



Statistics Division
United Nations



Department of Statistics
Malaysia

Seminar on
Data Dissemination: Emerging Trends and Issues
1- 3 August 2007
Kuala Lumpur, Malaysia

Empowering Users in Analysing Statistical Data

P.H. Khopkar

Deputy Director General

Shanker Lal Bodat*

Director, Computer Centre

Ministry of Statistics and Programme Implementation, India

August 2007

* Paper compiled and prepared by P. H. Khopkar, Deputy Director General and Shanker Lal Bodat, Director, Computer Centre; and is meant for discussion in the United Nations Seminar on Data Dissemination at Kuala Lumpur, Malaysia, 1-3 August 2007.

1. HISTORICAL BACKGROUND

The Ministry of Statistics and Programme Implementation has been collecting and compiling large volumes of statistical data of national importance. The processing of data collected on various subjects was done using main frame computers. The reports, prepared and made available to public, were based on processed information which contained largely meta-data and analytical tables. The users, due to non availability of computing facility with them were dependent on reports released by the various agencies including government agencies. In other words there was no significant demand for raw data by users.

1.1 In absence of resources for undertaking special studies requiring analysis of data in ways other than available in published reports, users were at freedom to demand tabulated information from the concerned Ministry/Department. Accordingly the Ministry used to generate and provide special tables as and when demanded by the users, both government/non-government, based on their requirements. However, at no time unit level data was either demanded by the user or provided by the government.

1.2 With the invention of Personal Computers the scenario has changed significantly. Now computing power is available with most of the users/researchers and it is felt by them that they be allowed to analyse data in the manner suitable to them. Further, availability of data gives them freedom to process data in various combinations which otherwise was not feasible in getting done through concerned department.

1.3 The changed scenario has demanded shift in government's policy of dissemination of statistical information. In 1999, Government of India introduced a national policy on "Dissemination of Statistical Data". Under this policy users of statistical data can now seek and obtain data at unit level on subject of interest subject to condition that data is not of sensitive/confidential nature. The data can now be given to any national or international users on payment of nominal cost which is worked out excluding cost of collection and validation of data. This approach of keeping cost at lower side is followed to encourage users for undertaking analytical studies/research.

2. CURRENT PRACTICES

2.1 The Ministry brings out various publications in the form of reports/publications based on surveys and also based on secondary data. In the past only hard copies of these publications were printed and provided to selected users on complimentary basis to the Central Ministries and relevant departments of the State governments and Universities etc. Some of the publications were also kept with the Directorate of Publications for the purpose of sale.

2.2 To enable a large segment of population to have access to the publications/reports and with the popularity of internet facilities, all the published reports of the Ministry were uploaded on the Ministry's website. To facilitate easy access and feasibility of processing as required by user, the reports/publications were provided in different formats such as; Excel, PDF, or ASCII. Making available soft copies of reports/publications resulted in faster dissemination apart from saving substantially due to reduced requirement for printing.

2.3 Introduction of IT and its use by various academicians, researchers etc provided a significant tool in their hand for processing of data in different manner than undertaken by the government agencies. This resulted in user to demand for unit-level data for undertaking their research work. Thus the government started providing soft copies of unit-level data in desired formats for use by researchers together with concepts and definitions of various terms involved. To facilitate appropriate analysis of data, technical guidance is also provided on methodology adopted for collection and compilation of data on request to the users both within and outside the country.

2.4 To cater to the needs of users, not having adequate processing facilities or do not wish to undertake processing of their own, the policy of providing processed/tabulated data on the selected variables/parameters is also followed by the government.

2.5 All the latest publications of the Ministry are now available in full on the web site. The users can access and download these reports free of cost, after registration. The government has also adopted Special Data Dissemination Standards of the IMF for timely release of various reports, indices etc. and are simultaneously uploaded on the Ministry's website. Press releases on Consumer Price Index Numbers, Index of Industrial Production and various estimates of GDP are uploaded on the website on the day of release itself.

2.6 The web site is an important portal detailing working of the ministry and a gateway for providing information to users. The ministry's website has been designed, developed and maintained by Computer Centre of the Ministry with the following data/information layers:

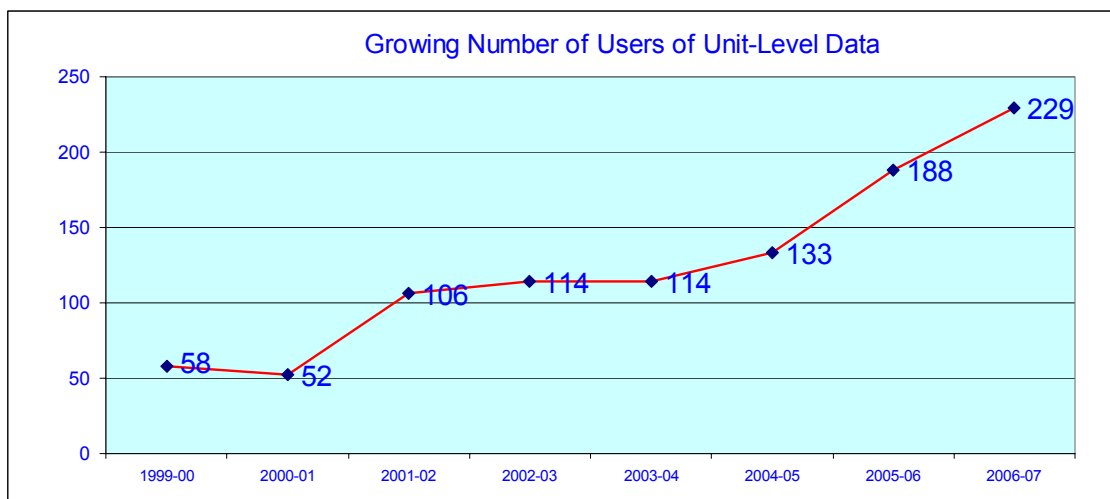
1. About the Ministry with Vision & Mission
2. Statistical Headline News
3. New Press release, PDF reports, Office Orders, important circulars
4. Search facility
5. National/International Conferences/Meetings
6. MPLADS(Member of Parliament Local Area Development Scheme)
7. Navigation to the website, Site Map
8. Data dissemination procedures, guidelines and price polices.
9. Procurement of Hardware/software/tendering
10. Flash Report on project costing more than Rs. 100 Crore.
11. Scheme for research
12. Other linkages:

- i) Database of Indian Economy (RBI)
- ii) Network of National and International sites
- iii) National Data Warehouse

2.7 The number of users registers for on line access/view/download of reports/publication during the last eight years is growing at a steady pace. As many as 717793 users have accessed ministry's website since 1999. Initially, number of users accessing the reports/publication was small, it has now reached to a level of about 1000 users per month.

The interest of researchers from across the world in data disseminated by the Ministry could be seen from the fact that about 27% of total users are from United States, United Kingdom, Singapore, Australia, Germany, China, Japan, Canada and France.

2.8 Unit level data in the form of soft copies are disseminated regularly to the users, both national and international, on demand on payment basis. The following chart presents the increasing interest shown by users in obtaining unit level data on various socio-economic subjects since 1999-00. The users includes, apart from individual(both national and international), universities, government and private organizations have increased many fold towards the use of data.



3. Changes in Information Technology - Concept of Data Warehouse

3.1 The mechanism (as detailed above) followed for supply of information to general public has its own limitations. The printed reports provide only summary information as visualised by the ministry concerned and need not meet user requirements. The processing of unit level data by the users also necessitates him/her to understand completely the methodology followed for gathering of data and complexities in processing. Often, it is beyond the capacity of user to easily

comprehend the complex procedures followed. Further, the user needed to contact subject matter Ministry (often more than one) for obtaining relevant data for his/her research work.

3.2 To add convenience, from the users point of view, and to facilitate use of data by a larger number of users it became essential to circumvent difficulties faced by users. This necessitated making it simple for the user to process data without even bothering to know exact methodology, concepts etc. The concept of data warehousing provides simple tools for the users to do the same.

3.3 In order to empower the users/researchers in analysing statistical data, the Ministry is now creating a National Data Warehouse of official statistics that would provide information to the specialized and general data users with easier access to published and unpublished data from one source.

3.4 The Ministry of Statistics has taken up a project on creation of the data warehouse. The proposed data warehouse would cover all statistical databases of the Central Ministries, State Governments and Public Sector Undertakings. The data warehouse is planned to be built in three phases. The first phase would cover all statistical databases maintained by the Ministry. The second phase would cover databases of all other central Ministries and State Governments. The PSUs databases would be covered in the third phase.

3.5 The prime objectives of National Data Warehouse are:

- To develop an integrated repository of current and historical data generated not only by Ministry of Statistics and Programme implementation, but all the Central Ministries and PSUs, that can provide various sliced/diced views of data across various dimensions of interest.
- It would enable the user to analyse the data by adopting the mechanism of drag & drop and generate information in the desired format.
- The definition, methodology, sampling procedures and other related information will be available by a click of mouse
- To build a state-of-the-art decision support infrastructure with online analytical processing (OLAP) capabilities, which includes providing a multi-dimensional and subject oriented view.
- To build user group specific data mart based on the central repository over various phases of implementation.
- To provide extensive integrated metadata support across all layers of data warehouse solution.

- It will provide freedom to the users from understanding complex processing needs.
- To enable web-based access to the Warehouse.
- Metadata will be provided with the better understanding to the data user with information related to quality issues, sampling methodology, classifications used, editing procedures that may impact on data quality and time series trends, survey coverage rules and changes etc. All of this information enables the data users to become better informed regarding statistical processes and the data series themselves. Ultimately this will result in a better-informed user community and consequently more sophisticated use of the available statistics.

It is expected that setting up of the data warehouse will go a long way in satisfying the data needs of :

- Data needs of specialized users, decision-makers, research and development as well as it will serve the general interest of the national and international level users.
- It would enable the users for data dissemination from one source.
- Enabling the users for on line analysis of data, generating the reports, tables and downloading the readymade reports.

4 Major Challenges in Dissemination of Official Statistics

4.1 Integration of formats

The issue of integration of various databases in data warehouse is significant from consideration that the user should be able to fetch information available from different databases without having to make multiple searches. The databases maintained by different agencies differ in the formats they are maintained and the parameters usually follow varying concepts/definitions. To bring at par for comparison purposes it is essential that the designing of data warehouse is done with due thought on interrelationship of parameters as also feasibility of their comparison across databases and over time.

4.2 Ensuring accuracy of data collected, and its availability

The state-of-art-technology can help to ensure accurate data collection. A national standard or agreed criteria for measurement can help to ensure accuracy, but this need not dictate the technology relied upon.

The availability of data to those responsible for its collection is clearly dependent on any number of issues, including time, financial and human resources, meteorological conditions, use of the right technology etc. According resources for these activities is a political issue dependent on national and local priorities.

4.3 Security issues: On the one hand the Internet opens new opportunities for the users while on other hand it poses challenges of its own, primarily in the form of data and system security, both of which are extremely important issues from integrity and authenticity point of view. Data security during transmission is very important, especially if unit level data are involved. Data may be sent in a coded form, password protected, or some other security system may be introduced to preserve confidentiality.

Further challenge relates to the security of the entire computer network within the national statistical office. The use of the Internet is most practical and efficient if the internal computer network of the national statistical office is directly connected to the Internet.

In order to protect the internal computer network and safeguard against the possibility of the introduction of computer viruses or unauthorized access to confidential files, it is recommended that "firewalls" be used; these are computer programs with the capability of controlling the flow of data and denying access to unauthorized users.

5. CONCLUSION

The need of the day is to persuade users undertaking studies, utilizing data gathered by the government agencies so as to assess the accuracy, applicability and need for continuity of various schemes launched by the government for the welfare of the general masses. Empowering user to attempt analysis of data in a simplest and most cost effective way provides an answer. The use of modern tool and technology can provide number of ways to achieve this target. The data warehouse, in the field of statistics can not only simplify efforts of the government in providing information to users but on users side provides simple tools for analyzing data in any possible permutation and combination. The data warehouse also helps user to get every information from his/her desktop without having to run from pillar-to-post for gathering information.

References :

1. Dr. S. M. Tam, First Assistant Statistics Information Management and Census Director, Australian Bureau of Statistics(ABS), Australia
2. Ghazy Mujahid, Advisor on Population Policies and Development, Developing a data dissemination policy, Viet Nam.