DESIDOC Bulletin of Information Technology, Vol. 17, No. 5, September 1997, pp. 3-10


# The Indian Database Scenario Portrayed Through INDAB

**SN Sur* and BG Sunder Singh****

**Abstract**

There is a marked shift from predominantly library based bibliographic information services to the generation of more factual, referral, numeric and multimedia information products. Yet, comparing with the global developments, we, the Indians are just at the starting line. In this article, portrayed through INDAB the indigenous database efforts, their growth & status, availability, opportunities for development, and related strategies have been analysed. It concludes that given some areas of strength and planned growth, India should have increased opportunities in information services in the years to come.

## 1. INTRODUCTION

In response to a growing demand for information services in the country in the 1960's, many library and information centres started providing various kinds of information services—of which the abstracting/indexing services were aimed at providing both current awareness and selective dissemination. These were predominantly produced in-house and by manual methods. By the 1980's, the diffusion of microcomputers, useful software and the necessary skill development in library/information centres had brought out improved production methodologies and delivery of these services in machine readable form. The provision of information services still remained, and continues to be so, in the domain of traditional providers such as libraries and information centres and largely accessible to in-house users.

By the early 90's, one could see changes in the national outlook. The changes in terms of liberalisation effect the way we have to do business, while change in terms of globalization stimulate how we generate knowledge and therefore wealth. As a result of these changes, there has been a perceptible increase in the use of computers and telecommunication facilities—not only in in-house data processing but also in providing information access and delivery to a wider user population. Apart from the diffusion of information technology at the workplace, there have been a number of efforts to organise indigenous information of utility and provide a competitive edge to individuals and institutions. Besides the government funded initiatives in different sectors, there is also a visible participation from the public as well as private sector institutions in providing information services. Such services are now accessible not only in local mode but also on a wider geographic area—partially through improved marketing and partially due to increased choices of information delivery. For instance, apart from the traditional bibliographic information services through print media, one now finds a migration to the digital

* *Institute for Social Analysis and Communication, New Delhi*
** *National Information System for Science & Technology (NISSAT), New Delhi*
(All correspondence may be addressed to Shri BG Sunder Singh at bgs@nissatd.ernet.in bgs@nissatd.ernet.in)

mode delivered over a network or on an optical medium such as CD-ROM and more recently, videodisc. Similarly, one could note a marked shift from predominantly library-based bibliographic information services to the generation of more factual, referral, numeric and multimedia information products. Yet, comparing with the global developments, we are just at the starting line. For instance, in spite of over 398 efforts known and recorded in INDAB, only one database, i.e., CMIE Database is available on DIALOG. Given some areas of strength and planned growth, India should have increased opportunities in information services in the years to come.

## 2. INDAB

Descriptions and discussions on the various aspects of indigenous information scenario may be found elsewhere.[1,2,3 & 4] The INDAB database is a conscious effort to identify potential database activities in the country through data obtained directly from the database producers; and where this was not possible, pertinent data had to be collated from the available sources. The intention was to identify the disparate efforts and on the basis of the information collected, nurture and develop such efforts into wider utilities and facilitate a planned development of information services on a national scale.

The reason to cover abstracting & indexing services, directories and databases together at this juncture is obvious. The traditional bibliographic/referral services are generated in machine readable form as by-products of their print equivalents. Similarly, those which were only in print media have been or are in the process of being converted to their machine readable forms.

Initially, INDAB1 attempted to collect information from several sources like:

- Questionnaires printed in NISSAT Newsletter
- Reconnaissance questionnaires
- Collation of bibliographic data from existing sources namely–
  - ☐ Directory of Science & Technology Information Systems. DST, 1990
  - ☐ Directory of Indian Scientific Periodicals. INSDOC, 1990.
  - ☐ National Union Catalogue of Scientific Serials in India (*NUCSSI*). *INSDOC, 1991.*
  - ☐ Directory of periodicals published in India. Sapra & Sapra,1989.
  - ☐ Ulrich's International Periodicals Directory, 1994.

In addition to these sources, responses from the database producers/sponsors obtained through direct questionnaires are also included in the update, i.e., INDAB2. Presently, INDAB2 has 398 records on the various information services mentioned.

The database content designation is after the Unesco's CCF/F [almost closer to The Gale's Directory of Databases 1995 in the choice of data elements], runs on Unesco's CDS/ISIS 3.07 software and under the MS-DOS.

### 2.1 The Size

*INDAB2 contains descriptions of* 398 'database' efforts–an increase by 105 entries from the earlier edition. Updated information on 200 abstracting & indexing services, databases and directories have been received directly from the contributing agencies. Information on 198 efforts need to be further updated.

### 2.2 Type of Services/Databases/Products

There are 113 efforts which are reported to be exclusively in their database form. Of the 202 abstracting/indexing services, and 83 directories listed, only 45 and 43 respectively are databases. Altogether 201 efforts have resulted in the generation of machine readable files/databases as shown in Table 1.

**Table 1. Type of services/databases/products**

| Type of Service | No. |
|---|---|
| Abstracting/Indexing Service | 202 |
| Database | 113 |
| Directory | 83 |
| Total | 398 |

Over the coming years, one could expect an increase in the number and variety of directories–not to leave out all those yellow

pages, pink pages, green pages and several referral tools in machine readable form. Similarly, one can also expect an increase in the full text and multimedia products—since the capabilities to generate CD-ROM products commercially are available in the country.

Further, indigenous information products/services are becoming increasingly available in a variety of modes and media—print, print-magnetic, digital, magnetic, optical and online modes.

## 2.3 Database Producers

Apart from the database activities in purely government departments/agencies (38.98%), one also notes an increased output from academic and national institutions (23.72%), CSIR labs (16.39%), private sector organizations (17.52%) and international institutions (3.39%). Actual number of each type of contributors is presented in Table 2.

**Table 2. Distribution of contributors**

| Government | | R&D Organisations | | | Univ/Inst | | Pvt |
|---|---|---|---|---|---|---|---|
| Central | State | Govt | CSIR | Priv. | Natl | Intl | |
| 34 | 4 | 31 | 29 | 12 | 42 | 6 | 19 |
| 38 | | 72 | | | 48 | | 19 |

[Total number of contributors = 177]

Government departments/agencies and R&D organisations form the larger chunk of contributors. Keeping in view dwindling governmental support for purely academic efforts, one finds the need for the development of databases with commercial value and wider accessibility as a stimuli for the information industry. Once again the economic necessity will compel joint efforts on the part of public funded agencies which generate enormous data and the commercial organisations that should find opportunities for profits and business.

## 2.4 Database Efforts by Broad Subject Areas

The maximum number of services/ databases/products are seen in the area of Engineering followed by Science & Technology, Agricultural Sciences and Medical Sciences. The first twelve groups along with the number of services/databases/products given in Table 3.

**Table 3. Distribution of services/databases/products**

| Subject Area | No. |
|---|---|
| Engineering | 61 |
| Science & technology in general | 48 |
| Agricultural sciences | 44 |
| Medical sciences | 38 |
| Social science | 22 |
| Chemical technology | 19 |
| Management, business & industrial | 19 |
| Environmental sciences | 14 |
| Biotechnology | 12 |
| Earth sciences | 11 |
| Leather industry | 11 |
| Textile industry | 11 |

The remaining groups are: metallurgy (9), chemical sciences (8), patents (8), applied sciences (7), library and information sciences (7), physical sciences (5). Other efforts are in botany, building Industries, science, organisations, assembled goods industries, biological sciences, mathematics, natural sciences, palaeontology, standards, geography, zoology, macromolecular materials, etc.

INDAB presently covers indigenous efforts limited to science & technology only. However, it would be necessary that information services pertaining to other fields—like management, economics, company financials, and so on may need to be included in future compilations. With increasing demands on trade, commerce and business information, the growth of mainly textual or factual databases would become essential.

A comparison of databases from INDAB1 and INDAB2 should provide an idea of the database efforts in some of the subject groups—mainly in science & technology, environmental sciences, medical sciences, engineering, agricultural sciences and management sciences (Table 4).

**Table 4. Some areas of databases in INDAB1 & INDAB2**

| Subject Area | INDAB1 | INDAB2 |
|---|---|---|
| Science & technology | 23 | 48 |
| Environmental sciences | 3 | 14 |
| Medical sciences (including biotechnology) | 33 | 50 |
| Engineering | 44 | 61 |
| Agricultural sciences | 20 | 44 |
| Management, business and Industrial | 8 | 19 |
| Other subjects | 0 | 15 |

With the availability of computers, many institutions are creating databases to meet their requirements. For example, universities are creating databases of their library collections as well as other services. These databases are of very generic nature comprising information on all kinds of documents.

## 2.5 Indigenous Database Efforts by Geographic Regions

State wise distribution of the services/databases/products shows that Delhi has the maximum number of efforts reported. The first seven states in order of number of efforts are: Delhi (155), Maharashtra (41), West Bengal (38), Karnataka (34), Tamil Nadu (34), Gujarat (22) and Uttar Pradesh (22). The actual area-wise (alphabetical order) efforts are shown in Table 5.

**Table 5. State-wise distribution of database Efforts**

| State | Number |
|---|---|
| Andhra Pradesh | 9 |
| Bihar | 8 |
| Delhi | 155 |
| Goa | 4 |
| Gujarat | 22 |
| Haryana | 9 |
| Himachal Pradesh | 1 |
| Jammu & Kashmir | 1 |
| Kerala | 2 |
| Karnataka | 34 |
| Madhya Pradesh | 3 |
| Maharashtra | 41 |
| Punjab | 8 |
| Rajasthan | 5 |
| Tamil Nadu | 34 |
| Union Territory | 2 |
| Uttar Pradesh | 22 |
| West Bengal | 38 |

## 2.6 City-wise Distribution

New Delhi leads with a maximum number of services/databases/products (155) reported. Other cities following New Delhi are Calcutta (37), Bangalore (30), Chennai (25), Lucknow (16), Ahmedabad (15), Nagpur (14), Pune (13), Mumbai (12), and Patiala (8). The remaining 73 services are shared among 36 locations. All the major cities have more number of efforts than reported earlier.

This pattern also suggests that it is the commercial, S&T or R&D intensive activities that might have a bearing on the need and hence generation of information services.

## 2.7 Growth of Databases

Information on year/start year of publication or database creation are available only for 314 efforts. Current year/start year information is not available for 83 of the efforts listed. Growth of the number of database efforts is listed in Table 6.

**Table 6. Growth of database efforts**

| Period | Number |
|---|---|
| 19 –50 | 10 |
| 1951–55 | 8 |
| 1956–60 | 10 |
| 1961–65 | 14 |
| 1966–70 | 29 |
| 1971–75 | 38 |
| 1976–80 | 24 |
| 1981–85 | 37 |
| 1986–90 | 68 |
| 1991–95 | 66 |
| 1996– | 10 |

Considering a span of five years, the breakdown of the number of services/databases/products is given, which indicates that there has been a spurt of information activities since 1980s. Perhaps, affordability of computers, skilled manpower,

software availability or development took place. This was the time when NISSAT started distributing CDS/ISIS and also helped diffuse standards like AACR II, and CCF. This once again is the period during which other commercial database management systems like dBASE III, FoxPro, Unify, Oracle, LIBSYS and non-commercial packages like MINISIS, CDS/ISIS were available. In addition, many institutions with their computer facilities and available manpower, developed local systems.

## 2.8 Frequency

Out of the total 398, only 294 efforts provided frequency of publications/services and of updation of the databases. Many organisations use the term regular, irregular, daily, once in 3 years and once in 5 years as the periodicity of database updates.

**Table 7. Frequency of databases/products and services**

| Frequency | No. | Frequency | No. |
|---|---|---|---|
| Daily | 13 | Bi-monthly | 1 |
| Weekly | 7 | 3/per month | 1 |
| Fortnightly | 5 | Monthly | 82 |
| Bi-monthly | 21 | Quarterly | 50 |
| 3/per year | 1 | Semi-Annual | 13 |
| Annual | 44 | Biennial | 1 |
| Regular | 42 | Irregular | 5 |
| Once in 3 years | 1 | Once in 5 years | 7 |

## 2.9 Status

In many cases the information about the status of the services is not mentioned, but as many of these appear in the currently published directories, they are taken to be continuing. Of the total 398, there are 368 efforts which are continuing; a small number, i.e., 19 efforts have ceased to publish. For 11 efforts, information on their status was not available at all.

## 2.10 Physical Form

The following table provides the physical form of the services/databases/products covered in INDAB in Table 8.

**Table 8. Databases by their physical form**

| Physical form | No. |
|---|---|
| In computerised form (both print & online mode included) | 262 |
| Available only in Printed form | 88 |
| Available in online mode only | 9 |
| Physical form not mentioned | 39 |

In addition to this, going by the trend the user may be seeing full text data increasingly in CD-ROM and hypertext forms and accessible on Internet as well.

## 2.11 Hardware & Software

Database generation depends a lot on the availability of the hardware and software to the producer. The pattern also reflects the extent of computer usage in the various institutions. Most of the database development has been on Microcomputers and in a predominantly library/information environment, CDS/ISIS is the major software used for this purpose. Indeed Tables 9 & 10 suggest that :

- A major share of the database activities are on microcomputers and CDS/ISIS software;
- Non-CDS/ISIS software include FoxBase, FoxPro, dBASE, MINISIS—which are relational database management systems; and also on software developed inhouse—perhaps some of these use a DBMS platform.

**Table 9. Hardware on which databases are created**

| Hardware Type | No. |
|---|---|
| Microcomputers | 159 |
| Minicomputers | 17 |
| Main frames | 10 |
| Mini & micro computers | 4 |
| Info not available | 81 |
| Total | 271 |

**Table 10. Software used in database creation**

| Type | No. |
|---|---|
| CDS/ISIS Only | 84 |
| CDS/ISIS and others | 25 |

| | |
|---|---|
| dBASE | 23 |
| FoxBase, ForPro | 9 |
| MINISIS | 6 |
| GWBasic | 6 |
| COBOL | 5 |
| FORTRAN | 5 |
| Libsys | 5 |
| Unify | 3 |
| In-house software | 11 |
| Others | 33 |
| Not available | 56 |

[Total 271 efforts]

## 2.12 Availability

Out of the 398 efforts listed, availability information is reported for only 293 databases. These may be available:

- ☐ Against exchange
- ☐ Free
- ☐ Against price or usage cost.

Under exchange category, there are 11 services listed; 152 databases/products are free. There are at least 130 products which are priced. This forms almost 33.3% of the total services recorded.

Pricing policy varies from institution to institution. Some adopt dual rates—one for individual users and another for institutional consumers. Noteworthy among these are the CSIR laboratories, Government institutions and ICSSR outfits.

Of the 105 efforts for which no availability terms were specified, the abstracting/indexing services account for 26 efforts; about 14 directories and 65 databases including bibliographic information are included in this category.

This is the picture that broadly emerges out from the INDAB2 data.

# 3. Opportunities for Indigenous Development

Going by the trend, certain observations can be made—the foremost is, that India is yet to emerge as a global contributor in the information industry. If the various database developmental efforts are any indication, then it would suggest that Indian databases would take a longer time to have their presence felt globally.

Secondly, several factors have contributed to the recent indigenous database development efforts. The hardware and software being no more a constraint, what would appear short of supply is the right kind of manpower, use of relevant standards and lack of adequate backup resources.

A pertinent question remains. Why indigenous databases? Why should one have an indigenous database development programme? One could offer any number of justifications, but these may be mainly under any of the following imperatives: largely, there is a need to

- Organise knowledge generated within the country
- Access Indian contributions published overseas
- Provide access overseas to Indian knowledgebases
- Aid our own national efforts in resources management and development.

What should be our priorities? Competition with global database providers is far away for India, but she should concentrate on areas of her primary importance like population control, food, shelter, health, agriculutre, and industries. To begin with, there is a need to look for opportunities and take appropriate initiatives in :

- Creation of databases in stronger areas—the areas of potential opportunities are too many.
- Areas of database generation where in depth treatment is wanting globally.
- Undertake activities of value addition and add Indian Information of significant interest to the global user.

Stronger points in India's favour in this venture would be:

- Manpower skills—in the subject areas apart from the area of database production, telecommunication, vending and so on.

- Low capital investments—considering the cheaper input possibilities, we need to identify potential areas of database efforts and with utmost commitment generate globally relevant products.
- High labour utilisation—as database activitiy is not only capital intensive, but also more labour intensive. Use of highly literate population skilled in computer use and English, it would be cheaper to take up such ventures.
- High availability of S&T manpower which could be used for knowledge capture, integration and value addition.
- Skill in use of English (to reach a global market)

On the other hand, there is a need to take note of weaker aspects. These are:

- The resources required for database production are disparate and are highly disorganised. Added to this, there is a general weakness in our primary/secondary sources. Even where they exist, there is a restricted access—not only to capture data but also to consult when required. In a procedural environment, resource sharing is at a premium. Where it is possible, procedural complexities damper any kind of 'database development'.
- Nationally, institutions and organisations suffer from low financial resources. Indeed accessing basic scientific resources is more a matter of money than anything. Added to the poor resource allocations, poor work culture and low marketing efforts hamper indigenous efforts. Further, the concept of Information as a saleable commodity is yet to percolate the society and therefore it becomes important to sensitise users and the service providers.

## 4. ALLOCATION OF RESOURCES

Database activity is not only capital intensive but also requires long time commitments. Besides, to grow commercially potential, larger inputs—in terms of expertise, manpower and most important of these, information resources are essential.

An aspect of database production should also consider its economic viability as against as a tool for welfare services. Most of the indigenous efforts started as a means to meet local requirements, suffer from limited or closed access and also poor delivery mechanisms. Added to these, data compilation entails huge investments in terms of time, efforts, and costs. For a wide, commercial access and delivery, the shortest way of providing access is to put the local, smaller databases on a web site and therefore on Internet. This would imply the skills to convert data or generate data using HTML and SGMLs need to be generated. Therefore, manpower development also has an implication in the activity.

## 5. STRATEGIES

Several strategies are possible. One is to take up databases in progress on print media. Another area of work would be to digitize the best of the Indian S&T periodicals both retrospective collections and current issues and offer the delivery choices in CD-ROM format.

In the government sector, a huge amount of data is collected, produced with very limited or restricted access. The cost–delivery balance is disproportionate and more importantly, it remains unused. Therefore, there is a need to organise the government tradeable information into a powerful, widely accessible resource.

Trade statistics is another area of greater importance. With the increased interest in EDI and the use of Industrial codes in commodity classification, appropriate trade statistics—both import–export can be generated.

There has been a spurt of commercial activities in the generation of various trade Directories and Yellow pages. Directories Today, Dunn & Bradstreet, Kompass India, Tata Yellow Pages are some of these whose utility is being increasingly felt. Apart from local access to the huge data through print medium, online access to these sources would be essential.

In certain areas, databases have to be developed *de novo*. While there is a limited interest in bibliographic information, efforts on factual information have to increase.

While the demand for appropriate information services would increase in future, there is a need to fill up gaps in commercial/business information and adopt a strong market approach to the developmental efforts: The industry would also need support from financial institutions for support and sustenance.

A database activity can be seen as a conscious effort to organise and provide effective access to the users of a knowledgebase. One may also view the growth of databases as a response to provide for advantages that individuals, institutions or nations may wish to gain. These advantages could be in academic pursuits, in the process of knowledge generation, in the protection of intellectual property rights, or gaining economic dominance, business competitiveness and political gains. Besides, these resources should help one's efforts towards a sustainable development.

## 6. ACKNOWLEDGEMENT

The authors are grateful to Dr A Lahiri, Adviser, Ministry of Science & Technology, Government of India and all the colleagues at NISSAT for encouragement and suggestions. They are also thankful to all the institutions and individuals who have provided inputs for the creation of INDAB.

## REFERENCES

1. Lahiri, A and Singh, BG Sunder. Bibliographic databases and networks: Indian scenario. *In* Bibliographic databases and networks. Edited by SS Murthy *et al.* Proceedings of the International Conference, New Delhi, India, 22-25 February 1989. Part 1. pp.41-65. Tata McGraw-Hill, New Delhi, 1990.
2. Sur, SN and Singh, BG Sunder. Indian abstracting & indexing services and databases in science & technology. *Information Today & Tomorrow*, 1995, 13(1), 11–24.
3. Vyasmurthy, P. Database industry in India: A current scenario. *DESIDOC Bulletin of Information Technology*, 1995, 15(4), 11–24.
4. Lahiri, A. Design of national information systems in developing countries: Concepts and considerations. Paper presented in the ASTINFO Consultative Committee, 2-9 May 1997, Tehran, Iran. 9p.