# Managing Cognitive Overload in the Flora of North America Project

Kay L. Tomlinson
Center for Botanical Informatics
Missouri Botanical Garden
tomlinso@mobot.org

J. Alfredo Sánchez
Department of Computer Systems Engineering
Universidad de las Américas-Puebla
alfredo@cca.pue.udlap.mx

Mark A. Spasser
Center for Botanical Informatics
Missouri Botanical Garden
mspasser@cbi.mobot.org

John L. Schnase
Center for Botanical Informatics
Missouri Botanical Garden
schnase@mobot.org

## Abstract

*The Flora of North America (FNA) is a large-scale collaboration involving over eight hundred scientists working together to create a 30-volume compendium of all naturally occurring plants in North America. The size and complexity of the project result in significant cognitive overload which comprises not just information or data but also work practices and activities. We view FNA as a distributed cognitive system where information, people, artifacts, processes, and expertise are functionally distributed among members of the community. Our Web-based work environment, the FNA Internet Information Service (FNA IIS), implements role-based views derived from the socially constructed roles that exist within FNA.. We propose that the concepts of role-based views and boundary object reification represent at least a partial answer to the problems of complexity management and information overload in large-scale collaborative systems such as FNA.*

## 1. Introduction

With more than 800 participating scientists, the Flora of North America project (FNA) comprises one of the largest collaborations currently funded by the National Science Foundation. Our current mandate is to become an entirely electronic flora project. In order to do so, we are developing and testing a series of Internet-based tools that we believe will radically alter the way the project works.

We view FNA as a distributed cognitive system [12]. The FNA idea space consists of nomenclature, descriptions, distribution, ethnobotany, illustrations, etc., of more than 20,000 plant species distributed over more than half a continent. It is clearly beyond the grasp of any single individual, and cognitive tasks are distributed among members of the group. Information overload presents a significant problem to a project of this size and complexity. But the cognitive power of the community and its ability to process this large volume of information are greatly enhanced if technologies that effectively mediate cognitive interactions can be introduced into the system.

The complexity of FNA lies not just in the amount of information but also in the work processes that are required to produce, manage, and coordinate this information. Thus, in this paper we refer to reducing or redistributing cognitive load (implying information plus process) over the entire project, not just information overload for the individual contributor. We believe that the key to managing the complexity of the project and reducing both cognitive and information overload is to deliver relevant information and functionality in a format tailored to the needs of individual users [21, 23]. The FNA Internet Information Service (FNA IIS) is being developed to provide this personalized information delivery [19].

In this paper we provide an overview of the FNA project, describe the FNA IIS, and discuss our early impressions of its effectiveness in dealing with cognitive and information overload.

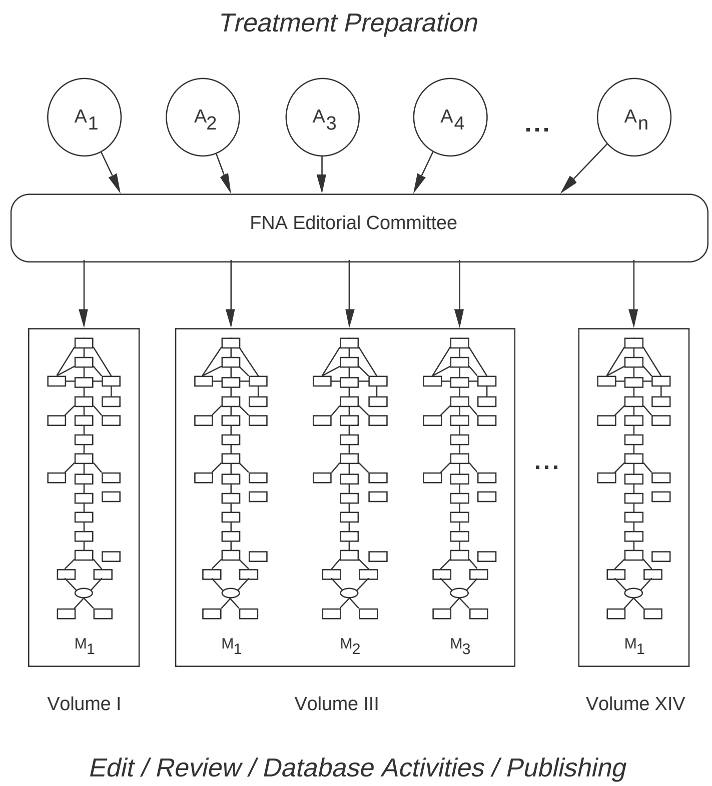


**Figure 1. Overview of the major work processes in the Flora of North America Project. Authors ($A_1$ - $A_n$) prepare treatments and submit manuscripts ($M_1$ - $M_n$ for each volume) that eventually make their way into printed volumes. Arrows depict major paths of information flow between the treatment preparation and database and publishing components of the project. (Adapted from Schnase et al., 1997)**

## 2. The Flora of North America

### 2.1 Background

The energy flow and balance of all life on earth depends on plants. It is not surprising, therefore, that information about plants is important to a wide range of disciplines, ranging from agriculture and medicine to government and industry. Botanical information tends to be idiosyncratic, scattered, and difficult to obtain, residing in rare book collections, private databases or card files, collections of pressed plant specimens preserved in herbaria large and small throughout the world, and of course in the minds of botanical specialists.

The Flora of North America project was begun in an attempt to gather information about the plants of North America north of Mexico. The original concept was that of a traditional, paper-based publishing effort to coordinate the activities of taxonomic specialists in many different groups of plants, create a 30-volume printed reference work containing the compendium of information, and also to create an electronic repository for that information [16]. In the ten years since the project's inception, many changes have occurred. The science of database publishing has grown up, the electronic world is vastly different, and the project has grown faster and larger than its creators expected.

In order to understand the significance of these changes, one must understand how traditional floristic work is done. Traditional flora projects, which are listings and descriptions of the plants occurring in a given area, have tended to be smaller in scope and managed by a relatively small group of contributors and editors. For example, the Flora of the Great Plains [10], the project after which FNA was originally modeled, was prepared by about 30 authors and editors, and covers all or part of only thirteen states. Part of the difficulty faced by the Flora of North America is that it has attempted to scale this traditional model of work. FNA now comprises more than 800 contributors distributed geographically throughout the world, and it deals with 20,000+ species of plants. Its geographic scope includes the continental United States, including the Florida Keys and Aleutian Islands, Canada, Greenland, and St. Pierre and Miquelon islands—an area of approximately 21,461,100 $km^2$. We have discovered that traditional project coordination, manuscript preparation, and book publishing methods simply do not scale to such a task [19].

## 2.2 The FNA Publishing Process

Figure 1 provides an overview of the FNA publishing process. Specialists are invited by the FNA Editorial Committee, the project's 35-member governing body, to prepare treatments describing various taxa; collections of taxonomic treatments, including distribution maps and illustrations, are then edited, reviewed, and assembled into printed volumes. Authors prepare "treatments" that provide data for the FNA database, electronic publications, and the printed flora. Treatments generally focus on species within a single genus, and each is prepared and reviewed by specialists. Authors study plants in the field, examine herbarium specimens, and review published reports of previous work.

Overall project activities are coordinated at Missouri Botanical Garden in the FNA Organizational Center. A total of five distinct review processes–-ranging from review of scientific content and stylistic reviews to evaluations of a taxon's conservation status–-are carried out on each treatment once it is submitted. Figure 2 shows the major steps involved in processing an FNA treatment once it arrives at the Organizational Center. These steps represent functional activities that are, for the most part, performed by geographically distributed referees and editors. At present, a mix of electronic and paper documents are used throughout; ongoing efforts for developing a common storage substrate and uniform, web-based user interfaces will allow authors to enter treatment data from remote locations into a relational database so it can be accessed online using diverse query options or extracted into various formats for electronic or paper distribution.

FNA's size and complexity result in a very large amount of information that must processed and managed by the various contributors to the project, especially the project’s editors. Processes include gathering and sorting through information; preparing treatments; and tracking deadlines, treatment status, author and reviewer contact information, reviewer's and technical editor's comments, etc. Editors are responsible for ensuring that the manuscripts in their assigned group are received and moved along in a timely manner, and that each event is documented and communicated to the appropriate participants. Much of the cognitive overload that exists within FNA occurs as a result of the many events associated with this edit and review process.

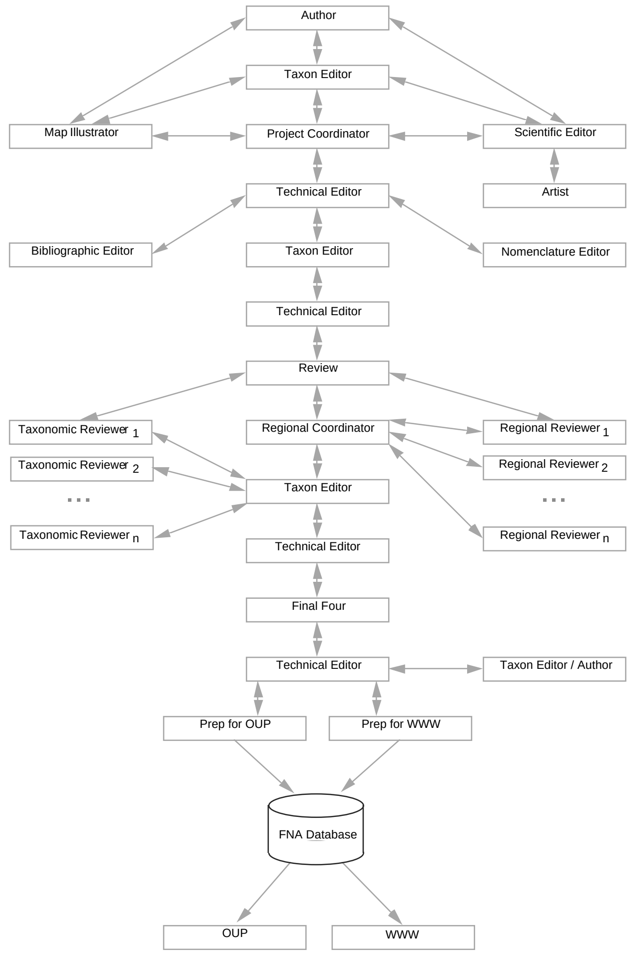


**Figure 2. Simplified manuscript flowchart for the FNA Project showing steps involved in the edit and review process for each manuscript. (Adapted from Schnase et al., 1997)**

## 2.3 Cognitive Overload in FNA

There are approximately 100 discrete events associated with the edit and review of a single manuscript, and each event must be tracked in some manner and coordinated with other activities. As many as three hundred manuscripts can go into a single printed volume, and it is necessary to coordinate progress across all volumes. This means that FNA must effectively manage a vast number of intra-document, inter-document/intra-volume, and inter-volume relationships among activities.

One way to understand the magnitude of the problem is to consider the number of dependencies associated with

FNA work practices. For a printed volume with 128 treatments, there are more than six million potential pairwise dependency relationships; for the entire FNA project, there are on the order of four billion [14, 19]. Clearly, the total cognitive load of the project is beyond the scope of any human being or any traditional project-management system, and even for an editor with only a few manuscripts to consider, the task of coordination is daunting. Cognitive overload is a serious concern.

## 3. FNA as a Distributed Cognitive System

The foregoing description shows that cognitive load in FNA comprises not just information or data but also work practices and activities. Our goal is to produce a system that will facilitate both cooperative and productive interactions and the interchange of ideas among different parts of the project while reducing or redistributing cognitive load. It is helpful, therefore, to view FNA as a *distributed cognitive system*.

### 3.1 Background

The practice of plant systematics has, for the past several hundred years, been primarily an individual activity. A specialist, usually working alone, develops ideas about the relationship of a particular group of plants to each other and the group's placement within a larger taxonomic scheme. Cognition, in this situation, occurs primarily at the level of the individual.

In contrast, the Flora of North America consists not just of information or data, but of people, expertise, work practices, artifacts, and processes—the generative and synthetic activities of the basic science and the processes we call edit and review. Cognitive artifacts, such as tools, people, work processes, etc., are distributed over a wide geographical range, and cognitive tasks are distributed among members of the community. Each person serves as a knowledge repository for their area of expertise, and no single individual can fathom the entirety of FNA. By accumulating information in these repositories, and by facilitating communication among them, the network increases the effective size of the community and its cognitive capacity [12]. We refer to FNA as a *cognitive ecosystem*, which conveys the dynamic, interactional, and evolving nature of the project as a whole.

Any community can be viewed as a cognitive system, some more effective than others, depending to a great extent on patterns of social organization such as group size, the patterns of communication between group members, and so on [12, 18]. Plant taxonomy, as traditionally performed, has created a relatively *inefficient* cognitive system because the *effective size* of the community is small and communication within the group is limited: for the most part, only small subsets within the community share information on a regular basis.

The FNA project, as originally conceived, increased the botanical community's effectiveness by coordinating group interactions and promoting communication among community members. FNA's adoption of Internet technology has further increased the communication potential and the likelihood, speed, amount, and quality of interaction within the community. In addition, the Internet expands even further the effective community size by permitting interaction not just among taxonomists but among taxonomists and other groups, such as computer scientists, foresters, agriculturists, government agencies, etc., that are traditionally separate from the network of botanical scientists.

### 3.2 Boundary Objects

What, then, are the mechanisms whereby subunits within a community communicate in order to become a networked, cognitive system? We find it helpful to think of this assembly process from the perspective of boundary objects [20, 22]. Boundary objects are "...those scientific objects which both inhabit several intersecting social worlds ... *and* satisfy the informational requirements of each of them" [22, p. 393] (emphasis in the original). They are interpretationally flexible information entities that allow members or stakeholders of different groups to simultaneously share the same information, while enabling them to maintain quite different perspectives on, or views of, that information.

The idea of boundary objects fits nicely within the notion of distributed cognition; indeed, boundary objects must exist within a distributed cognitive system in order for communication within the system to occur [11]. Boundary objects exist in areas where social roles or territories overlap. Distributed cognition presumes the division of cognitive tasks; therefore, cognitive roles and "territories" of interest will differ. Boundary objects are indispensable to such a system because they provide the "common coin" or means by which different cognitive units within the system communicate and share information.

## 4. Managing Cognitive Overload: Role-based Views in the Flora of North America Internet Information Service

We set out to design a Web-based computing environment to reduce cognitive overload within FNA. We call the system the FNA Internet Information Service

(FNA IIS). Since no individual can comprehend the entire FNA ecosystem, it follows that there is no single *viewpoint* from which it can be conceptualized (cf. the poverty of the "view from nowhere" in Suchman [24] ). This leads us to the idea of managing information overload by means of dynamically constructed activity-and-information spaces, or *role-based views*. Role-based views are derived from the socially constructed roles that already exist within the project and consist of boundary objects that reify the various information requirements, tasks, and responsibilities required in order for the FNA cognitive system to work.

As shown in Table 1, our initial set of boundary objects includes roles such as taxon (lead) editor, editor, reviewer, and author. Information artifacts include, among other things, lists of authors, editors, and taxa; data structures that map the assignments of authors to taxa, editors to taxa, etc.; taxonomic treatments, which include textual descriptions, keys, maps, and illustrations of the plants being described; and other material such as informal notes, bibliographies, pointers to relevant resources, etc. (A complete specification for the FNA IIS will be presented in future publications.)

Activities can be applied to various roles and information artifacts or events defined within the system. Taxon editors, for example, can be given the responsibility, and system functionality, to add a new author to the system or make an author/treatment assignment. Manuscripts being developed by each author can be viewed by the author; the editor responsible for that author and their treatments can view relevant information on the author's activities and associated treatments; editors view and manage the treatments assigned to them; and the taxon editor can view the entire enterprise.

The various role-based views are delivered through dynamically constructed, personalized home pages that can be accessed using any current Web browser. A profile database maintains state information, and activities are implemented by a library of CGI scripts and simple form interfaces. Personalized pages are built in response to user logins.

**Table 1. Initial set of boundary objects currently managed within the FNA IIS.**

| Roles | Activities | | Information Artifacts | |
|---|---|---|---|---|
| Taxon Editor | Add | - author | List | - authors |
| Editor | | - editor | | - editors |
| Reviewer | | - reviewer | | - reviewers |
| Author | Delete | - author | | - taxa |
| | | - editor | Map | - author/taxa |
| | | - reviewer | | - editor/taxa |
| | View | - list | | - reviewer/taxa |
| | | - map | Treatment | - text |
| | Assign | - author/taxon | | - key |
| | | - editor/taxon | | - map |
| | | - reviewer/taxon | | - illustration |
| | Unassign | - author/taxon | | |
| | | - editor/taxon | | |
| | | - reviewer/taxon | | |
| | Set | - due date | | |
| | | - notify event | | |
| | Unset | - due date | | |
| | | - notify event | | |
| | Write | - treatment | | |
| | Edit | - treatment | | |
| | Review | - treatment | | |

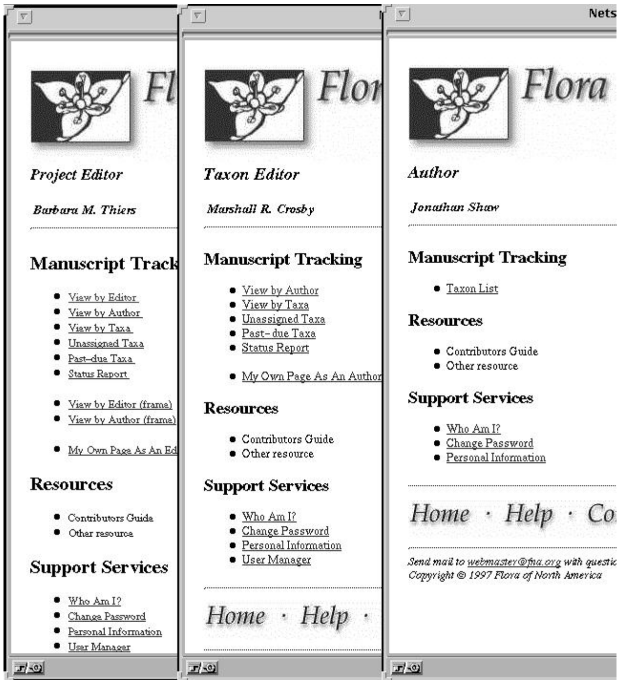


**Figure 3. Role-based views for three different users.**

Figure 3 shows side-by-side comparisons of the views presented to three different users. The rightmost view is that available to a treatment author, who can keep track of taxa for which he is expected to submit manuscripts. A taxon editor's view (Figure 1, center) allows the user to maintain an up-to-date view of the taxa under his responsibility via various mouse-selectable actions. Finally, the leftmost view illustrates the view of a project editor. In addition to the options available to taxon editors, the project editor can also initiate a number of actions to keep track of work related to the taxon editors for which she is responsible. As shown in the figure, a project participant with appropriate privileges may choose to change roles and see different views within the system. In the example, both the Project Editor and the Taxon Editor are able to change their current view to that of a treatment author. E-mail communication among participants can be initiated directly from any view.

Figure 4 illustrates a view by author as obtained by the Project Editor, whereas Figure 5 shows a taxonomically organized view of the same data. Overdue items are brought to the user's attention by highlighting them using different colors. Icons are used to refer to the status of each treatment component (textual morphologic description, taxonomic key, illustrations and distribution maps). At any given time, multiple users can be accessing the FNA information space and performing different actions. The current prototype allows authors, for example, to upload documents that are part of taxonomic

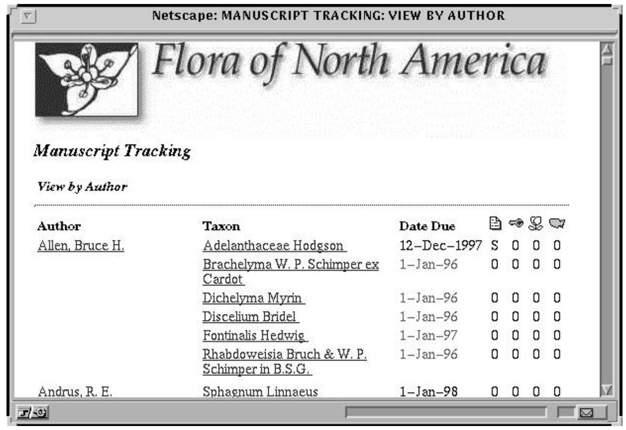


**Figure 4. View by author**

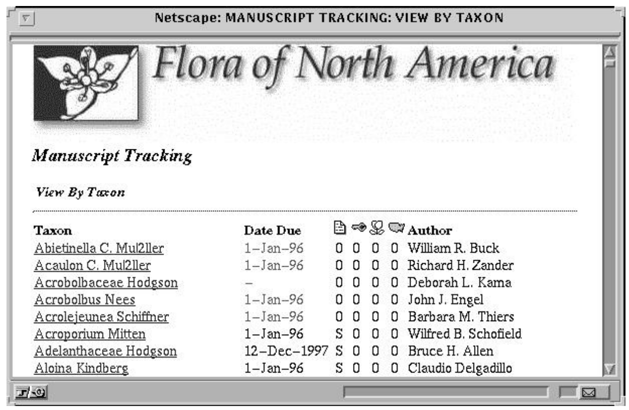


**Figure 5. Taxonomic view**

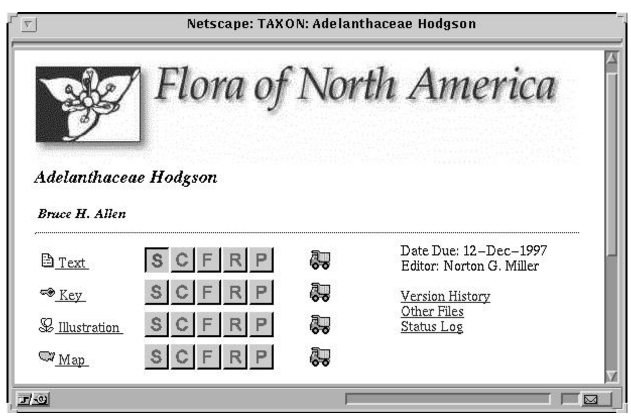


**Figure 6. Some operations available to authors and editors.**

treatments from remote sites using the interface shown in Figure 6.

## 5. Discussion

FNA IIS is a distributed system, one that interrelates minds; behaviors or activities; artifacts, tools, and objects; and arrangements of socio-physical space as constitutive elements [13]. In the traditional publishing mode, individual botanists participating in FNA bear cognitive load in an overlapping and redundant manner, resulting both individual overload and systemic inefficiencies.

FNA IIS accomplishes three things which are significant in terms of distributed cognition and reducing information overload:

First, FNA IIS *lessens individual cognitive load by delegating information and task organizational duties to an external representational structure*, i.e., the interface; it does the organization "behind the scenes" and presents to the user external representations that do two things: "(1) they organize information so that it is in the right place at the right time; and (2) they encode information in a form that is more explicit and thus easier to use" [8, p. 273].

This precomputation of information-task organization essentially removes invariant aspects of the task from the manipulations required of users. Simply stated, there are many different information-worlds within FNA IIS, all of which map onto the same information space; which information-world a particular user needs to see depends on his or her role. This mapping function is enormously powerful—the computational constraints and requirements of the user's interaction with the information space have been built into the structure of the interface-as-tool [12]. Rather than simply amplifying the cognitive abilities of individual botanists, information structurizing tools such as FNA IIS transform the task the botanist confronts by representing it in such a way that the user can readily see exactly how to perform it. "What you see is what you need to do."

The second important accomplishment of FNA IIS is that it *enhances system performance by enabling massively parallel simultaneous use of a large information space* via mapping of permissible views plus suites of operations onto individual and group knowledge resources and capabilities. Hutchins reminds us that distributed cognitive systems achieve their computational or information-processing power, which can be construed as their ability to effectively redistribute cognitive load, by superimposing several kinds of representations, or representational structures, on a single framework. In our case, the framework is a single, very large information space.

Finally, FNA IIS *objectifies the concept of boundary objects by instantiating these role-based views or boundary objects* over the information space. It structures not just the information but also the tasks: it simultaneously affords and constrains opportunities for the user to interact with the information. Concentrating information organizational complexity in boundary objects is a particularly efficient way to facilitate simultaneous and tailored access to, and use of, information over a single, very large information space.

## 6. Related Work

The idea of views as a means of constraining perspectives on a large body of information is not new. The concept appears in the database [2, 3] and hypertext [1, 4, 5, 6] literature. What particularly appeals to us is the traditional sense of view combined with activities, which is a concept that appears less frequently. Our work is influenced primarily by closely related research in the areas of computer-supported cooperative work and social informatics.

### 6.1 Computer-Supported Cooperative Work (CSCW)

Development of the FNA IIS has been informed by related work in the CSCW arena. Systems such as Oval, gIBIS, Coordinator, and Information Lens provide varying capacities to create personalized work spaces for collaborative activities [7]. Oval in particular has shown the effectiveness of "radically tailorable" environments for building cooperative work systems [15]. These systems, however, are monolithic, whereas with FNA IIS, we are working to provide similar capabilities within an open, Web-based framework. We also place special emphasis on FNA IIS as an information service—where the applications used by its participants are in a sense irrelevant (any type of document or program can be shared) and where the natural social protocols of the project provide some constraint over the service's tailorability.

### 6.2 Social Informatics

The hybrid nature of role-based views, as implemented in the FNA IIS, is captured nicely by Fujimura's [9] concept of standardized packages. She coined the term to describe a theory plus methods of assemblage that was used to further the development of proto-oncogene research. In the case of FNA IIS, the theory is that of distributed cognition; the methods or tools are the web pages that are built for each role.

Another perspective on the FNA IIS's role-based views is Paepcke's [17] information compound, which unites pieces of information into a collection that can either solve a problem (e.g., reduce information overload) or be used as an end product for a task (e.g., a task-relevant database interactional view). Paepcke [17] states that fully-developed information compounds help users specify task- and role-based selections of information from very large information spaces [23]. Specifically, FNA IIS binds different views to different compounds, depending on whether the user is a project editor, reviewer, editor, and/or author. Crucially, from an information overload perspective, information compounds function as powerful filters both by limiting and structuring information elements that are retrieved and actions that are permitted, and by presenting them in an intuitively meaningful way.

## 7. Summary

The existence of large, distributed cognitive systems like the Flora of North America, particularly when viewed from the standpoint of boundary objects, raises interesting questions about the changing nature of scientific endeavor. We propose that the concepts of role-based views and boundary object reification represent at least a partial answer to the problems of complexity management and scalability in these collaborative systems. The combination offers the advantage of distributing task accomplishment in a socio-material and socio-technical system: individual users can be assigned tasks that they can do best, while the software tools and processual orderings accomplish steps less easily performed by individuals. This approach appears to (re)distribute total system cognitive load and reduce the load borne by any single user-botanist. And, like habitats in a natural ecosystem, role-based views can evolve over time as the project itself changes.


### Acknowledgements

This work is supported in part by grants from the Andrew W. Mellon Foundation and the National Science Foundation (DEB-9505383).


### References


[1] Association for Computing Machinery. Special Issue on Hypertext. *Communications of the ACM*, Vol. 31, No. 7. 1988.

[2] Association for Computing Machinery. Special Issue on Heterogeneous Databases. *ACM Computing Surveys*, Vol. 22, No. 3. 1990.

[3] Association for Computing Machinery. Special Issue on Next-Generation Database Systems. *Communications of the ACM*, Vol. 34, No. 10. 1991a.

[4] Association for Computing Machinery. Special Issues on Collaborative Computing. *Communications of the ACM*, Vol. 34, No. 12. 1991b.

[5] Association for Computing Machinery. Special Issue on the Dexter Hypertext Reference Model. *Communications of the ACM*, Vol. 37, No. 2. 1994.

[6] Association for Computing Machinery. Special Issue on Designing Hypermedia Applications. *Communications of the ACM*, Vol. 38, No. 8. 1995.

[7] Baecker, R. M. *Readings in Groupware and Computer-supported Cooperative Work: assisting human-human collaboration*. San Francisco: Morgan Kaufman. 1993.

[8] Flor, N. V., & Hutchins, E. L. "Analyzing distributed cognition in software teams: a case study of team programming during perfective software maintenance." In R. M. Baecker (ed.), *Readings in groupware and computer-supported cooperative work: assisting human-human collaboration.* San Francisco: Morgan Kaufmann Publishers, Inc. 1993, pp. 272-286.

[9] Fujimura, J. H. "Crafting science: standardized packages, boundary objects, and 'translation.' " In: A. Pickering (ed.), *Science as practice and culture.* Chicago: University of Chicago Press. 1992, pp. 168-214.

[10] Great Plains Flora Association. *Flora of the Great Plains*. Lawrence, Kansas: University Press of Kansas. 1986.

[11] Haraway, D. "Situated Knowledges: the science question in feminism and the privilege of partial perspective." In: D. Haraway, *Simians, cyborgs, and women*. New York: Routledge. 1991, pp. 183-201.

[12] Hutchins, E. L. *Cognition in the wild*. Cambridge, MA: MIT Press. 1996.

[13] Keller, J. D. "Review symposium: Cognition in the Wild. E. Hutchins." *Mind, Culture, and Activity*, 3(1). 1996, pp. 46-50.

[14] Malone, T. W., and Crowston, K. "The interdisciplinary study of coordination." *ACM Computing Surveys*, 26. 1994, pp. 87-119.

[15] Malone, T. W., Lai, K-Y., and Fry, C. "Experiments with Oval: a radically tailorable tool for cooperative work." Available: http://ccs.mit.edu/ccswp181/index.html. 1994.

[16] Morin, N. R., Whetstone, R. D., Wilken, D., and Tomlinson, K. L. "Floristics for the 21st Century." *Monographs in Systematic Botany from the Missouri Botanical Garden*. Vol. 28. 1989.

[17] Paepcke, A. “Information needs in technical work settings and their implications for the design of computer tools.” *Computer Supported Cooperative Work*. 1996.

[18] Roberts, J. “The self-management of cultures.” In W. Goodenough (ed.), *Explorations in Cultural Anthropology: Essays in honor of George Peter Murdock*. McGraw-Hill. 1964.

[19] Schnase, J. L., Kama, D. L, Tomlinson, K. L., Sánchez, J. A., Cunnius, E. L., and Morin, N. R. “The Flora of North America digital library: A case study in biodiversity database publishing.” *Journal of Network and Computer Applications*. Vol. 20, 87-103. 1997.

[20] Star, S. L. “The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving.” In M. Huhs, & L. Gasser (eds.), *Readings in distributed artificial intelligence 3*. Menlo Park, CA: Morgan Kaufmann. 1989, pp. 37-54.

[21] Star, S. L., ed. *The Cultures of Computing*. Cambridge, MA: Blackwell Publishers. 1995.

[22] Star, S. L., & Griesemer, J. R. “Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Musuem of Vertebrate Zoology, 1907-39.” *Social Studies of Science*, 19. 1989, pp. 387-420.

[23] Star, S. L., and Ruhleder, K. “Steps towards an ecology of infrastructure: design and access for large information spaces.” *Information Systems Review*, 7(1). 1996, pp. 111-134.

[24] Suchman, L. “Working relations of technology production and use.” *Computer Supported Cooperative Work (CSCW)*, 2. 1994, pp. 21-39.