop Project      | China                 | CAS Taxonomy is the extension of IFRS Taxonomy and fellows the same architecture                   | Implemented |
| 15     | China Galaxy Securities XBRL Implementation Project                      | China                 | Use XBRL formatted information disclosure to provide investing advices                             | Implemented |
| 16     | China Listed Company Information Disclosure Taxonomy Framework Project   | China                 | Use of XBRL by regulators to collect information from those they regulate.                         | Implemented |
| 17     | China Mutual Fund Information Disclosure Taxonomy Project(including IPO) | China                 | The purposes are both for information disclosure and off-line supervision of mutual fund products. | Implemented |
| 18     | Colombian Ministry of Finance Pilot                                      | Colombia              | Use of XBRL by regulators to collect information from those they regulate.                         | Development |
| 19     | Commission Bancaire Financier Assurance XBRL Project                     | Belgium               | Use of XBRL by regulators to collect information from those they regulate.                         | Implemented |
| 20     | Companies  | Ireland               | Use of XBRL by regulators to collect   | Development |

| S. No. | Project Name   | Country        | Brief Description   | Status      |
|--------|--|----------------|---|-------------|
|        | Registration Office (Ireland) XBRL Project               |                | information from those they regulate.   |             |
| 21     | CONTAEP  | Spain          | Ministry of Finance and Public Administrations, the General State Comptroller promotes the use of CONTAEP to report the annual accounts and other information to the Spanish Court of Audit.  | Implemented |
| 22     | Corporate Social Responsibility (CSR)                    | Spain          | The XBRL-RSC reports represent the use of XBRL to prepare and transmit Corporate Social Responsibility reports. XBRL-RSC reports include 491 indicators grouped into eight major areas of stakeholders: 'General information about the company ', 'Governing Bodies', 'Employees', 'Customers', 'Suppliers', 'Community ', 'Environment ' and 'Competition'. Most of them are quantitative. | Implemented |
| 23     | Corporation Tax Online - HMRC UK                         | United Kingdom | Use of Inline XBRL by tax authority to collect supporting information from corporate tax filers.  | Implemented |
| 24     | Danish Commerce and Companies Agency (DCCA) Project      | Denmark        | Use of XBRL by regulators to collect information from those they regulate. GAAP.  | Implemented |
| 25     | Department of Enterprise, Trade and Employment (Ireland) | Ireland        | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 26     | Deposit Insurance Corp of Ontario Credit Union Project   | Canada         | Use of XBRL to collect financial information of credit unions.  | Implemented |
| 27     | Deutsche Borse AG Project                                | Germany        | Use of XBRL by regulators to collect information from those they regulate   | Implemented |
| 28     | Deutsche Bundesbank Project                              | Germany        | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |

| S. No. | Project Name  | Country       | Brief Description   | Status      |
|--------|---|---------------|---|-------------|
| 29     | Developed a filing system for financial companies                     | Korea         | Developed a filing system using a XBRL Taxonomy which was self-developed to collect information from financial companies.   | Implemented |
| 30     | Direct Reporting of Foreign Investments                               | Italy         | Use of XBRL to collect information on foreign investments made by Italian firms. Data are used for the compilation of Balance of Payment  | Implemented |
| 31     | Dubai Stock Exchange (DIFX) XBRL Project                              | UAE           | Use of XBRL to collect financial information of listed companies.   | Development |
| 32     | Dutch SBR (was Dutch Taxonomy Project)                                | Netherlands   | Use of XBRL for reporting to government.  | Development |
| 33     | EBA XBRL Project  | Europe        | Use of XBRL by National Supervisory Authorities, coordinated by the European Banking Authority, to collect information from those they regulate.  | Implemented |
| 34     | ECCBSO XBRL Project   | Europe        | Use of XBRL by regulators to collect information from those they regulate   | Development |
| 35     | Electronic reporting of general data identification (GDI)             | Spain         | This taxonomy allows Electronic reporting of general data from entities, individuals and general information structures associated as well as general information of interest according to several Spanish official institutions. | Implemented |
| 36     | Examination Tracking System-Supervisory Applications Generating Exams | United States | Use of XBRL and iXBRL to modernize FDIC's bank examination application program.   | Development |
| 37     | Federal Service Finance - Tax   | Belgium       | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |

| S. No. | Project Name  | Country       | Brief Description  | Status      |
|--------|---|---------------|--|-------------|
|        | administration  |               |  |             |
| 38     | Federation des Experts Comptables Européens XBRL Initiative | Europe        |  | Development |
| 39     | FFIEC Call Report Modernization                             | United States | Use of XBRL by the Federal Financial Institutions Examination Council (FFIEC) to define, collect, validate, and distribute quarterly financial data (Call Reports) from 7300 regulated financial institutions.   | Implemented |
| 40     | Financial Supervisory Authority of Norway XBRL              | Norway        | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |
| 41     | Financial Supervisory Service (Korea) XBRL Project          | Korea         | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 42     | Gepsio  | United States | Gepsio is a .NET-based document object model for XBRL documents. Load your XBRL document with the XbrlDocument class and work with your XBRL document exposed as a set of .NET classes with a variety of properties and methods. Loaded XBRL documents are automatically validated against the information against the XBRL specification, and exceptions are thrown when invalid XBRL documents are loaded. | Development |
| 43     | GRI Taxonomy development project                            | Europe        | The Global Reporting Initiative (GRI) produces the world's most comprehensive sustainability reporting guidelines. In June 2011 GRI and Deloitte Netherlands started work on a new project that aims to publish an XBRL taxonomy covering both the G3 Guidelines and GRI's latest G3.1 Guidelines  | Development |
| 44     | iDataPlatform   | International | Use of XBRL to enable collaboration of analytical models and presentation  | Implemented |



| S. No. | Project Name  | Country       | Brief Description  | Status      |
|--------|---|---------------|--|-------------|
|        |   |               | templates.   |             |
| 45     | IFRS Taxonomy Creation                                    | International | The IFRS Foundation XBRL Team is responsible for developing and maintaining the XBRL representation of the IFRSs, known as the IFRS Taxonomy.                            | Development |
| 46     | India's Listed Companies Filing Using XBRL                | India         | Use of XBRL to collect financial information of listed companies.  | Development |
| 47     | Inland Revenue Department (New Zealand) Project           | New Zealand   | Inland Revenue Department (New Zealand) Project  | Development |
| 48     | Insurance Regulatory and Development Authority (IRDA)     | India         | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 49     | Interactive Data to Improve Financial Reporting           | United States | Collect financial information from approximately 15,000 public companies and 8,000 mutual funds who are regulated by the SEC.  | Implemented |
| 50     | IRAN Securities and Exchange XBRL Project - phase 1       | Iran          | An investigation into XBRL project to determine the vision, goals, strategies and list of action plans to address XBRL successful implementation in Iran capital market. | Implemented |
| 51     | IRAN Securities and Exchange XBRL Project - phase 2       | Iran          | XBRL Taxonomy for listed and unlisted companies  | Prototype   |
| 52     | Irish Financial Service Regulatory Authority XBRL Project | Ireland       | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 53     | Israel Securities Authority XBRL Project                  | Israel        | Use of XBRL to collect financial information of listed companies.  | Implemented |
| 54     | Italian Government XBRL Reporting Requirements            | Italy         | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |

| S. No. | Project Name  | Country       | Brief Description   | Status      |
|--------|---|---------------|---|-------------|
| 55     | KOSDAQ XBRL Expansion Project to Company Level Taxonomy | Korea (South) | Use of XBRL to collect financial information of listed companies.   | Implemented |
| 56     | LENLOC/PENLOC   | Spain         | Since 2007, the General Secretariat for Regional and Local Coordination has received XBRL reports with budget implementation data for municipalities and local authorities (LENLOC), and in 2009 it started receiving budget preparations (PENLOC). | Implemented |
| 57     | Minister of Finance (Brazil) Project                    | Brazil        | Information supply chain which collects information from microfinance institutions.   | Development |
| 58     | Ministry of Company Affairs (India) XBRL Project        | India         | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 59     | Ministry of Corporate Affairs                           | India         | Taxation and Accounting   |             |
| 60     | Monte deiPaschi di Siena - Italy                        | Italy         | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |
| 61     | Mutual Fund Risk Return Summary Taxonomy                | United States | Taxonomy for use by 8,000 mutual funds for reporting to the SEC   | Implemented |
| 62     | National Bank of Poland                                 | Poland        | COREP and FINREP project  | Implemented |
| 63     | National Institute for Statistics                       | Belgium       | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |
| 64     | National Stock Exchange of India XBRL Project           | India         | Use of XBRL to collect financial information of listed companies.   | Development |

| S. No. | Project Name                                   | Country       | Brief Description   | Status      |
|--------|--|---------------|---|-------------|
| 65     | National Tax Agency (NTA) (Japan) XBRL Project | Japan         | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |
| 66     | New Zealand SBR Project                        | New Zealand   | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 67     | New Zealand Stock Exchange Project             | New Zealand   | Use of XBRL to collect financial information of listed companies.   | Development |
| 68     | NOFCAC2010                                     | Spain         | NOFCAC2010 (Spanish GAAP 2007) is used by those companies required to report their annual financial statements to the Business Register according to the Preparation of Consolidated Financial Statements | Implemented |
| 69     | Norway Companies Registrar XBRL Project        | Norway        | Use of XBRL by regulators to collect information from those they regulate.  | Implemented |
| 70     | Office of the Controller (Nevada) XBRL Project | United States | Use of XBRL to transfer information relating to collections of receivables.   | Development |
| 71     | Oslo Stock Exchange XBRL Project               | Norway        | Use of XBRL to collect financial information of listed companies.   | Implemented |
| 72     | Pension Fund Project                           | South Africa  | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 73     | PGC2007  | Spain         | Annual collection of financial reports according to the Spanish GAAP.   | Implemented |

| S. No. | Project Name  | Country       | Brief Description  | Status      |
|--------|---|---------------|--|-------------|
| 74     | Record of Credit Ratings Taxonomy                                   | United States | Completed taxonomy, developed by XBRL US under contract with SEC   | Implemented |
| 75     | Reserve Bank of India XBRL Project                                  | India         | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 76     | Revenue Commissioners XBRL Project                                  | Ireland       | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 77     | Shanghai Stock Exchange XBRL Project                                | China         | Use of XBRL to collect financial information of listed companies.  | Implemented |
| 78     | Shenzhen Stock Exchange XBRL Project                                | China         | Use of XBRL to collect and share reporting information of listed companies.  | Implemented |
| 79     | SIIF- Financial Information Exchange System                         | Spain         | Use of XBRL by regulators to collect information from those they regulate.   | Implemented |
| 80     | Solvency II Implementation @ Autoritatea de Supraveghere Financiara | Romania       | An XBRL compatible Information System - Intema - to be used by the National Supervisory Authority (ASF) for data collection, validation and aggregation from local undertakings, according to Solvency II taxonomy | Pilot       |
| 81     | South African Revenue Service Project                               | South Africa  | Use of XBRL by regulators to collect information from those they regulate.   | Development |
| 82     | Spanish Banking Association (AEB)                                   | Spain         | Publication of selected reports  | Implemented |
| 83     | Spanish Credit Cooperatives (UNCC)                                  | Spain         | Publication of selected reports  | Implemented |
| 84     | Spanish Savings Banks Association                                   | Spain         | Publication of selected reports  | Implemented |

| S. No. | Project Name   | Country            | Brief Description   | Status      |
|--------|--|--------------------|---|-------------|
| 85     | Spanish Securities Commission (CNMV)                           | Spain              | The Spanish Securities Commission (CNMV), as a pioneer in XBRL technology, was the first public institution that massively received and published information in XBRL format. In July 2005, CNMV made mandatory the reporting in XBRL for listed companies. Since 2008, CNMV receives reports in XBRL format from the mutual fund managers, and, since 2009, from the securitization fund management companies. | Implemented |
| 86     | Stock Exchange of Thailand XBRL Project                        | Thailand           | Use of XBRL to collect financial information of listed companies.   | Implemented |
| 87     | Superintendencia de Bancos de Panamá (SBP) - SBP_PA-2012-06-30 | Republic of Panama | The taxonomy is based on the requirements of Superintendency of Banks of Panama for the regulatory reporting in Panama in accordance with IFRS.   | Pilot       |
| 88     | Superintendencia de Valores y Seguros XBRL Project             | Chile              | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 89     | Taiwan Stock Exchange Voluntary Filing Project                 | Taiwan             | Use of XBRL to collect financial information of listed companies.   | Development |
| 90     | Taiwan Stock Exchange XBRL Project                             | Taiwan             | Collect financial information of approximately 2,000 public companies regulated by Financial Supervisory Commission.  | Implemented |
| 91     | The Company House  | United Kingdom     | Developing a taxonomy for the system supporting processes of collecting and publishing financial statements based on Polish GAAP and IFRS   | Development |
| 92     | The Dutch Bank   | Netherlands        | Use of XBRL by regulators to collect  | Implemented |

| S. No. | Project Name  | Country       | Brief Description   | Status      |
|--------|---|---------------|---|-------------|
|        | XBRL Project  |               | information from those they regulate.   |             |
| 93     | Tokyo Stock Exchange XBRL Project                               | Japan         | Use of XBRL to collect financial information of listed companies.   | Implemented |
| 94     | UAE XBRL  | UAE           | The project is meant to develop e-filing for UAE listed companies' financial statements, UAE brokers' prudential reporting and Mutual Funds' application submission | Development |
| 95     | UFOCatcher  | Japan         | Gather XBRL data from JFSA EDINET and TSE TDnet and provide web API and analysis environment over the web.  | Implemented |
| 96     | US GAAP Financial Reporting Taxonomy                            | United States | Completed taxonomy; initial releases created by XBRL US; FASB has responsibility for support and maintenance ongoing  | Implemented |
| 97     | Wacoal ERP Integration  | Japan         | Integration of several ERP systems  | Implemented |
| 98     | Warsaw Stock Exchange Project                                   | Poland        | Use of XBRL by regulators to collect information from those they regulate.  | Development |
| 99     | XBRL and Public Sector Financial Reporting: Oregon CAFR Project | United States | Financial reporting by 88,000 state and local governmental entities within the US.  | Pilot       |
| 100    | XBRL Challenge  | United States | Contest to encourage development of analytical applications that consume XBRL-formatted financial statement data from the SEC EDGAR system                          | Prototype   |
| 101    | XBRL Financial Reporting Prototype                              | International | Prototype to text XBRL for financial reporting in the UK.   | Prototype   |
| 102    | XBRL Financial Statement for Johannesburg Stock Exchange        | South Africa  | Use of XBRL to collect financial information of listed companies.   | Prototype   |

| S. No. | Project Name   | Country | Brief Description  | Status      |
|--------|--|---------|--|-------------|
| 103    | XBRL Project at The Swedish Companies Registration Office (Bolagsverket) | Sweden  | Use of XBRL by regulators to collect information from those they regulate. Actually for SMEs under the Swedish GAAP  | Implemented |
| 104    | XBRL Taxonomy for Banks  | India   | Development of taxonomy to enable banks to prepare their financial statements, i.e., Profit and Loss Account, Balance Sheet and Cash Flow Statement, in XBRL | Development |

\*Source: XBRL International site

## Annex VI

### Quality Assurance Processes – Study of select central banks and other major organizations

#### ECB

ECB has detailed quality assurance procedures in its document “Quality Assurance Procedures within the ECB Statistical Function” . The ECB statistical function’s compilation procedures include several quality checks on the national contributions received and the euro area aggregates compiled. The aim of these checks is to detect problems in the national data which may affect the quality of the euro area aggregates. Additionally, they ensure that the ECB statistical function’s internal processes function in a way that guarantees the accurate compilation and dissemination of euro area aggregates.

These quality checks are grouped into seven main categories:

##### 1. Completeness checks

Completeness checks are carried out to detect missing information. In the event of gaps, representatives of the country involved are contacted and asked to transmit the missing data as quickly as possible.

##### 2. Internal consistency

DG-S verifies that all linear constraints are correctly fulfilled in the data received, for instance whether balance sheets balance and sub-totals add up to the totals.

##### 3. Consistency across frequencies of the same dataset

The ECB statistical function ensures that there is consistency across frequencies of the same dataset, checking, for instance, whether the sum of monthly transaction values equals the quarterly values, or whether the end-year stocks are equal to the end-December stocks

##### 4. External Consistency

The ECB statistical function carries out various checks to assess the consistency of the data received with other datasets. For instance, monetary financial institution (MFI) balance sheet statistics received by the ECB statistical function on the cross-border

positions of euro area banks are compared with similar data collected by the BIS. In the case of balance of payments statistics, the gross flows of the goods account are compared with statistics on external trade in goods as published by Eurostat.

##### 5. Revision Studies

Revision studies are carried out for all types of statistics, although the revision policy is not the same for all datasets. For instance, while MFI statistics (balance sheet and interest rates statistics) may be revised with each new data transmission, balance of payments statistics and financial accounts data are revised according to



a predetermined schedule. All significant revisions are automatically detected by the system and lead to further investigations in cooperation with the representatives of the country concerned. The ECB statistical function also performs a regular in-depth revision analysis for certain euro area aggregates. For instance, it closely monitors the magnitude of data revisions affecting monetary aggregates and their counterparts. In the same vein, the stability of the balance of payments data is assessed annually by analyzing the extent to which the first assessments of these data differ from the final assessments. The results of this analysis are published.

#### 6. Plausibility checks

Plausibility checks aim to detect outliers in the reported data. This is accomplished by reviewing the time series of the variable concerned; for instance, for statistics with a pronounced seasonal pattern, the most recent figure is compared with the data reported for the same period in previous years. Values which markedly deviate from the usual pattern of the series are isolated and analyzed further. In the case of MFI balance sheet statistics, data compilers use ARIMA models to assess the plausibility of new national data received. This approach is also applied to euro area aggregates. Balance of payments data compilers apply some statistical tests to all series (e.g. the identification of outliers through the standardization of the data), but also use tailor-made tests for specific series (e.g. the comparison of portfolio investment flows with leading market indicators).

#### 7. Regular Quality Reporting

Whenever data are received and compiled, quality assessment reports that summarize the results of all the above-mentioned quality checks are circulated to ECB internal users and to the NCBs

#### Process Management

The ECB statistical function has introduced a formal process management framework with the following main objectives: to assess and improve the overall efficiency of statistical processes within the ECB statistical function; to document all statistical processes used within the ECB statistical function in a consistent manner; to implement appropriate risk management and change management procedures

## Bank of England

Bank of England has formulated the Data Quality framework. Quality is one of the key principles set out in the Bank's Statistical Code of Practice. In the Code, quality is defined as the fitness for purpose of published data for users and it encompasses a broad range of criteria. The Code requires that: Quality standards for statistics will be monitored, and reliability indicators for individual data series will be progressively developed and published to assist users, Indicators of quality are being progressively developed for key statistical outputs, Statistical publications and publicly accessible databases will indicate where information on data quality may be found.

The Data Quality Framework is designed to enable users of the Bank's published statistical data to be better informed about aspects of the quality of those data. Information on data quality can have a wide scope: it consists of explanatory material describing the relevance of data, how statistics are compiled, the use of imputation methods, and any other information (including quantitative measures) useful to users in their understanding of what the data represent and how they are constructed. The Framework includes a number of references to explanatory articles relevant to data quality previously published by the Bank's Statistics Division(2) and other bodies. It describes how the Bank's approaches to statistical data quality conform to international standards on quality assessment. Quantitative indicators relating to the coverage and revisions aspects of data quality are defined in the framework.

In April 2013 the Statistics Division took over responsibility for the production of a wide range of regulatory data on banks, building societies, insurance companies, investment firms, credit unions and friendly societies previously undertaken by the Financial Services Authority. The Statistics Division is in the course of developing a similar quality assurance Framework for the regulatory data it processes for the Bank's own and other official uses.

The new dual responsibility of the Statistics Division for both statistical and regulatory data collections opens up possibilities for using data more widely and efficiently across the Bank's several functions: for monetary policy, financial stability and prudential supervision of financial institutions. Data Quality Framework adopts the European Statistical System (ESS) dimensions of statistical data quality, in order to address the requirement in the Bank's Statistical Code of Practice that the Statistics Division should progressively develop and publish quality standards for its statistics. The various data quality dimensions include relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence. Various detailed processes are followed to facilitate achievement of these individual dimensions.

For example, for the parameters on coherence, examples of investigations which promote coherence include:

- Cross-form plausibility checks on interest receivable or payable applied to the effective interest rates return, form ER, and the profit and loss return, form PL.
- Reconciliation of MFIs' loan transfers data with those reported by Specialist Mortgage Lenders and with market intelligence.
- Checks against commercial data sources for quoted interest rates and capital issues data.
- Reconciliation of the inter-MFI difference: ie consistency of the aggregate measures of MFI's assets and liability positions with respect to each other.
- Reconciliation of sectoral and industrial analyses of lending and deposits.
- Comparison of UK Debt Management Office data with MFIs' government debt data.

## OECD

OECD has drafted the “Quality Framework and Guidelines for OECD Statistical Activities”. Given the work already done by several statistical organizations, the OECD drew on their experience and adapted it to the organization’s context. Since several statistical organizations have already identified the dimensions of quality, these have also been adapted to the OECD context. Thus, the OECD views quality in terms of seven dimensions: relevance; accuracy; credibility; timeliness; accessibility; interpretability; and coherence. Another factor is that of cost-efficiency, which though is not a quality dimension, is still an important consideration in the possible application of one or more of the seven dimensions cited previously to OECD statistical output. Detailed processes are built in regard to each of these dimensions.

From example, coherence perspective reflects the degree to which they are logically connected and mutually consistent. Coherence implies that the same term should not be used without explanation for different concepts or data items; that different terms should not be used without explanation for the same concept or data item; and that variations in methodology that might affect data values should not be made without explanation. Coherence in its loosest sense implies the data are "at least reconcilable." For example, if two data series purporting to cover the same phenomena differ, the differences in time of recording, valuation, and coverage should be identified so that the series can be reconciled. Coherence has four important sub-dimensions: within a dataset, across datasets, over time, and across countries.

An overview of procedures for assuring the quality of proposed new statistical activities and for reviewing the quality of the output of existing statistical activities. In addition, the promotion of best practice used in-house and elsewhere is designed to help OECD statisticians adopt the most effective approaches to data and metadata collection, management and dissemination.

### **Procedure for assuring the quality of new activities**

The main steps in the development of a new statistical activity were defined as:

- a) definition of the data requirements in general terms;
- b) evaluation of other data currently available;
- c) planning and design of the statistical activity;
- d) extraction of data and metadata from databases within and external to OECD;
- e) implementation of specific data and metadata collection mechanism;
- f) data and metadata verification, analysis and evaluation; and
- g) data and metadata dissemination.

For each step the quality concerns and the instruments available to help in addressing them were identified. In particular, a set of guidelines and concrete procedures have been prepared for each step, taking into account good existing practices within the OECD and in other statistical agencies. In order to minimize the

burden placed on activity managers, a simplified version of the procedure would be appropriate for statistical activities planned to be once rather than repeated.

### **Procedure for reviewing the quality of existing activities**

The procedure for reviewing the quality of existing statistical activities conducted across the OECD takes into account the fact that the review will be carried out on a rotation basis over a number of years. The stages envisaged are as follows:

- a) identification by the OECD Statistical Policy Group (SPG) of the statistical activities for review during the course of the year, following a biannual rolling calendar;
  - b) self-assessment by the statistical activity manager and staff, resulting in a report that includes a brief summary of quality problems and a prioritized list of possible improvements, together with an assessment of additional resources required for their implementation.
  - c) review of and comments on the self-assessment report by major users;
  - d) review of and comments on the self-assessment report by statistical, information technology, and PAC dissemination staff, co-ordinated by an expert designated by the SPG;
  - e) preparation of the final quality report, combining all comments, jointly by the activity manager and designated expert, and tabling of the report to the SPG;
  - f) discussion and resolution of any concerns about the report by the SPG, and transmission of the report to the relevant director;
- assignment of resources for selected quality improvement initiatives by the directors and through the Central Priorities Fund;
- h) feedback by the Chief Statistician to stakeholders on the quality improvement initiatives proposed and the plans for their implementation.

## **UNECE**

An important international initiative concerning the production of official statistics has been the development of the Generic Statistical Business Process Model (GSBPM), which is sponsored by the United Nations Economic Commission for Europe (UNECE) under its statistical metadata initiative, known as METIS.

| Quality Management / Metadata Management |   |  |                                     |                                     |                                    |   |                                 |
|--|---|--|-------------------------------------|-------------------------------------|------------------------------------|---|---------------------------------|
| Specify Needs                            | Design                                      | Build  | Collect                             | Process                             | Analyse                            | Disseminate                                     | Evaluate                        |
| 1.1<br>Identify needs                    | 2.1<br>Design outputs                       | 3.1<br>Build collection instrument               | 4.1<br>Create frame & select sample | 5.1<br>Integrate data               | 6.1<br>Prepare draft outputs       | 7.1<br>Update output systems                    | 8.1<br>Gather evaluation inputs |
| 1.2<br>Consult & confirm needs           | 2.2<br>Design variable descriptions         | 3.2<br>Build or enhance process components       | 4.2<br>Set up collection            | 5.2<br>Classify & code              | 6.2<br>Validate outputs            | 7.2<br>Produce dissemination products           | 8.2<br>Conduct evaluation       |
| 1.3<br>Establish output objectives       | 2.3<br>Design collection                    | 3.3<br>Build or enhance dissemination components | 4.3<br>Run collection               | 5.3<br>Review & validate            | 6.3<br>Interpret & explain outputs | 7.3<br>Manage release of dissemination products | 8.3<br>Agree an action plan     |
| 1.4<br>Identify concepts                 | 2.4<br>Design frame & sample                | 3.4<br>Configure workflows                       | 4.4<br>Finalise collection          | 5.4<br>Edit & impute                | 6.4<br>Apply disclosure control    | 7.4<br>Promote dissemination products           |                                 |
| 1.5<br>Check data availability           | 2.5<br>Design processing & analysis         | 3.5<br>Test production system                    |                                     | 5.5<br>Derive new variables & units | 6.5<br>Finalise outputs            | 7.5<br>Manage user support                      |                                 |
| 1.6<br>Prepare business case             | 2.6<br>Design production systems & workflow | 3.6<br>Test statistical business process         |                                     | 5.6<br>Calculate weights            |                                    |   |                                 |
|  |   | 3.7<br>Finalise production system                |                                     | 5.7<br>Calculate aggregates         |                                    |   |                                 |
|  |   |  |                                     | 5.8<br>Finalise data files          |                                    |   |                                 |

(Source:UNECE)

The GSBPM attempts a comprehensive itemization of all stages in the production of good official statistics from, for example, determining users' information needs to imputation, validation and archiving of output releases, and to the post-evaluation action plan. In all there are 47 of these stages under nine headings, as set out in the figure below. But the GSBPM is not a rigid framework in which all steps must be followed in a strict order; rather, it is an elaboration of all the possible steps and the interdependencies between them. Hence, the GSBPM is applied and interpreted flexibly.

## **Annex VII**

### **Suggestions on timeframe for implementation of recommendations**

#### **Short term:**

##### **(I) Data standards**

- 1) Regulatory data reporting standards - XBRL may be the reporting standard for all regulatory reporting of structured data.
- 2) The well known Statistical Data and Metadata eXchange (SDMX) is recommended as a standard for exchanging statistical data.
- 3) System of National Accounts (SNA) 2008 can be used for classification of institutional categories and International Standard Industrial Classification (ISIC)/ National Industrial Classification (NIC) codes for economic activity being financed by a loan may be incorporated.
- 4) In order to take up data element/return standardisation through standardising or harmonising definitions, efforts of earlier working groups(Committee on rationalisation of returns and Committee on harmonisation of banking statistics) can be consolidated by setting up an inter-departmental project group within RBI which can work in a project mode so as to ensure comprehensive and effective implementation of standardisation and consistency of data element definitions across complete universe of returns/data requirements of RBI.

##### **II. Key components of Data governance architecture in banks**

- 1) Issuing guidelines on key components of data governance architecture in banks with focus on the various aspects recommended by the Committee.
- 2) Guidance on Best practices on data governance and information management can be formulated by IDRBT.

##### **III. Data standardization in regulatory reporting–Commercial banks, UCBs, NBFCs**

- 1) Robust internal governance structure needs to be set up in regulatory entities with clear responsibilities and accountabilities to ensure correct, complete and timely submission of regulatory/supervisory returns.

#### **IV. ADF implementation by banks**

- 1) **Use of ADF for Internal MIS** - The RBI Approach Paper highlighted usage of ADF platform for generating internal MIS as one of the key benefits of ADF. In this regard, banks may explore aligning the platform for generating internal MIS and other uses, if not done already. Indicatively some aspects include :
  - NPA Management Automation Module
  - Automation of SLBC Returns through the same platform
- 2) Detailed survey can be carried out by RBI to ascertain the status of ADF implementation by banks. Feedback may also be obtained from DBS regarding any issues relating to ADF implementation obtained during AFI/RBS examination process. Independent assurance on the ADF central repository mechanism in individual banks may also be verified. This would enable assessment of the quality and comprehensiveness of ADF implementation by individual banks. Any specific issues may be taken up with concerned banks for remediation.
- 3) Banks may also port the necessary details required by RBI under Guidelines on “Framework for Revitalizing Distressed Assets in the Economy - Guidelines on Joint Lenders' Forum (JLF) and Corrective Action Plan (CAP)” in February, 2014 under ADF central repository platform.
- 4) Depending on the requirement of RBI regarding granularity of data, ADF system needs to be suitably updated to provide for the requisite granular data fields at the central repository level.

#### **V.XBRL Project of RBI**

- 1) Similar forms can be taken together within/ across the departments of RBI and thus common reporting elements can be arrived at. Rationalisation /Consolidation of returns before taking up the returns pertaining to a department must be done. The rationalisation / consolidation of returns may be examined and reviewed on a periodic basis.
- 2) An Inter-Departmental Data Governance Group (DGG) for the RBI as a whole may be formed, so that the process of rationalization regarding data elements, periodicity, need for provisional returns can be carried out in a concerted manner. All future returns to be prescribed by any department may be routed through the DGG, to avoid duplication.
- 3) As part of its data governance activities, the DGG may also pro-actively identify any data gaps in the evolving milieu and prepare plan of action to address the gap.
- 4) The XBRL taxonomy must include data definitions so as to completely leverage the utility offered by XBRL.

- 5) The XBRL taxonomy should be designed flexibly so as to take care of the anticipated changes in the format of return, i.e., addition and deletion of data elements.
- 6) The XBRL based submission by financial companies to MCA should be shared across the regulators as required.
- 7) Since new tools/software are developed for leveraging XBRL, there needs to be process of continuous monitoring of new developments so as to examine their utility and possible value addition

## **VI. Recommendations on data quality assurance process**

- 1) Exclusive data quality assurance function can be created under the information management unit of RBI.
- 2) A data quality assurance framework may be formulated by RBI detailing the key data quality dimensions and systematic processes to be followed. The various key dimensions include relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence. The framework may also be periodically reviewed.
- 3) Various validation checks like sequence check, range check, limit check, existence check, duplicate check, completeness check, logical relationship check, plausibility checks, outlier checks are among the key checks which need to be considered and documented for various datasets with assistance from domain specialists.
- 4) Whenever data are received and compiled, quality assessment reports that summarize the results of various quality checks may also be generated internally.

## **VII. System-wide Improvements:**

- 1) A separate standing unit Financial Data Standards and Research Group may be considered with involvement of various stakeholders like RBI, IBA, banks, ICAI, IDRBT, SEBI, MCA, NIBM, CAFRAL etc for looking at the financial data elements/standards and to try to bring them into holistic data models apart from mapping with applicable international standards.

## **VIII. Future trend and developments**

- 1) Committee recommends that research/assessment of new developments in technology and financial data/technology standards need to be made a formal and integral part of the information system governance of banks and the regulator.



## **Medium Term**

### **(I) Data standards**

- 1) The ISO 20022 standard is recommended to be the messaging standards for the critical payment systems.
- 2) In regard to standardisation of coding structures, in accordance with international standards, the ISO 4217 currency codes and ISO 3166 country codes can be used.
- 3) Given the LEI initiative, efforts to facilitate LEI for legal entities involved in financial transactions across financial system needs to be expedited to maximise coverage over the medium term.
- 4) The utility may also need to be effectively leveraged to map the corporate group hierarchy.
- 5) While presently LEI caters to legal entities involved in financial transactions, ultimately LEI or similar system needs to be made broad-based to incorporate other categories of customers like partnership firms and individuals.

### **II. Key components of Data governance architecture in banks**

- 1) Committee also identified an illustrative list of key data aspects relating to credit function that would need to be addressed by banks to facilitate standardization and data comparability across banks.
- 2) Apart from providing enhanced focus during AFI/RBS, data governance mechanisms in banks may also be examined intensively through focussed thematic reviews by DBS of RBI. Based on outcome of thematic reviews, detailed guidance may be issued to banks to address issues identified during review.
- 3) While specifying key regulations, RBI may also endeavour to specify any key system related validation parameters and details of data quality dimensions expected from concerned regulated entities.
- 4) Banks can also endeavour to establish a centralised analytics team as a centre of excellence in pattern recognition technology and artificial intelligence (AI) to provide cutting edge analysis and database tools or information management tools to support business decisions.
- 5) RBI may facilitate creation of Data Governance Forum under the aegis of IBA or learning institutions like CAFRAL or NIBM with other stakeholders like IDRBT, RBI, IBA, banking industry technology consortiums, banks, to assist in

development of common taxonomies/data definitions/meta data for banking system.

- 6) Bank Technology Consortia under the aegis of IDRBT and other stakeholders like banks can validate critical banking applications like CBS and provide guidance on expected minimum key information requirements/validation rules and address to the extent possible different customizations across banks.

### **III. Data standardization in regulatory reporting—Commercial banks, UCBs, NBFCs**

- 1) XBRL platform may be gradually expanded across the full set of regulatory returns.
- 2) Regulatory reporting - Commercial Banks:
  - a. Adoption of uniform codes among different returns of RBI will reduce inconsistency among returns
  - b. The BSR codes need to be updated based on latest NIC 2008 classification. The BSR codes may be reviewed periodically and updated. Further, it should be possible to establish one-to-one mapping of sector/ industry codes in various other regulatory returns from the same.
  - c. The nature of returns are generally dimensional in nature, consisting of various components like measures, concepts, elements, attributes, dimensions and distributions. A suitable data model may be generated to facilitate element-based, simplified and standardised data collection process by RBI under a generic model structure that is suitable for both primary and secondary data.
  - d. There is a need to ultimately move over to “data” centric approach from the current “form” centric approach
  - e. The values of various attributes and dimensions should be standardised to enable the collation of data from different domains.
  - f. Suitable data sets with varied nature like hierarchical, distributional or dimensional can be created to facilitate submission of data in summarised or granular form as the case may be from the central repository of the banks.
  - g. Phased implementation of various standardised data definitions can be commenced based on elements which were already standardised.
- 3) NBFCs/UCBs:
  - a. Rationalisation of returns needs to be attempted for NBFCs and UCBs.
  - b. The Committee recommends an online data collection mechanism for larger NBFCs and Tier II UCBs.
  - c. In due course, after rationalisation exercise, data element based return submission may also be initiated.

d. Suitable data model and robust meta data system may be developed.

#### **IV. ADF implementation by banks**

- 1) Banks may evaluate and take steps to enable the ADF Platform to cater to the Risk Based Supervision(RBS) data requirements by suitably mapping the RBS data point requirements.
- 2) Banks may also port the necessary details required by RBI under Guidelines on “Framework for Revitalizing Distressed Assets in the Economy - Guidelines on Joint Lenders’ Forum (JLF) and Corrective Action Plan (CAP)” in February, 2014 under ADF central repository platform.
- 3) Depending on the requirement of RBI regarding granularity of data, ADF system needs to be suitably updated to provide for the requisite granular data fields at the central repository level. The ADF system of the banks should be designed flexibly to accommodate any anticipated changes in the format of return, i.e., addition and deletion of data elements.

#### **V.XBRL Project of RBI**

- 1) For granular account level data and transactional multi-dimensional data, RBI may develop and provide specific details of RDBMS/text file structures along with standardised code lists and basic validation rules so that banks can run the validation logics to ascertain that the datasets are submission-ready. In this connection, XBRL based data element submission may also be explored.
- 2) In due course, moving from return based approach to data element based approach needs to be considered.
- 3) It is expected that banks would generate the instance document from the Centralised Data Repositories (CDR) and submit the same to RBI without manual intervention. The banks should validate the generated instance documents based on the XBRL taxonomy and validation rules before sending them to the Reserve Bank. Thus, the present approach of spreadsheet(Excel) based submission of returns needs to be given up ultimately.
- 4) The XBRL taxonomy should be designed flexibly so as to take care of the anticipated changes in the format of return, i.e., addition and deletion of data elements.
- 5) Therefore the existing Data Ware House needs to be revamped with Next Generation Data Ware House capabilities. Big Data solutions also need to be explored for enhancing analytical capability in the new data paradigm which would be of particular use in areas like banking supervision.

## **VI. Recommendations on Automating data flow from banks to RBI**

- 1) Using secure network connections between the RBI server and the bank's ADF server, the contents of the dataset can be either pulled through ETL mode or pushed through SFTP mode and loaded onto the RBI server automatically as per the periodicity without any manual intervention. Pushing of data by banks could enable easier management of the process at RBI end. An acknowledgement or the result of the loading process can be automatically communicated to the bank's ADF team for action, if necessary.
- 2) The validation schemes may also be expressed in XBRL/XML form so that the systems at banks automatically understand the requirement, accordingly process their data and return the data to RBI, without any manual intervention. This would enable a fully automated data flow from banks to RBI even with dynamic and changing validation criteria.
- 3) While the traditional RDBMS infrastructure in place in RBI may be used for storage and retrieval of aggregated and finalized data, Big-data solutions may also be considered for micro and transactional datasets given their high volume, velocity and multi-dimensional nature.
- 4) The enterprise-wide data warehouse (EDW) of RBI should be made the single repository for regulatory/supervisory data pertaining to all regulated entities of RBI with appropriate access rights. Any unstructured components pertaining to RBS data may be maintained in EDW using new tools available for such items.
- 5) As a key support for risk based supervision for commercial banks, internal RBI MIS solution needs to seamlessly generate two important sets of collated information: (i) Risk Profile of banks (risk-related data – mostly new data elements), and (ii) Bank Profile (mostly financial data – DSB Returns and additional granular data) based on data supplied by banks.
- 6) Once the system stabilises, the periodicity of data can be reviewed to examine obtain any particular set of data at shorter intervals or even up to near real time.

## **VII. Recommendations on data quality assurance process**

- 1) Deployment of data quality tools as part of the data warehouse infrastructure could also provide for comprehensive assessment of data quality dimensions.

## **VIII. System-wide Improvements:**

- 1) Inter-regulatory forums could help facilitate improvements in data/information management standards across the financial sector to benefit all stakeholders and furthering collaboration with international stakeholders.
- 2) A separate standing unit Financial Data Standards and Research Group may be considered with involvement of various stakeholders like RBI, IBA, banks, ICAI, IDRBT, SEBI, MCA, NIBM, CAFRAL etc for looking at the financial data elements/standards and to try to bring them into holistic data models apart from mapping with applicable international standards.
- 3) Regulators like RBI, SEBI, MCA are in the process of undertaking various XBRL projects. Given the benefits offered by XBRL and its usage across the globe by regulatory bodies, all the regulators may explore possibilities of commonalities in taxonomy and data elements and protocols and formats for sharing of the data among themselves.
- 4) In respect of knowledge sharing and research, various measures recommended include (i) Research by IDRBT regarding ways and means of leveraging new data technological platforms like XBRL for enhancing overall efficiencies of banking system (ii) conducting of pilot for enhancing leveraging of technologies like XBRL for internal uses by banks.
- 5) As the leveraging of machine readable tagged data reporting increases, the audit and assurance paradigm also need to get re-engineered to carry out an electronic audit and electronic stamp of certification using digital signatures.
- 6) Committee recognizes that coordinated efforts are being carried out by various organizations which have developed standards like FIX, FpML, XBRL, ISD etc for laying the groundwork for defining a common underlying financial model based on ISO 20022 standard. Costs of migration and inter-operability would be key factors going forward.
- 7) There is also need to incorporate training and education on the new technologies like XBRL by various academic bodies as also training/learning institutions so as to help in capacity building and to improve the availability of trained resources.

## **IX. Future trend and developments**

- 1) Banking technology research institute IDRBT may carry out research on new technologies/development and serve as a think tank in this regard.
- 2) Banks may explore Big Data solutions for leveraging various benefits of the new paradigm concerned with volume and velocity of data.

## **Long Term**

### **I. Data standardization in regulatory reporting—Commercial banks, UCBs, NBFCs**

- a. NBFCs/UCBs - In due course, after rationalisation exercise, data element based return submission may be initiated.

### **II. Recommendations on Automating data flow from banks to RBI**

- 1) The enterprise-wide data warehouse (EDW) of RBI should be made the single repository for regulatory/supervisory data pertaining to all regulated entities of RBI with appropriate access rights. Any unstructured components pertaining to RBS data may be maintained in EDW using new tools available for such items.
- 2) Once the system stabilises, the periodicity of data can be reviewed to examine obtaining any particular set of data at shorter intervals or even up to near real time as required.

### **III. System-wide Improvements:**

- 1) Ultimately, from a banking system perspective full benefit would arise by enabling transactional and accounting systems in banks to directly tag and output data in formats like XBRL to maximize efficiency and benefit. Thus, there is need for integration of standard formats like XBRL in internal applications/accounting systems of banks. The present scope of XBRL data definitions have to be further extended to cover in depth data definitions covering almost all data elements that are required to carry banking business.
- 2) Standard Business Reporting, which involves leveraging technologies like XBRL by Government for larger benefits beyond the field of regulatory reporting, is being implemented in various countries like Australia and Netherlands. The same may be explored in India by Government of India in a phased manner.
- 3) While the G-20 Data Gap initiative is a work in progress, there is also need for international stakeholders to analyse and examine how technologies like XBRL can help facilitate ease of comparability of data as also to identify differences between countries in respect of financial reporting rules in an automated manner.
- 4) Akin to initiatives in US, financial instrument reference database could be explored with focus on key components relating to ontology, identifiers and metadata and valuation and analytical tools.

- 5) A single Data Point Model or methodology at international level can be explored for the elaboration and documentation of XBRL taxonomies.

## References

1. FSB and IMF, The Financial Crisis and Information Gaps (2009)
2. Formalizing and Securing Relationships on Public Networks by Nick Szabo <http://szabo.best.vwh.net/formalize.html>
3. Data-driven Validation Rules: Custom Data Validation Without Custom programming by Don Hopkins, Ursa Logic Corporation, Durham, NC <http://analytics.ncsu.edu/sesug/2003/DM02-Hopkins.pdf>
4. Universal Data Models for Financial Services By Len Silverston [http://www.univdata.com/Portals/9/Publication\\_Articles\\_7\\_02\\_Financial\\_Serv.pdf](http://www.univdata.com/Portals/9/Publication_Articles_7_02_Financial_Serv.pdf)
5. Literature Review of Data Validation Methods <http://www.prepared-fp7.eu/viewer/file.aspx?fileinfoID=215>
6. Automatic Exchange of Information What it is, How it works, Benefits, What remains to be done, OECD
7. Improving transparency in financial and business reporting — Harmonisation topics (CEN/WS XBRL) (2013)
8. Reserve Bank of India (2010a): Approach Paper on Automated Data Flow from Banks.
9. Reserve Bank of India (2010b): Report of the High Level Committee for Preparation of the Information Technology Vision Document 2011-2017
10. Investment Roadmap(2010) – FIX, FpML, ISITC, XBRL, SWIFT, SIIF
11. IMF Working Paper: Why are the G-20 Data Gaps Initiative and the SDDS Plus Relevant for Financial Stability Analysis? - Robert Heath (2013)
12. SDMX (2011): SDMX-ML: Schema and Documentation, version 2.0; Available at [http://sdmx.org/docs/2\\_0/SDMX\\_2\\_0\\_SECTION\\_03A\\_SDMX\\_ML.pdf](http://sdmx.org/docs/2_0/SDMX_2_0_SECTION_03A_SDMX_ML.pdf)
13. Working Group CWA1 Draft Experts: Ignacio Santos (Bank of Spain), Roland Hommes (Rhocon), Katrin Heinze (Deutsche Bundesbank) - DPM2MDM Workshop Group at <http://www.wikixbrl.info/index.php?title=DPM2MDM>
14. Semantic Web at [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)
15. W3C, <http://www.w3.org/standards/semanticweb/>
16. BIS - Principles for effective risk data aggregation and risk reporting (2013)
17. (Margarita S.Brose, Mark D.Flood, Dilip Krishna and Bill Nichols- Ed) - Handbook of Financial Data and Risk Information – Cambridge University Press (2014)
18. Report of the Committee on Data and Information Management in the Reserve Bank of India (2014), RBI (Chairman- Shri.Deepak Mohanty)
19. Report of the Working Group on Harmonisation of Banking Statistics, RBI, 2014 (Chairman – Shri.A. B. Chakraborty)
20. OECD – Guidance Note – Standard Business Reporting (2009)
21. OECD - Data and Metadata Reporting and Presentation Handbook (2007)

22. Charles Hoffman, CPA and Raynier van Egmond – Digital Financial Reporting – Using XBRL based model
23. KPMG , Information enabled Banking - Information Management Survey - South based private sector banks (2013)
24. KPMG, Improving regulatory reporting – realizing the benefits of XBRL (2004)
25. Margarita S. Brose, Mark D. Flood, David M. Rowe - Risk Management Data and Information for Improved Insight (2012)
26. Josephine Wapakabulo Thomas- Data-exchange standards and international organizations : adoption and diffusion - Information Science Reference (an imprint of IGI Global) (2010)
27. Srilatha Kappagantula, Infosys Thought Paper, Messaging Standards in financial industry (2012)
28. Economist Intelligence Unit – Big Data as key to better risk management (2014)
29. Oliver Wyman - Managing Complexity – State of the Financial Services Industry (2015)
30. Office of Financial Research, USA, Annual report (2014)
31. BIS, Progress in adopting the principles for effective risk data aggregation and risk reporting(December, 2013)
32. Drejer, P. A. (2013): Optimizing Checking of Statistical Reports, IFC Working Papers No 11, Irving Fisher Committee on Central Bank Statistics, Bank for International Settlement.
33. European Central Bank (2008): Quality Assurance Procedures within the ECB statistical function, ECB website.
34. <http://www.ecb.europa.eu/pub/pdf/other/ecbstatisticsqualityassuranceprocedure200804en.pdf>
35. Bank of England(2014), Statistics and Regulatory Data Division: Data Quality Framework
36. UNECE (2013),The Generic Statistical Business Process Model, available at
37. [www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model](http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model).
38. FSB—Understanding Financial Linkages: A Common Data Template for Global Systemically Important Bank (2011)
39. Senior Supervisors Group (2010), Observations on Developments in Risk Appetite Frameworks and IT Infrastructure
40. XBRL India, Training material on XBRL at [www.mca.gov.in/XBRL/pdf/training\\_material.pdf](http://www.mca.gov.in/XBRL/pdf/training_material.pdf)