Chapter 3

RESEARCH METHODOLOGY

The objective of the current project was to develop a tool for the organisation of information in the area of mycology which could be used on the Web, would cater to the needs of local users from different disciplines and could handle factual, textual and image information. The obvious choice was the use of a database management system.

The global scenario of data management practices in the area of biodiversity is changing. Advanced computer and communication technologies make it possible to retrieve biodiversity information from online databases on the Internet. The main problem is the availability of several databases to choose from, each catering to a special user group and with different types and levels of information. The user invariably has to search several databases to satisfy his information needs on single species.

Further, with a view to help users from different disciplines working on fungi, the current research sought to study the user information needs and approaches to information. Based on this understanding of user needs, a database prototype was to be designed and tested. The project, thus involved both research and development components.

3.1 Research design

The objective of the research component was to identify the types of information about fungal species that was needed by scientists and farmers. The collected information would be used for database design. The survey method was used in the present study to assess the information needs of the users. A questionnaire was designed and used as the primary data collection instrument, although some informal interviews were also conducted.
Survey of Scientists

Pre-questionnaire discussions were held with four experts in the field of mycology, library & information science and database design. During the course of discussions it became clear that information needs could vary in mycology according to different parameters. These included the discipline from which scientists approached the subject (biotechnology, microbiology, plant pathology, etc.), the specific group of fungi they worked with (basidiomycetes, ascomycetes, etc.) and the activities they were involved in (maintaining herbarium, identify and classification of species, etc.) and the use to which their research was geared (agriculture, medicine, pharmaceutical industry, etc.). All mycologists are involved with fungi information, its collection, organisation, preservation, and documentation. These are basic tasks they have to perform regularly. Keeping in mind all these parameters the first draft of the questionnaire for scientists was developed.

Questionnaire for scientists

The questionnaire was in five sections. It included questions relating to the users and their information needs, their research activities, data collection methods, and their problems in seeking the information. The questions were categorized in following areas:

A. **General information**: The general information section focused on information about the scientists viz. name, designation, organisation, address, phone, fax, e-mail and web site etc.

B. **Research activities**: The second section sought to find out areas of subject specialization, the subject expertise of scientists, and group of fungi they worked on. Scientists from the areas like biotechnology, microbiology, and plant pathology all contribute to fungal research. There are various groups of fungi and it was necessary to know the different groups of fungal species, which formed the subject of the scientists work. This would also be useful in understanding the user needs of different groups.
Part B of the questionnaire, which dealt with research activities, had originally listed eight activities, which were as follows:

1. Naming a new species
2. Producing molecular sequences
3. Maintaining herbaria
4. Specimen listing (for e.g. Preparation of exsiccata)
5. Industrial applications
6. Use of environmental biotechnology, genetic engineering
7. Bioinformatics
8. Others

C. Information needs: In this section respondents were requested to give their information requirements as per the activity in which they were involved. The respondents were asked to give an illustration of the information required against each activity:

- Specimen collection: name of collector, date, etc.
- Nomenclature details: Fungus name, family, group, sub group, etc.
- Geographical locations: name of the country, district, area, etc.
- Host parasite relationship: name of host, nature of infection and related information.
- Spore data: morphology of spores
- Publications: bibliographic information
- Image data: metadata of image, individual specimen images, etc.
- Prevention: use of fungicides, resistant varieties, etc.

D. Data activities: This part of the questionnaire dealt with

i) The data collection tasks and practices of mycologists, the difficulties they faced during data collection and the size of the data.

ii) Data storage practices and other details of data documentation.
Chapter 3: Research Methodology

An open-ended question asked respondents to share the difficulties they faced viz. data are scattered, not available etc. This part of the questionnaire was asked to identify the difficulties they encountered, during data collection.

Information about the total number of species collection, number of genera, etc. was also asked. This section sought to find out the kind of data documentation of the species maintained by scientists. Questions were asked on the use of computers and appropriate software used. In the data documentation sub section, scientists were asked whether they were familiar with the concept of metadata and whether they have provided metadata to their collection and if so which standard has been used.

E. Data usage: This part of the questionnaire dealt with the possible utility of collected specimen data for others. It specifically asked if others were aware of the collection, would the data be useful for others, whether they had actually shared the data and the possible usage to which it could be put. Further this section also requested scientists to provide data updation frequencies, funding agency, and data ownership details. The last open-ended question was for general remarks about the survey. Mycologists were requested to provide their opinions about the present survey.

Pre test
To assess the suitability of the data collection tool the draft was shown to 2-3 experts and then tested with 10 scientists. The pilot study helped in identifying questions that tended to be misunderstood by the participants and required changes. These changes were in the nature of giving explanation or detailing out the questions.

Final questionnaire
Results of the pre test indicated that the research activities (naming a new species, producing molecular sequences, maintaining herbaria and specimen listing), should be separated from data application areas (industry, environmental biotechnology, genetic engineering). This suggestion was checked with subject experts who agreed that this would add to the clarity of the information collection. Accordingly this list
was divided into two sections viz. research activities and application areas as follows:

**Research activities**
- Maintaining herbaria
- Specimen listing
- Classification and identification of species
- Producing molecular sequences
- Bioinformatics

**Applications in**
- Industry
- Agriculture
- Industry and Agriculture

Further, it was noticed that researchers are involved in more than one activity such as maintaining herbaria, identifying and classifying the species. They are also involved in producing molecular sequences and bioinformatics applications. Also the application areas are multiple. Therefore respondents were asked to prioritise their research activities and application areas as high, medium and low as follows:

<table>
<thead>
<tr>
<th>Research activities</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintaining herbaria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specimen listing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification and identification of species</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Producing molecular sequences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioinformatics</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The revised questionnaire was discussed with 2 scientists and was finalised (Annexure I).

**Sample**
Potential users of the proposed database were a worldwide group of scientists, students and farmers. The questionnaire described above was meant for mycologists.
The universe of Indian mycologists is not well defined. Work in the area of mycology is carried out in CSIR institutions, universities, medical colleges, pharmaceutical industries, agricultural institutions and several small research institutes across the country. There are two mycological societies namely the Indian Society of Mycology & Plant Pathology and the Mycological Society of India. Their membership lists were requested for and obtained.

However, it was believed that the complex nature of the subject and the questionnaire would require the researcher to be present to clarify any doubts that the scientists had. Hence it was decided to focus on local scientists. A total of 60 local scientists were identified from the above lists. All these scientists were consulted, the project explained and an appointment with them was sought. Initially 30 agreed to meet the researcher to fill up the questionnaire.

In addition, to supplement the local sample, 25 questionnaires were sent out to different parts of the country. No response was received from these scientists.

Further to extend the reach of the questionnaire, it was also put on the Net at http://bioinfo.ernet.in/library/fungi/survey.html and distributed via the following listservs:

1. Taxacom: a biological systematics discussion group of 1470 life scientists
2. Fungi and you: a group of 315 mycologists and amateurs
3. Indianenvironmentonline: a group of 1070 Indian environmentalists
4. Network of Indian Environmental Professional (niep): a group of 1062 members

Response from five international scientists was received through e-mail as a result of dissemination of the questionnaire through the listserv.

The organisation of two national conferences in Pune by the two mycological societies provided an opportunity for contacting a large number of scientists including a few from outside the city. Contact with research scientists was also facilitated in the course of a few special invited lectures organised by National
Chemical Laboratory on 15th October 2004, and Department of Botany on University of Pune on 24th March 2004 where the researcher was given an opportunity to explain the project.

In all 50 responses were received from scientists including 40 from Pune, 5 from other parts of India and 5 from various parts of world. They were collected during the period October 2003 - December 2004. In Pune city, mycologists from different disciplines working in different applications were available and hence the information collected, though local, was considered representative.

**Analysis of questionnaire**

The analysis of the questionnaire was done using Microsoft Access and Statpac software. Questionnaires were fed into the relational tables. Various relationships were created among different tables viz. research activities (maintaining herbaria, producing molecular sequences, etc.) and their applications (agriculture, industry, etc.), and the information needs, etc. This helped in quantitative analysis of the data.

The discipline of the scientists and the group of activities they were involved in were used as independent variables. The type of fungi and the application area could not used as variables, as the sample distribution was very small. Similarly the area of the application also not be used because of the statistically small sample.

**Questionnaire for farmers**

A separate questionnaire was prepared in the local language – Marathi for farmers. The initial prototype included thirty questions related to name, address, land in acres, number of people working in the field, water management, infection on plants, fungicides used, information collection habits, information sharing practices, etc. These thirty questions were not categorised in any specific order. The draft was shown to 2-3 experts from fungicide companies and plant pathologists and pre tested with 10 farmers.
Final questionnaire

On the basis of the results of the pre-test it was decided to revise the questionnaire (Annexure II) and divide it into four sections. Table 3.2 shows the sections and type of information requested.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Sections</th>
<th>Type of information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General Information</td>
<td>Personal details, land ownership, size of farm, number of people working, water management, etc.</td>
</tr>
<tr>
<td>2</td>
<td>Crops and fungal infection</td>
<td>Number of crop farmer’s harvest in a year and the fungal infections on it.</td>
</tr>
<tr>
<td>3</td>
<td>Care of crops</td>
<td>Care of crops, fungal infections and identification of infection, name of fungal diseases, use of fertilizers, fungicides, pesticides, etc.</td>
</tr>
<tr>
<td>4</td>
<td>Use of fungicides</td>
<td>Use and selection of fungicides, application of fungicides, and types of fungicides used.</td>
</tr>
<tr>
<td>5</td>
<td>Information collection and seeking habits</td>
<td>Habits of information seeking, resources used by farmers viz. audio, video, images, etc., information needs and information sharing practices.</td>
</tr>
</tbody>
</table>

Sample

This was perhaps a first attempt of involving the very end user in designing the contents of a scientific database. Data collection from farmers would be different from collection of data from scientists, as language, status and environment would make communication difficult. Therefore it was decided to work through two intermediaries who were familiar with farmers and had already established a rapport with them. Since the questionnaire was structured, the chances of misinterpretation and bias were less. Further, the objectives of the survey and the questionnaire were explained in detail to the intermediaries. Since this was an exploratory study, no specific sampling method was used and the questionnaires were distributed on a convenience basis by the intermediaries. A total of 50 questionnaires were distributed, 30 were received back, making the response rate 60%.
3.2 Development component

Database development was the second part of the project. This part of the work was done in close collaboration with a database expert. The researcher provided the guidelines and parameters at every stage of the development process and the database expert actually implemented the task. His input was to provide advice on various issues such as the choice of operating systems, database management systems, server, interface design language and other tools.

Final decisions at each step were taken through a process of consultation between the LIS professional (the researcher) and the computer professional. At these consultations, the pros and cons of each decision were considered and wherever possible, alternatives were physically demonstrated. After each decision was jointly taken, the database expert implemented the step.

The development of Fungal Species Information System – FSIS was divided into the following steps explained diagrammatically in Figure 3.1

1. **Development of database (first prototype):** Analysis of the survey of information needs helped to identify the fields to be created in the database. These fields were organised into broad clusters and a prototype was created.

2. **Experts’ comments on first prototype:** Copies of this prototype in electronic format were sent to five experts from different countries viz. USA, UK, Australia, Netherlands and India for comments. These experts were mycologists involved in biodiversity informatics activities and creation of fungal databases. Comments from different experts were considered and the database prototype suitably modified.

3. **Populating the database:** The database was populated with fifteen records to test the prototype. Different groups of fungi were represented. The information was collected through various ways such as from Ph.D. theses.
submitted by mycologists to the University of Pune, Internet and online databases.

4. **Laboratory testing with local scientists**: The second prototype of the database was tested, to check its usability for data entry and information retrieval. The database was loaded on local machines in the laboratory and scientists were asked to enter data in and/or search on FSIS. They were then asked to fill in a small questionnaire giving their feedback (Appendix III). The necessary changes as per their suggestions were made in the prototype and prototype III was developed.

5. **Web hosting for testing and feedback**: The third prototype of the database was then hosted on the Web. Scientists who had responded to the questionnaire on information needs and others who had expressed an interest on the project were informed about the database. They were asked to register, enter data, search and provide feedback. The earlier feedback form was used. This request was also sent to the five listservs (bionet.mycology, taxacom, fungiforyou, etc). It was made available for comments for a period of three months.

6. **Finalization**: After taking into account all the comments received and examining the data entered, changes were again made before finalising and freezing FSIS prototype IV.

### 3.3 Limitations of the research design

The project envisaged a number of sequential activities—needs assessment, specifications of requirements, database design, testing, modifications etc. Ideally each phase should have been given sufficient time in order to cover a wide base. However, since it was part of a doctoral research, it had to be accomplished in a certain time frame. Many of the phases of the project had therefore to be compressed to a smaller / shorter scale. Thus the sample of scientists and farmers taken for the
needs assessment studies was small. The sample entries used to populate the database were few. The testing phase was short.

**Challenges**

Striving to strike a balance between a participatory design and maintaining quality control was a major challenge. Users from different disciplines collect different kinds of data and document them in different ways. In order to enable them to contribute to the building up of FSIS, the database was kept fairly flexible. It was recognised that this would pose questions of standardisation of terms and quality control, and ultimately affect retrieval. For example, information about seasons could be represented by different scientists in different ways for e.g. summer, March to May, end of winter etc.

However, standardisation practices set up by the database administrator would hinder the flexibility which scientists may require. It was decided that at this initial stage flexibility was more important as it would encourage scientists to participate and facilitate in capturing digital data. The attempt to create vocabulary control could be taken up by the administrator at a later date when a substantial amount of data had been collected. In the entire project the researcher’s role was that of an intermediary, interacting with scientists of different disciplinary to understand their information needs and then translating these needs in terms of database specification, which was then communicated to the computer expert. The challenge here lay in understanding concepts and dealing with the special language and terminology of different disciplines. An understanding of the domain acquired at graduation and further developed over a decade and a half of dealing with its literature and users was of considerable help. Similarly working with databases, experimenting with several open source software, participating in the development of a city network and a course undertaken in the design of user interfaces of information systems all helped the researcher in speaking to the computer expert in his own language.

Though language problems were overcome with a little effort, the difficulty of being labelled ‘a librarian’ did create a few difficulties. Scientists assumed that when the
word information was used, the researcher was only interested in bibliographic information. It took some effort to convince them that the researcher’s interest went beyond that. This attitude surfaced several times at interviews, at lectures and at national and international mycology conferences.

The next two chapters describe the two components of the project viz. User Needs Assessment and Database Development.
Chapter 3: Research Methodology

Study of Information Needs
- Questionnaire
- Interview
- Observations

FSIS Prototype I

User

FSIS Prototype II

Experts: Mycologists and Taxonomists

Error Reports

Lab Testing

FSIS Prototype III

User Feedback

FSIS Web Hosting

FSIS Prototype IV

- Mycologists
- Agriculturists
- Pharmacists
- Bioinformaticians
- Biotechnologists

Figure 3.1 Steps involved in FSIS
References

Directory of Members, (2002). Mycological Society of India, Delhi


Taxacom (url) Peter Rauch, Berkeley, California and Jim Beach, University of Kansas Natural History Museum and Biodiversity Research Center, Lawrence, Kansas.