An Experimental System for Navigating Statistical Meta-Information - The Meta-Stat Navigator

Noboru Ohsumi
The Institute of Statistical Mathematics, Department of Statistical Methodology
4-6-7, Minami-Azabu, Minato-ku,Tokyo 106, Japan

1. Introduction

In research organizations handling statistical information, the volume of stored information resources, including research results, materials, and software, is increasing to the point that conventional separate databases and information management systems have become insufficient to deal with the amount. Increasing diversification in the media used these days interferes with the rapid retrieval and use of the information needed by users. A new system that realizes a presentation environment based on new concepts is needed to inform potential users of the value and effectiveness of using the vast amount of diverse data.

In particular, the information handled by the statistical sciences lies on the boundaries of various other sciences and clarifies the relationships and nature of information that joins these sciences. Development of a system that fully organizes and integrates strategic information is essential. The system introduced in this paper is one of many large-scale projects being carried out at the Institute of Statistical Mathematics (ISM) of Tokyo, Japan, by whom the author is employed. Studies of various design concepts for the new system began in the middle of last year and development is currently underway.

Several major objectives are built into the design concepts. One is to further the integration of stored statistical information resources. Another objective is the facilitation of sharing of the integrated resources and their provision to general users. A third objective is the construction of an integrated presentation system that uses a hyper-media environment to realize the first two objectives.

The "Statistical Meta-Information System" being introduced today, to which we have given the name "Meta-Stat Navigator," is intended as the prototype of a hyper-media environment for presenting statistical research results in a variety of formats.

1.1 Statistical information resources

To clarify the object of this discussion, it is necessary to first discuss what I mean by statistical information resources. The Institute of Statistical Mathematics is a type of cooperative research facility. Its essence lies in its accumulated know-how in the actual, practical analysis of statistical data born of research conducted in cooperation with researchers in other fields. At the same time, as a research facility, the Institute stores an enormous number of research results. Broadly classified, these information resources consist of the following:

• Cooperative research catalogs	Results obtained from cooperative research carried out together with other organizations (universities, research facilities, businesses)
• Library databases	Libraries classified according to the Institute's classification system
• Statistical software	Registration information on software accumulated by the Statistical data Analysis Center
• Published materials	English and Japanese journals, copies of the *Institute Annual*, newsletters, research memoranda, computer monographs, etc.
• Social survey data	Data sets from the many opinion surveys, questionnaire surveys, etc., conducted by the Institute itself
• EJDA materials	Data registered and distributed by the *Electronic Journal for Data Analysis*
• Miscellaneous	Various announcements (some in electronic form) introducing research facilities, computer-related information (on networking, systems, etc.)

Despite the large volume of statistical information resources available, insufficient consideration has been given to inter-relating these materials when they are used. The situation can be said to be much the same in other research facilities. These statistical information resources tend to share the following common traits: (1) the information is recorded on a wide variety of media, (2) the information is not integrated or is not managed in a common format, (3) the information is not unified or is stored in scattered locations, and (4) the information is fragmented or its relationship with other information is not necessarily clear.

For example, some information may exist only on paper, while other information may be distributed as a statistical database on computer. Some information has existed only in files managed by individuals, some of which may have been referenced by electronic mail. To fully realize the state of this information is to understand the crying need for studies of system design policies that can utilize these statistical information resources effectively.

1.2 Objectives of the Statistical Meta-Information System: system design and development policies

When we consider the four characteristics of statistical information resources listed above, it seems very natural to use the concepts of hypertext and hypermedia. In other words, if the nature of statistical information is that it is stored on a variety of media, then the use of a hypermedia environment is most suitable. In addition, converting the information to hypertext is essential to the integration of scattered statistical information. Finally, it is highly desirable that the statistical information be absorbed and integrated in a hypermedia environment that retains as far as possible the features of the many existing information management systems. This involves more than simply integrating existing systems. It entails the design of a system based on the concept of meta-information that recognizes hyperlinks between information and that uses suitable methods to convert various types of information to new information.

As a separate problem, the immense volume of data to be handled makes it difficult to create such a system at once. It would require a tremendous effort to implement at one time all of the functions of the numerous systems currently used by individuals. Thus, it would be best to continue using existing information collection methods and user functions while gradually migrating to a new system.

For this reason, development of the system's functions must be divided into a number of steps. We believe that there are two major stages in the process, as follows.

1) Functions that permit an overview of the activities of an organization and of the entire contents of its statistical information resources (that is, browsing through an overall image of the organization and its statistical information resources), and
2) Functions that permit detailed views of various information resources and "navigation" along the mutual relationships among the resources.

The former functions constitute a guide to the overall image of the organization according to some classification of its activities and statistical information resources. The user is presented with an overview of the information space and can search through individual modules in greater detail. The interface functions used at this stage include hierarchical menus and graphic information such as fish-eye views and web views that permit the user to navigate through the space. These functions are needed to allow the user to confirm the information contained in the information space and view summaries of the contents [2, 6].

The latter functions are designed to permit the user to select any desired reference path through the concrete information. Important also are support functions for cases in which the meaning of information or terms displayed on the screen is unclear. The first set of functions requires the establishment of a net to hold the basic information. The second set requires retrieval of information details or retrieval in some order, and moreover, the linking of information.

1.3 Navigation system hardware

Examining the user environment from the viewpoint of hardware, we see that the statistical information resources that we have been discussing require the management of various types of media. Some information exists only on paper, while other information is stored as text-based data or data in individual databases, both on computer. Providing an overall grasp of this fragmented information will require proposals for new system design concepts that manage a multimedia environment (using CD-ROM, optical disk, video cassette recorder, laser disk, etc.), link this information organically, and provide a "hyperdata library" in which the media used for a particular piece of information is transparent. Such an experimental proposal is the purpose of this paper.

To implement such a system requires more than simply connecting a number of computers. To provide a suitable user environment, careful consideration must be given to the databases (specifically relational databases), server-client networks, distributed system architectures that are independent of computer type, and so forth.

2. Concepts behind the Meta-Stat Navigator

As the preceding discussion has indicated, the potential users of the Meta-Stat Navigator work in various fields with computers with different features and performance. It follows that the single most important design policy for the Meta-Stat Navigator is the provision of an easy-to-use human interface. The interface must arouse the user's curiosity and make him or her want to use the Navigator. A second important policy is the development of technologies and applications that can firmly grasp the storage forms of existing statistical information resources and convert them to meta-information and hypertext. Thirdly, the use of relational databases is assumed, but it will also be necessary to construct a networking system (which we might call a hypernetwork) that will allow the converted statistical hyperdata to be retrieved and referenced freely. These three design policies are discussed in more detail below.

2.1 The human interface to the hypernetwork

As explained above, the interface to the system consists of directory functions that provide an overview of the activities of the organization and navigation functions that will permit the user to search for and reference information in greater detail. The objective of these functions is to allow the user to determine easily who has what information where and in what form. The interface must be designed using a variety of simple, intuitively understandable representations such as icons. This is explained in more detail below.

2.1.1 Entrance area (cyberspace platform functions)

The entrance area is a collection of functions that present the user with a virtual research center, allowing him or her to access statistical information in a very intuitive, non-intimidating manner. The following interfaces are provided.

1) Virtual research organization setup (organization charts, visual access, etc.)
2) Introduction to the navigation system (basic configuration)
3) Display of entrances and their relations (in practice, and introduction to the relationships among information modules and the functions for accessing them)

The human interface for entering the system will be provided in the form of the entrance browser. In addition to obtaining an overview of the organization and its activities, the user be provided with visual information on the basic system features and the user environment.

2.1.2 Reception area (multiphase navigation functions)

These functions take the user into the virtual research center and list the available statistical information resources. For example, the user is actively guided through organization charts, research departments, centers, libraries, research fields, research results, and so forth. This information is interrelated and the user can navigate freely from any location to another. Representative functions are listed below.

1) Checking the overall organization and accessing individual modules
   An overall view of the organization can be displayed and individual research departments, research topics, etc., can be retrieved and displayed on the screen. Visual indexing (pop-up lists, hot text, and hot spots) will be used and the desired research topics and keywords can be retrieved or entered and explained. As far as possible, images will be used at this point. In addition, the relationships among research topics can be retrieved using keywords and diagrams.
2) Visual guide functions
   Proper usage and guidance will be displayed visually on the computer screen using visual help and online help functions.

2.2 Conversion of statistical information resources to hypertext

Technologies and applications that can firmly grasp the storage forms of existing statistical information resources and convert them to meta-information and hypertext will be required. For example, the following functions will be needed to provide a common form for research reports, cooperative research information, library and document databases, and other information that is recorded on different physical media.

2.2.1 Functions for converting existing statistical information resources to formats that can coexist in a database

Applications will be required to convert existing information to new forms in a database regardless of its storage format (that is, while maintaining existing data collection functions). These functions will provide object-type data structures for existing statistical information and create statistical meta-information that can handle not only character data, but also image data (still or animated) and voice or sound data.

2.2.2 Data registration and update functions

It is in the nature of statistical information that it will periodically be added to or updated. Thus, there will be a need for entry functions that provide intelligent hypertext structures and are sufficiently easy for users and researchers to register and edit information of which they are the providers. This hypertext interface that permits information providers themselves to register and update their own information also insure that the system users will always be able to retrieve and reference the latest information.

2.3 Management functions for statistical meta-information in a network environment

Functions, called "intelligent consulting" functions, that facilitate cross-referencing among statistical meta-information converted to hyperdata (character and numeric data, still images, animated images, voice and sound data, etc.) regardless of its actual format or physical storage media will be essential. Thus the methods selected for managing character and numeric databases and image and sound databases will be the deciding factor in improving the efficiency of the system.

We are studying the adoption of a dual management method that uses two database servers. The system would consist of a relational database that would provide overall management of text and numeric data and its corresponding server and a separate server for image and sound databases. These two networks would be linked. Clients would then access these servers. However, since the two servers are running in tandem, links between the information that fulfills a client's request would be displayed automatically. Of course, these activities would occur in the background, out of the client's sight. From the client's viewpoint, she or he would appear to be using only a single computer, and complex, intimidating operations would be avoided.

The use of this method also provides flexibility in editing and updating the statistical information. To prevent statistical information from becoming outdated, it must be possible to maintain and update the statistical information easily and rapidly using a simple interface. At the same time, these functions must maintain the interrelationships among the data. For example, the functions must provide hypertext entry that will permit image data to be added to text and numeric data and then to further add data in some other format. This method would be most efficient for maintaining and updating statistical meta-information in the complex hypermedia environment that we have been discussing. Functions can also be provided for freely creating electronic reports (note-taking functions) of retrieved and cross-referenced statistical information. Finally, it would not be impossible to provide additional functions (self-organizing functions) that use these note-taking functions to add user-created links as new information in hypertext format.

3. Advantages of the Meta-Stat Navigator

More time will be required before the system introduced in this paper is completed. However, the following advantages can be obtained from using the system in its current state.

1) Increased user support from ease of operation
   The system receives increasing support from a wide range of users because of the emphasis given to the desktop metaphor and the ease with which users can connect to the system and retrieve and view statistical information.
2) Visual presentation
   Research results and reports are presented in a graphic, visual manner; the multiphase index permits cross-sectional retrieval by research field, researcher, and research topic.
3) Full tutorial functions
   Learning curves are reduced for researchers, cooperative researchers, students, and others. Information retrieval, extraction, input, and update can be conducted easily because explanations of various functions are displayed interactively from an image database in the computer.
4) Increased information interchange
   The use of software technologies that make use of multimedia environments will increased dissemination of the results of high-level statistical research. This can be expected to increase the level of intra- and inter-disciplinary research.
5) Extensibility
   The adoption of graphical user interface (GUI) concepts and of object structures for all of the system's functions allows users with little computer experience to use the system. Application of the system can be expected to expand into other fields. Specifically, the design concepts remain even when the contents are replaced with other types of information, making the system usable as a navigation system for other types of organizations and disciplines.

4. Conclusion

When it is completed in several years, the Meta-Stat Navigator will integrate statistical information resources and provide a support environment for users wishing to "borrow" the accumulated know-how of statistical science. The ultimate objective of the system is networking with international statistical information systems and to provide this statistical information for use by researchers in many countries.

References

[1] Barrett, E. (ed.) (1989). *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*, The MIT Press.
[2] Brathwaite, K.S. (1991). *Relational Theory: Concept and Application*, Academic Press.
[3] *Communication of the ACM*, Special Issue: Hypertext, **31**, 7.
[4] Conklin, J. (1987). Hypertext: An Introduction and Survey, *IEEE, Computer*, September, 17-41.
[5] Frisse, M. (1988). From Text to Hypertext, *Byte* (Special Issue for Hypertext), October, **13**, 247-253.
[6] Nielsen, J. (1990). *HYPERText & HYPERMedia*, Academic Press.