

Ensuring public access to data

N. Vittal

Satarkata Bhawan, General Pool Offices Complex, Block-A, INA, New Delhi 110 023, India

Availability of data is a major problem. Reliability of the data is another. The third problem is objectivity of the data. The fourth is whether the data is up to date. The fifth is whether the data is accessible. If we are thinking of modern information technology and putting everything on the Internet, then we must ensure the availability of computers, connectivity, content of data and above all the cost aspect are taken care of. One dimension of the issue of public access to data relates to the issue of logistics of provision of data. This involves three stages – data compilation, data transmission and data usage. Every organization has a method of compiling data. If only it takes into consideration who are the users and what exactly is the type of data and in which form it is needed, it will go a long way in making the data compilation worthwhile. In Government of India, the Indian National Science Documentation Centre (INSDOC) and Defence Science Documentation Centre (DESIDOC) are major compilers of data. It will be worthwhile to see the cost of data compilation and how the data compiled by these two dedicated bodies pass the test of specific parameters mentioned above.

It is a cliché to say that we are living in the age of information. In fact, information forms part of the quartet, which moves from data to information, knowledge and wisdom. Data is the starting point for the exploration of the subject. From the data, intelligence and to some extent imagination will help us to derive information. Further mental processing of information leads to knowledge. It is finally wisdom, which enables us how to use that knowledge. Many a time we can also moan with the poet T. S. Eliot 'Where is the wisdom we have lost in knowledge and where is the knowledge we have lost in information'.

As we start exploring the subject of ensuring public access to data, there are different dimensions of data about which we have to be clear about. What do we mean by public? As I see it public represents two things. One is the fact that the data is not kept secret and is in the public domain and is available to anyone who is interested in the data. The other thing that the term public means representing a group of people or communities. There will be different publics. The scientific researchers as a public may be interested in scientific data. Social researchers may be interested in data related to social parameters. Economists may be inter-

ested in data relating to the performance of the economy. The same will be true for each sector. Each sector of users represents one sector of the public. It is therefore obvious that the data also will have to be fine-tuned to the needs of the different sectors.

Availability of data itself is a major problem. This is the first complaint. A classic example of how data when required was not available was when Rajiv Gandhi launched the various technology missions. Sam Pitroda who became Advisor for technology mission at that time found that many a time critical data about drinking water or education were not available at all. Perhaps one of the early challenges of the technology mission was to collect the data itself.

Reliability of the data is the next major complaint. The third is the objectivity of the data. The fourth is whether the data is up to date. The fifth is whether the data is accessible. Then as we go on, we get into the various issues of the content of the data. If we are thinking of the modern information technology and putting everything on the internet, then we come into the series of Cs. The first is the availability of Computers, then the Connectivity, then the Content of the data and above all, the Cost. So, one dimension of the issue of public access to data will relate to the whole logistics of provision of data.

In our country, the government is the biggest generator of data and perhaps also the most inefficient generator and user of data. Every government office goes on sending a series of what are called periodical returns. Many a time when these returns are not sent, then there are reminders. If we make an analysis of the information contained in the periodical returns, study them and see how much use is made of them, we will find that there is tremendous shortage in the usage of data. I would guess as a ball park figure, hardly 20% of the information and data generated by the government organizations is utilized.

The entire issue of access to data has three stages – data compilation, data transmission and data usage. It is the manner in which the data compilation takes place that decides whether the data will be used. After all the term user-friendly has been made popular by the electronic industry. Many times, in practically every sector, we hear complaints about the appropriate information or data not being available. The success of the Centre for Monitoring Indian Economy (CMIE) which depends only on published information and producing data sys-

tematically over the last 25 years or more shows how a little imagination and perseverance can help in making virtually the available data into a mine. The problem with the government agencies connected with data is that they are always behind time. There is no reason why in this age of information technology and with the emphasis on India becoming an IT super power, we should not look to making maximum use of IT for speedy compilation of data and ready access. In fact one great feature of IT is that it helps to process massive volumes of data and make it also available easily. So in the context of a grand strategy to make data of public interest available to the public, the first step should be looking at the compilation of data.

Here perhaps every organization has a method of compiling data. If it only takes care as to who are the users and what exactly is the type of data and in what form it is needed by the users, it will go a long way in making the effort at data compilation, worthwhile. A classic example is the information kiosks, which have been put in the government of India offices in Delhi thanks to the energetic Cabinet Secretary T. S. R. Subramaniam. I do not know how much of the data that is available in the information kiosks is being used. Many government departments have started their web sites and what I hear is that these web sites contain government junk. Many a time the information is not up to date. So, it will be worthwhile if certain simple principles could be evolved by which any agency or organization compiling data is made to follow certain standards like quality, objectivity, up to date-ness, accessibility through the computer networks, and, above all, design the data in form which will be readily useful for the consumer.

In Government of India, we have, so far as scientific knowledge is concerned, two major data compiling organizations. One is the Indian National Science Documentation Centre (INSDOC) and the other is Defence Science Documentation Centre (DESIDOC). It will be worthwhile to see the cost of data compilation and how the data compiled by these two dedicated bodies pass the test of the specific parameters I have mentioned above. From their experience we can learn and see how we can improve on the data compilation and dissemination side.

When we talk about access to data, the most important aspect is the level of secrecy to be maintained. I would suggest that there should be a very small negative list of data, which may be kept secret. This may include strategic information like atomic energy or data that have a bearing on national security and so on. All other data should be in the public domain. A lot of exercise has been made to see that as against the Official Secrets act we have also a Freedom of Information Act. It is high time that we frame this Freedom of Information Act. If necessary we may even issue an Ordinance

and get this going instead of endlessly going around in circles. A good model to follow may be the United States where perhaps the impact of the First Amendment and Freedom of Information has made the data availability access and compilation much easier.

When it comes to data in India, particularly industrial and economical data, there is another aspect to be borne in mind. It is nice to blame government that they are obsessed with secrecy. A classic example is how Survey of India maps are treated as secret whereas these maps could perhaps be got on payment of dollars from United States more easily. But if you ask private sector how many of them are willing to give real data, few will come forward. Thanks to our tax system and tendency among industry to cheat on taxes, many industries and institutions in the private sector maintain dual accounts and as a result, we never get access to real data.

The permit licence raj, which started dying from 1991, had at least one benefit. Thanks to DGTD's omnivorous desire for data and deciding on the capacity and technology of every small industry in the country, DGTD became a good source for collecting data regarding industry. But now that liberalization has come, there is not pressure on the industry to give data. I understand that in the United States, there is legislation about making the data available. Perhaps in addition to Freedom of Information Act, we should also have an Act which will force the industry and others who are generating data in the private sector of the public sector, to give correct, objective, timely data to the authorities compiling the data. As we are in the age of private enterprise, such access rights must be available even to the private sector who take data compilation as their profession. Especially with the growth of IT, databases and data mining are becoming common place. Further, they are also throwing open new areas of employment. According to an estimate made by Vishwanathan of INSDOC, three lakh to thirty lakh jobs can be created in the area of bibliographic database alone in our country. So, in the context of generating new employment and also giving a boost to the economy, it is high time we have a law which will make making information available a legal requirement.

Technology for collection of data also varies and keeps on changing. Thanks to satellite imagery, it is possible today to develop accurate maps containing a wealth of data. Geographical information systems (GIS) are becoming very vital documents whether it is relating to rural planning and development or to urban management. It will be necessary to make the wealth of information available in these maps readily available to the public at large. It is good that the Government of India has thought of making India into an IT super power. The difficulties experienced in this area were also gone into by the National Task Force on Information Technology which had made specific recommendations

about the clearances to be given by the Defence Ministry so far as propagation of information regarding GIS is concerned. I hope new drills have been evolved which will ensure that denial by delay does not become a method of obstructing the spread of IT.

Another important feature of our country so far as data is concerned is that there is not enough networking among the various organizations. With the explosive growth of the Internet and improvement in telecommunication it should be possible to establish networks of various sources of data. In fact, it will be worthwhile to have a grand project for listing all the data producing organizations or sources of data. This sort of a master compilation itself will generate new ideas and sources for lateral thinking as well as cross fertilization of ideas. So far as scientific ideas are concerned, DESIDOC and INSDOC experiences can be the building blocks. But when it comes to other areas, be it the social sector or the industrial sector of the economy, we will have to see whether a massive effort could be made to start information network and then focus on the three specific aspects of data compilation, dissemination and usage.

While most of the time we are talking about the data which is current or which is generated in different ongoing activities, there are also data which have a bearing on the rich cultural tradition of India. I understand for example in Kerala and Tamil Nadu alone there are 24,000 manuscripts in different archives, of which only about 7000 have been published. So if we are thinking about public access to data, should we not also look into data mining of a different kind or going into the archeo-

logical aspect of data so that we use modern aspect of technology to see if it can help us to reach ancient wisdom?

There are also other activities, which today are not finding a place in the data in an accessible form. For example, in IIM, Ahmedabad, Misra has for the last eight years, running an NGO, which compiles innovations made by villagers, which never find publicity. In fact these are all practical solutions by illiterate people to practical problems faced in village environment. If we can bring this skill, information and innovations on a much wider basis, we will be able to also contribute to the rapid development and improvement in the quality of life of the people, especially away from the cities and in the rural areas. The same is equally true of another type of sub-terranean skill, which Yashpal has been highlighting. This relates to the knowledge of traditional skills like the potters, the handloom weavers or the metal smiths in their various handicrafts. In the age of information and knowledge, it is necessary that we find means of tapping such data which in turn can go through the spectrum of information to knowledge to wisdom and bring them also in an integrated fashion throughout the country if not international so that data compilation, dissemination and usage get a more added vigorous dimension exploring new horizons.

I am glad that the Indian Academy of Sciences has started this exercise of launching a detailed discussion on public access to Indian geographical data and also public access to data of various kinds and in particular scientific data. I hope the ideas I have mentioned may be of some use in this nationally significant discussion.