CHAPTER 3
THE STUDY: OBJECTIVES AND METHODOLOGY

3.1 RATIONALE OF THE PRESENT STUDY

The use of information is rapidly taking centre stage in society and it has become imperative to devise methods for the quantification of the parameters through which the developmental status of ‘information’ in a society can be assessed. It is natural to expect that the initiative in this direction should be taken up by the LIS profession.

The LIS professionals regularly measure several aspects at the individual libraries level namely books available, average price of the book, circulation figures, etc. as part of their general assessment of the library growth and development. At the macro level very little attempt has been made to measure library parameters at the state, national or global level, although librarians find favour with statements such as “libraries are the lifeblood of our communities” (Piero, 2011 p.7). Studies by LIS professionals have been generally restricted to librametrics, bibliometrics, scientometrics, and webmetrics. In India, unfortunately attempts to measure library related parameters have been extremity limited. In his editorial Sen (2001) raised important questions such as how many libraries, librarians, and library science students are there in India? How many books and journals are published? While much is desired in the estimates of tangibles such as libraries, books, circulation figures etc. in India, the challenge of measuring intangibles such as information or knowledge remains a bigger challenge.

However information and its related parameters have been included in various indexes developed in the last decade to measure socio economic development, technological growth and information society. Some major efforts are as follows.

At the national level attempts to measure socio economic development as well as ICT have also been made. These reports are excellent in terms of the data collected, methodologies employed and analysis. The National Human Development Report-India 2001 by the Planning Commission, Government of India, on lines similar to
HDRs of UNDP has indexed human developmental level of 15 major Indian States. NHDR sought to measure socio economic development through indicators of education, health and spending power of individuals. The impact of the UNDP studies and the NHDR encouraged many states to further analyze human development levels within their respective areas, e.g., the Gujarat HDR was published in 2004 and compared the HD levels in 25 districts. In fact, a few districts of Gujarat have been motivated to compile their own HDRs describing the HD levels at the Taluka level, e.g., Jamnagar, Surendranagar, Banaskantha, Dangs, Surat, Panchmahal and Narmada Districts. (Marvania, Personal communication, December 10, 2010) These documents have been extremely valuable for policy making at their respective levels.

The Department of Information Technology, Ministry of Communications and Information Technology, Government of India, and National Council of Applied Economic Research (NCAER) have collaborated in producing the India e-Readiness reports since 2003 in an attempt to evaluate the e-Readiness of State Governments. Since then it has been updated annually. The latest e-Readiness Report 2006 is the fourth in the series. The report measures the readiness of the government departments, markets and individuals to accept ICT. This report covers state level data but sub state level attempts have not been made. Thus for example while one can get an idea of how e-ready Gujarat state is, it is not possible to find out how e-ready Rajkot district is.

From the perspective of the LIS profession the e-Readiness Index, though helpful in accessing the electronic channels of information and communication, was not fully satisfactory. It did not attempt to cover the entire information scenario, particularly the traditional information resources and channels including the libraries, and the current and potential information users. It is with the specific objective of bridging this gap, which the LIS discipline experiences, that the present study is being undertaken.
3.2 OBJECTIVES OF THE STUDY

The present study is an attempt to develop an Information Index to measure the information infrastructure of a given region. This will include both the traditional and technology based indicators of information infrastructure.

More specifically the objectives of the present study are to:

1. Review the attempts made to measure Information Infrastructure and related concepts like e-Readiness, Information Society, etc.
2. Identify the indicators used in various attempts
3. Select appropriate indicators for an Information Index in the Indian context
4. Develop a methodology for construction of an Information Index to measure information infrastructure of a society
5. Construct an Information Index for various Indian States
6. Compare the constructed Information Index with National Human Development Index to observe the relationship between the two.

3.3 SCOPE OF THE STUDY

The present study seeks to examine various attempts at index constructions, select appropriate indicators of information, gather secondary data on them and decide a suitable methodology for compiling a composite index based on the data. This methodology for constructing an index would be applied to the various states and union territories of India. The result would be a ranking of states according to their level of information infrastructure development. If possible a similar exercise would be undertaken for the districts of Gujarat. The theoretical and empirical issues faced in the exercise would be discussed.

3.4 METHODOLOGY

3.4.1 Basic Concepts and Model

The information index was planned to measure the information infrastructure and its present and potential use in a society to leverage development. The first step in
designing a research study is to clearly specify the questions to be answered and to operationalise the concepts involved.

▪ What exactly will be meant by information?
▪ How will the concept of information be operationalised?
▪ Where will be the data collected from?
▪ How will the data be compiled and converted into an index?

Initially the concept of ‘information’ itself proved to be a challenge. As explained in an earlier chapter the word neither has an agreed meaning nor an accepted unit of measure.

Among the various definitions examined in the earlier pages the UNISIST definition was found to be fairly comprehensive and satisfactory from the viewpoint of the LIS profession. To bring it into the context again UNISIST defined information as

“Information is made up of symbolic elements, communicating scientific and technical knowledge, irrespective of their nature (numerical, textual, graphic, etc.), material carriers (paper-print, microform or machine readable form), form of presentation, etc. It refers both to the substance or contents of documents and to the physical existence; the term is also used to designate both message (substance and form) and its communication (act)”.

Information thus understood has several dimensions such as generation, communication, use, quality, quantity, cost, characteristics etc. The next challenge which needed to be addressed was to decide which dimensions of information were to be selected. Libraries exist to support the generation, communication and use of information for education, research, culture and recreation. This is recognised by the National Knowledge Commission (2007, p.5). It states in its report “A library is not a building stacked with books – it is a repository and source of information and ideas, a place for learning and enquiry, and for the generation of thought and the creation of new knowledge.” The three primary functions of LIS are to support and promote these three activities of generation, communication and use of information. Hence it was decided to limit the index to these three dimensions.

Generation, communication and utilization of information require extensive support of technology, particularly, ICT, and a conducive political, economic and social environment. Hence technology and social environment both have to be kept in mind
while developing an Information Index. Fig 3.1 depicts a model of ‘information infrastructure’ developed and used while undertaking the present study. The creation, communication and use of information with appropriate technical support and a facilitating socio economic environment are considered to be critical factors in the development of a region. These factors together are termed as information infrastructure for the present study.

Figure 3.1: The model of Generation, Communication and Utilization of Information

The model identifies the three critical components of information in a society with which LIS professionals are concerned. Libraries and information centres are repositories of knowledge created over the centuries and are significant intermediate agencies in the information flow process. They enable scholarly communications which is a continuous spiral of knowledge acquisition, creation and publication that happens in the library through the knowledge resources available. The three aspects of information generation, communication and use are further described in the following sub sections.
3.4.1.1 Information Generation

Information is generated as a product of, different types of human activities, events, or incidents undertaken by individuals or by organisations to achieve specific objectives. A considerable amount of information is generated through research and development, government activities, business and industries. These are briefly outlined as follows.

Some activities are undertaken with a specific objective of creating or generating information; they are referred to as research or R & D activities. It is creative work undertaken on a systematic basis in order to increase the stock of knowledge and use it for the benefit of society. It is a highly organized activity throughout the world which continuously creates a large mass of new information. There has been a dramatic increase of both research institutions and researchers in all branches of knowledge as more funds are being allocated for conducting research. The progress of a nation is often judged by the percentage of national income that is spent on R & D activities.

Research organisations in science, technology, social sciences and humanities specifically work for this purpose. Apart from research institutions, academic institutions like universities also undertake and consider research as one of their major tasks. Industries and commercial agencies too, conduct research as part of their effort to create new products, reduce costs, and identify potential markets.

Government and its multiple agencies are perhaps the largest supporters of research to improve defence capability, promote science and technology, and formulate appropriate policies. They perform their tasks as a matter of routine. Their routine activities in turn generate information as a by-product. A major part of the information needed by the social science researchers and decision-makers in governments, business and industries emanates from government sources. Besides these administrative organs of the government, the legislative and judicial bodies also contribute to the growth of information. In addition, the government also conducts specific activities to collect large scale data through organisations set up for this purpose, such as Census of India, National Sample Survey, Central Statistical...
Organisation etc. This data helps the government in policy making and is extensively used in social science research.

Business and financial research activities conducted by associations, commercial houses, evaluative agencies, banks and individual consultants also result in valuable and large scale information particularly related to economic, financial and industrial development.

3.4.1.2 Information Communication

Communication is integral to the very etymology of the word ‘information’ and contains within it two concepts which suggest the major lines of meaning leading out from this core. ‘To inform’ carries with it the concept of communication of the process by which knowledge is transmitted and disseminated within society. The second concept of ‘being in a form’, that is of being formed into a shape or structure carries with it connotation to do with the various processes by which knowledge is shaped, packaged or organised.

In the first meaning the communication may be with a known or unknown individual or sets of people. One to one communication can take place through face to face conversations, personal letters, e-mails, or through telephone calls. Where as one to many communication, particularly with an unknown audience takes place via lectures, radio and television broadcast etc. The second meaning relates to printers publishers, producers and transmitters of the various carriers or containers of information, and the containers themselves, e.g. books, films, newspapers, etc. Repositories of the various carriers such as archives, museums and libraries support the information communication process.

3.4.1.3 Information Use

If information is the product of all activities, its use is also an important input to every type of human activity. The need and uses to which information is put are quite literally countless, hence the power of those metaphors which link the human and the social organisms in describing information as the “lifeblood of society” (Martin,
Information has been a significant element in the life of all societies, from the stone age to the present-day modern society. It is common for each culture or society to have its own characteristic information and knowledge base. In the present world the question of access to, and exploitation and control of, information are of increasing significance for the government and citizens alike.

Without the constant use of this vital resource, society as we know it, would quickly run into difficulties, with business and industry, education, leisure, travel and communications, national and international affairs all vulnerable to disruption. In more advanced society this vulnerability is heightened by an increasing dependence on the enabling power of information and communication technologies. Everybody from the common citizen to the highly qualified professionals needs and uses information for some purpose or the other. The generators of information are also its users, thus creating a spiral of information generation, communication and use.

Citizens need and use information for their various day to day activities like vocational, administrative, educational, socio-cultural, leisure, entertainment etc. Information used by students relate to the prescribed syllabi for pursuing academic studies. Teachers use information for imparting education to the students; and researchers or scientists use information on a continuing basis and are considered the biggest consumers of information. They use information for three purposes: to keep up with new developments in their area of interest; to get acquainted with the state-of-the-art; and to gather specific pieces of data and information needed at different stages of their work.

Professionals like medical and legal practitioners need and use information to pursue their professional work. Physicians cannot ignore the new developments in the medical sciences. Similarly legal practitioners must keep in touch with the latest cases, laws, and judicial verdicts to ensure fair justice. Engineers and technologists need and use information for solving technical snags faced by them on the shop floor. Managers and executives of business and industrial organisations use information to enable them to take appropriate decisions relating to issues having both short-term and long-term implications.
Government officials also need and use information for decision making and legislators use information for arguing a point on the floor. Thus information is truly the lifeblood of any society. It is needed in all types of activities and is used by a cross-section of people.

3.4.2 Operationalisation

It is against the above mentioned model that the three concepts of information generation, communication and use were operationalised. Since the three selected concepts are directly unobservable they had to be represented by indicators which could be measured. This process is known as operationalisation and is very crucial in the entire research process. There is no exact congruence between any concept and indicator; one concept may have several indicators while the same indicator may represent more than one concept. Thus for example, information generated could be represented by the number of research studies completed, the number of researchers, number of research institutes, the number of patents generated and many other. Similarly the indicator ‘number of Internet users’ may have bearing on information communication, information use, level of ICT development, Government policy etc. To overcome this problem several indicators are used to represent one concept. Proper selection of indicators is the key to good research design.

At the very beginning of the project a list of indicators of information generation, communication and use was prepared by the researcher based on her knowledge of the field. During the same period five experts, one from the field of library and information science and four from social sciences were consulted and their opinion was sought on the probable indicators. From these discussions two major points emerged, namely the importance of socio economic environment and the political situation. Socio economic factors that needed to be included were for example number of industries, employment, and e-commerce, while the political situation included various policies, voter awareness and e-governance.

Taking into account these suggestions a revised list of indicators was prepared. This list contained 21 indicators broadly divided into 6 major categories. To verify these indicators an opinion poll was conducted along the line of Delphi technique.
3.4.2.1 Delphi Method

The Delphi technique, mainly developed by Dalkey and Helmer (1963) at the Rand Corporation in the 1950s, is a widely used and accepted method for achieving convergence of opinion concerning real-world knowledge solicited from experts within certain topic areas. Predicated on the rationale that, “two heads are better than one, or...n heads are better than one” (Dalkey, 1972, p. 15), the Delphi technique is designed as a group communication process that aims at conducting discussions of a specific issue for the purpose of goal setting, policy investigation, or predicting the occurrence of future events. Common surveys try to identify “what is,” whereas the Delphi technique attempts to address “what could/should be” (Miller, 2006). In the literature, Delphi has been applied in various fields such as program planning, needs assessment, policy determination, and resource utilization. Theoretically, the Delphi process can be continuously reiterated as often as required.

A group of 50 experts were identified; they included professors of library science and leading practitioners in the LIS field, experts from the field of telecommunication, professors of economics and sociology, researchers and consultants, and a few citizens who were known for their active role in the society. The opinion sheet (Annexure A.) containing the list of indicators was sent to them by e-mail or given personally to gather their views. The experts were asked to rate these variables according to their importance to the information infrastructure and also to suggest new indicators which may not have been included. In all 38 experts responded. The scores obtained for each of the variables were summed up and the mean value was obtained, which is shown in table 3.1.

The experts placed a greater weightage on government policy and research & development, followed by information flows and channels. The economic and education environment factors came next, whereas ICT penetration was placed last brought down by low score given to the ‘number of landline telephones’. This opinion poll helped in the reaffirmation of the indicators to be included.
Table 3.1: Indicators of Information Infrastructure

<table>
<thead>
<tr>
<th>Indicators of Information Infrastructure</th>
<th>Mean Value</th>
<th>Group Mean Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Literacy Level</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>· Primary Level Enrollment</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>· Secondary Level Enrollment</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>· Higher Education Enrollment</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>· Professional Higher Education</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>18.13</strong></td>
<td><strong>3.626</strong></td>
</tr>
<tr>
<td><strong>Economic Capacity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Per Capita Income</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>· Employment rate</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>7.33</strong></td>
<td><strong>3.665</strong></td>
</tr>
<tr>
<td><strong>ICT penetration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· No. of Computers</td>
<td>3.66</td>
<td></td>
</tr>
<tr>
<td>· No. of Internet users</td>
<td>4.46</td>
<td></td>
</tr>
<tr>
<td>· No. of fixed line telephones</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>· No. of mobile phones</td>
<td>3.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>14.45</strong></td>
<td><strong>3.6125</strong></td>
</tr>
<tr>
<td><strong>Information Flow/Channels</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· No of newspapers and magazines</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>· No. of Radio sets</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>· No. of TV sets</td>
<td>3.86</td>
<td></td>
</tr>
<tr>
<td>· Postal services</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>· No. of Libraries</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>19.12</strong></td>
<td><strong>3.824</strong></td>
</tr>
<tr>
<td><strong>Research &amp; Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· No of researchers</td>
<td>4.73</td>
<td></td>
</tr>
<tr>
<td>· No of patents</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>8.93</strong></td>
<td><strong>4.465</strong></td>
</tr>
<tr>
<td><strong>Government policy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Education Policy</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>· ICT policy</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>· S &amp; T Policy</td>
<td>4.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>14.06</strong></td>
<td><strong>4.686</strong></td>
</tr>
</tbody>
</table>

In the mean while review of relevant literature and an examination of other indexes had continued. The indicators used in these indexes were compiled into one long list.
of 269 number of items. The indicators were grouped into 12 broad categories as per the Annexure B. These were (1) Demography and health; (2) Economy and labour; (3) Infrastructure; (4) Political climate; (5) Education and research; (6) Newspaper and publishing; (7) Libraries; (8) Telephones (Landlines & Cellular); (9) Computers; (10) Internet; (11) Media and (12) General ICT. This exercise resulted in some definite learning. A wide variety of indicators was necessary. Further the same item could be measured differently resulting in different indicators; Internet use could be measured through number of Internet hosts, number of Internet users per 1000 people, Internet users according to online tenure, Internet Users per host etc. It was also interesting to note that only 19 (boldfaced in Annexure B) indicators had been used in three or more than three indexes constructed, perhaps because they were measuring varying concepts from human development to technology development to information society. Seven most recurring indicators were adult literacy, number of computers, teledensity, number of Internet use and users, Internet hosts, Internet access tariff, and royalty and licence fees.

Finally 40 distinct indicators representing the three concepts of information generation, communication and use, the technical support and the socio economic environment in which they function were selected from the list and are given in table no 3.2

### Table 3.2 Indicators of Information Generation, Communication, Use

<table>
<thead>
<tr>
<th>Technical Support and Socio Economic Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Generation</td>
</tr>
<tr>
<td>1. Number of Books Published, Per Unit of Population</td>
</tr>
<tr>
<td>2. Number of Journal Articles Published, Per Unit of Population</td>
</tr>
<tr>
<td>3. Number of Patents filled, Per Unit of Population</td>
</tr>
<tr>
<td>4. Number of Research Institutes, Per Unit of Population</td>
</tr>
<tr>
<td>5. Number of Research Scientist, Per Unit of Population</td>
</tr>
<tr>
<td>6. Royalties and License fees received, Per capita</td>
</tr>
<tr>
<td>7. Expenditure on R &amp; D, Regional equivalent of GDP</td>
</tr>
</tbody>
</table>
### Information Communication

1. Number of Libraries, Per Unit of Population
2. Postal traffic, Per Unit of Population
3. Number of Radio Sets, Per Unit of Population
4. Teledensity (LL & Cellular), Per Unit of Population
5. Number of TV Sets, Per Unit of Population
6. Internet access tariff, Per unit of time
7. Degree of Internet access through mobile phones, % of Internet Access
8. Broadband households, % of Internet Access
9. Wireless Connections, % of Internet Access
10. Average monthly household expenditure on telephone, mobile and internet, as % of average total household Expenses

### Information Use

1. Gross enrolment ratio (Tertiary)
2. Number of Professionals, Per Unit of Population
3. Number of Library Users, Per Unit of Population
4. Number of Workers in Service Sector, Per Unit of Population
5. Number of Universities, Per Unit of Population
6. Number of Internet Users, Per Unit of Population
7. News Paper Circulation, Per Unit of Population

### Technical Support

1. Number of Computers, Per Unit of Population
2. Number of Internet Host, Per Unit of Population
3. Schools with computer labs, % of-
4. Schools with Internet access, % of-
5. IT qualified teacher to total teachers, % of-
6. Computer Engineers to total engineers, % of-
7. Electricity consumption, Per Unit of Population
8. Government expenditure on ICT/NSDP, as unit of Regional GDP

### Socio Economic Environment

1. Adult literacy rate (%)
2. Gross enrolment ratio-(%)(Primary, Secondary, Tertiary)
3. Life expectancy at birth
4. Employment rate
5. GDP per capita (PPP $US)
6. Education, Expenditure on-% of Regional Equivalent of GDP
7. Status of e-Governance, (Transaction on Government Websites,) Per Unit of Population
8. Transportation statistics (Road, Railway, and Air Traffic)

### 3.4.3 Data Collection

Just as the selection of proper indicators determines whether one is really measuring what is to be measured, the quality of data determines how correct this measurement is. There are several ways by which any research study collects data. These are broadly divided into primary and secondary data and also into qualitative and quantitative data. The nature of research determines which type of data is required.
Data observed or collected directly from first-hand experience or data never gathered before is primary data whereas published data and the data collected in the past or by other parties is called secondary data, i.e. pre-existing data not gathered for purposes of the current research.

Primary data whether quantitative or qualitative, are collected by using traditional research methods, namely experiments, surveys, observations and ethnographic methods. The very nature of these methods limits the size and scope of the study, and the time span to be covered. Research in history, philosophy and literature depend heavily on secondary sources of information. Literature constitutes an important source for them. In a project which involves creation of an index, such as the present study, one requires wide scale comparative data from different sources. Hence, by its very nature one is compelled to use secondary data sources.

In the present study secondary data was collected from datasets compiled by Census of India, various ministries, NCAER, CSIR, UGC, etc and hence was assumed to be authoritative. In spite of the best efforts datasets for some of the above indicators could not be obtained hence the indicators had to be dropped e.g. number of books, number of journal articles, research scientists etc. For some other indicators proxy indicators had to be used e.g. number of households with radio had to be used instead of number of radio sets per unit of population. Alternative methods of data collection had to be considered for some other indicators for which direct datasets could not be obtained, for example computer and Internet penetration, cellular and fixed line telephone users, etc. Since data of these indicators were already included in the e-Readiness Index; it was decided to use the index values directly from the e-Readiness Index. Similarly, since the NHDR had developed an index to measure human development in different states of India, it was decided to directly accept these index values in the information index.

In the final analysis a Basic Index consisting of 12 indicators for which data were available was constructed. The values of both NHDI and e-Readiness Index were added to this index to constitute the final Information Index. Care was taken to avoid duplication of indicators across the indexes. This Basic Index consisted of those indicators which were considered important in the construction of the Information
Index and not covered by ERI. Data sources, their publishers, and their publication year, the year for which they were used in the compilation of the Basic Index are given in Annexure C. The annexure also provides a brief note on the data sources used by the NHDI and ERI. In a few cases different sources were used to get the complete dataset. Thus for example, in counting the number of libraries in a state data for various types of libraries had to be compiled together. Details of the datasets used, their source, definitions, and methodology of calculation are described in Chapter 6

3.4.4 Statistical Techniques Used

In the entire process of constructing the Information Index four basic statistical techniques were used: unitization, scaling, normalization, and finding the correlation coefficient values.

3.4.4.1. Unitization and Scaling

To be comparable data has to be expressed as either a percentage or in some unitized scale. Some of the data were available as percentage, e.g., number of households with radio and TV sets. Many others had to be expressed as a unit to get a comparative and clearer picture. Thus for example, the number of patents filed in a state had to be converted into the unit of number of patents per lakh population.

While adding up the values of the three components of the Information Index, it was required to scale down the Basic Index. Values of NHDI and ERI are shown on a scale of .001 to .1. The values of the Basic Index were in a scale of 0 to 10 hence this figure had to be divided by 10 to be brought on a comparable level.

3.4.4.2. Normalization

The indicators in the list of Basic Index span over different ranges of values, all variables are normalized from 0 (weakest) to 10 (strongest) and 15 major Indian states were ranked on an ordinal scale. The normalization procedure (Chen and Dahlman, 2005, p.17) used in the information index is the standard one and is as follows:
a. The actual data (u) is collected from various national and state level datasets for all the variables and states.

b. Ranks are allocated to the states based on the absolute values (actual data) that describe each and every one of 12 indicators (rank u). States with the same performance are allocated the same rank. Therefore, the rank equals 1 for a state that performs the best among the states in the study on a particular indicator (that is, it has the highest score), the rank equals to 2 for a state that performs second best, and so on.

c. The number of states with higher rank (Nh) is calculated for each state.

d. The following formula is used in order to normalize the scores for every state on every indicator according to their ranking and in relation to the total number of states in the sample (Nc) with available data:

\[
\text{Normalized (u)} = 10 \times (1 - \frac{N_h}{N_c})
\]

e. The above formula allocates a normalized score from 0 to 10 for each State.

3.4.4.3. Correlation

A correlation is a single number that describes the degree of relationship between two variables. Since the correlation is nothing more than a quantitative estimate of the relationship, we would expect a positive correlation. It means that, in general, higher scores on one variable tend to be paired with higher scores on the other and that lower scores on one variable tend to be paired with lower scores on the other. The formula for measuring the correlation is:
\[ r = \frac{N \sum xy \times (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \]

Where:

\( N \) = number of pairs of scores

\( \sum xy \) = sum of the products of paired scores

\( \sum x \) = sum of x scores

\( \sum y \) = sum of y scores

\( \sum x^2 \) = sum of squared x scores

\( \sum y^2 \) = sum of squared y scores

We use the symbol \( r \) to stand for the correlation. The \( r \) will always be between -1.0 and +1.0. If the correlation is negative, we have a negative relationship; if it is positive, the relationship is positive.

### 3.4.5 Index Generation

As described above the study has used values of two available indexes and compiled a third one to create a composite Information Index. The Basic Index consisted of 12 indicators grouped into the three key aspects of information generation, communication and use, not covered by both the available indexes. In the three groups each indicator was first unitized and then normalized as discussed in the earlier section. The scores were added up to form three sub indexes and these scores were again added up to construct the scaled Basic Index. Finally the values of NHDI, ERI and the Basic Index were used to create the Information Index. In this process the ERI and BI were given a weightage of one each while the NHDI was given a weightage of 0.5. This represented the final Information Index.
3.4.5.1 Reaffirming the Information Index

In order to examine the validity of the Information Index the rankings of the 15 major Indian States were shared with the experts. A comment sheet was sent out to all those 38 experts who had replied in the preliminary opinion poll, and they were asked to comment whether they agreed with the rankings. The comment sheet is attached in Annexure D.

3.4.5.2 Comparison of Information Index and NHDI

The last objective of the research was to compare the constructed Information Index with National Human Development Index to observe the relationship between the two. As discussed in the introduction there is some definite evidence to indicate that information and development are closely interrelated. Although the constructed information index had included NHDI values in the final score and given it a weightage of 20%, the two indexes were compared by calculating their correlation coefficient.

3.5 LIMITATIONS OF THE STUDY

The objective and methodology of the study were heavily dependent on statistical data and their availability which limited the study. Another major limitation also arose from the quality and the quantity of the data available.

The concept of information infrastructure was visualized to include information generation, communication and use. In the operationalisation of these concepts several indicators which were identified had to be excluded because of non-availability of relevant statistical data e.g. number of books published or number of library users etc.

Further the data used was of secondary level. No effort has been made to collect primary data. All such attempts of constructing composite indexes are generally based on reliable secondary data sources. Moreover it is only the government agencies that have the ability to collect this data and it is beyond the reach of an
individual researcher. Although all attempts have been made to ensure that wherever possible, the data used was collected by an authoritative government agency, it is fairly well known that no data is comprehensive or completely reliable. Thus the limitations of the secondary data become limitations of the present study as well. Another limitation was the non availability of all datasets of one particular year; hence the used datasets belonged to different time series.

The original intended scope was to cover all the states and union territories of India to create a national level Information Index and to collect data for the districts of Gujarat to create a Gujarat State Information Index. However the non availability of data for all states and union territories and the non availability of district wise data for Gujarat State limited the geographic scope of the study.

These limitations notwithstanding during the design and implementation of the research study maximum care was taken to ensure that the methodology was sound and the results valid. The study was able to fulfill its intended objective of constructing the Information Index for India.
REFERENCES


Primary data: [http://www.businessdictionary.com/definition/primary-data.html](http://www.businessdictionary.com/definition/primary-data.html)

