

'Banking on Quality Data'¹

Dr. Bhaskaran, CEO, Indian Institute of Banking and Finance, delegates attending this Conference, ladies and gentlemen. I am grateful to Dr. Bhaskaran for inviting me to inaugurate this Conference. The theme 'data quality management' is topical and of prime importance to the financial sector as banking increasingly gets embedded with technology, from micro devices to stacks of Mainframe computers.

Introduction

2. Before I go on to the core theme of my speech, let me cite four examples suggesting the magnitude of data:

3. When the Sloan Digital Sky Survey² ("Data data everywhere"- The Economist) started work in 2000, its telescope in New Mexico collected more data in its first few weeks than had been amassed in the entire history of astronomy. Now, more than a decade later, its archive contains a more than whopping 140 terabytes of information. A successor, the Large Synoptic Survey Telescope, due to come on stream in Chile in 2016, will acquire that quantity of data every five days.

4. Such astronomical amounts of information can be found closer to Earth too. Walmart, a retail giant, handles more than 1m customer transactions every hour, feeding databases estimated at more than 2.5 petabytes—the equivalent of 167 times the books in America's Library of Congress.

5. Facebook, the social-networking website, is home to more than 40 billion photos.

¹ Inaugural address by Shri G. Padmanabhan, Executive Director, RBI at the Conference on Data Quality Management organised by the Indian Institute of Banking and Finance on October 18, 2013. Assistance provided by Smt Nikhila Koduri, Shri A. Madhavan and Smt Radha Somakumar is gratefully acknowledged.

² A special report on Managing information- February 27, 2010 edition in "The Economist"

6. Decoding the human genome involves analysing 3 billion base pairs—which took ten years the first time it was done in 2003, can now be achieved in one week.

7. All these tell the same story: that the world contains an unimaginably vast amount of digital information, and is expanding rapidly. This provides an opportunity if the huge amount of data which is raw is processed, organised, structured and in short converted into information. The all encompassing digital world makes it possible to do many things that previously could not be done: spot business trends, prevent diseases, combat crime and so on. Managed well, the data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account. But, if innovation is not used with technology based analytics, then the vast amount of collected data becomes a digital heap, a sheer information overload.

8. You would all appreciate that banking is a data-driven business, and banks are the trusted custodian of a plethora of customer and corporate data. In a marketplace that is highly competitive, commoditized, heavily regulated, and facing continued economic challenges, success to banks comes from differentiation enabled by the ability to do better analysis and conversion of data to error-free information in a time-bound manner. Modern banking cannot be run without storing all banking sensitive transactional data in standard error free formats, deploying strategically designed analytics, and extracting decision supporting inferences through technology tools. Such an approach alone would differentiate a true market leader from the rest of the pack.

9. Understanding the importance of data, I am reminded of Arthur Conan Doyle's Sherlock Holmes when he says 'It is a capital mistake to theorize before one has data'.

10. In my address today, I would like to draw your attention to two facets of data management - one related to information and the other that is often taken for granted i.e. related to technology. I would attempt to give an overview of these aspects and their significance in the present day context. While covering these topics, I would touch upon the initiatives taken by the RBI in this regard, the future road map for banks and conclude posing a few questions that may be deliberated further.

11. One issue that banks face is the single-version of the truth - a single representation of critical data such as customers, products, assets, that is unique, complete, and consistent, which becomes the most reliable and authoritative information for the entire bank. With data that is duplicated, inconsistent, incomplete, and spread in silos across disparate systems, special techniques are required to recognize (disparate data), resolve (into a single version of the truth), and relate (them to derive some meaning) the data. In addition to this, you would also need to persist with a single version of the truth, store the complete history of all changes, and lineage of how duplicate data has been merged. In banks, the fact that data is crucial to decision-making and planning is taken as given but what about its quality? And why is it important to have data of good quality? Well the answer is simple though the solution may not be. The answer is: we need good quality data which is reliable, available, auditable and correct to help us in taking appropriate timely decisions. In modern governance, the regulatory returns have acquired different significance. Unlike the bygone era, the returns are not merely used for tracking a

bank's performance but also as forerunners for understanding market trends. It is this macro-prudential analysis that has gained global attention. But the end technological solution to achieve this is a long climb!

Dilemma faced by banks

12. It is frustrating for banks when they do not see the kinds of returns generally expected from heavy investments in IT/IT enabled initiatives like the BI or CRM. In most of the cases, it is the data that is often put to blame. In most cases, banks may not be even aware of the magnitude of the problem and the criticality of the issue. When the identification and correction of the problem does not take place early, the risk of defective data contaminating critical information assets is very high, leading to escalation of costs, jeopardising customer relationships, and most importantly, imprecise forecasts leading to improper decisions.

Do banks manage to give the right information?

13. Before I speak on the procedures adopted by banks to produce those lucid documents which make a good reading, I am reminded of a quote of Ronald Coase, a Nobel Laureate in the field of Economics (1991), who said 'Torture the data, and it will confess to anything', which may have a humorous connotation to some, but has a deeper meaning when we analyse the sentence. Do we indulge in 'torturing the data'? By stretching the definition, can I interpret 'torturing the data' as data massaging? Sometimes this process is also referred to as "ETL" meaning "Extract, Transform, Load". Massaging the data is the "transform" step, but it implies ad-hoc fixes that is done to smoothen out problems rather than transformations between well-known formats; but data quality can become a casualty if such transformation

meanders into a mundane mechanical exercise. IDRBT is compiling a Handbook on 'Data Quality in Indian banking Industry: Issues, remedies and impacts' dealing with issues relating to data quality and has suggested a new framework towards achieving this; this may prove beneficial for banks.

14. Though one may be tempted to think that a term like Data Management is more apt for technologically agile sectors like IT, the truth is that good data management practices can bring a huge benefit to the banking industry as well. Further, the benefits arising out of this are not generic in nature, but rather deal with the core challenge today's financial industry is facing – regulatory compliance.

Defining Data Quality

15. The broad components of data quality are precision, consistency, completeness, reliability and accessibility of standardised data across variety of platforms. But how do we define Data Quality? The meaning of the word 'quality' depends on the context in which it is applied. We generally use it to indicate the superiority of a manufactured good or atleast to a high degree of craftsmanship or artistry. However, it is difficult to define quality for data. Unlike manufactured products, data does not have the physical characteristics that allow quality to be easily assessed. Quality has therefore to be seen as a function of certain intangible properties such as 'completeness' and 'consistency'.

16. The question that logically follows is what is it that makes the quality of data a suspect? There could be variety of reasons such as duplication of data, maintenance of multiple data sets/standards, sluggish response and time to retrieve, polish and

make it presentable, creation of redundancy within systems etc. The poor data quality always necessitates reconciliation, manual process of neutralisation or 'torture the data' before it could be used. This low reliability, on account of poor quality of data often leads to time and cost over runs to reaffirm the reliability. The poor quality of data not only emanates from IT but from business streams as well. This poor quality is further sustained by ambiguous roles and responsibilities in data management governance and oversight. This cost, both in monetary and time terms, impairs the business decision or the missed opportunity ceded to a better prepared competitor.

17. The basic issue as regards the computerisation of the Indian banking system was that we migrated from the branch banking software to a core banking solution in which every activity/transaction processing is centralised. At the time of migration, since a variety of platforms were used for capturing different business activities, the data was naturally non-standard and when this data was imported ipso facto, as part of core banking solution, with a bit of data sprucing, the non-standard data needed some 'treatment' before it could be used across all activities. The problem became more complex as the banks were migrating to a non data centre environment; without any business process re-engineering. Therefore, the bulk of data could not be used straightaway in a seamless Decision Support System (DSS). Having done this now, and also subsequently having established a robust data warehouse model, the natural attention is now to deploy a state of art Business Intelligence (BI) software that could double up as DSS.

Data Quality and Basel II

18. Data quality is a global issue and not just restricted to any geographical or economic jurisdiction. A survey focusing on data quality and approaches to data management sponsored by the Risk Management Association and Automated Financial Systems, Inc. (November 2012, The RMA Journal) revealed that only 46% of those surveyed rated the quality of data within their institutions as either excellent or above average. Other important findings of this survey were:

- Regulatory requirements were more of a driver for large institutions' data cleanup initiatives, while community banks were driven more by risk management and quality initiatives.
- Banks around the world were beginning to assign data management responsibilities to senior-level position.
- All banks, regardless of size, reported that accurate data is essential for more timely identification of emerging problems and for reducing regulatory and legal risks.
- the lack of integration of risk and financial-related reporting, as well as the lack of flexibility to extend the current systems, created the most concern for large institutions across the board.

19. The Pillar II Accord of Basel II vests the responsibility of data quality and data management on the financial institutions. Banking business must at all times ensure accuracy of credit risk exposure calculations. This becomes all the more challenging for international banks that have global footprints. Several jurisdictions have made it a requirement to certify the data quality. Banks under certain jurisdictions need to self-certify the accuracy, completeness, and appropriateness of Basel-critical data.

20. Reserve Bank on its part, had issued a notification laying down a time schedule for all scheduled commercial banks operating in the country for implementation of the advanced approaches for the regulatory capital measurement under Basel II framework. In its detailed IRB guidelines released in December 2011, the Bank had laid stress on the data part for most of the parameters that go into the calculation metrics, be it sovereign exposure, bank exposure, mapping to external data for PD estimation etc.

21. One of the seminal guidelines in the recent years has been brought out by the European Central Bank (ECB) on data quality management framework for Centralised Securities Database (CSDB) (September 2012). ECB has prescribed a data quality framework in order to ensure the quality of data output. The ECB states “This Guideline establishes a framework for DQM in the CSDB, the aim of which is to ensure completeness, accuracy and consistency of output data in CSDB by consistently applying rules on quality standards for such data”. The framework deals in greater detail the DQM targets, implementation, attributes and DQM threshold basis.

22. Given these parameters, let us try to understand Data Quality Management and its significance. An important step in positioning for efficiency in bank operations is to evaluate the quality of data and the data-management practices in use. In order for a data quality management initiative to succeed, what is most important is a strong partnership between technology and business groups. Having won half the battle, Indian banks can now work towards finalising the finer details of the approach i.e. putting in place its reactive and proactive elements. While the proactive elements would include establishment of an entire governance structure, identification of roles

and responsibilities, creation of quality expectations as well as supporting business strategies and implementation of a technical platform that facilitate these business practices; the reactive elements would include management of issues in data located in existing databases.

23. We must understand that DQM is not an event by itself but a journey which needs to be sustained. Only then can we have improved data in a holistic, cross functional way. A systematic roll out of Data Quality strategies and processes may help banks to deliver trustworthy, strategic and tactical information to its users. What is also needed is creating cadres of data scientists like data miners, data architects, data stewards, data quality managers etc. thus developing a cadre of information officers just as we develop a cadre of technology officers.

24. Switching gears, allow me to go into another realm-- the work done by RBI with regard to data, standards and dataflow.

RBI's initiatives

Data Quality and Data Standards

25. In its IT Vision 2011-17, the Reserve Bank of India has emphasised the importance to both quality and timeliness of data and its processing into useful information for MIS and decision making purposes. For achieving this objective, it is pertinent that uniform data reporting standards are put in place. We recognise that use of uniform reporting standards for data collection process will effectively reduce the reporting burden, ease validation, and improve overall efficiency. To ensure smooth flow of quality data in a timely manner to the users, we have mandated that:

- Uniform data reporting standards be developed

- Data flow be automated from the source systems of banks to their MIS server and submitted to the Reserve Bank in an automated manner without any manual intervention.

XBRL and ISO 20022

26. Reserve Bank has also made some pioneering efforts in the field of reporting standards. Basically two significant initiatives i.e. XBRL project and adoption of ISO standards for RTGS messages. Within the Reserve Bank, XBRL has been viewed as a natural evolution of its existing Online Returns Filing System (ORFS). While ORFS does the job of data capturing and transmission of returns from banks to the Reserve Bank, it incorporates no in-built standardization. XBRL enables standardization and rationalization of elements of different returns using internationally recognised best practices in electronic transmission. In the process, XBRL also facilitates rationalisation of number of returns to be submitted by the banks, thus reducing their reporting burden. As far as the second initiative is concerned, the new RTGS that is being launched tomorrow is using ISO 20022 standards. This is first instance across the globe for high value payment systems.

Automated Data Flow (ADF)

27. Another effort that has borne fruit is the ADF Project. Banks have undertaken to bring returns to be submitted to the Reserve Bank under Automated Data Flow (ADF). A majority of banks have implemented suitable solutions to generate all the returns to be submitted to the Reserve Bank. By implementing this, commercial banks are able to automate the flow of data from their CBS or other IT systems to the Reserve Bank in a straight through manner without any manual intervention. Banks have adopted different strategies for putting in place systems and processes

to achieve the above. Reserve Bank is closely monitoring the progress of implementation of this Project. This will benefit the banks as also the regulator by enabling more informed policy decisions and better regulation. Over and above the regulatory benefits, there are many other benefits as well to the internal operations, efficiency and management of banks. All these would fructify if only the data quality is assured at every stage by every stakeholder at every point of time.

Technology story of data

28. The other facet of data management relates to the technology aspect, which is generally taken for granted. It is here that I want to highlight a few issues which are of prime importance to business continuity of applications. I am sure you all must be aware of the Principles of Financial Market Infrastructure (PMFI). Principle 17 states that Business continuity management should aim for timely recovery of operations and fulfillment of the FMI's obligations, including in the event of a wide-scale or major disruption. It speaks of having in place a robust BCP, ensure good crisis management and communication, have an adequate secondary site as also review and testing of business continuity arrangements. Therefore what is required is a plan that ensures continuity of application and the associated database. A golden copy perhaps!

29. Here I would like to quote a recent incident regarding an internal application of an organisation well known to me. All applications in that organisation are housed at the Data Centres with complete data backup at the DR sites; the timing of the backup depending on the criticality of the application. In the instant case, due to an unexpected error (from the outsourced vendor personnel), the entire data from the primary site was deleted during a routine housekeeping operation. However all

efforts to ensure restoration from the DR site were not successful as there was a flaw in the replication rules. This meant downtime of an entire application for almost a week and redoing the data entry. Thus even keeping a golden copy was not of use. What I want to highlight is that the presence of a “golden copy” which implies a record that epitomises the highest quality information, on the application could not prevent the downtime of the application.

30. There is work underway in the payment space (by the CPSS) on ‘Cyber security in FMIs’ relating to how cyber attacks can bypass well designed preventive and detective measures by exploiting a large surface of IT vulnerabilities, and the weakest link, the human factor, by social engineering. These attacks have the potential to seriously disrupt the operations of a FMI and can present unique challenges to its operational risk management and business continuity plans, with implications that may threaten financial stability. Principle 17 also requires that FMIs should resume operations within two hours following disruptive events; in extreme cases an FMI should complete settlement by the end of the day of the disruption. In such attacks, recovery requires different integrity verification checks than recovery from a natural disaster or data centre failure. The time needed to find out the root cause and be operational once again within a period of two hours can be a great challenge; the operational impact would be that it may cause financial market disruption. This also raises questions relating to the definition of RTO in cases where the data/system integrity is lost.

Way Forward

31. There are exciting times ahead of us! Let me just flag a few areas which have the potential to create efficiencies and values in the long run. Today, banks have that rare opportunity to reinvent themselves again—with data and analytics. Major decisions to drive revenue, to control costs, or to mitigate risks can be infused with data and analytics. Analytics can be used by banks to gain a competitive edge—by improving risk assessment, fraud prevention and predicting customer behaviour. In fact in its IT Vision, the Reserve Bank has stressed upon the usage of analytics for improvement of Customer Relationship Management (CRM), risk management and fraud detection / prevention.

32. With the world economy opening up and businesses covering many different geographies and economies, the need for a common reporting format has never been more urgent. Accounting and auditing are done differently in different places, which has given rise to the present-day push for implementing IRFS, a common financial reporting standard.

33. But that said, the first step towards regulatory compliance is the need to get control over the various scattered sources of data and integrate them into a centralized system. That is what we hope the ADF has achieved. But I also hope banks will use this opportunity to develop MIS beyond the regulatory prescriptions as well.

34. The other technology that is being talked about and has a lot of potential is Big Data and its capabilities. Big Data refers to large repositories of corporate and external data to uncover trends, statistics, and other actionable information to help

decide on the next move. Huge amounts of data is being created by banks on a daily basis- both structured and unstructured. Not only are banks struggling to keep rate and pace of this data but also are unable to use it in a meaningful way to improve services and customer experience.

35. The Reserve Bank of India has set up the Data Warehouse which has the potential to meet the MIS and DSS requirements – both within the Bank and outside.

It is working towards further enhancing the usefulness of the Data Warehouse for internal decision and policy making and external dissemination which would be especially beneficial for the banks.

36. The Reserve Bank as also other banks have adopted technologies that can handle data warehousing. But the increase in the amount of data is resulting in complexities. This is leading to a trend wherein banks are moving from basic reporting to analytics and are heading towards predictive analytics. And to enable predictive analytics, IT Departments should be integrating Big Data with already stored data.

37. Although decision-makers are realising that there is value in Big Data, getting to that value is not so easy for most businesses. Here IT can help by offering services that empower researchers to delve through large data stores to perform analytics and discover important trends. The use of predictive analytics would enable banks to provide more value for their customers and understand their views and opinions. This is an area of statistical analysis that deals with extracting information from data and using it to predict future trends and behavior patterns. This is a technology that will give banks a competitive edge over others.

Final thoughts

38. Before I conclude, I would like to pose a few questions to the delegates of this Conference for further deliberations:

(i) What strategy should a bank/banks adopt for moving towards data driven decision making? Would it be sustainable?

(ii) Given the focus on systemic risk, banks are being pushed by the regulators to demonstrate better understanding of data they possess, transform the data into information which supports business decisions and manage risks more effectively. Thus each such request has major ramifications on data collection, governance and reporting. Banks should start transforming their business models today to comply with a radically different regulatory environment. Are they getting ready for this?

(iii). How do we give an assurance to all stakeholders that the data that we possess is of high quality?

(iv) Traditionally technologists look at data dimensions largely from a storage perspective. Looking ahead, this needs to be addressed more from a business perspective and hence information has to belong to business domain. Is it not time that Information and Technology be seen as distinct disciplines thus necessitating the need to handle them accordingly; ensuring that Information shall be part of business domain?

39. My final thought as I prepare to conclude my address on this important issue is that given the current state of things in our country, and looking at non-uniform approaches adopted by the financial sector with regard to data governance

framework, data quality management procedure with a board driven enterprise-wide data architecture standard, is a more proactive regulatory intervention for data quality assurance - a la ECB style- called for in our country as well?

Here's wishing you fruitful discussions in the Conference.

Thank you.