# Key Issues in Managing and Utilising IT as a Strategic Resource for NSOs: India Paper

Dr. Pronab Sen
Chief Statistician of India

The greatest challenges that Statisticians have always faced are early and accurate processing of data and quick dissemination of the processed information. The invention of machine computing has dramatically changed the scope and reach of statistics as a as a fundamental tool of analysis and decision-making. The single most significant step which contributed extensively to quick processing of data was the development of punched card based "unit record system" and the Hollarith code. However, with the increase in the timeliness and accessibility of Statistical data, policy analysis and policy advocacy has proliferated rapidly and has led to further demands for new data within ever-decreasing timeframes. In fact, today the phrase "shortest possible timeframe" has become a moving target for a Statistician. Every new innovation in IT hardware and software gives rise to the expectation that existing data should be available quicker and new data should be available with increasing frequency.

With the replacement of the "unit record system" by computers, the nature of man-machine interface changed in a fundamental manner. While it certainly speeded up data processing, it required not only a completely new skill set, but also a considerably higher degree of technical discipline. It was not an easy task for NSOs, especially in developing countries, to adopt the new technology in view of lack of trained manpower coupled with hardcore technology assimilation issues involved in integrating computer based data processing in initial years of computerization. The problem was compounded by fear among staff about redundancy and obsolescence, arising mainly from improper knowledge about the power of computer and the manner in which it could be used. The huge of cost of computerization also prohibited many countries of the developing region in go for latest technology. Thus, at least in the initial years, it was not possible for many NSOs to switch over to Computers and, as a result, the Statistical System in these countries could not quite keep pace with the standards which were being evolved and implemented by the front-runners.

The issues of "affordability" and "adoptability", the main constraints which created a kind of "divide" among "developed" and "developing" countries in adoption of information technology (IT), were ameliorated significantly with the advent of personal computers and the possibility of decentralized computing. Development of user friendly operating systems coupled with other application software like development of RDBMS, Electronic Spreadsheet made a sea change in the use of computer by the NSOs.

With the advent of "networking" technologies, and more particularly the "Internet", a completely new dimension of IT applications emerged which, while taking care of a number of aspects of NSOs' erstwhile problem areas, especially relating to dissemination, opened up a fresh set of expectations that NSOs are now expected to meet.

## *The Statistical System in India*

The Indian Statistical System is a decentralized system reflecting the federal principles on which the Constitution of India is based. It consists of: (a) the Central Statistical System, comprising of the Ministry of Statistics & Programme Implementation (MoSPI), the Census Commissioner and Registrar General of India (RGI), the Reserve Bank of India (RBI) and the statistical wings of the various Central Ministries/Departments; and (b) the State Statistical System, which is also spread over several State departments and whose structure varies from State to State. There are well-defined boundaries in terms of allocation of duties and responsibilities of the various component parts of this system, and each component has considerable autonomy within its defined sphere. However, by and large, they all work within a common set of concepts and definitions in order to ensure harmonization and inter-state comparability of data. Many of these standards are laid down in international agreements to which India is a signatory and others have been evolved within the country itself. The Central Statistical Organisation (CSO), which is one of the wings of MoSPI, is the nodal agency responsible for developing, adopting, propagating and ensuring conformity with these standards.

The vision of the Ministry of Statistics and Programme Implementation is to be recognised as an efficient apex organisation for timely disseminating reliable and credible statistics consistent with international standards to all users, and ensuring efficient use of national resources through effective monitoring of programmes and projects. The Ministry has adopted following mission to achieve its vision:

- To maintain statistical standards, and reorient the processes and priorities in the realm of official statistics, in tune with the changing technological and economic environment and needs.
- To collect, coordinate, collate and disseminate credible and timely statistics for informed decision making and debate, within and outside the Government.
- To promote human resource development in the official statistics and encourage research and development in theoretical and applied statistics.
- To facilitate and monitor infrastructure and large central projects for minimizing the cost and time overrun and for achieving high performance levels through systems improvement, and adoption of best practices.
- To facilitate the efficient implementation of Member of Parliament Local Area Development Scheme (MPLADS).
- To monitor the Twenty Point Programme.
- Coordination and policy issues relating to National Common Minimum Programme excluding sectoral policies allocated to other Ministries/Departments

A Commission set up by the Government in January 2000 under the Chairmanship of Dr. C. Rangarajan reviewed the statistical system and the entire gamut of Official Statistics in the country. One of the key recommendations of this Commission was to establish a permanent National Commission on Statistics to serve as a nodal and

empowered body for all core statistical activities of the country, evolve, monitor and enforce statistical priorities and standards and to ensure statistical co-ordination among the different agencies involved. In line with the above recommendations, the Government of India established a permanent National Statistical Commission (NSC). The NSC has been entrusted with the following responsibilities:

(a) to identify the core statistics, which are of national importance and are critical to the development of the economy;
(b) to constitute professional committees or working groups to assist the Commission on various technical issues;
(c) to evolve national policies and priorities relating to the statistical system;
(d) to evolve standard statistical concepts, definitions, classifications and methodologies in different areas in statistics and lay down national quality standards on core statistics;
(e) to evolve national strategies for the collection, tabulation and dissemination of core statistics, including the release calendar for various data sets;
(f) to evolve national strategies for human resource development on official statistics including information technology and communication needs of the statistical system;
(g) to evolve measures for improving public trust in official statistics;
(h) to evolve measures for effective co-ordination with State Governments and Union Territory Administrations on statistical activities including strengthening of existing institutional mechanisms;
(i) to exercise statistical co-ordination between Ministries, Departments and other agencies of the Central Government;
(j) to exercise statistical audit over the statistical activities to ensure quality and integrity of the statistical products;
(k) to recommend to the Central Government, or any State Government, as the case may be, measures to effectively implement the standards, strategies and other measures evolved under clauses (c) to (h);
(l) to advise the Government on the requirement of legislative measures on statistical matters including the statute for the National Statistical Commission;
(m) to monitor and review the functioning of the statistical system in the light of the laid down policies, standards and methodologies and recommend measures for enhanced performance.

The Ministry has three major wings, viz. Central Statistical Organisation (CSO), National Statistical Organisation (NSSO), and Programme Implementation. The Central Statistical Organisation is responsible for coordinating statistical activities in the country, and for evolving and maintaining statistical standards. The Computer Centre, an attached office of CSO, was set up in 1967 as an attached office of the then Department of Statistics with three second generation Computer Systems to cater to the data processing needs of not only the Department of Statistics but other Ministries/Departments of the Union government as well.

The National Sample Survey Organisation (NSSO), established on the basis of a proposal from Prof. P.C. Mahalanobis to fill up data gaps for socio-economic planning and policy making through sample surveys, conducts nation-wide, large-scale socio-economic surveys. The NSSO also undertakes field work for the Annual Survey of Industries and follow-up surveys of Economic Census, sample checks on area enumeration and crop estimation surveys and prepares the urban frames used in drawing of urban samples, besides collection of price data from rural and urban sectors. Both household surveys and establishment surveys are conducted by the NSSO to cover various aspects of need for data for policy purposes.

The Programme Implementation Wing of the Ministry is the management arm of the Government of India. It provides management services by providing latest information, by adopting the mechanism of Online Computerised Monitoring System (OCMS), in a global manner on implementation of projects/programmes costing Rs. 200 million and above and on the performance of infrastructure sectors.

The Office of the Census Commissioner and Registrar General of India is a statutory body primarily concerned with conducting the population census and with registration of births and deaths. However, in keeping with federal principles, it has only limited field presence and operates largely through State government formations. Much of the vital statistics of the country is gathered through the Sample Registration System (SRS) operated by the RGI.

The Reserve Bank of India is responsible for all data on the monetary and financial sectors in the country, but also collects data on a number of other areas such as trade values and corporate investment flows, both domestic and international. Most of the data collected by RBI are either record-based or returns-based.

Some of the Ministries of the Central Government are charged with the responsibility for collecting and maintaining data related to their remit. The most important of these are Ministries of Agriculture, Industry & Commerce, Labour & Employment, Human Resource Development and Health & Family Welfare. Most of them do not have field formations, and rely largely upon either administrative records, the State government machinery or the surveys conducted by the NSSO.

The main source of much of the primary data in India is the State Statistical system, which is supposed to provide all the information required by State governments for their planning and policy purposes and are expected to cover data needs at the sub-State levels as well. However, there is large variation between States in terms of the capability of their statistical systems. This causes enormous problems in deriving national aggregates from the data provided by States

## *Status of Information Technology in India*

The use of IT in NSOs cannot be seen in isolation from the spread of ICT in the country. One necessarily impinges on the other. The ICT industry and its use, in this region, are at various stages of development and during last two decades this scenario has undergone an amazing transformation. In fact the ICT industries of this region, in particular the IT software and services, have not only managed to catch up with their more technology savvy global leaders, but they are also being actively sought by companies worldwide for their onsite, offshore expertise and wealth of manpower resources. Thus, we should be able to meet our own IT demands for Statistical activities since the socio-economically point of view, we have uniform statistical data needs. If this agreed to then should we not share how we manage our system through use of IT as a potential tools for development and improvement of our Statistical system ?

Use of Personal Computers has tremendously increased from 5.4 million PCs in 2001 to 14.5 million in 2005. As on today, only every hundredth person has a personal computer, which is much less compared with any developed country.

**PC population and in use per 100 population**

| Year (March Ending) | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|
| PC population (in million) | 5.40 | 6.00 | 8.00 | 11.00 | 14.5 |
| PC in use per 100 population | 0.53 | 0.58 | 0.77 | 1.04 | 1.34 |

Source: Ministry of Communications and Information Technology

'Internet Users per 100 populations' is one of the indicator of Millennium Development Goals. Though we have a rapid positive trend for this indicator, compared to the developed countries, we are at the infant stage. Even the 200th person is not an internet subscriber in India. However, every 35th person is using internet in India.

**Status of Internet Subscribers**

| As on | Internet Subscribers (million) |
|---|---|
| 31.03.1999 | 0.230 |
| 31.03.2000 | 0.943 |
| 31.03.2001 | 2.909 |
| 31.03.2002 | 3.239 |
| 31.03.2003 | 3.500 |
| 31.03.2004 | 4.050 |
| 31.03.2005 | 5.300 |
| 30.06.2005 | 5.556 |

Source: Ministry of Communications and Information Technology

Indian ITES-BPO sector industry continues to grow from strength to strength, witnessing high levels of activity - both onshore as well as offshore. The export revenues from ITES-BPO grew from US $ 2.5 billion in year 2002-03 to US $5.1 billion in the year 2004-05. The ITES-BPO sector has become the biggest employment generator amongst young college graduates with the number of jobs almost doubling each year. the number of professionals employed in India by IT and ITES sectors has grown from 284,000 in 1999-2000 to 1.05 million in 2004-05, growing by over 200,000 in the last year alone. the pace of recruitment picked up for IT serves; while ITES-BPO companies were recruiting in large numbers through the year. It is estimates that the ITES-BPO sector hired 400 personnel every working day of the year.

The Government vision is to use Information Technology as a tool for raising the living standards of the common man and enriching their lives. Towards this end an ambitious programme of PC and Internet penetration to the rural and under served urban areas has been taken up. The Department of Information Technology has initiated a programme to establish State Wide Area Network (SWAN) upto the block level with a minimum Bandwidth of 2 MBPS to provide reliable backbone connectivity for E-Governance. Further, in order to bring about a substantially increased proliferation of .IN Internet domain name, a new .IN policy framework and implementation plan has been formulated and announced in October 2004. It aims to put in place a more liberal, efficient and market friendly domain name registration system. It aims to ensure that the internet traffic, which originates within India and also has destination in India, remains within the country, resulting in improved traffic, reduced cost and better security.

The National Policy of the government recognizes the potential of E-Governance not only to improve governance but also to facilitate people's access to government services. We are working on a National E-Governance action plan that seeks to lay the foundation and provide impetus for a far more pervasive spread of E-Governance to reach the Common man, particularly in far-flung areas. Seeking to do so we are putting together various elements that are needed for leveraging the enormous power of ICT for the economic development of our country and enable the common man to access Government services in an efficient, convenient and cost effective manner.

To bridge the imbalance between urban and non-urban areas, provide connectivity at grass-root level, and to facilitate the spread of benefits of Information and Communication Technology to all and to accelerate the socio-economic development of these areas, the Department has set up Community Information Centres (CICs) in hilly, far-flung areas of North-East and J&K. It is also proposed to set up CICs in other hilly, far-flung areas of the country.

Broadband services greatly contribute in the growth of GDP through applications including e-education, e-health, tele marketing, e-governance, entertainment, etc. However, the current level of Internet and broadband access in India is abysmally low. In the Broadband Policy announced in October, 2004, the broadband connectivity has been defined as "an always on" data connection that is able to support interactive services including internet access with a minimum download speed of 256 kbps to an individual

subscriber. The policy visualizes creation of infrastructure through various access technologies for providing broadband services. It is expected that the number of broadband subscribers would be 3 million by 2005, 9 million by 2007 and 20 million by 2010".

Other initiatives taken by the Government in the IT Sector include announcement of the Information Technology Act 2000 for copyright protection, the Internet Service Providers (ISPs) Policy permission to private ISPs to set up international gateways and internet access through cable TV infrastructure among others. With a view to improving timeliness of collection of data from establishments and promoting use of ICT, the schedule of the Annual Survey of Industries was modified in 2000-01 to include questions on availability of computerized accounting system with the manufacturing units, their ability to provide information on electronic media, their e-mail IDs etc. Availability of computerized information from establishments would also help in developing an on line data capturing mechanism which would further help in utilizing resources effectively.

## *Experiences and Lessons*

India moved relatively early in adopting IT for statistical purposes. It introduced the use of the computer in 1952 when a HEC-2M system was installed at the Indian Statistical Institute, Kolkata, mainly for research purposes. However, the NSO had to wait for another 15 to 20 years to get the full advantage of computerization. The processing of data collected through various administrative sources and surveys conducted by the NSSO, was initially done in house at the Computer Centre of the Ministry. To cater to the need of increased work load for processing of data as also to improve timeliness, six zonal data processing centres were set up by the Ministry. The Ministry, in the field of ICT, has taken a number of initiatives for improving the reliability and timeliness of statistical information from the stage of collection of data to processing and dissemination of information to the end users. The future development plans, in respect of use of ICT, include setting up of an All India Statistical Network and Ministry-wide INTRANET connecting all offices spread over entire country.

Given the long history of the use of IT in the Indian Statistical System, a number of lessons have been learnt which have a bearing on the strategic use of IT. In order to structure the issues, it may be desirable to classify under the various activities which comprise the courses of generating statistics in national statistical offices.

**Data Collection and Transcription:**

There are three broad methods of data collection which are (a) record based; (b) returns based and (c) survey based. Although IT can be used productively in all the three methods, its utility and effectiveness differ significantly. In the record based system, which is used primarily for capture of administrative data, the statistical system is dependent upon the extent of IT usage by the concerned government agencies. There is substantial variation in this regard with different Ministries and agencies operating at

different levels of IT sophistication. To make matters more difficult, in the federal system that exists in India the State Governments organize their administrative records in the non-standardized manner therefore, even if there is extensive IT use, it becomes very difficult to collate such data in a comparable manner. The solution clearly lies in standardizing the record keeping systems along the major functional lines across ministries and states. This is however strenuously resistant as it is perceived to impinge upon the autonomy of these agencies. More over, in a country as large as India there are regional specificities which do not readily permit exact standardization.

The most convenient way of capturing record based data would be through some form of electronic data interchange (EDI). However, any rational EDI system requires absolute standardization, which can happen only if there is a single agency or centralized system which is responsible for that particular data collection. This has been possible in limited spheres such as trade data and perhaps other centralized systems such as income tax. In other cases, a considerable degree of work would need to go in to developing record keeping systems which would be amenable to EDI. In the lower formations of administration, such as at the block or village level, lack of infrastructure coupled with absence of trained manpower can prove to be an insurmountable hurdle.

With the advent of the internet, IT use in returns based data collection has become much simpler and more user friendly. E filing of records is now becoming common place in a number of activities particularly those which relate to the corporate sector. However, in the Indian context, where there are a large number of non corporate enterprises there are limitations to the extent to which E filing can be used. More importantly, the experience over the years has been that returns based system for data collection are not particularly effective except in cases where there are statutory provisions and strong penalties for non compliance. In almost all other cases the response rates are usually very poor. For enterprise surveys, non-availability of Accounts in e-format has become a major bottle-neck in India where there are more than 42 million enterprises. Therefore, the use of returns based data collection in India is extremely limited, confined only to the corporate sector, and the IT component of what exists is even less.

There has been considerable expectations that the use of IT in collecting survey data would simplify and speed up the data collection process. In India there have been a number of experiments in this regard in trying to use hand held devices for electronic data entry at the field level. By and large the experience has not been entirely positive. The main problem arises from the fact that Indian surveys use highly sophisticated and lengthy schedules which cannot fit in present hand-held devices. Moreover, there are technical and environmental problems relating to battery life and recharging facilities, dust and moisture and maintenance in remote villages. Besides, India is a multi lingual country and simple electronic devices are unable to effectively capture the linguistic nuances which are necessary to ensure that comparable data is being collected across the country. As far as more complex devices are concerned, it is not clear that the cost associated with a large number of such devices are justified by the end result. This is particularly true in view of the fact that many data collection activities are periodic in nature and as a result the devices tend to be severely under utilized over their life span.

The one experiment which appears to have worked has been the use of "intelligent character recognition" technology in capturing census information through scanning. This combines the flexibility of a paper based system with the speed of electronic technology. The learning from this experience has been that the ICR technology becomes really useful when the schedule to be canvassed is completely standardized in every manner. The language issue in India does create problems in this regard simply because the character sets have to be programmed for each language separately. It has also been found that the use of this technology requires considerable discipline on the part of the enumerators while filling up the schedules.

**Data Validation :**

By and large, the checking and validation of data has been a manual function, and continues to be so even today. While it is certainly true that certain standard machine validation processes, such as use of logical relations, identifying outliers and howlers, are already widely used in India , these only limit the domain of manual verification and do not substitute it in its entirety. It is felt however that the power of IT has not really been harnessed in this particular dimension. It seems that it should be possible to carry out much more sophisticated statistical checks especially where time series data are available and through use of "artificial intelligence". These kinds of applications of IT for validation would appear particularly important in cases where estimates are being made either on the basis of proxies or incomplete data sets.

**Data Storage & Retrieval :**

This is one area where IT has completely revolutionized the official statistical systems. In the Indian NSO, practically no paper record is maintained and almost all data are stored in electronic form. The situation however is different in other data collections agencies, particularly as far as administrative records are concerned. Conversion of paper based records in the electronic form is a process which has begun in India, but the pace of progress is relatively slow. In recognition of this fact, the Government of India has launched a major nation wide initiative called the National E Governance Action Plan (NEGAP) through which computerization and networking will be taken down to the lowest level of the administrative structure. Once this is done, it would be possible to computerize the existing data set into electronic form in a decentralized manner and all future record keeping could be done electronically. There are two main issues in this area. The first is security. If electronic records are to be utilized effectively, they would have to be accessible through a network, which would make them vulnerable to unauthorized access and use. Given the decentralized nature of the Indian Statistical System and the huge variation in the quality of the manpower involved, it is not clear whether highly sophisticated data protection systems can be implemented at all levels. The second problem relates to data corruption. Although back ups are standard, there appears to be a tendency to treat electronic backups as if they were permanent. This is clearly not the case. There is need for all agencies which maintain data to make periodic

assessment of the entire data set and to migrate to more secure media as and when they become available. Unfortunately, this is rarely done.

As far as retrieval is concerned, while IT has made it enormously simpler than before, there are limitations which need to be recognized. First of all, efficient retrieval would require a degree of standardization of software and operating systems, which may otherwise neither be desirable nor possible. Second, there is a historical dimension to data storage, since such data were never stored in RDBMS, which limits the application of improved retrieval systems without developing data marts. Therefore, unless there is a continuous process of upgrading the format of the stored data, retrieval systems can become unmanageably complex.

**Data Processing :**

This is perhaps the one area in which IT has completely transformed the operation of the NSO. Today in India almost all data processing is computerized through use of RDBMS under client-server architecture, which has not only reduced processing time to a fraction of what it used to be but has also led to greater depth and coverage of official statistics. There are two main issues that need to be considered in this regard. The first is that computerized data processing demands a level of technical discipline that needs to be inculcated in the concerned manner. Errors in machine processing are not uncommon, but there is a tendency for people to assume that the computer processed data is error free, especially when some software packages are used without knowing their limitations. The fact that human errors can creep in at all stages of the process appears to be neglected. One of the ways of handling this problem would be to subject the processed data to statistical checks on a routine basis, and that such checks would have to be separately identified and applied to different kinds of processing work.

The second issue relates to the fact that complex data sets may not be amenable to standard statistical packages, and customized programmes have to be developed from time to time and occasion to occasion. The development of such software requires the availability of highly skilled programmers who are not only capable of developing appropriate algorithms and instruction sets but also are fully familiar with the nature of the data and the manner in which it is stored. This is rapidly developing into a major constraint, particularly in the Indian case since in large-scale survey data processing such software are generated in-house and cannot be outsourced due to confidentiality restrictions.

**Data Dissemination :**

It is a fact that until the advent of ICT, access to official data was fairly restricted simply because there was a limitation on the hard copies that could be printed and circulated. Today however with the advent of the internet and wide availability of personal computers official statistics can be accessed by a much wider spectrum of people. Gradually, the use of printed documents is giving way to electronic dissemination techniques. In order to provide key statistical data pertaining to all major

sectors of economy through a single window to all the users including planners and researchers simultaneously, the Centre has taken up the pilot project of creation and maintenance of National Data Warehouse of Official Statistics from the data generated through sample surveys, censuses, enquiries etc. Under this project, the Centre will preserve data generated by various Central Ministries and Departments and Public Sector Undertakings on electronic media, organize the data in the form of databases, create data Warehouse and provide remote access facilities to end-users through a network.

There are two dimensions that need to be kept in mind while recognizing this significant development. First, electronic data sets cannot assume that all potential users will have access to the requisite software and that it becomes the responsibility of the data provider to ensure that the data is made available in a accessible form. This either means the standardized format, which may not be possible, or providing the software capability within the electronic media for accessing the data set.

The second problem is some what more acute. Since the ability of different users to access electronic data differs, specially in developing countries, there is a danger of adding yet another dimension to the digital divide if paper based dissemination is discontinued. Moreover, since processing capabilities will also be very different, ensuring a minimum level of uniformity of information will require the NSO to provide critical information in more processed forms. This would require the NSOs to increase the degree of involvement with the user community so as to ensure that the information divide is unwittingly not made even wider.

## *Strategic Issues in Use of IT*

There are two main strategic issues that are being experienced in the use of IT in the Indian Statistical System. The first issue relates to manpower. In the early years of introduction of IT in the country, the statistical system was the major employer and also the training ground for IT professionals. With the rapid growth of information technology as a commercial activity in the country it is becoming increasingly difficult for the public statistical system to retain not only IT professionals but also statisticians. The private sector in India has been facing a severe shortage of manpower with strong quantitative skills during the past few years, and the statistical system provides a pool of skilled manpower which is relatively easy to poach upon. Things have come to such a pass that not only has it become extremely difficult to recruit young professionals with high level of quantitative skills, there is increasing attrition from the existing manpower as well. Although training of manpower is not an issue in itself, retention of the manpower so trained has become a major concern. As a part of the government, it is very difficult for the NSOs to offer compensation packages which can match those given by the private sector while keeping within the broad structure of public sector salaries. Consequently, it becomes necessary to evolve a strategy to make the best use of the manpower that is recruited and trained while recognizing that there will be a high rate of turn over. The solution is not obvious, particularly since the manpower will need to be familiar not only with the use of IT but also with the structure and nature of official

statistics. Unfortunately this combination takes time to develop, which is precisely where the problem arises.

The second issue has to do with the fact that almost all NSOs collect significantly larger volumes of data and are usually put out in their dissemination programme. Much of this is the outcome of the limitations faced by NSOs when the only form of dissemination was through publications. This limitation caused standard data sets to be developed which usually involved the production of highly aggregated information. With the advent of IT, and its spread among the user community there is increasing demand for release of more disaggregated data. Since information is a critical element in decision making and efficiency, it would appear that the NSOs should make available data in as disaggregated a form as possible. However, there is a limit to which this can be done keeping in mind the fact that confidentiality restrictions need to be observed meticulously. Moreover, in a competitive market environment, it is the responsibility of the NSO to ensure that it does not inadvertently create large discrepancies in the level of information that is accessible to different players in the market. This conflict can only be resolved through a dynamic process of ascertaining both the minimum information that the NSOs should put out and the maximum data that it can release so that market distortions are kept within manageable limits.