to appear in *Journal of Management Information Systems*
Winter 1991/92 Special Issue

Building an Electronic Community System

Bruce R. Schatz

Community Systems Laboratory, Life Sciences South
University of Arizona, Tucson, AZ 85721 USA
schatz@cs.arizona.edu

Abstract

An *electronic community system* encodes and manipulates the range of knowledge and values necessary to function effectively in a community or organization. The knowledge includes both formal data and literature and informal results and news. The manipulation includes both browsing through the available knowledge and recording and sharing interrelationships between the items. A large-scale experiment is underway to build an electronic community system for the community of scientists studying the nematode worm *C. elegans*, a model organism in molecular biology. This paper discusses a model for community systems and previous such systems in science, the biology experiment and a previous system, the enabling technology for handling the knowledge, the enabling mechanisms for handling the values, the state of the prototype, and speculations on future applications in supporting organizational memory.

Key Words and Phrases: electronic community systems, electronic communities, scientific applications, information spaces, telesophy, organizational memory

Introduction

To function effectively in an organization, one needs access to a wide variety of knowledge. This knowledge includes not only the archives of financial data about products and technical information about designs, but also test results, meeting reports, and other informal sources. To integrate this knowledge, one needs to understand the relationships between the available items in the context of the company values and culture. Productivity would be greatly enhanced if a substantial fraction of this integrated knowledge was easily available from the employee's desktop computer. Traditional management information systems have only supported a small portion of this functionality.

It is now possible to build substantial prototypes of information systems which handle such a wide variety of integrated knowledge. These can be termed *electronic community systems*, which encode and manipulate formal and informal knowledge and their interrelationships. A large-scale experiment is being carried out to encode the knowledge of a community of scientists and build a software environment to manipulate this knowledge from their laboratories. The community is those scientists studying the nematode worm *C. elegans*, a model organism in molecular biology. This paper discusses a model for community systems and previous such systems in science, the biology experiment and a previous system, the enabling technology for handling the knowledge, the enabling mechanisms for handling the values, the state of the prototype, and speculations on future applications in supporting organizational memory.

A Model for Electronic Community Systems

The word "community" is closely allied with the words "common", meaning "the same", and "communicate", meaning "to exchange information". Originally, **community** referred to the people residing in some small physical location and more generally to their shared values. The meaning of the word has been extended to groups of people with common interests and shared values who reside in geographically separate places. Thus, one refers to the scientific community or the physics community or the relativity community. This section discusses a general model for a community and its support by a computer system.

To support a community electronically, it is necessary to encode as much of its knowledge as possible. Figure 1 illustrates the range of possible knowledge which might be supported by a computer system. To live effectively within a community, one must have available both the formal archival material and the informal transient folklore. This includes the fundamental items of data for the community, e.g. as maintained by database management systems, and the intermediate results, e.g. as contained in electronic mail messages. This includes the archival literature for the community, e.g. as maintained by information retrieval systems, and the intermediate news, e.g. as contained in electronic bulletin boards. Finally, it includes support for the shared values as well as the common interests. The mores of the community can be supported via a variety of mechanisms for recording the relationships between the data and information, e.g. by providing hypertext documents.

An **electronic community system** is a computer system which encodes the knowledge of a community and provides an environment which supports manipulation of that knowledge. Different communities have different knowledge but their environment has great similarities. The community knowledge might be thought of as being stored in an electronic library. Much of the material originates within external sources. The environment must accordingly provide software for building a library to access these sources, e.g. convenient mechanisms for encoding and browsing what is available. But, unlike existing physical libraries, a community library is dynamic and the members will actively add items to it. The environment must also provide software for updating this library, e.g. convenient mechanisms for refereeing and sharing added items. The environment thus provides support for both the knowledge and the mores of the community.

The functionality of an electronic community system can be motivated by considering the use of a physical library. Consider the analogy of doing research in a physical library in order to write a book. You start with references from a paper or colleagues, look the references up in the card catalog, go to an appropriate section of the library, and scan your eye along the titles on the spines of the books. If any books look relevant, you pull them off the shelf for detailed examination of the pages. If some pages look relevant, you make a copy for later use. After some scanning and examining, you go to another section of the library, often using references found in the previous section. When the research phase is finished, you write your book, utilizing references to copied pages, and submit your book to be published (and subsequently placed itself into the library).

In the model of a community library, the books are distributed multimedia objects. There are three basic stages in the interaction process: browsing, filtering and sharing. In *browsing*, the user can rapidly examine the items in the library. This can be accomplished via search, by giving an associative specification and viewing matching items, or via navigation, by following the connections from a given item. The results can be displayed at a variety of summary levels. Multiple searches and navigations can be issued and cross-compared to located desired items.

The next stage is *filtering*, where the user culls the items located by browsing into some desired set, relevant to the current need. If the browsing speed is sufficiently fast, user view of displays may be sufficient to manually select relevant items. If the items are too numerous or too complex, manual examination may not be sufficient. In this case, a set of selected items can be passed into an external analysis program for automatic filtering. Such a program might sort the items by date or perform a complicated computation to determine rank ordering against some similarity metric.

After a set of desired existing items has been found, the user may wish to add this set with comments back into the library. *Sharing* is the support for publishing in the electronic library. A variety of mechanisms are supported within an electronic community system for grouping the items to record their relationships, e.g. storing a set with a description of its relationship or forging a connection link between related items. Mechanisms are also supported for writing hyperdocuments, which incorporate other items into the text via embedded links. Once a sufficiently important new group or document has been composed, facilities are available for releasing this to the community. A variety of mechanisms are supported to provide editorial and privacy control of the release process. These mechanisms are the attempt to encode the mores of the community, by permitting members of the community to control the quality of the material in the library and who may view the material.

Electronic Scientific Communities

The business of scientific research is an unusually good domain to investigate the development of an electronic community system. Practicing scientists need access to a wide variety of knowledge to carry out their research. Much of this knowledge resides in formal published literature, but much also resides in informal community knowledge. Some of this informal community knowledge, such as preliminary results, will eventually become published, but other knowledge, such as details of methods and the "lore" of experimental systems, never reaches the formal literature. The scientist who shares a community's current informal knowledge and has rapid access to the formal knowledge can do better research. This is particularly true in the biological sciences, which are largely data-driven, because the choice and design of experiments depends on familiarity with the most current methods and knowledge. As the pace of science increases, only a small number of insiders who lead each field, the "invisible college" [1], have enough knowledge to perform seminal research efficiently. If the informal community knowledge could be captured and disseminated more widely, the quality and efficiency of scientific research would improve. This is because the invisible college would be larger and because it would be open to scientists from diverse disciplines which would encourage novel interdisciplinary research.

The existence of nationwide networks has fostered electronic scientific communities. The ARPANET was the first nationwide computer network widely available to the scientific community in the United States. It was constructed during the late sixties and reached its potential throughout the seventies. The original motivation for the network was support remote access to the large "supercomputers" constructed and purchased by ARPA-funded researchers. However, what emerged as the most important service was the new facility of *electronic mail*, which provided a new communications medium. The ability to convey informal information rapidly caused a new feeling of closeness among the researchers on the network and the emergence of the first

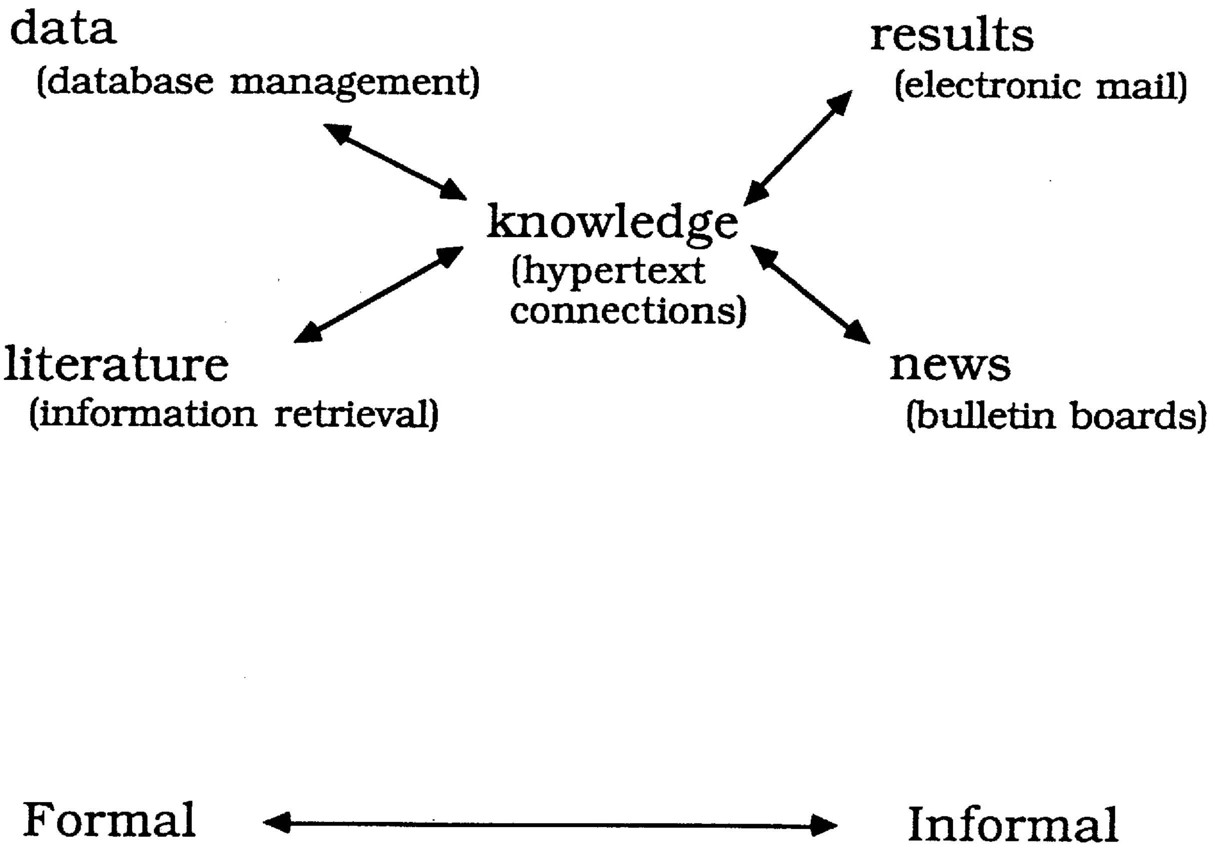


Figure 1. An electronic community system records and manipulates all the knowledge of a community or organization.

widespread electronic community, the ARPANET community [2]. Researchers on the ARPANET could get essential information more quickly than those not on the net and many collaborations took place without the collaborators ever physically meeting.

Early users of electronic mail in the ARPANET noted the convenience of being able to send items to many others on a distribution list. Standardized lists became established to distribute messages to people interested in a wide variety of topics. These lists evolved into the next generation of community system, the *electronic bulletin board.* An illustrative current-day bulletin board system is Netnews [3] which distributes messages over USENET [4]. USENET is not a centrally planned and maintained network but instead is a loose collection of computers running the Unix™ operating system connected by a wide variety of physical transmission lines from high-speed leased lines to ordinary telephone lines. It operates today with more than 250,000 users on more than 10,000 machines spread throughout the world. Netnews contains more than 650 boards across a wide variety of topics, ranging from comments about existing computers to technical science to popular culture to job positions to movie reviews to cooking recipes. The software functionality has evolved to support streamlined posting to the appropriate boards, comments on previous messages, reading of selected boards, and saving of selected messages.

Everyday operation of Netnews shows the benefits of community sharing. When you post a technical question, you often get a detailed technical answer from somewhere out in the Net, often from a place you would never have thought of looking. Frequently, your posting stimulates a series of postings, each illuminating the problem from a different point of view. It is common to see the understanding of a problem evolve over a week through comments from a series of different postings from different parts of the world. For example, a recent query by the author concerning radio interference on laptop computers elicited helpful comments and critiques from people at sites in Massachusetts, California, Wisconsin, Oregon, Hawaii, Ontario, Germany, and Sweden. There is a real feeling of community interaction to solve shared problems on the many responsible boards. As with the ARPANET boards, Netnews tends to be self-policing. The users tend to be responsible individuals, who understand that the system is supported by the generosity of their employers. People who abuse the Net (by posting inflammatory or irresponsible messages) are quickly dealt with by peer pressure. There are elaborate documents on appropriate Net etiquette: what boards are suitable for what topics, what content is appropriate for a posting, how to be terse and polite, and so on. Different board types have evolved a spectrum of editorial control, ranging from boards where anyone can post anything to ones where all messages pass through a human moderator for topic and quality control.

As the speed of networks increases, so does the range of information that can be effectively encoded within an electronic community. The dream of an all-encompassing science information system is an old one, since the possibility of being able to sit in front of a computer and be able to access all the knowledge you need for your research is so attractive. This dream has resurfaced periodically whenever the computing and communications technology makes a dramatic increase in functionality. See, for example, the "future of libraries" study after the advent of minicomputers in the sixties [5], the "world scientific information system" study after the advent of computer networks in the seventies [6], and the "national collaboratory" study after the advent of workstations and supercomputer networks in the eighties [7]. The forthcoming NREN (National Research and Educational Network) will provide network speeds fast enough to support interactive manipulation of a wide variety of material across the national scientific network. This leads to the possibility of realizing an all-encompassing information system with the next generation of community systems.

Building an Electronic Community System

Community systems of the near future will support the complete range of knowledge and functionality discussed in the Model section above. They will support a wide variety of database management and information retrieval functions to support a wide variety of formal experimental data and literature information. In addition, they will support a wide variety of electronic mail and bulletin board functions to support a wide variety of informal results and news.

The Community Systems Laboratory is building an electronic community system in the domain of scientific research and evaluating its use within the community as a large-scale experiment. The resulting system is meant to serve a wide variety of communication needs within the community, both retrieval and analysis, as well as rapid sharing of knowledge with others. It will permit researchers, who have common interests and shared values but are geographically dispersed, to browse and share the community knowledge. The scientific community chosen for this experimental project is the community of molecular biologists united by their common use of a model organism, the nematode worm *Caenorhabditis elegans* [8].

The Worm Community

Building an electronic scientific community in today's largely non-electronic world requires a specialized community with an appropriate set of characteristics. It must have a large amount of data, both formal and informal, and a real need to manipulate this data extensively. The data must be freely available and already largely in electronic form. There must be many interested users who are willing to experiment with new technology and who have adequate computer equipment and network connections. There must be real support for data administration and software development, which implies that the community must be an important one scientifically so that adequate funding is available.

A scientific community which exhibits these properties is the **worm community**, the molecular biologists who utilize the nematode worm *Caenorhabditis elegans* . Molecular biology is a largely data-driven experimental science and, due to such efforts as the Human Genome Initiative, its data is growing rapidly and being stored in databases. Communities in molecular biology often form around organisms, rather than techniques or problems. *C. elegans* is a non-parasitic worm found in the soil, which has been extensively studied, with a wide range of experimental data available on its genetics, anatomy, and development [9]. The "worm" has become a primary model organism and will likely become one of the first to be completely sequenced.

The worm community itself is young but growing rapidly, with more than 500 researchers at the last large meeting in June, 1991. It has a close-knit and communicative group of "insiders", the postdoctoral fellows of Sydney Brenner who initiated the modern genetics of *C. elegans* in the late sixties. It has a strong community tradition of free sharing of unpublished data, unlike the competitive nature of many other areas of science. Substantial bodies of data are already in electronic form, and there is an adequate level of computer literacy and interest in electronic communication.

Two examples of widely shared unpublished data within the worm community illustrate its suitability as a cooperative experimental electronic community. The *Worm's Breeder Gazette* is a newsletter analogous to a moderated electronic bulletin board, which consists of short research items and has been published several times a year for more than 10 years in an unrefereed open format. The physical map database is an electronic recording of an ordered set of cloned DNA fragments which constitute the genetic material of the worm [10]. Its curators map new fragments and distribute the database as a service to the community; the easy availability has dramatically facilitated molecular analysis of genes.

The worm community is an excellent candidate for an electronic scientific community due to a number of unusual characteristics. It is the right size: big enough so that newcomers no longer know the pioneers directly but small enough so that fierce competition has not yet set in. It is the right age: old enough so that there is extensive knowledge already discovered but young enough so that the insiders still remember the original days and want to preserve their closeness. It is the right importance: significant enough so that discoveries make a scientific difference but offbeat enough so that researchers do not hoard their data.

Finally, the worm community has always had a special tradition of sharing knowledge. Reasons include that many members of the worm community were trained in the phage group where openness was encouraged and that there has always been a primary goal of understanding everything about the worm. The worm community's informal network of communication is becoming inadequate as the community grows, and there is concern among the insiders about losing the community's unique flavor. Thus there is an immediate need for an electronic worm community and a favorable set of conditions for an experiment in electronic communication.

Worm Community Knowledge

One major advantage of the worm community as a community system testbed is the large amount of important data available. The most significant available sources are listed below.

The categories for these materials span the range of editorial control: published literature is refereed, archival data is edited (checked for quality), informal information is moderated (checked for topic), and unpublished data is posted (no checking). Archival data is typically maintained by a central curator, whereas unpublished data is maintained by individual researchers. Similarly, much of the informal information is unrefereed literature.

- *Archival Data*

Gene Descriptions. These are text descriptions of the phenotypes of worm mutations and other genetic effects. About 1200 of perhaps 5000 genes are known.

Genetic Map. This represents the relative positions of genes on the chromosomes, based on how often two genes recombine during sexual reproduction. The display is a line drawing with marks denoting the gene positions.

Physical Map. This represents cloned fragments of worm DNA and how they overlap to form the chromosomes. Genes known to be within a fragment thus have a known absolute position. The display is a drawing with many overlapping lines and located genes. About 11,000 clones nearly cover the chromosomes.

DNA Sequences. These are the nucleotide codes for the genes, with only a few known at present. They display as strings of ACGT.

Cell List. This table and text identifies every one of the 959 cells and its location in the worm.

Wiring Diagram. This table gives the complete connection pattern of all 302 neurons. It is displayed as a line drawing.

Cell Lineage. This table and line drawing gives the complete developmental history of every cell, i.e. which cell develops into which other cells during growth.

- *Formal Literature*

Bibliography. This text gives citations for most of the *C. elegans* literature. The current list of some 1400 articles is maintained by a central curator.

Abstracts. Text abstracts for most of these articles are available in on-line literature databases such as Medline, Biosis, and Agricola.

Full-text. Text of the complete contents of all the papers would be preferable for a complete community library. If copyright permission was available, it would be possible to scan and recognize the characters in worm journal articles.

Page images. The most desirable storage would be complete article contents including, in addition to text, drawings, figures, tables, graphs, and equations. The most practical approach is to digitally scan page images of the non-textual material. After an article is located by searching the full-text, the text can be displayed in formatted form along with the images of the figures.

Worm Book. This is the standard reference book *The Nematode Caenorhabditis Elegans* [9], which contains review articles on all aspects of *C. elegans* biology as well as the existing data, nearly 700 pages total.

- *Informal Literature*

Newsletters. The *Worm Breeder's Gazette* has been published for 10 years and is a rich source of one-page notes about items of interest from experimental results to methods to news.

Conferences. The bi-yearly C. *elegans* Meeting publishes one-page abstracts of presented papers which can be cited as personal communications.

- *Unpublished Data*

Lab Directory. This is the list of current researchers with their addresses.

Strain List. This is the list of the worm mutations available from the centralized stock center.

There is also useful unpublished data maintained in the laboratories of individual researchers. They include: experimental methods (text), genomic maps (drawings), micrographs (images), and strain lists (text).

- *Analysis Software*

There are a variety of programs available for analyzing the biological data, e.g. comparing sequence similarity. The environment provides a facility for selecting a set of items and passing it into an external analysis program. Some of these programs also provide sophisticated displays of the data, e.g. genetic and physical maps.

- *Community Lore*

Much of the knowledge about the worm is not currently recorded anywhere. The Worm Community System will support facilities for entering annotations, specifying relationship links, and writing documents. This new material will be added to the shared community library.

The Telesophy System

As an introduction to the technology required to implement an electronic community system, its predecessor will be briefly discussed. Then the specific support for technology and for sociology will be described. Finally, the existing prototype will be discussed, along with future plans.

During the sixties, Douglas Engelbart's NLS project carried out a pioneering effort to build a tool to "augment the human intellect", a single computer system that enabled a researcher to interactively manipulate all their knowledge [11]. The resulting system could manipulate collections of documents consisting of a hierarchy of paragraphs, which were interconnected on the basis of similar words. There was extensive support for collaboration, e.g. a shared journal which kept a record of annotations and revisions [12], and support for remote access over ARPANET. The project gathered a group of devoted users, eventually totaling several hundred, and attempted to support a few small specialized communities [13]. Typical users were information specialists whose jobs involved examining and writing large formal bureaucratic documents with detailed hierarchical structure.

During the eighties, the present author carried out a project to investigate whether sufficient technology was available to build a complete community system [14,15]. Much infrastructure had already matured, e.g. hardware technology such as large-memory graphics workstations and high-speed fiber networks and software technology such as bibliographic information search and object-oriented programming systems. This project was called **telesophy**, or wisdom at a distance, to indicate that the system was intended to support transparent manipulation of knowledge across computer and communications networks. The concept of telesophy as an all-inclusive network system was intended to be analogous to the concept of telephony, with the ultimate goal of supporting transparent manipulation of "all the world's knowledge" just as the telephone system supports transparent connection of "all the world's telephones".

As part of the Telesophy project, a prototype was constructed of a complete community library, including both data and software [15,16]. The system forms a distributed digital library, which enables fast browsing across a wide range of data physically distributed across a network. Data is collected from external sources and transformed into uniform objects, which can be manipulated by a uniform set of commands, regardless of the original physical data type. The set of objects is called the *information space* and may be distributed across many machines within a network. Retrieval takes place transparently, regardless of the actual physical location, and is done either by associative search or by following links between objects. The software runs on Sun workstations and has been distributed to more than 45 sites. The software contains a custom window manager, object system, network handler, and text searcher.

In the main configuration, a wide variety of data was collected and transformed into units in the information space. This collection represented a good sample of what is currently available electronically. It also spanned the range of different media types, from text to graphics to image to video, as a test of the system's ability to support type transparency. The prototype space consisted of some 300,000 items from some 20 data sources. The informal material were short text messages which included bulletin boards, electronic mail, wire services, and notes. The formal material included bibliographic citations with abstracts covering computing from INSPEC™ and biology from Medline™, and also full-text of magazine articles and movie reviews. The pictorial material included line drawing graphics, black and white images, color magazine figures, glossy photographs, and videodisc stills. Finally, the video material included played segments from a variety of educational and entertainment videodiscs. The data thus spanned the range from informal to formal material, as well as including material such as pictures and graphics. A few materials with code were also collected, including playing of videodisc segments and stored queries which are executed on-the-fly to provide a different result each time.

The software supports transparent, distributed information retrieval. Every data item is searchable by combinations of phrases on all the text associated with the item, e.g. abstract, body of text, picture captions. Searches can be done across all the different databases and all matching items returned. The databases can be physically distributed across machines in a network. In the prototype configuration, the 20 databases were spread across 3 large file servers connected by a building Ethernet so that any appropriate workstation (i.e. a Sun-3) could access them. The database searching takes place on the file servers, while the user interface and manipulation takes place on the user's workstation. The Telesophy System concentrated on supporting associative search as in information retrieval systems rather than following of links as in NLS or hypertext systems; [17] discusses tradeoffs between search and navigation.

The software implementation is tuned to support fast browsing. The data is fully indexed so that processing a query typically takes 1-2 seconds. The resulting items are then downloaded from the remote file server over the network to the local workstation. The interaction is "instantaneous" (less than 1 second) for displaying and page flipping one-line summaries of query results or zooming into the complete items. This same speed has been maintained for the vast majority of the items (text and line drawings) across a variety of physical networks: building Ethernet, campus Ethernet, and a WAN (wide-area network) consisting of two building Ethernets 40 miles apart connected by a private T1 line.

In addition to supporting library browsing, the Telesophy System also supports kinds of community sharing. All of the external data items are represented as information units, the collection of which forms the information space. There is a single set of commands for basic manipulation of any information unit, independent of type and location. The user can perform an exploratory browsing session, issue multiple queries, and save a collection of selected items as a new information unit which can be indexed and placed back into the space. These collections are a simple form of "meta-level" grouping, of classifying sets of items of different types from different databases into new semantic groupings, which can arise by saving the results of a simple query or from the results of considerable searching and analysis. Since all users, regardless of their physical location, access the same information space, these new composite IUs, which form regions in the space, are automatically shared with other members of the community.

Proving the viability of a new communications medium requires demonstrating its implementation is technically feasible and its deployment causes a sociological change. The Telesophy System demonstrated the technical feasibility of building a community system, but failed to achieve widespread usage due to the difficulty of obtaining suitable data in electronic form for the needs of the user community, electrical engineers in an industrial research laboratory. Experience with physical libraries has shown that one of their most important features is complete coverage, i.e. essentially all materials on the covered subjects are available. Coverage is even more important in an electronic community library since a key feature is rapid annotation of existing material. During the course of the Telesophy project, it became clear that demonstrating the requisite sociological change would require carrying out a large-scale trial with a specialized community, thus prompting the beginnings of the Worm Community System project.

Enabling Technology

There are a number of technologies required to effectively implement an electronic community system. This section discusses one of the most important -- the representation for the knowledge in the community library. This builds upon the experience from the Telesophy System.

Information Spaces

The data model for a community system must support uniform commands for browsing and sharing across the complete spectrum of community knowledge. This requires supporting features not well supported by the models underlying traditional database management systems.

Community knowledge spans a wide range of types, each requiring its own operations for search and display. Community knowledge is interconnected and needs an efficient representation for making relationship links between items. Community knowledge exists across many sources, which can be distributed across a network. The relational model, for example, cannot easily support multiple types or arbitrary links between arbitrary groupings. An appropriate object-oriented model can, since each type of object can have its own set of operations and each object its own set of pointers to other objects. A community system uses a particular kind of federated heterogeneous distributed object-oriented database, called an *information space.*

Information spaces support uniform manipulation of heterogeneous data items by transforming them into homogeneous information units. The generation of an information space begins with data already existing in some external source. The format of this data is administratively transformed into a canonical internal representation called an **information unit or IU**. An information unit is an encapsulated object, in the sense of an object-oriented programming language, which has an associated set of operations to provide manipulation capability for its particular data type. Every "database" thus has a set of transformation routines and every "data type" has a set of data operations. Once the data items have become information units, there are a set of generic operations available for performing on them. These generic operations support uniform commands at the user level for such functions as search, display, and grouping. Thus a user of an information space need only learn one set of commands to manipulate information units, which operate uniformly across a wide range of external data types. Each information unit may be connected to other units to represent a semantic relationship and collections of information units may be grouped into new composite units. An **information space** is a set of information units and their connections. Logically, it is a single uniform graph structure, although physically it may be composed of many different sources of data of many different types stored on many different machines in many different locations spread across a network.

There are several levels of representation in an information space. Data exists in the external sources and is transformed into information within the space. Knowledge, in the sense of **community knowledge,** is represented by the different components of information units. Any IU can be annotated; a typical *annotation* is a note stating some additional feature of the encapsulated data, e.g. this gene may encode this function. Any two IUs can have a relationship specified between them; a typical *connection* is a link to another IU supplying additional information, e.g. this article discusses this gene. Any collection of IUs can be grouped into a single composite IU which forms a region in the information space; a typical *region* is a set of IUs on the same topic, e.g. all genes coding for mechanosensory deficiencies. Since every IU has a unique identification within the entire space, it is possible to implement a uniform mechanism for forging and maintaining these groupings, even across sources. As discussed below, every IU also has specification to provide publication control over the sharing of these groupings.

Forming the Space

Anything accessible may potentially be incorporated into the space. That is, all data reachable via the underlying network for which appropriate transformation routines exist can reside logically within the information space. The time when administration is done to bring data physically into the space depends on ease of reliability and maintenance. In many cases, maintaining the data directly in an external database is the most convenient; in this case, data items are transformed into information units only when actually retrieved (and then only temporarily during use so that any updates must be written back into the database itself). If the data is to be maintained directly in the information space itself, the data items are transformed once into information units when they are brought into the space and then any updates are performed by operations within the space. Since maintaining consistency and correctness of large amounts of data requires considerable system support, initial implementations of information spaces will likely rely on existing database management systems to provide maintenance, transforming external data items into internal information units on-the-fly or periodically whenever the database has been significantly updated.

In the worm information space, for example, there are a variety of methods for incorporating external data and software. The support for these may be handled internally (as objects brought into the system) or externally (as objects existing outside the system). Some external data is read in from text files, then handled by internal software. For example, the gene list is a text description kept in a file then supported by the built-in text display. Some external software is invoked as a separate process with arguments. For example, the sequence map display is called as an external program. Some external software is invoked with objects passed in and out. For example, a sequence analysis program is passed sequences in a canonical textual format and returns text which is transformed back into sequence objects. Finally, some external software supports its own classes which are directly communicated with, providing internal software with direct interactive access to external objects. For example, the genetic map displayer is an external program which implements an annotation command that invokes the internal support for annotating the objects belonging to the external program.

The major generic operations built into the system, as part of the IU class definition, are the support for grouping. These include connection links and region sets. Other operations, which provide support for the uniform user commands, are implemented at the individual subclass level, e.g. those for search and display. This enables the system to support many different types of search, e.g. text and sequences, and of display, e.g. text and maps. Some of the type classes are available in essentially every community, e.g. an atomic class for text and a composite class for some kind of hyperdocument. Other types are specific to individual communities, e.g. an atomic class for gene and a composite class for genetic map containing gene positions. The object structure of information units enables an electronic community system to be extensible, with a base set of classes that can be augmented by specific classes for a specific community.

Enabling Sociology

The above discussions have indicated it is technically possible to collect a significant amount of community knowledge and make this easily available to community members. Insuring their active participation in this electronic experiment requires resolution of the following sociological problems, among others.

Editorial and Quality Control

Published literature typically goes through a careful refereeing process. This is also true of the archival data, where there is typically a trusted central administrator who performs editorial quality control. With informal information or unpublished data, especially when entered by the users, quality control becomes significant.

The solution to quality control in the printed literature is to have a range of editorial review which leads to a spectrum of documents ranging from lab notes to working documents to internal memoranda to newsletter announcements to conference papers to journal articles to research monographs to text books. A similar spectrum has emerged in electronic bulletin boards. In public boards, anyone can post any message. In moderated boards, all messages go first to a moderator who eliminates those on wrong topics, redundant, or inflammatory. In edited boards, the editor passes judgement not only on topic but also on quality and format. There has been talk of true refereed boards with long articles, but not many examples exist.

A community system should provide a mechanism for "levels of editorial release", i.e. how carefully checked an item is before it is released to the community. Following on the experience of electronic bulletin boards, the spectrum of editorial control should include: posted, moderated, edited, refereed. The system, however, does not determine the policy of which level an author chooses for a particular item or who performs the function of the editor for which items. An appropriate set of conventions will have to evolve for the electronic community library, just as it has already for electronic bulletin boards. Based on experience with the worm community in the past, editors will emerge who can provide appropriate levels of quality control for each data source and who are sufficiently respected by the community so that their blessing of this data is trusted.

The level of editorship should be recorded on each item in the information space, because this is of interest to the researchers who are evaluating the suitability of particular information units for their current purposes. This is a form of policy which permits the individual users to choose for themselves whether they are currently interested in refereed facts and data or in rumors and notes.

Privacy and Reward Considerations

Another problem in extending the community library beyond formal data is whether the members are willing to share this data before it has been formally published. The tradition of freely sharing unpublished data is a primary reason for choosing the worm community for the initial experiment. But there is a significant problem in any scientific community establishing credit and priority, particularly as competition becomes more intense.

The community system should provide the mechanism of "levels of privacy release", i.e. who is permitted to view which material. Sample levels include: private (user only), colleagues (local), colleagues (global), community. As with the editorial release, the policy for each item is individually determined by the author and can be changed as the item evolves in maturity and quality. Each researcher can also set who is permitted to view each level of release, e.g. who their colleagues are.

Conversely, for searching purposes, each researcher can use the privacy level to help determine the appropriateness of those items in the information space that they have permission to access. It should be noted that the privacy levels enable the community system to support services equivalent to electronic mail and bulletin boards.

An issue related to privacy is rewards. What reward does the author of an information unit receive? It will be a long time before the prestige of making a connection in information space rivals that of publishing a paper in a journal. The system can provide the mechanism of a super citation index, by keeping track of the frequency that an item is retrieved and the number of times a connection is made to it. Hopefully, these usage statistics will aid in establishing policies for electronic publishing. Also note that every information unit has complete attribution of its creation, e.g. author and date. In fast moving fields with extensive electronic coverage, this could provide a method for establishing priority and credit.

Prototype Community Systems

The first release of the Worm Community System was completed during the summer of 1991 and is now in the labs of the initial test users. They are using it to browse the data and beginning to add annotations.

The current community knowledge spans the potential range. It includes fairly complete archival data, e.g. the list of gene descriptions, the genetic map, the physical map, and many DNA sequences. It includes abstracts of most of the archival worm literature. The worm newsletter has been completely scanned and the text recognized, so that articles can be searched for then displayed as formatted text with accompanying images for the figures. Unpublished data is available, such as standard strains and a person directory, plus a sampling from individuals.

The software functionality also spans the potential range. Searches can be done across all the sources for text phrases. An extensive set of links has been made between information units by a variety of automatic and manual means. These links can be followed from any IU to the related set of IUs. Sets of IUs can be selected and grouped. Several external analysis programs can be called to provide displays of worm data for the genetic map and the sequence coding map. Finally, an annotation facility is available which permits a note to be added to a set of IUs giving additional information about the group. This note may include embedded links as well as text. When saved, annotations are released into the information space, where they can be manipulated as ordinary information units.

Sample Session

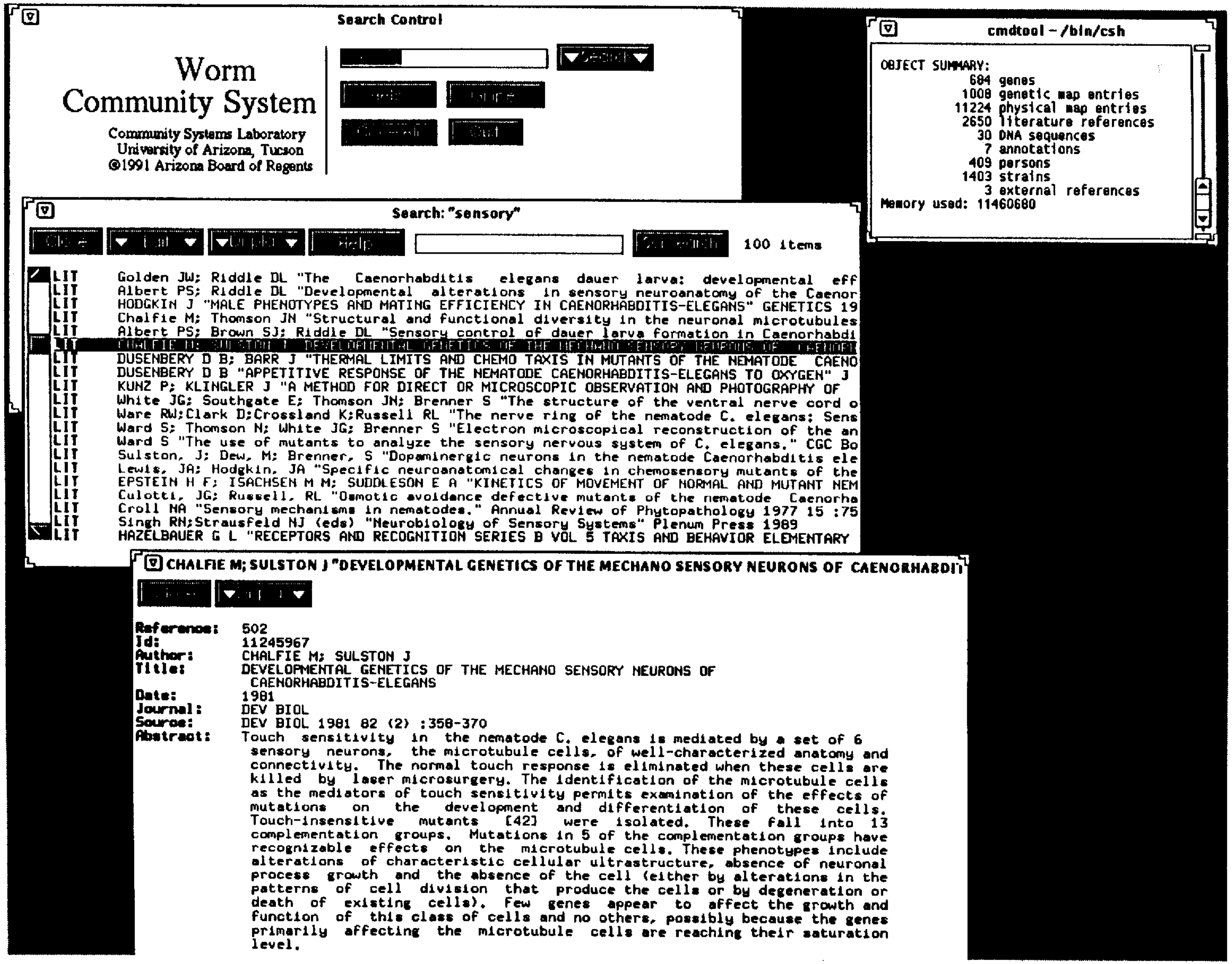
Figure 2 is a screendump from a sample session with the Worm Community System. This session is a summary of the interaction with a biologist, interested in the sensory neurons, who is attempting to discover which genes in *C. elegans* control the sense of touch, mechano-sensation. The information space enables the biologist to rapidly locate all such known genes and retrieve information about them.

The session starts with the user entering a search for the keyword “sensory” as shown in the topmost Search Control window. The search is performed across all objects of any type contained in the information space; the number of objects is shown in the upper right. The window below Search Control labelled Search: "sensory" contains a summary of the set of objects in the worm information space matching this keyword (containing that text string). Each object has a one-line summary (uniform for all types) which can be zoomed into by pointing with the mouse and double-clicking. The selected object is displayed in the bottom window. it is a literature object containing a citation and abstract from the journal articles about the worm.

In addition to associative search, units in the information space are interconnected and the user can follow links to navigate to related units. For the worm space, literature objects are linked to all genes described in the article. Figure 3 shows a link following. In the window labelled Search: "Traversal Set", the user has requested all objects linked to the selected literature object and the system displays one-line summaries of the set of matching gene objects. The user selects one gene “mec-3” and zooms into its description, which is displayed in the bottom window. This description shows that mutations in the gene indeed make the worm insensitive to touch.

The user now wants to see where the gene is located physically on the DNA of the worm and issues the “show physical map” command on the selected gene in the Traversal Set. Figure 4

Figure 2. Search



Search Control
Worm
Community System
Community Systems Laboratory
University of Arizona, Tucson
©1991 Arizona Board of Regents
cmdtool - /bin/csh
Search: "sensory"
100 items

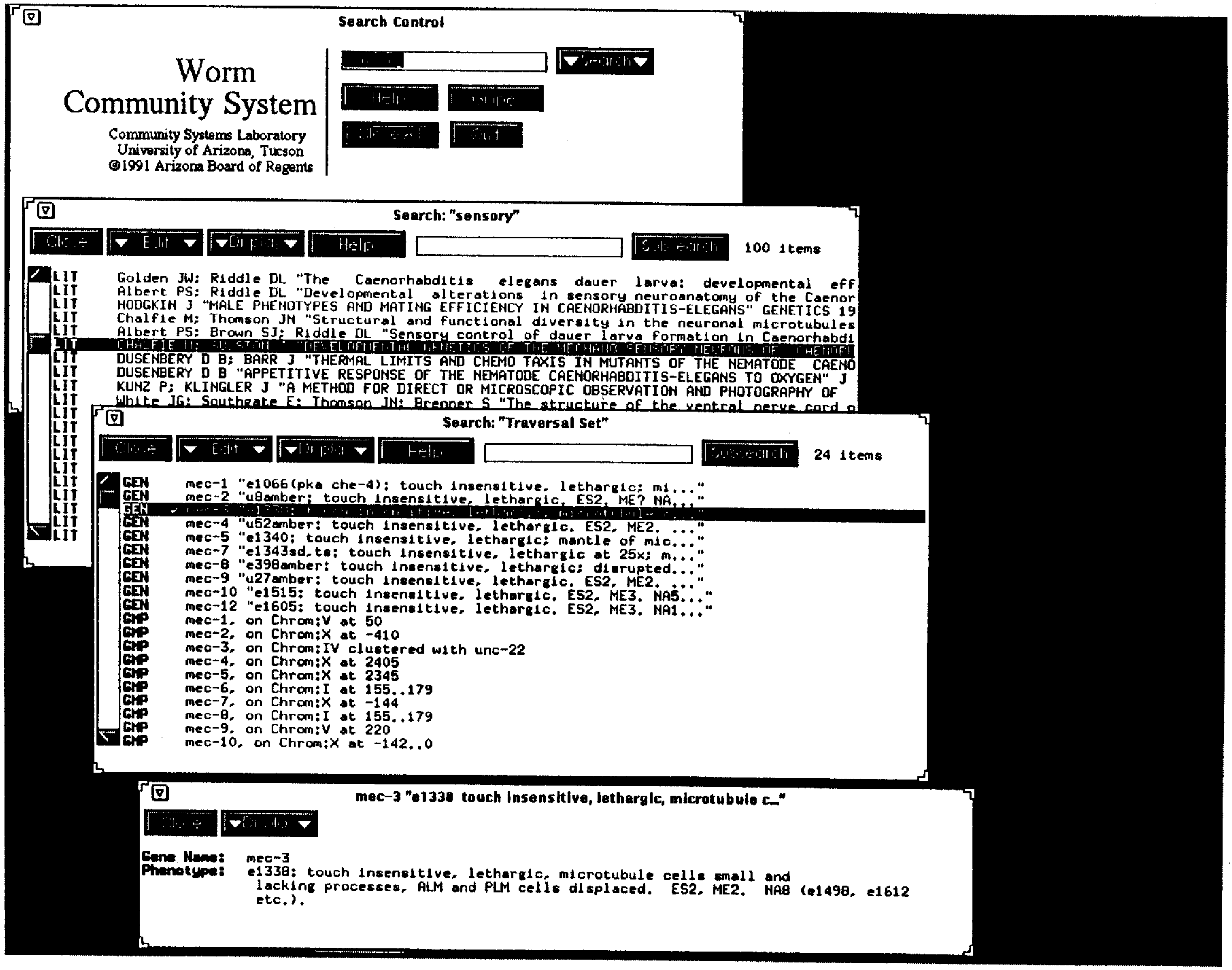
```
OBJECT SUMMARY:
        684 genes
       1008 genetic map entries
      11224 physical map entries
       2650 literature references
         30 DNA sequences
          7 annotations
        409 persons
       1403 strains
          3 external references
Memory used: 11460680
```

```
LIT   Golden JW; Riddle DL "The  Caenorhabditis  elegans  dauer  larva:  developmental  eff
LIT   Albert PS; Riddle DL "Developmental  alterations  in sensory neuroanatomy of the Caenor
LIT   HODGKIN J "MALE PHENOTYPES AND MATING EFFICIENCY IN CAENORHABDITIS-ELEGANS" GENETICS 19
LIT   Chalfie M; Thomson JN "Structural and functional diversity in the neuronal microtubules
LIT   Albert PS; Brown SJ; Riddle DL "Sensory control of dauer larva formation in Caenorhabdi
LIT   CHALFIE M; SULSTON J "DEVELOPMENTAL GENETICS OF THE MECHANO SENSORY NEURONS OF  CAENORH
LIT   DUSENBERY D B; BARR J "THERMAL LIMITS AND CHEMO TAXIS IN MUTANTS OF THE NEMATODE  CAENO
LIT   DUSENBERY D B "APPETITIVE RESPONSE OF THE NEMATODE CAENORHABDITIS-ELEGANS TO OXYGEN" J
LIT   KUNZ P; KLINGLER J "A METHOD FOR DIRECT OR MICROSCOPIC OBSERVATION AND PHOTOGRAPHY OF
LIT   White JG; Southgate E; Thomson JN; Brenner S "The structure of the ventral nerve cord o
LIT   Ware RW;Clark D;Crossland K;Russell RL "The nerve ring of the nematode C. elegans: Sens
LIT   Ward S; Thomson N; White JG; Brenner S "Electron microscopical reconstruction of the an
LIT   Ward S "The use of mutants to analyze the sensory nervous system of C. elegans." CGC Bo
LIT   Sulston, J; Dew, M; Brenner, S "Dopaminergic neurons in the nematode Caenorhabditis ele
LIT   Lewis, JA; Hodgkin, JA "Specific neuroanatomical changes in chemosensory mutants of the
LIT   EPSTEIN H F; ISACHSEN M M; SUDDLESON E A "KINETICS OF MOVEMENT OF NORMAL AND MUTANT NEM
LIT   Culotti, JG; Russell, RL "Osmotic avoidance defective mutants of the nematode  Caenorha
LIT   Croll NA "Sensory mechanisms in nematodes." Annual Review of Phytopathology 1977 15 :75
LIT   Singh RN;Strausfeld NJ (eds) "Neurobiology of Sensory Systems" Plenum Press 1989
LIT   HAZELBAUER G L "RECEPTORS AND RECOGNITION SERIES B VOL 5 TAXIS AND BEHAVIOR ELEMENTARY
```

CHALFIE M; SULSTON J "DEVELOPMENTAL GENETICS OF THE MECHANO SENSORY NEURONS OF CAENORHABDI

```
Reference:  502
Id:         11245967
Author:     CHALFIE M; SULSTON J
Title:      DEVELOPMENTAL GENETICS OF THE MECHANO SENSORY NEURONS OF
             CAENORHABDITIS-ELEGANS
Date:       1981
Journal:    DEV BIOL
Source:     DEV BIOL 1981 82 (2) :358-370
Abstract:   Touch sensitivity in the nematode C. elegans is mediated by a set of 6
             sensory neurons, the microtubule cells, of well-characterized anatomy and
             connectivity. The normal touch response is eliminated when these cells are
             killed by laser microsurgery. The identification of the microtubule cells
             as the mediators of touch sensitivity permits examination of the effects of
             mutations on the development and differentiation of these cells.
             Touch-insensitive mutants [42] were isolated. These fall into 13
             complementation groups. Mutations in 5 of the complementation groups have
             recognizable effects on the microtubule cells. These phenotypes include
             alterations of characteristic cellular ultrastructure, absence of neuronal
             process growth and the absence of the cell (either by alterations in the
             patterns of cell division that produce the cells or by degeneration or
             death of existing cells). Few genes appear to affect the growth and
             function of this class of cells and no others, possibly because the genes
             primarily affecting the microtubule cells are reaching their saturation
             level.
```

Figure 3. Navigation



Search Control

Worm
Community System

Community Systems Laboratory
University of Arizona, Tucson
©1991 Arizona Board of Regents

Search: "sensory"

100 items

```
LIT    Golden JW; Riddle DL "The  Caenorhabditis  elegans  dauer  larva:  developmental  eff
LIT    Albert PS; Riddle DL "Developmental  alterations  in sensory neuroanatomy of the Caenor
LIT    HODGKIN J "MALE PHENOTYPES AND MATING EFFICIENCY IN CAENORHABDITIS-ELEGANS" GENETICS 19
LIT    Chalfie M; Thomson JN "Structural and functional diversity in the neuronal microtubules
LIT    Albert PS; Brown SJ; Riddle DL "Sensory control of dauer larva formation in Caenorhabdi
LIT    CHALFIE M; SULSTON J "DEVELOPMENTAL GENETICS OF THE MECHANO SENSORY NEURONS OF  CAENOR
LIT    DUSENBERY D B; BARR J "THERMAL LIMITS AND CHEMO TAXIS IN MUTANTS OF THE NEMATODE  CAENO
LIT    DUSENBERY D B "APPETITIVE RESPONSE OF THE NEMATODE CAENORHABDITIS-ELEGANS TO OXYGEN" J
LIT    KUNZ P; KLINGLER J "A METHOD FOR DIRECT OR MICROSCOPIC OBSERVATION AND PHOTOGRAPHY OF
LIT    White JG; Southgate E; Thomson JN; Brenner S "The structure of the ventral nerve cord o
LIT
LIT
LIT
LIT
LIT
LIT
LIT
LIT
LIT
LIT
```

Search: "Traversal Set"

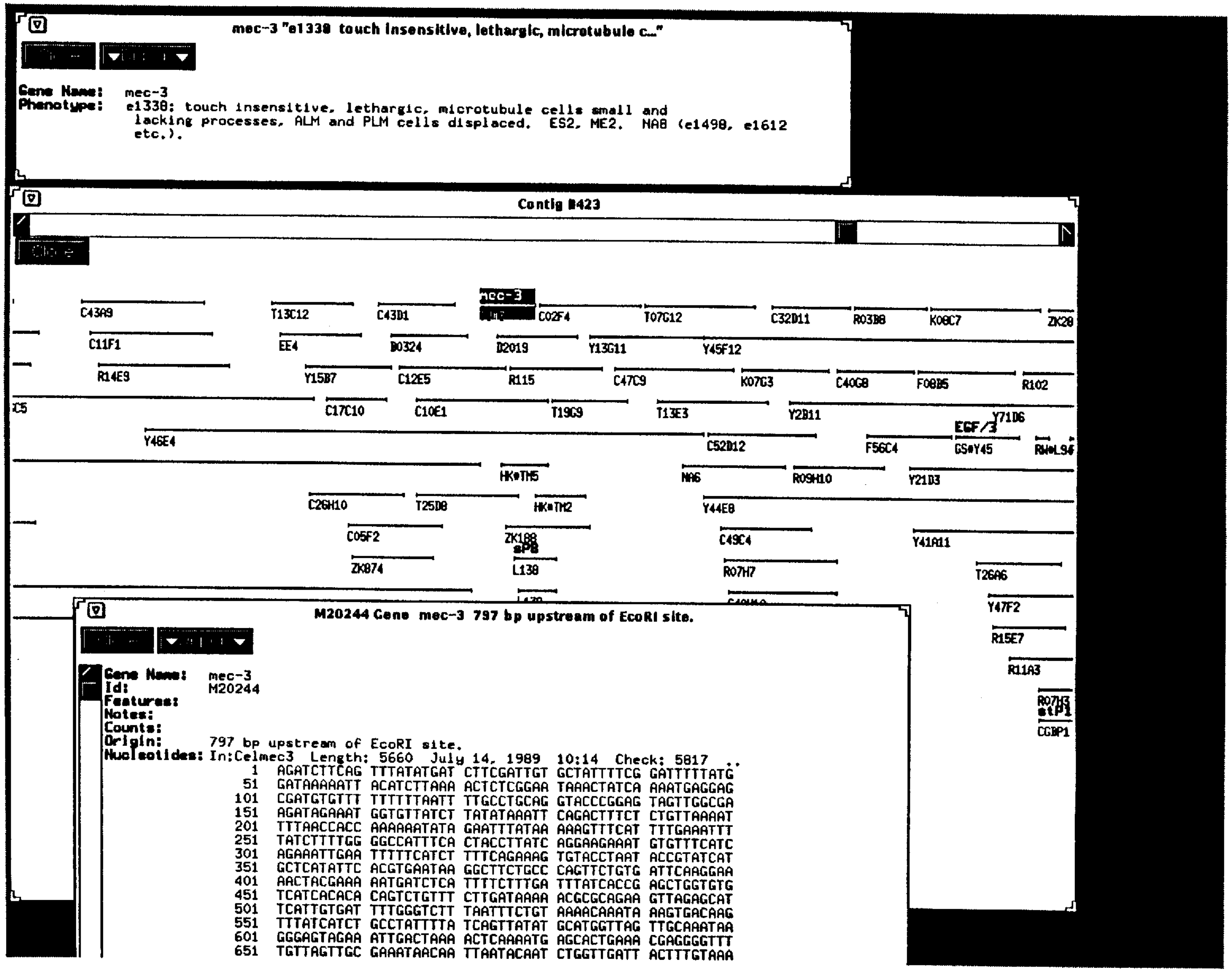
24 items

```
GEN    mec-1 "e1066(pka che-4): touch insensitive, lethargic; mi..."
GEN    mec-2 "u8amber; touch insensitive, lethargic, ES2, ME? NA..."
GEN  ✓ mec-3 "e1338: touch insensitive, lethargic, microtubule c..."
GEN    mec-4 "u52amber: touch insensitive, lethargic. ES2, ME2. ..."
GEN    mec-5 "e1340: touch insensitive, lethargic; mantle of mic..."
GEN    mec-7 "e1343sd,ts: touch insensitive, lethargic at 25x; m..."
GEN    mec-8 "e398amber: touch insensitive, lethargic; disrupted..."
GEN    mec-9 "u27amber: touch insensitive, lethargic. ES2, ME2. ..."
GEN    mec-10 "e1515: touch insensitive, lethargic. ES2, ME3. NA5..."
GEN    mec-12 "e1605: touch insensitive, lethargic. ES2, ME3. NA1..."
GMP    mec-1, on Chrom:V at 50
GMP    mec-2, on Chrom:X at -410
GMP    mec-3, on Chrom:IV clustered with unc-22
GMP    mec-4, on Chrom:X at 2405
GMP    mec-5, on Chrom:X at 2345
GMP    mec-6, on Chrom:I at 155..179
GMP    mec-7, on Chrom:X at -144
GMP    mec-8, on Chrom:I at 155..179
GMP    mec-9, on Chrom:V at 220
GMP    mec-10, on Chrom:X at -142..0
```

mec-3 "e1338 touch insensitive, lethargic, microtubule c..."

```
Gene Name:   mec-3
Phenotype:   e1338: touch insensitive, lethargic, microtubule cells small and
              lacking processes, ALM and PLM cells displaced.  ES2, ME2.  NA8 (e1498, e1612
              etc.).
```

Figure 4. Data Displays



mec-3 "e1338 touch insensitive, lethargic, microtubule c..."

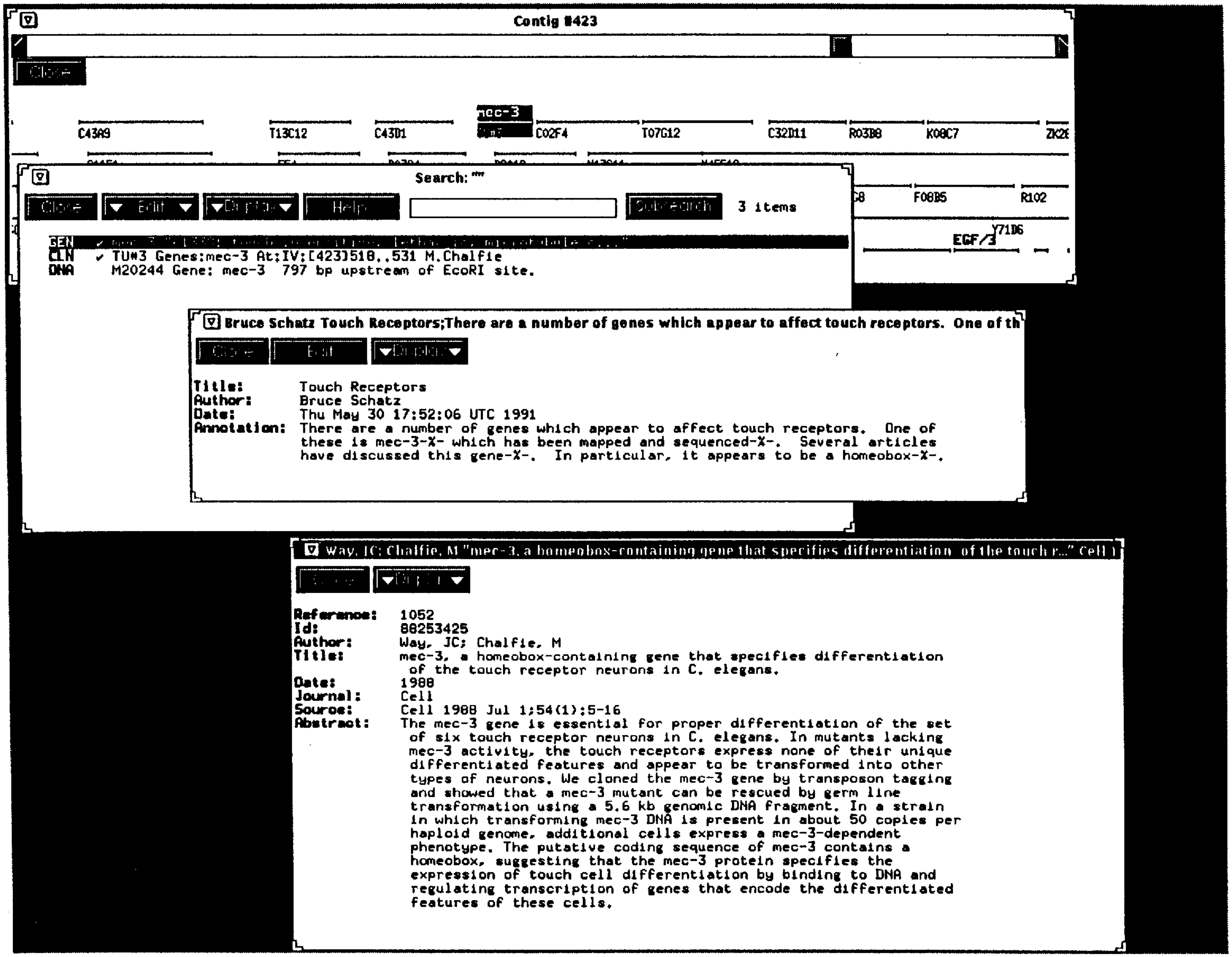
```
Gene Name:  mec-3
Phenotype:  e1338: touch insensitive, lethargic, microtubule cells small and
             lacking processes, ALM and PLM cells displaced.  ES2, ME2.  NA8 (e1498, e1612
             etc.).
```

Contig #423

M20244 Gene mec-3 797 bp upstream of EcoRI site.

```
Gene Name:   mec-3
Id:          M20244
Features:
Notes:
Counts:
Origin:      797 bp upstream of EcoRI site.
Nucleotides: In:Celmec3  Length: 5660  July 14, 1989  10:14  Check: 5817  ..
                  1   AGATCTTCAG TTTATATGAT CTTCGATTGT GCTATTTTCG GATTTTTATG
                 51   GATAAAAATT ACATCTTAAA ACTCTCGGAA TAAACTATCA AAATGAGGAG
                101   CGATGTGTTT TTTTTTAATT TTGCCTGCAG GTACCCGGAG TAGTTGGCGA
                151   AGATAGAAAT GGTGTTATCT TATATAAATT CAGACTTTCT CTGTTAAAAT
                201   TTTAACCACC AAAAAATATA GAATTTATAA AAAGTTTCAT TTTGAAATTT
                251   TATCTTTTGG GGCCATTTCA CTACCTTATC AGGAAGAAAT GTGTTTCATC
                301   AGAAATTGAA TTTTTCATCT TTTCAGAAAG TGTACCTAAT ACCGTATCAT
                351   GCTCATATTC ACGTGAATAA GGCTTCTGCC CAGTTCTGTG ATTCAAGGAA
                401   AACTACGAAA AATGATCTCA TTTTCTTTGA TTTATCACCG AGCTGGTGTG
                451   TCATCACACA CAGTCTGTTT CTTGATAAAA ACGCGCAGAA GTTAGAGCAT
                501   TCATTGTGAT TTTGGGTCTT TAATTTCTGT AAAACAAATA AAGTGACAAG
                551   TTTATCATCT GCCTATTTTA TCAGTTATAT GCATGGTTAG TTGCAAATAA
                601   GGGAGTAGAA ATTGACTAAA ACTCAAAATG AGCACTGAAA CGAGGGGTTT
                651   TGTTAGTTGC GAAATAACAA TTAATACAAT CTGGTTGATT ACTTTGTAAA
```

Figure 5. Annotation



Search: ""

3 items

GEN
CLN ✓ TU#3 Genes:mec-3 At;IV;[423]518,.531 M.Chalfie
DNA M20244 Gene: mec-3 797 bp upstream of EcoRI site.

Bruce Schatz Touch Receptors;There are a number of genes which appear to affect touch receptors. One of th

Title: Touch Receptors
Author: Bruce Schatz
Date: Thu May 30 17:52:06 UTC 1991
Annotation: There are a number of genes which appear to affect touch receptors. One of these is mec-3-X- which has been mapped and sequenced-X-. Several articles have discussed this gene-X-. In particular, it appears to be a homeobox-X-.

Way, JC; Chalfie, M "mec-3, a homeobox-containing gene that specifies differentiation of the touch r..." Cell)

Reference: 1052
Id: 88253425
Author: Way, JC; Chalfie, M
Title: mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans.
Date: 1988
Journal: Cell
Source: Cell 1988 Jul 1;54(1):5-16
Abstract: The mec-3 gene is essential for proper differentiation of the set of six touch receptor neurons in C. elegans. In mutants lacking mec-3 activity, the touch receptors express none of their unique differentiated features and appear to be transformed into other types of neurons. We cloned the mec-3 gene by transposon tagging and showed that a mec-3 mutant can be rescued by germ line transformation using a 5.6 kb genomic DNA fragment. In a strain in which transforming mec-3 DNA is present in about 50 copies per haploid genome, additional cells express a mec-3-dependent phenotype. The putative coding sequence of mec-3 contains a homeobox, suggesting that the mec-3 protein specifies the expression of touch cell differentiation by binding to DNA and regulating transcription of genes that encode the differentiated features of these cells.

displays a section of the physical map of the chromosome showing the known locations for a variety of cloned DNA fragments. The window labelled Contig #423 displays the region containing mec-3. This window is not just a line drawing but a live first-class graphical display of a composite object "physical map" which contains many sub-objects of type "clone". Thus the individual objects can be manipulated and interacted with. In this session, the user zooms into the clone on the map containing mec-3 and displays its DNA sequence in the window at the bottom of the screen. An external analysis program might now be invoked (but not yet in this prototype) to compare this gene controlling touch in worms to a library of the sequences for genes in all organisms to identify similar genes in humans.

Finally, the user checks whether other community members have added any informal information about the gene of interest. Figure 5 shows the result of zooming into the physical map entry for mec-3. Three IUs are displayed, corresponding to the gene, the clone, and the DNA sequence. The gene has a checkmark by its summary line indicating that an annotation is available. Issuing "show annotations" brings up the window discussing "Touch Receptors". This has a number of embedded references denoted by -%-. Zooming into the reference stating that mec-3 is a homeobox retrieves the paper in the bottom window. The user has thus made good use of the value-added informal knowledge to find a relevant specific paper concerning the gene of interest.

Future Plans

The current prototype system is written in GNU C++ and runs under the Unix™ operating system, typically on a Sun SPARCstation™. The external sources are maintained in files in text form and transformed into information unit objects when the system starts up. All the software is custom written, including the object manipulation. The current system runs all in memory and takes about 11 megabytes when loaded. This comprises some 18,000 objects, of which the bulk of the objects is physical map entries and the bulk of the size is literature items. Some of the test sites run the system on a local Sun workstation directly. Since the display uses X-windows, others run it remotely on an Apple Macintosh™ II running MacX™, with acceptable response over a local area network.

This first release is now in the labs of the initial users, on the order of 10 laboratories. Initially the goal is to recruit enough users to support a "fair test" of this kind of system. These users must be enthusiastic enough to use a preliminary system and influential enough to have their reactions be taken seriously by community members. In addition, geographical distribution is important since the experiment is a test of a nationwide electronic scientific community.

The feedback from this release is being used to design the second release. This version will be a distributed system with separate modules for database searching, for object manipulation, and for user interface. Computer scientists associated with the project will be experimenting with what caching and protocol technology is necessary to provide interactive retrieval across the NSFNET. The plan is for this version to be a fully featured system, which is propagated to a significant fraction of the worm community. Sociologists associated with the project will be investigating the usage to understand the effects on the community's communication patterns.

In the longer-term, as the system for supporting the worm community becomes functional, the software will be made available to other communities. The next electronic communities will probably be molecular biologists whose communities are also organized around experimental organisms, including the bacterium *E. coli,* the fruit fly *Drosophila,* the weed *Arabidopsis,* the slime-mold *Dictyostelium,* the alga *Chlamydomonas,* yeast, mouse, and man. The problem for many of these communities will be the lack of coverage of available data in electronic form, but much of the software should prove transferable.

As attempts to build more electronic community systems begin, more will become known about the characterization of communities. Many factors play a role in the suitability and usefulness of such a system to a community. These include: data (extent of coverage and vitalness of need), maturity (size and age of the community), competitiveness (readiness to share, pace and stakes), sophistication (computer literacy, tolerance for new technology), and many others. It may eventually prove possible to tailor an electronic community system to be more effective for a given set of community characteristics.

Towards Electronic Systems for Organizational Memory

An organization is in many respects similar to a community. It consists of people with common interests and shared values. So the knowledge in an organization is similar to the knowledge in a community. This knowledge might be termed **organizational memory**, i.e. the knowledge that enables the organization to continue functioning effectively. This is the permanent knowledge, as opposed to the transient knowledge generated during meetings. As with communities, organizational memory includes not just the tangible information in designs and memoranda, but also the intangible information in company procedures and values. The permanent memory in an organization that can enable it to outlive its founders is contained in both the company products and the company culture. Recording this memory and making it easily accessible electronically would clearly be of enormous use to organizations functioning.

The knowledge encoding and software system techniques developed for an electronic community system will likely be relevant to building electronic systems for organizational memory. An industrial organization, for example, has similar knowledge to the scientific community described in this paper. There is archival data, such as design specifications and product evaluations, and intermediate data, such as test results and market surveys. There is formal literature, such as technical memoranda and progress reports, and informal literature, such as design notes and meeting minutes. An organization also has similar needs to manipulate this knowledge. There is a need to browse the knowledge, to filter out selections relevant to the problem, then to share annotations of these selections with other members of the organization.

The sociology of an organization tends to be somewhat more rigid than a scientific community. Thus, although the faster pace will likely require the fast distribution of the knowledge, greater control over its dissemination is likely to be important. A finer granularity of specification and tracking of the editorial and privacy controls is likely to be necessary. For example, a design must be approved by a specified series of people and product information is available only on a need-to-know basis. There may be other controls to support the policies and procedures, and to enforce the flow of information within the organizational structure. Finally, there may be other types of functionality necessary to capture some degree of the company culture. For example, there may be precise style and content constraints on hyperdocuments in the organization's information space.

Large-scale experiments in real organizations will be necessary to access the value of electronic community systems for supporting organizational memory. Preliminary evidence indicates that this technology will be valuable to the scientific community, so great potential exists for its value to the business community as well.

Acknowledgements

This paper is partially based upon a grant proposal written with Samuel Ward, who introduced me to the worm community and has supported me at the University of Arizona. Many people worked on the implementation of the prototype, including Terry Friedman, Ed Grossman, Scott Hudson, Kevin Powell, and Andrey Yeatts. Mark Edgley and John Sulston provided much of the worm data. John Wooley provided helpful advice. Planning for this project was supported by a Small Grant for Exploratory Research DIR-9003540 from the National Science Foundation. The project itself is now supported by NSF IRI-9015407 through the Special Initiative in Coordination Theory and Collaboration Technology.

References

1. D. Crane, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, University of Chicago Press, 1972.
2. J. Licklider, R. Taylor, and E. Herbert, "The Computer as a Communication Device", *Science and Technology*, April, 1968, pp. 21-31.
3. B. Reid, "The USENET Cookbook -- An Experiment in Electronic Publishing", *Electronic Publishing* , vol 1, 1988, pp 55-76..
4. J. Quarterman, *The Matrix: Computer Networks and Conferencing Systems Worldwide*, Digital Press / Prentice-Hall, 1989.
5. J. Licklider, *Libraries of the Future*, MIT Press, 1965.
6. UNESCO, *UNISIST: Study Report of the Feasibility of a World Science Information System*, United Nations Educational, Scientific, and Cultural Organization, Paris, 1971.
7. National Science Foundation, "Towards a National Collaboratory", Report of an Invitational Workshop, Mar, 1989, Directorate for Computer and Information Science and Engineering.
8. L. Roberts, "The Worm Project", *Science* , vol. 248, June 15, 1990, pp. 1310-1313.
9. W. Wood (ed), *The Nematode Caenorhabditis elegans*, Cold Spring Harbor Laboratory Press, N.Y, 1988.
10. A. Coulson, J. Sulston, S. Brenner, and J. Karn, "Towards a physical map of the genome of the nematode *Caenorhabditis elegans*", *Proc National Academy Sci*, Vol. 83, 1986, pp. 7821-7825.
11. D. Engelbart and W. English, "A Research Center for Augmenting the Human Intellect", *Proc AFIPS Fall Joint Computer Conf*, vol 33, 1968, pp. 395-410.
12. D. Engelbart, "Collaboration Support Provisions in AUGMENT", *Proc AFIPS Office Automation Conf*, Los Angeles, Feb, 1984, pp. 51-58.
13. D. Engelbart, "Coordinated Information Services for a Discipline- or Mission-Oriented Community", R. Grimsdale(ed), *Computer Communications Networks*, NATO Series vol 4, Noordhoff, 1975, pp. 89-99.
14. B. Schatz, *Telesophy*, Technical Memorandum TM-ARH-002487, Bell Communications Research, Aug 1985, 74pp.
15. B. Schatz , "Telesophy: A System for Manipulating the Knowledge of a Community", *Proc IEEE Globecom '87*, Tokyo, Nov, 1987, pp. 1181-1186.
16. B. Schatz , "A Prototype Information Environment", *Proc 2nd IEEE Workshop Workstation Operating Systems*, Pacific Grove, CA, Sep, 1989, p. 118-124.
17. B. Schatz, "Searching in a Hyperlibrary", *Proc 5th IEEE Int Conf on Data Engineering*, Los Angeles, Feb, 1989, pp. 188-197.

Author's Biography

BRUCE R. SCHATZ is Director of the Community Systems Laboratory, Arizona Research Laboratories, University of Arizona, and an associated faculty in the Department of Management Information Systems. He spent 10 years in industrial research laboratories at Bell Labs and Bellcore, as a lead architect on a variety of research and development projects in information and communications systems. He came to the University of Arizona in 1989 to found a laboratory to construct novel science information systems and propagate them to large-scale communities. Its funding includes a grant, on which he was Principal Investigator, that was one of the winners of the NSF National Collaboratory competition. He earned an MS in artificial intelligence from MIT and a PhD in computer science from Arizona. His research interests include community systems, electronic libraries, computational biology, and network applications.

Address:
Dr. Bruce R. Schatz
Community Systems Laboratory
Life Sciences South
University of Arizona
Tucson, AZ 85721 USA

schatz@cs.arizona.edu
(602) 621-9174
Fax (602) 621-3709