

---

---

**Report of  
Expert Group  
on Internet Deployment of  
Central Database Management System**

---

---



---

**RESERVE BANK OF INDIA  
MUMBAI  
SEPTEMBER 2004**

---

Date: 7<sup>th</sup> September 2004

Dr. Rakesh Mohan  
Deputy Governor  
Reserve Bank of India  
Central Office  
Mumbai –400 001

Dear Dr. Mohan,

Sub: Report of the Expert Group on Internet Deployment of Central Database Management System (CDBMS)

I am pleased to submit the Report of the Expert Group on Internet Deployment of Central Database Management System (CDBMS) appointed vide the RBI memorandum dated January 3, 2004.

With regards

**Yours sincerely,**

**Sd/-  
(A. Vaidyanathan)  
Chairman**

## CONTENTS

### APPENDIX

Section	Topic
Appendix I	Office Memorandum: Expert Group on Internet Deployment of CDBMS
Section II	Review of Central Database Management System (CDBMS) data contents
Section III	Data Dissemination by RBI - Current Practices and Group's recommendation
Section IV	CDBMS - tasks ahead
Section V	Summary of Recommendations

## List of Abbreviations

ADR	American Depository Receipts
BOP	Balance of Payments
BPSD	Balance of Payment Statistics Division, division of DESACS
BSA	Balance Sheet Analysis
BSB	Banking Services to Bank
BSD	Banking Statistics Division, division of DESACS
BSE	Bombay Stock Exchange
BSR	Basic Statistical Return
CAS	Central Accounts Section
CD	Certificate of Deposits
CDBMS	Central Database Management System
CDBMSi	Central Database Management System on Internet
CF	Capital Formation
CFD	Company Finance Division, division of DESACS
CIBIL	Credit Information Bureau India Ltd.
CM	Currency Management
CMIE	Centre for Monitoring Indian Economy
CP	Commercial Papers
CPI	Consumer Price Index
CRAR	Capital to Risk Weighted Assets Ratio
CSD	Corporate Studies Division, division of DESACS
CSO	Central Statistical Organisation
DAD	Deposit Accounts Department
DBOD	Department of Banking Operations and Development
DBS	Department of Banking Supervision
DBS	Division of Banking Studies, division of DESACS
DEAP	Department of Economic Analysis and Policy
DESACS	Department of Statistical Analysis and Computer Services
DGBA	Department of Government and Bank Accounts
DGCI&S	Directorate General of Commercial Intelligence and Statistics
DIF	Division of International Finance, division of DEAP
DIT	Division of International Trade, division of DEAP
DNBS	Department of Non Banking Supervision
DRI	Differential Rate of Interest
DSBB	Dissemination Standards Bulletin Board
ECB	External Commercial Borrowing
EPW	Economic and Political Weekly
EPWRF	Economic and Political Weekly Research Foundation
FCNR	Foreign Currency Non Resident Accounts
FDI	Foreign Direct Investment
FEM	Foreign Exchange Management
FFMC	Full Fledged Money Changer
FI	Financial Institution
FID	Financial Institution Division, divisions of DBS
FII	Foreign Institutional Investors
FMA	Financial Market Analysis
FSSA	Financial Sector Stability Analysis
GCF	Gross Capital Formation
GDDS	General Data Dissemination Standard
GDP	Gross Domestic Product
GDR	Global Depository Receipts

GFCE	Government Final Consumption Expenditure
GFCF	Gross Fixed Capital Formation
GIC	General Insurance Corporation
HUDCO	Housing and Urban Development Corporation Limited
ICRISAT	International Crop Research Institute for the Semi Arid-Topics
IDBI	Industrial Development Bank of India
IDMD	Internal Debt Management Department
IECD	Industrial and Export Credit Department
IFCI	Industrial Finance Corporation of India
IIBI	Industrial Investment Bank of India
IIP	Index of Industrial Production
IMF	International Monetary Fund
KCC	Kisan Credit Card
LIC	Life Insurance Corporation
MEA	Macro Economic Analysis
MMM	Macro Monetary Management
MPD	Monetary Policy Department
MRO	Mumbai Regional Office
NABARD	National Bank for Agriculture and Rural Development
NBFC	Non Banking Financial Companies
NCDC	National Cooperation Development Council
NEER	Nominal Effective Exchange Rate
NHB	National Housing Bank
NIC	National Informatics Centre
NOF	Net Owned Funds
NPA	Non Performing Assets
NRE	Non Resident (External) Rupee Accounts
NRNR	Non Resident Non Repatriable Rupee Accounts
NSE	National Stock Exchange
OD	Overdraft
P/L a/c	Profit and Loss a/c
PDO	Public Debt Office
PLR	Prime Lending Rate
RBI	Reserve Bank of India
RBIFS	RBI's Financial Statements
REER	Real Effective Exchange Rate
RMC	Restricted Money Changer
RNBC	Residuary Non banking Companies
RPCD	Rural Planning and Credit Department
SACP	Special Agricultural Credit Plan
SBI	State Bank of India
SCB	Schedule Commercial Banks
SDDS	Special Data Dissemination Standard
SDP	State Domestic Product
SDR	Special Drawing Rights
SEBI	Securities and Exchange Board of India
SGSY	Sawarnjaynti Gram Swarojgar Yojana
SIDBI	Small Industry Development Bank of India
SJSRY	Swarna Jayanti Shahari Rojgar Yojana
SLRS	Scheme for Liberation and Rehabilitation of Scavengers
SSI	Small Scale Industries
SSS	Small Saving Scheme
UBD	Urban Banks Department

UCB	Urban Cooperative Banks
UN	United Nations
UTI	Unit Trust of India
WB	World Bank
WMA	Ways and Means Advance
WPI	Wholesale Price Index
WSS	Weekly Statistical Supplement

## **Executive Summary**

The Central Data Base Management System (CDBMS), the corporate data warehouse has been operational for internal users within Reserve Bank of India, for the last one and half years. Presently it covers data series relating to four subject areas viz. Macro Economic Aggregates, Financial Market Analysis, Financial Sector Stability Analysis, and Macro Monetary Management. These data have been collected from various Departments of RBI as well as from Government Statistical agencies. In November 2003, Bank management decided that as much of the data in CDBMS as possible should also be made available to researchers and analysts outside the Bank through Internet deployment of CDBMS, to be called CDBMSi, in parallel with promoting internal usage. The Bank constituted an expert group in January 2004 for an independent expert review to advice on the scope, content and detail of the data elements that can be placed in the public domain. The Expert Group met three times since its formation. Group members reviewed the details regarding the items of data currently incorporated in the system, their content, frequency, level of dis-aggregation and periods covered. The basis for determining the items to be made accessible to outsiders was also extensively discussed.

The Group feels that building CDBMS in itself is an impressive and commendable achievement. However, the system is still evolving and teething problems are inevitable. The Group has noted the issues raised by users regarding the shortcomings of current implementation of CDBMS. It is difficult to judge whether these are an adequate explanation for the slow pick-up in the internal usage of the system. Perhaps one has also to look at the extent to which higher levels of the Bank's management demand databased analyses and the extent to which heads of Departments and Divisions stimulate and lead this effort. Nevertheless study of the current data coverage and the way data are presented in the system also corroborate that some of these problems do exist and need to be addressed.

The Group found the coverage of subjects of interest to in-house users, for which RBI is the primary source, to be quite comprehensive. Coverage of some important segments of the financial sector on which RBI does not directly collect much data, will require inputs from other institutions like NABARD, IDBI, SEBI and the stock exchanges. Data compiled and digitized by non-governmental organizations (like, the EPW Research Foundation, CMIE, ICRISAT), after due scrutiny and review, may be incorporated in the CDBMS.

The usefulness of data series provided will be limited unless the sources of data, concepts, coverage and methodology of estimation and changes therein are properly documented. That this has not been done systematically for all data sets is a major lacuna in the CDBMS as it now stands. Revisions in estimates as well as changes in scope and classifications are

not always promptly reported to CDBMS. This gives rise to apparent inconsistencies between data in the CDBMS and other sources. Where secondary sources are used, authentication is far more difficult.

The Group has reviewed current data dissemination practice of the Bank. It is observed that the main channels of dissemination are publications - periodical and ad hoc, press releases, Governor's bi-annual announcements of monetary and credit policies, speeches, reports, circulars, manuals and the Bank's public website ([www.rbi.org.in](http://www.rbi.org.in)). Most of the information so disseminated consisted of processed data. Access to primary data was restricted to 'authorized' users within the organization.

The Group feels that in order to be useful it is essential to ensure that the data are validated by source agencies and promptly updated; inform users about the scope, concepts and methodology of the sources of information posted in the system; and indicate the extent of comparability of data in the above respect for any given subject area across regions and over time and changes that may have occurred in these respects. In CDBMSi it would be useful to classify the information into data sets based on frequency of observation (like monthly, quarterly, etc.), and functional categories (like money supply, state budgets, etc.).

The Group strongly recommends that the task of incorporating the detailed metadata, particularly conceptual underpinning and compilation methodology of different data series, be given high priority by RBI management and RBI should back it with adequate funds and staff. Data in CDBMSi should be made available in tabular form and it should also be capable of being downloaded in Excel and pdf format.

It is understood that the RBI has taken an internal decision that the Departments of RBI that are forwarding data to CDBMS would be treated as owner of data and the departmental in-charge would frame the access policy with respect to data provided by that Department of RBI. However, a review of the Departments' negative list suggests that they would like to limit the data to be made available to the public through CDBMSi only to those data which are already being published by RBI. This falls far short of the stated government policy of freedom of information and transparency. Enlarging the scope and detail of information on the operations of the financial and real sectors in the public domain enables outside experts to undertake analyses of trends and policy issues of wider scope and provides independent assessments on various issues of interest to policy makers in the Bank and the government than is possible with in-house expertise.

The negative list must, in Group's view, be limited to items of information that (a) are collected on the assurance that informants' identity and information relating to them will be kept confidential; and (b) give information advantage to the RBI as regulator. **Confidentiality** can be taken care of in three ways 1) Aggregation/consolidation 2) Masking the identity of units and 3) Releasing data after a time gap. It appears that the Departments have not identified any threshold time beyond which a data element would lose its sensitivity and can be made available to public. The Expert Group is of the opinion that unless data are obtained with explicit commitment to maintaining its confidentiality, it should be fit for publication if it does not directly impact the functioning and performance of market players. Even in the later case, after a hiatus of reasonable time such data should be fit for publication. Further, information that cannot be disseminated on considerations

of sensitivity and secrecy should not be perpetually buried in official files. A system should be established whereby essential information is distilled, analyzed and put out in the form of periodic studies.

The Expert Group is of the opinion that the Departments'/Divisions' lack of confidence on the quality of the data should not be the grounds for it being placed in the negative list. Provided that data quality disclaimers are given, and provided that appropriate qualifications are given in the data definitions, there is no harm in providing it for the public. The first page of the website therefore should contain a strong qualifier that the data are only indicative and may be changed/corrected in the future.

The following measures are suggested to strengthen the CDBMS. Technological infrastructure has been provided by the CDBMS but the harder part of making available annotated detailed and long time series data are yet to be accomplished. It needs to be ensured that the series included in the CDBMS are crosschecked for consistency with the data published by the original source; revisions made by the source agency are promptly incorporated; and current data compiled by various Departments/ affiliates of the RBI are promptly and automatically reported and incorporated in the CDBMS data repository. The data producing Departments need to be sensitized about releasing the data as soon as it is ready, including any revisions thereof. If the data producing entities disseminate the data through multiple channels, then it might be difficult to keep data integrity. In this connection there is a need to focus on Weekly Statistical Supplement and the RBI Bulletin. Producers of the data should be responsible for data integrity and synchronization of data flow to CDBMS with other modes of release of data by the Bank and other agencies outside RBI. The main point that the group would like to strongly underscore on this issue is that the objective of building such a massive database with required annotations involves a gigantic task and cannot be undertaken by one small single unit. It must be undertaken jointly by all relevant Departments/Units within the RBI.

The staff strength and budgetary allocation for the CDBMS unit and related activities have to be increased in a measure commensurate with the task it is expected to perform. It is also important to institutionalize mechanisms to ensure greater interaction and collaboration among CDBMS as well as the sources and users of the system. Further, to strengthen this process, as well as to bring outside researchers' perspective regarding data content and format of CDBMSi on an ongoing basis, the RBI may consider formation of a Standing Advisory Group including outside experts.



## Section - I

### Introduction

#### I.1 Genesis of the Group

The Reserve Bank has recently established a state-of-the-art computerized decision support system for use within the Bank. This system, called Central Database Management System (CDBMS), has been designed to provide a comprehensive and readily accessible collation of current and historical data on various aspects of real and monetary sectors of the economy needed by the Bank staff for operational and policy decisions. Though primarily meant for the internal use of the Bank, much of these data are also of interest to researchers, analysts and other users. The system was planned in such a way that, in due course, the same platform could be used to disseminate information to the public through the Internet. The Bank management decided that Internet deployment of data in CDBMS (to be called CDBMSi) to outsider users be taken up in parallel with promoting internal usage. Since the warehouse includes data of a confidential and sensitive nature, all of it cannot be made available to outsider users. The Bank management decided to commission an independent expert review to advice on the scope, content and detail of the data elements that can be placed in the public domain, as well as directions in which the warehouse could be extended to better serve the needs of users within and outside the Bank.

Accordingly, an Expert Group on Internet Deployment of CDBMS consisting of the following members was constituted in January 2004. Subsequently, Prof. Mihir Rakshit, Director, RBI Central Board was associated with the Expert Group as a **special invitee**. The composition of the Expert Group is given below.

Dr. A. Vaidyanathan, Emeritus Professor, Madras Institute of Development Studies and Former member, Planning Commission	Chairman
---	----------

Dr. R.B. Barman Executive Director Reserve Bank of India	Member
--	--------

Dr. S.L. Shetty Director EPW Research Foundation Mumbai-400101	Member
---	--------

Dr. Vaskar Saha Add. DG CSO Ministry of Statistics Planning and Implementation Government of India New Delhi	Member
--	--------

Dr. Surjit Bhalla Managing Director Oxus Research and Investment	Member
--	--------

New Delhi

Dr. S Gangopadhyay Director India Development Foundation (IDF) Haryana	Member
Dr. Laveesh Bhandari Indicus Analytics New Delhi	Member
Dr. A.K Nag Adviser, DESACS Reserve Bank of India	Member-Secretary
Prof. Mihir Rakshit Director, Monetary Research Project, ICRA	Special Invitee

## **I.2 Terms of Reference**

The terms of reference of the Expert Group were:

- (a) To review data contents of the publishable segment of CDBMS in terms of its usefulness to researchers and analysts outside the Bank and suggest fresh data items that could be included in CDBMS for public dissemination subject to its feasibility.
- (b) To provide a user requirement perspective to the Bank in respect of the Bank's current data dissemination policy and suggest changes, if any, in the light of best international practices in this regard for the Bank's consideration.
- (c) To review the current data flow system for CDBMS in terms of its timeliness, reliability and coverage and suggest improvement, if any
- (d) To suggest a technically feasible user interface for accessing CDBMS data over Internet so that it is made available publicly.
- (e) To provide guidelines for metadata creation for end users of CDBMS over Internet.
- (f) Any other issue considered relevant for dissemination of data through CDBMS.

The memorandum authorized the Expert Group to co-opt members depending on the subject under consideration and may constitute technical groups to report on specific technological issues (Appendix-I).

### **I.3 Functioning of the Group**

The Expert Group met twice in Mumbai on 20th January, 29<sup>th</sup> March 2004 and once in New Delhi on 14<sup>th</sup> August 2004 and had extensive discussions with the CDBMS group and other senior officials of the Bank. Details regarding the items of data currently incorporated in the system, their content, frequency, level of dis-aggregation and periods covered, were reviewed in detail at these meetings. Members also had an opportunity to examine contents of the website first hand. Members were also given an idea of the meta-data and the extent to which they are available for incorporation in the system. The basis for determining the items to be made accessible to outsiders was also extensively discussed. We also had the benefit of detailed comments based on their experience in using CDBMS from various user Departments of the Bank. Participation of the Governor and the Deputy Governor in the meeting held on 29 March 2004 gave the Committee a clear idea of the Management's concept of the role of CDBMS and policies regarding making it accessible to the public. Based on these deliberations and background work undertaken by the CDBMS team, the Group finalized its report at the third meeting.

### **I.4 Organization of the Report**

The report is divided into five sections including the introduction. Section II reviews the scope, coverage, and content currently incorporated in CDBMS, indicates the problems encountered in getting timely and validated data and identifies some of the major gaps. In section III, a brief account of recent developments in data dissemination policies and practices of the IMF at the international level and the RBI in India, is followed by a discussion of and suggestions regarding (a) items, level of details, and forms in which the contents are to be deployed on the internet and (b) data structuring, format, user interface and related aspects to make the system user friendly. The Group's suggestions about the steps needed to ensure required coverage of CDBMS, its timeliness, quality and building of metadata are spelt out in section IV. The last section gives a summary of the Group's recommendations.

### **I.5 Acknowledgement**

The group expresses its sincere thanks to Dr. Y. V. Reddy, Governor, RBI and Dr. Rakesh Mohan, Deputy Governor, RBI for sharing their views and visions about CDBMS and to Prof. Mihir Rakshit for his valuable contribution to the Group's deliberations.

The Group acknowledges with thanks the efforts made by the regional offices of DESACS for organizing various meetings of the Group and specially New Delhi Regional Office for organizing detailed demonstration of CDBMS for benefits of the Group members. The efforts made in this regard by Shri Pradip Maria, Director DESACS, New Delhi Regional Office and his team is gratefully acknowledged. Thanks are also due to all the user Departments of RBI, especially DEAP, for providing user feedback and preparing negative list of data series.

The Group would like to record their appreciation for the work done by Dr. Nag and his CDBMS team in the DESACS. They did an excellent job of facilitating and supporting our work by compiling and circulating background information, providing clarifications sought by members, and taking care of logistics of meetings. Special mention may be made of Shri Anil Kumar

Sharma, Director, Shri Anujit Mitra, Assistant Adviser, Shri Indranil Das, Assistant Adviser and Shri Dibyendu Bhaumik, Research Officer for their contribution to the Group's work. Besides leading his team both in building the CDBMS and providing able secretariat support to the Group, as Member Secretary of the Group, Dr. Nag's contribution deserves a special mention. Besides participating actively in our substantive discussions, and clarifying several issues, he also helped prepare the draft report. The Group deeply appreciates his valuable contribution to our work.

## **Section - II**

### **Review of Central Database Management System (CDBMS) data contents**

#### **II.1 Coverage**

The CDBMS was designed primarily to provide ready and easy access to a large variety of data needed by officials within the Bank. The following eight major subject areas are planned to be covered by the system in a phased manner:

1. Macro Economic Analysis (MEA)
2. Macro Monetary Management (MMM)
3. Financial Market Analysis (FMA)
4. Financial Sector Stability Analysis (FSSA)
5. Banking Services to Banks (BSB)
6. Foreign Exchange Management (FEM)
7. Currency Management (CM)
8. RBI's Financial Statements (RBIFS)

Currently CDBMS covers the first four areas. The subject areas, specific data items, periodicity and other particulars of data included in the system were decided after extensive discussions with different Departments in the Bank and reflect their requirements for monitoring and analytical purposes. The following is a brief description of the scope and content of the data in respect of these four areas.

#### **1. Macro Economic Analysis**

This subject area provides time series data, mostly on an annual basis and at the national level, essential for analysing the emerging trends in the national economy as a whole and in its broad sectors. They cover all components of national income accounts estimated by the Central Statistical Organisation; considerable amount of disaggregated data on commodity trade obtained from DGCI&S and wholesale and consumer price indices from the source agencies are included in the system; some disaggregated data for agriculture and rural economy are also available. Data on external sector, corporate finances and capital markets draw on estimates published by the RBI and are quite detailed. The system also includes some state level data on population, gross domestic product, agricultural production, credit and procurement and off take of food grain/commodities.

The main dimension of classificatory attribute for this subject area is time and, therefore, all time series data are available on weekly, fortnightly, monthly, quarterly and yearly basis, depending on the lowest frequency at which a particular data

item is available in the source system. The CDBMS application software allows users to aggregate a lower frequency data to higher frequencies say, from weekly data to monthly data and so on.

## **2. Macro Monetary Management**

This subject area comprises information on monetary and liquidity management and debt management. It provides various liquidity indicators and its composition, and finances of Central and State Governments. The data included in the subject area supports analysis of liquidity indicators, utilisation of ways and means advances (WMA) and overdrafts (OD) vis-à-vis limits, borrowing requirements of individual governments and their impact on monetary management. The data on monetary aggregates include various measures of money supply like M0, M1, M2, M3, NM1, NM2, NM3, daily bank's balance with RBI; data on liquidity indicators covers liquidity supports provided to financial system on day-to-day basis by RBI, resource management of banks. Data on Government's debt management covers Government's daily balance like WMA, OD, Receipts and Payments, etc. The major sources of the above data series are various Departments of RBI. Since these aspects require constant monitoring the majority of data series are reported on a daily basis, some at fortnightly and a few at monthly intervals.

## **3. Financial Market Analysis**

This subject area includes various data series pertaining to volume and prices of various financial instruments traded in different segments of the financial market. They cover call money rates, different interest rates prevailing in the money market, spot and forward rates in the foreign exchange market; stock market indices for equity and debt instruments; and prices and volume data relating to government securities in the primary and secondary markets. These again are monitored and reported mostly on a daily basis.

## **4. Financial Sector Stability Analysis**

This is by far the largest section of CDBMS. It is meant to help in the analysis of the systemic integrity, financial performance and health of banks, financial institutions and non-banking financial companies and the development of the financial system. It includes data relating to defaulters, compliance with statutory norms, performance and balance sheets of Indian and foreign banks, non-bank finance companies (NBFCs) and industrial and export credit. Most of the data in this section are based on information compiled and reported by various Departments of the Bank from statutory and off-site surveillance returns (like Basic Statistical Returns (BSR), Banks' Balance sheets, Section 42 (RBI Act) returns, priority sector lending and NBFC returns) received by RBI from various financial institutions. Detailed information on spatial distribution of bank credit and deposit ownership pattern and composition of bank employees are also available. These data are sourced from various Departments of RBI.

### **II.2 An assessment of the current status of CDBMS**

The design and construction of CDBMS involving collation of a massive amount of diverse data from numerous sources, structuring them to meet users' requirements and designing an operational system to ensure that they can be readily accessed and used individually and various combinations for specific descriptive or analytical purposes is a huge complex undertaking demanding a high degree of professional skill and sophistication. Specialised inputs from external consultants

no doubt played an important role in working out hardware configurations and developing the software appropriate to the specific requirements of CDBMS. But equally important is the role and contribution of the CDBMS internal group in all stages of this process. That all this has been successfully done and the system made operational is an impressive and commendable achievement.

The system is still evolving. Teething problems are inevitable. In-house users of CDBMS from different Departments have indicated the problems they face in using the system. Some are administrative problems. The substantive issues relate to (a) gaps in the data posted in the system, differences from the original source, omission of units of measurement, clarity of terminology and authentication of data sources; (b) inadequacy of posted data in relation to users' requirements because of time lags in updating, lack of time series for a sufficiently long period, time frequency being too long and non-inclusion of data for cross country comparisons; (c) user friendliness of current system for access and use of the data. Several users want the system to provide more details and dis-aggregation and/or long time series with proper annotation of sources and method to be incorporated.

It is difficult for us to judge whether these are adequate explanations for the slow pick-up in the internal usage of the system. Perhaps one has also to look at the extent to which higher levels of the Bank's management demand databased analyses and the extent to which heads of Departments and Divisions stimulate and lead this effort. Nevertheless our review of study of the current data coverage and the way data are presented in the system also corroborate that these problems do exist and need to be addressed.

Our assessment and suggestions on the difficulties cited earlier are given below.

### **II.2.1 Coverage**

It is our impression that coverage of subjects of interest to in-house users for which the RBI is the primary source is quite comprehensive. However, coverage of some important segments of the financial sector on which RBI does not directly collect much data leaves considerable scope for improvement. This will require inputs from other institutions like NABARD, IDBI, SEBI and the stock exchanges. RBI should explore the possibility of getting these institutions to furnish, and regularly update, basic data in a format appropriate for incorporating in the CDBMS. Providing a link to their web sites through the CDBMS will enable users to access to additional or more detailed information they may need relating to their domains.

Access to data for cross country comparisons or analysis can also be facilitated by providing links to websites of international organisations (such as UN, World Bank and IMF) and references to websites where more detailed, country specific information is available.

There is also a felt need to expand the coverage of macro-economic indicators. The extent to which this is possible depends on the content and format in which the data collected by various government and quasi government agencies are made available to the public. Very few of them – with the significant exception of the CSO – maintain properly designed and regularly updated websites through which their data can be electronically accessed by the public. The NIC web site, which is supposed to have detailed of socio economic data by states and districts and open to the public, is not easily accessible nor are they always up-to-date. However several other organisations – for example, the Registrar General and DGCI&S –

have detailed data in digitised form. Others, including several central government ministries, National Sample Survey and Annual Survey of Industries, put out a considerable amount of processed data in print. Some of it has been compiled and digitised by non-governmental organisations (like the EPW Research Foundation, CMIE, ICRISAT). These can, after due scrutiny and review, be incorporated in the CDBMS. In many cases incorporating them in the CDBMS will require a considerable amount of effort to persuade the concerned agencies to cooperate by locating, assembling and digitising the data and agreeing to making the data available to the public.

### **II.2.2 Time Series**

Meaningful analysis and interpretation of trends requires long time series. CDBMS has been able to meet this requirement only to a limited extent. For data collected and processed by units within RBI, long time series dating back to early 60s and, in some cases, even earlier are available for several items. All of these are not yet loaded into CDBMS. Data on key financial indicators (bank deposits, bank credit, corporate finance, and external accounts) have been published regularly since long but are available only in print form. It is understood that compiling some of these series for loading to CDBMS is in process. There are however some items for which such series are yet to be compiled. For most of the macro-economic indicators CSO has time series estimates. The basis and methodology of estimation and price base are revised from time to time. It has also published in print form long-term series for most items of national accounts; but not all are available in their web site. All these details from 1950-51 have been collated and published by EPWRF.

A more comprehensive compilation of long term time series of overall and state-wise GDP, agricultural and industrial outputs by commodity groups and states, as well as series on public finance, BOP, banking and prices from 1950 to 1990 is available in Chandok and The Policy Group (1990). These compilations need to be reviewed and checked with original sources and converted to a common price base (where necessary) for inclusion into the system. But their usefulness will be limited unless the sources of data, concepts, coverage and methodology of estimation and changes therein are properly documented. That this has not been done systematically for all data sets is a major lacuna in the CDBMS as it stands now.

### **II.2.3 Metadata**

The data posted in the CDBMS is selectively culled out and processed from the much larger and more detailed primary information collected by sourcing agency. Many of the detailed and specialized requirements of users can be met, provided the contents of the primary data are made known and the CDBMS permits and enables access to the meta data or a means of getting the meta data processed in the light of specific requirements. While this is recognized, the system gives very little information on meta data for which the RBI is the source agency. For most items in the system information on meta data are reported as not available. Getting information on the contents of meta data from other government and quasi government source agencies is far more difficult partly because higher level management in most of these agencies do not attach high priority, or allocate sufficient resources, to data collection and analysis or partly because of their reluctance to make the data available to others even within government.

#### **II.2.4 Authentication and Validation**

In most cases exact sources are not indicated. It is observed that in many cases data are taken as published by the source agency or as reported by it to CDBMS. In these cases the presumption is that the agency/ internal units in RBI have verified and authenticated the accuracy of the published/supplied data. However, even in these cases, there are revisions in estimates as well as changes in scope and classifications, which are not always promptly reported to CDBMS. This gives rise to apparent inconsistencies between data in the CDBMS and other sources. Where secondary sources are used authentication is far more difficult.

#### **II.2.5 User friendliness**

CDBMS provides ready access to a large database on a variety of subjects and pre-formatted reports based on these data for meeting routine requirements of in-house users. It is also intended to enable users to explore the database to answer a very wide range of questions of fact and interpretation beyond the preformatted reports. Non-routine and ad hoc in house users have pointed to several difficulties in using CDBMS “as single independent platform for data compilation, numerical analysis, and geometry (i.e., various kinds of charts, graphs, etc.)”. They in effect are asking for a facility by which the system not only provides data but also allows basic statistical data analysis (calculation of growth rates, trends, correlations, etc.), preparation and formatting of charts in single unified environment. While it is technically possible to customize the system to meet these requirements, it is necessary to consider whether the cost involved for incorporating this facility would be commensurate with the benefits.

### **Section – III**

#### **Data Dissemination by RBI - Current Practices and Group’s recommendation**

##### **III.1 Current Data Dissemination Process of the Bank**

The Reserve Bank has a long and rich tradition of dissemination of information. It disseminates information relating not only to monetary and banking areas which fall directly under its purview, but also to the economy and finance for which it releases information collected from primary sources to give a comprehensive picture of India's economy and financial sector. The Bank disseminates information, data analyses and views. The main channels of dissemination are publications - periodical and ad hoc, press releases, Governor's bi-annual announcements of monetary and credit policies, speeches, reports, circulars and manuals. Besides the paper based channels, the Bank's public website ([www.rbi.org.in](http://www.rbi.org.in)) is also being utilized for disseminating information in electronic form, since August 1996.

Most of the information so disseminated consisted of processed data. Access to primary data was restricted to ‘authorized’ users within the organization. Outsiders, especially non-governmental researchers, were given access only on a highly selective basis under special circumstances. Even in such cases access was subject to the condition that the data should be used only for the approved purpose and should not be made available to any others. Civic society, non governmental



organizations generally and researchers in particular have long argued that free access to information data collected by public agencies is necessary for informed discussion on issues of public policy and for ensuring efficient, responsive and accountable governance. This has led to a gradual change in government policy on information sharing.

The National Policy on Dissemination of Statistical Data, adopted by Central Government in September 1998, envisages the creation of 'Data Warehouse' for disseminating data as per the convenience of users and which will encourage the research studies and in depth analysis. The policy measures include sharing of data not only in respect of the hitherto published data usually but even in respect of un-published data collected by the Government agencies. With some significant exceptions – notably the CSO's website and liberalized access to primary data from the National Sample Survey Organizations - implementation of this policy has been very uneven and slow. Against this background the RBI's initiative to set up the CDBMS as an archive to include, progressively, all the data in its possession and incorporate key items of information on the real economy and its decision to make this archive accessible to the public is welcome and highly commendable.

### **III.2 International Experience**

International organizations have been actively involved, since a long time, in developing and constantly improving conceptual framework, design and methodologies for collecting data on several aspects of monetary and financial sectors and balance of payments. This has helped to evolve international standards for collection and reporting of key economic statistics by national statistical agencies. These standards are spelt out for example in *Balance of Payments Manual*, *Manual on Government Financial Statistics and Draft Manual on Monetary and Financial Statistics*, *UN System of National Accounts*. Most countries accept these standards. Those that are not in position to comply with these standards, use them in improving the coverage and quality of their statistics.

International standards are meant to ensure comparability of data on key aspects of economic structure and performance across countries. National data and estimates compiled in that framework used to be published in print form by various international organizations and have been accessible to the public the world over on their web sites.

Although these standards are comprehensive in terms of coverage, quality and periodicity, they did not address the issue of timeliness and access to public. The crisis in Mexico in 1994-95 brought to the fore the need to ensure transparency in the wake of increased integration between international financial markets. In order to address these issues, the International Monetary Fund established the Special Data Dissemination Standard (SDDS) in April 1996 to enhance availability of timely and comprehensive statistics on real, fiscal, financial, external sectors as well as socio demographic data on population. Till August 2004, there have been 57 countries subscribed to SDDS including India. The SDDS emphasize on the following four dimensions of data dissemination:

- Data coverage, periodicity and timeliness;
- Access by the public;
- Integrity of the disseminated data; and
- Quality of the disseminated data.

In December 1997, IMF established the General Data Dissemination System (GDDS) applicable to all members of IMF, with a view to widening the scope of dissemination standards. The GDDS covers socio-demographic data such as population, health, education and poverty in addition to the four sectors real, fiscal, financial and external covered by SDDS. The main emphasis of the General System is to improve the national systems of data compilation with main focus on data quality on four characteristics of data namely 1) data coverage, periodicity and timeliness 2) quality 3) access and 4) integrity.

The establishment of dissemination standards was aided to a great extent by the Internet technology. The Internet has removed the distinction between the various types and locations of data users. The data are available simultaneously to users across the world. The IMF established the Dissemination Standards Bulletin Board (DSBB) on the IMF's web site providing information about the standards, the subscribers and the dissemination practices of the subscribers. DSBB also provides links to the summary pages of national compilers providing link between the SDDS metadata and the actual country data.

### **III.3 CDBMS experience**

CDBMS is similar in concept but seeks to incorporate a much wider range of data, in greater sectoral and spatial detail, provide long time series on several aspects and indicate the location, content and accessibility of meta data. The task of improving data formatting, software design and other technical aspects of the CDBMS to make it more user friendly are already in progress. We have also highlighted the fact that in order to be useful it is essential to (a) ensure that the data are validated by source agencies and promptly updated; (b) inform the user about the scope, concepts and methodology of the sources of information posted in the system; (c) indicate the extent of comparability of data in the above respect for any given subject area across regions and over time and changes that may have occurred in these respects. Improvements are needed in all these respects.

The data content and software in the CDBMS were designed to meet in-house users' routine requirements. It would need substantial expansion and restructuring to make it useful for research purposes both within and more so for outsiders. Currently, published as well as unpublished and sensitive data of various Departments of RBI are available in CDBMS. However, only published data of CSO, DGCI&S and other government agencies are included. It is necessary to try and expand the coverage of the latter data sets to include unpublished data in as much detail as possible. It is further necessary to provide (a) long time series with proper annotations of sources, concepts and methodology and changes therein; (b) higher levels of dis-aggregation by region and sector in respect of several indicators ensuring as far as possible as per standardized classifications; (c) information on metadata relating all major data series and incorporating them to the maximum extent possible in the CDBMSi. All these would require considerable strengthening of CDBMS both in terms of budgetary resources and human capital.

CDBMS should also be designed to enable and facilitate its users, especially researchers, to go beyond the pre-formatted tables and analyses incorporated in the CDBMS. It should be possible for them to combine different data elements, generate different types of tabulations and cross tabulations, and try out models of different scope and complexity. For this purpose it would be useful to classify the information being presented into data sets based on frequency of observation (like monthly, quarterly etc.), and functional categories (like money supply, state budgets etc.). While different groups of variables may be related, researchers find it easier to work with smaller and well-grouped data.

At present it is not possible to have access to the detailed metadata, particularly conceptual underpinning and compilation methodology of different data series. This is because details of the content and availability of these primary data from the source agencies are yet to be compiled. Getting them incorporated in the archives and making them accessible will take time. The process should be relatively easy and quick in respect of current (and past) data collected by the RBI and preserved in a digitized form. We strongly recommend that this task be given high priority by RBI management and back it with adequate funds and staff.

The task will be much more difficult to accomplish where – as is the case with most macro economic indicators – the source agencies are outside the RBI. Getting these source agencies to give information even on their current meta data (not to speak of past data), obtaining their consent to incorporate them in CDBMSi and to actually digitize them is likely to take a great deal of time and effort. This is a task, which cannot be done by RBI. It calls for pro-active and sustained effort by the CSO (or the national statistics commission when it is established). As the CSO has already initiated steps to build a Data Warehouse on Official Statistics, RBI's experience in designing and building CDBMS will be useful to CSO to build the same. The RBI can play a useful role in persuading the CSO to undertake this effort and collaborate with it in that task.

There should be two modes of data retrieval from CDBMSi. In the standard mode, data would be available through standard, intuitively meaningful tabular and pre-determined format. In this mode, a user would be able to select time slice (from and to selected periods) for which data can be downloaded in Excel and pdf format. In the ad-hoc query mode, users would be given the choice of creating his/her own basket of downloadable data series. It would also help users considerably if they could visualize data in graphical format in the CDBMSi environment itself. The Group suggests that there should be two basic documents available to CDBMSi users - one showing the definitions and concepts of data and the other the hierarchies of the available data series in tree structure in pdf format.

The CDBMS unit is aware of these problems and need to address them in a phased manner. The process will take time and effort in conjunction with the source agencies within and outside the Bank. While urging that this process should be strongly supported, helped and facilitated by management, we are of the opinion that the data already available in CDBMS would be of considerable value to general researchers outside RBI and should be put on CDBMSi for access to the general public. The scope, content and quality of data can be improved in a phased manner.

Finally, the Group also deliberated on the issue of levying user charges for supply of data through CDBMSi. Some members suggested that while pre-formatted historical data can be made available free to users, current data and special request for customised data series should be charged. However, considering the various technical and procedural

difficulties in administering such usage specific charges, the Group has decided not to make any specific recommendation on this at this juncture.

#### **III.4 Negative List**

Making CDBMS accessible to the public through the Internet raises question as to which data and in what detail should be thrown open to the public. It is understood that the RBI has taken an internal decision that the Departments of RBI that are forwarding data to CDBMS would be treated as owner of data and the departmental in-charge would frame the access policy with respect to all data provided by that his/her Department. Accordingly, each data providing Department was requested to identify the data series supplied by it which in its view **can not** be made available to public through CDBMSi because of sensitive and confidential nature of the information.

Our perusal of this negative list suggests that most of the Departments are in favour of limiting data made public through CDBMSi mostly to what are already being published by RBI through its various communication media like RBI bulletin, Weekly Statistical Supplement etc. Incorporating them in the CDBMSi will of course facilitate easier access for outside users. But this falls far short of the stated government policy of freedom of information and transparency. Enlarging the scope and detail of information on the operations of the financial and real sectors in the public domain enables outside experts to undertake analyses of trends and policy issues of wider scope and provides independent assessments on various issues of interest to policy makers in the Bank and the government data than is possible with in-house expertise. It also helps more informed discussion of policy issues in the public domain. It is apparent that some of the Departments/Divisions within the RBI have not given sufficient weight to these considerations in suggesting the negative lists.

The negative list must, in our view, be limited to items of information that (a) are collected on the assurance that informants' identity and information relating to them will be kept confidential; (b) give information advantage to the regulator, and (c) are of uncertain quality. These issues are discussed below.

**Confidentiality** can be taken care of in three ways

- Aggregation/consolidation: wherever the information is collected with confidentiality agreements, the data can be released *immediately* after aggregation/consolidation.
- Masking the identity of units: in most cases unit level data can be released after removing information relating to their identity and precise location.
- Releasing data after a time gap: If found necessary, the unit level data can be released without the identifiers after a specified period- say 6 months or a year at the most.

**Maintaining information advantage** to the regulator: Maintaining an informational advantage with the RBI enables it take corrective action before the market(s) can react.

This consideration may be important in some highly sensitive areas where the organization under question is large enough to have a significant impact on the financial sector. This may also be important in areas such as foreign exchange

transactions where markets are known to react unduly to new information. Completely eliminating the information availability will only harm the regulatory and corrective actions.

It appears that the Departments have not identified any threshold time beyond which a data element would lose its sensitivity and can be made available to public. The Expert Group is of the opinion that unless data are obtained with explicit commitment to maintaining its confidentiality, it should be fit for publication if it does not directly impact the functioning and performance of market players. Even in the latter case, such data should be fit for publication after the lapse of a reasonable time. Those actively involved in studying the operations of the financial and foreign sectors should decide, in consultation with various concerned Departments within the RBI, which item should be released within a day or week or month. The release of data on a regular basis would also force the concerned entities within RBI to act rapidly.

All the data that is to be released with or without a lag (there should be no permanent negative list) should be released on a pre-determined calendar basis. The RBI already has a good system in place where it releases such information. This should be expanded to cover all areas.

**Quality of underlying data:** In many instances the underlying data goes in for repeated changes/updates from estimates, to provisional, to actual. In some other cases, certain errors sometimes creep in due to human errors and it may not be possible to cross-check all the data elements. In such cases it is the natural response of any data providing entity to avoid putting in the information altogether.

The expert group is however of the opinion that the Departments'/Divisions' lack of confidence on the quality of the data should not be the grounds for it being placed in the negative list. Provided that data quality disclaimers are given, and provided that appropriate qualifications are given in the data definitions, there is no harm in providing it for the public.

The first page of the website therefore should contain a strong qualifier that the data are only indicative and may be changed/corrected in the future. This will ease the pressure on RBI and more importantly, reduce the negative attitude of the various Departments.

By way of illustration, the above issues in relation to some specific data categories are addressed below.

The Group is of the view that all data, which are already available in public domain in print or in websites of government Departments, RBI and other data generating agencies must be made available to general public through CDBMSi. This would cover most of the macro economic indicators and annual series under MMM, FMA and FSSA. CDBMS also has a large amount of detailed information on the portfolios and the transactions of various types and institutions on a daily, weekly and fortnightly basis. Some of it – like defaulters' lists, daily receipts, payments and balances of central and state governments – might be considered too sensitive to be made public. But since for the most part, data posted in the system are by category of institutions and transactions, the risk of revealing the identity of individual institutions and their transactions would seem minimal.

Thus there is no reason why bank group-wise consolidated returns for all the characteristics (including details of investments in equities and bonds, loans against shares, foreign currency assets and liabilities, and balances held abroad and debit balances in NOSTRO accounts) should not be released for public consumption on a quarterly basis.

In some cases, (for example the cash, ways and means and overdraft positions of individual states and stoppage of payments by the RBI on occasions) there may be a case for not revealing to the public certain information contemporaneously, but consideration may be given to releasing such information after the lapse of a month or so.

Further, when a mass of information is available with the system and it cannot be disseminated on considerations of sensitivity and secrecy, such information should not be perpetually buried in official files. A system should be established whereby essential information is distilled, analyzed and put out in the form of periodic studies. Two distinct examples of this are: (i) details of government receipts and payments as are available in the books of the RBI based on Accountant General (Central Revenues) and the RBI's own Nagpur office; and (ii) mass of information collected by the RBI on individual banks and financial institutions (FIs) under supervisory and regulatory returns. Studies and reports should be prepared and published almost on a statutory basis so that the public at large comes to know about the performances of individual banks and institutions.

## **Section – IV**

### **CDBMS - Task Ahead**

#### **IV.1 Measures to Strengthen CDBMS**

Collecting data from diverse sources and collating them in an integrated and structured framework for analytical use by business analysts and researchers is a difficult and complex task by any measure. The CDBMS has accomplished this task to an appreciable extent in so far as it has laid the technological foundation for achieving this. However, the technological foundation that the CDBMS has laid out is a first step towards building a rich data repository that would meet the needs of researchers, analysts and policy makers. What is required is to build compatible and comparable data series in different subject areas of interest to the bank in a systematic way. This is the hard part of the job and the Reserve Bank, being the central bank of the country, is in a unique position to accomplish this in collaboration with other source agencies, provided a clear long-term strategy is put into place for achieving this goal. The Group deliberated on this strategy in detail and its views/ suggestions are outlined in the following paragraphs.

In the first place, an institutionalized procedure must also be put in place to make sure that (a) the series included in the CDBMS are crosschecked for consistency with the data published by the original source; (b) revisions made by the source agency are promptly incorporated; and (c) current data compiled by various Departments/ affiliates of the RBI are promptly and automatically reported and incorporated in the CDBMS data repository. It is also desirable to enter into a dialogue with the source agencies to work out mechanisms and modalities for prompt reporting of current estimates, revisions of earlier estimates and changes in scope, concepts and methodology used.

The data producing Departments need to be sensitized about releasing the data as soon as it is ready, including any revisions thereof. If the data producing entities disseminate the data through multiple channels, then it might be difficult to keep data integrity. They should ideally put the data first in CDBMS and then generate the statements from this platform. In case there is unavoidable lag in updating data in CDBMS that should be properly annotated. In this connection, there is a need to focus on Weekly Statistical Supplement and the RBI Bulletin. Producers of the data should be responsible for data integrity and synchronization of data flow to CDBMS with other modes of release of data by bank and other agencies outside RBI.

Then there is the task of including long time series along with Metadata and annotations in the CDBMS. To begin with, the internal project team of CDBMS should, in consultation with user Departments, work out a programme for compiling time series for all the indicators for which RBI is the source agency from 1990 onwards and draw up a phased programme for back-casting. The RBI Departments, which are responsible for generating various monetary, banking and other financial sector data on a day-to-day basis may be assigned the task of constructing back series as well as that of providing descriptions of concepts and definitions, data sources and data limitations for disseminating through the CDBMS.

For example, the various functional divisions within Department of Economic Analysis and Policy (DEAP) of RBI use practically every aspect of macroeconomic data. At present, they obviously compile data and analyse them for policy purposes or otherwise, though on a selective scale. The RBI may achieve some economies of scale by entrusting the officers of the various DEAP divisions with the task of constructing historical series in their respective areas of interest. Such a division of labour between the two research Departments is necessary for the successful execution of the CDBMS programme. Capital Market Division, DEAP may take care of capital market data series, Rural Economics Division may construct agricultural statistics, Industrial Division- industry data and National Income Division - the entire data set concerning national accounts and state domestic product. Likewise, the external sector data may be cleaned and made available to the CDBMS unit by the Division of International Finance (DIF) and Division of International Trade (DIT) in the DEAP. They should also collect information about the availability of longer time series on these subjects and make arrangements to compile and incorporate them in the CDBMS in a time bound programme. As part of this process, in particular, details of concepts, scope and methodology used as well as changes therein may be obtained from the agencies responsible for collection and/or estimation of the relevant data series.

As regards macro economic and cross country data from non-RBI sources, component units of DEAP which use them most and are familiar with the sources, should be in a good position to locate and cull out the relevant old series in areas of interest to them. They should be entrusted with this task and supplying the relevant material to the CDBMS over a period of time.

The primary responsibility of compilation of metadata should lie with the agencies that collect the primary information. Departments and units within the Bank that collect the primary data should be required to furnish details of their meta data for the system. Locating, accessing and digitising metadata from non-RBI sources is too big a task for the RBI to undertake. Many of the problems in accessing government data can be overcome only through the Central Statistical Organisation and the proposed National Statistics Commission. It is very important that respective Divisions which are regular users of these outside sources of data on different subjects take an active interest and serve as catalytic agents in compiling time series

from such sources as well as persuading them to compile the meta data or facilitate its compilation by commissioning researchers and institutions to undertake this job. The CDBMS team should also be associated with this process as this would help in building a knowledge base within internal project team.

As all this would take time and resources, it is necessary to prioritise the data series to be covered and prepare and implement a phased, time-bound programme. This would require more funds and personnel, if need be by creating a special unit specifically dedicated to this task. We need not, however, wait for the completion of this task to open the data already available in CDBMS for public access through Internet. Improvement in the coverage and quality of CDBMS will necessarily be a gradual and phased process. The main point that the Group would like to strongly underscore on this issue is that the objective of building such a massive database with required annotations involves a gigantic task and cannot be undertaken by one small single unit. This task must be undertaken jointly by all relevant Departments/Units within the RBI. Further, to strengthen this process, as well as to bring outside researchers' perspective regarding data content and format of CDBMSi on an ongoing basis, the RBI may consider formation of a standing advisory group including outside experts.

The proposed division of labour within the RBI notwithstanding, the residual responsibility left with the internal project team of CDBMS will remain substantial and, therefore, the unit should have sufficient amount of expertise in all matters concerning data management. Hence, the unit has to be substantially strengthened with qualified research staff so that they can handle queries of any type and also be sure of what they are disseminating through the CDBMS. The staff strength and budgetary allocation for the CDBMS unit and related activities have to be increased in a measure commensurate with the task it is expected to perform. It is also important to institutionalise mechanisms to ensure greater interaction and collaboration among CDBMS as well as the sources and users of the system.

## **Section V**

### **Summary of Recommendations**

Based on the above detailed analysis of current data content of CDBMS, the group has the following recommendations to make in regard to improving its coverage, quality, timeliness, presentation and access of data and metadata by public, so as to make it useful to the researchers and analysts within and outside RBI. Specific recommendations are also made for strengthening organizational infrastructure of CDBMS

#### **V.1 Coverage**

There is a considerable scope to improve the coverage of some important segments of the financial sector on which RBI does not directly collect much data. This will require inputs from other institutions like NABARD, IDBI, SEBI and the stock exchanges. RBI should explore the possibility of getting these institutions to furnish, and regularly update, basic data in a format appropriate for incorporating in the CDBMS. Providing a link to their web sites through the CDBMS will enable users to access to additional or more detailed information they may need relating to their domains. (Section II.2.1)



Access to data for cross country comparisons or analysis can also be facilitated by providing links to websites of international organisations (such as UN, World Bank and IMF) and references to websites where more detailed, country specific information is available. (Section II.2.1)

There is also a felt need to expand the coverage of macro-economic indicators. The extent to which this is possible depends on the content and format in which the data collected by various government and quasi government agencies are made available to the public. These data, which are published in print and digitised form can, after due scrutiny and review, be incorporated in the CDBMS. In many cases, incorporating them in the CDBMS will require a considerable amount of effort to persuade the concerned agencies to cooperate by locating, assembling and digitising the data and agreeing to making the data available to the public. (Section II.2.1)

Meaningful analysis and interpretation of trends require long time series. CDBMS has been able to meet this requirement only to a limited extent. It is recommended to include long time series data in CDBMS as per its availability. (Section II.2.2)

The data posted in the CDBMS is selectively culled out and processed from the much larger and more detailed primary information collected by sourcing agency. Many of the detailed and specialized requirements of users can be met provided the contents of the primary data are made known and the CDBMS permits and enables access to the meta data or a means of getting the meta data processed in the light of specific requirements. (Section II.2.3)

## **V.2 Quality & Timeliness**

It is observed that in many cases data are not obtained from the original source. In these cases, there are revisions in estimates as well as changes in scope and classifications, which are not always promptly reported to CDBMS which in turn lead to inconsistency in data. It is essential to source data from the original source of data so that necessary authentication can be done by the data-generating agency. (Section II.2.4)

In order to be useful it is essential to (a) ensure that the data are validated by source agencies and promptly updated; (b) inform the user about the scope, concepts and methodology of the sources of information posted in the system; (c) indicate the extent of comparability of data in the above respect for any given subject area across regions and over time and changes that may have occurred in these respects. (Section III.3)

## **V.3 Presentation of Data**

There should be two modes of data retrieval from CDBMSi. In the standard mode, data would be available through standard, intuitively meaningful tabular and pre-determined format. In this mode, a user would be able to select time slice (from and to selected periods) for which data can be downloaded in EXCEL or pdf format. In the ad-hoc query mode, users would be given the choice of creating his/her own basket of downloadable data series. It would also help users considerably if they could visualize data in graphical format in the CDBMSi environment itself. The Group suggests that there should be two basic documents available to CDBMSi users - one showing the definitions and concepts of data and the other the hierarchies of the available data series in tree structure in pdf format. (Section II.3)

#### **V.4 Metadata**

It is further necessary to provide (a) long time series with proper annotations of sources, concepts and methodology and changes therein; (b) higher levels of dis-aggregation by region and sector in respect of several indicators ensuring as far as possible as per standardized classifications; (c) information on meta data relating all major data series and incorporating them to the maximum extent possible in the CDBMSi. All these would require considerable strengthening of CDBMS both in terms of budgetary resources and human capital. (Section III.3)

#### **V.5 Data Access by Public**

The negative list (i.e. Data not to be made available to public) must, in Group's view, be limited to items of information that (a) are collected on the assurance that informants' identity and information relating to them will be kept confidential; (b) give information advantage to the regulator, and (c) are of uncertain quality. **Confidentiality** can be taken care of in three ways 1) Aggregation/consolidation 2) Masking the identity of units and 3) Releasing data after a time gap. It appears that the Departments have not identified any threshold time beyond which a data element would lose its sensitivity and can be made available to public. The Expert Group is of the opinion that unless data are obtained with explicit commitment to maintaining its confidentiality, it should be fit for publication if it does not directly impact the functioning and performance of market players. Even in the later case, after a hiatus of reasonable time such data should be fit for publication. Further, when a mass of information is available with the system and it cannot be disseminated on considerations of sensitivity and secrecy, such information should not be perpetually buried in official files. A system should be established whereby essential information is distilled, analyzed and put out in the form of periodic studies. The expert group is of the opinion that the Departments'/Divisions' lack of confidence on the quality of the data should not be the grounds for it being placed in the negative list. Provided that data quality disclaimers are given, and provided that appropriate qualifications are given in the data definitions, there is no harm in providing it for the public. The first page of the website therefore should contain a strong qualifier that the data are only indicative and may be changed/corrected in the future. (Section III.4)

#### **V. 6 Maintenance of CDBMS**

In the first place an institutionalized procedure must also be put in place to make sure that (a) the series included in the CDBMS are crosschecked for consistency with the data published by the original source; (b) revisions made by the source agency are promptly incorporated; and (c) current data compiled by various Departments/ affiliates of the RBI are promptly and automatically reported and incorporated in the CDBMS data repository. It is also desirable to enter into a dialogue with the source agencies to work out mechanisms and modalities for prompt reporting of current estimates, revisions of earlier estimates and changes in scope, concepts and methodology used. (Section IV.1)

The RBI Departments, which are responsible for generating various monetary, banking and other financial sector data on a day-to-day basis may be assigned the task of constructing back series as well as that of providing descriptions of concepts and definitions, data sources and data limitations for disseminating through the CDBMS. (Section IV.1)

As all this would take time and resources, it is necessary to prioritise the data series to be covered and prepare and implement a phased, time-bound programme. This would require more funds and personnel, if need be by creating a special unit specifically dedicated to this task. We need not, however, wait for the completion of this task to open the data already available in CDBMS for public access through Internet. Improvement in the coverage and quality of CDBMS will necessarily be a gradual and phased process. The main point that the Group would like to strongly underscore on this issue is that the objective of building such a massive database with required annotations involves a gigantic task and cannot be undertaken by one small single unit. This task must be undertaken jointly by all relevant Departments/Units within the RBI. Further, to strengthen this process, as well as to bring outside researchers' perspective regarding data content and format of CDBMSi on an ongoing basis, the RBI may consider formation of a standing advisory group including outside experts. (Section IV.1)

## **Appendix - I**

### **Office Memorandum: Expert Group on Internet Deployment of CDBMS**

The Bank has established a Central Database Management System (CDBMS), an enterprise wide data warehouse, built around an integrated repository of current and historical data, which is a state of the art decision support system within the Bank. The system has been available for access and analysis to users in the Bank over the corporate closed user group network since December 2002. As CDBMS contains a rich repository of data sourced from different Departments of the Bank as well as Government, it is proposed to place the publishable part of the CDBMS in the public domain for the convenience of researchers, analysts and other users. To guide this process of placing relevant data in CDBMS in public domain and to dovetail it to the data requirements of the user community outside the Bank within the overall framework of data dissemination policy of the Bank, it has been decided to constitute an Expert Group.

The Group will have the following terms of reference:

- a) To review data contents of the publishable segment of CDBMS in terms of its usefulness to researchers and analysts outside the Bank and suggest fresh data items that could be included in CDBMS for public dissemination subject to its feasibility.
- b) To provide a user requirement perspective to the Bank in respect of the Bank's current data dissemination policy and suggest changes, if any, in the light of best international practices in this regard for the Bank's consideration.
- c) To review the current data flow system for CDBMS in terms of its timeliness, reliability and coverage and suggest improvement, if any
- d) To suggest a technically feasible user interface for accessing CDBMS data over Internet so that it is made available publicly.

- e) To provide guidelines for metadata creation for end users of CDBMS over Internet.
- f) Any other issue considered relevant for dissemination of data through CDBMS.

The constitution of the Expert Group would be as follows

Dr. A. Vaidyanathan, Former member, Planning Commission	Chairman
Dr. R.B. Barman Executive Director Reserve Bank of India	Member
Dr. S.L. Shetty Director EPW Research Foundation Mumbai-400101	Member
Dr. Vaskar Saha Add. DG CSO Ministry of Statistics Planning and Implementation Government of India New Delhi	Member
Dr. Surjit Bhalla Managing Director Oxus Research and Investment New Delhi	Member
Dr. S Gangopadhyay Director India Development Foundation (IDF) Haryana	Member
Dr. Laveesh Bhandari Indicus Analytics New Delhi	Member
Dr. A.K Nag Adviser, DESACS Reserve Bank of India	Member-Secretary

The Group could co-opt members depending on the subject under consideration and may constitute technical groups to report on specific technological issues. The bank will reimburse expenses towards travel, transport and incidentals for non-official members for attending the meetings of the Expert Group. The Expert Group may submit its report within three months of its first meeting.

DESACS would provide secretarial assistance to the Expert Group

Sd/-  
(Rakesh Mohan)  
Deputy Governor

Mumbai

Date: January 3, 2004