

# genSpace: Exploring Social Networking Metaphors for Knowledge Sharing and Scientific Collaborative Work

Christian Murphy, Swapneel Sheth, Gail Kaiser, Lauren Wilcox  
Department of Computer Science, Columbia University, New York NY 10027  
{cmurphy, swapneel, kaiser, wilcox}@cs.columbia.edu

## Abstract

*Many collaborative applications, especially in scientific research, focus only on the sharing of tools or the sharing of data. We seek to introduce an approach to scientific collaboration that is based on the sharing of knowledge. We do this by automatically building organizational memory and enabling knowledge sharing by observing what users do with a particular tool or set of tools in the domain, through the addition of activity and usage monitoring facilities to standalone applications. Once this knowledge has been gathered, we apply social networking models to provide collaborative features to users, such as suggestions on tools to use, and automatically-generated sequences of actions based on past usage amongst the members of a social network or the entire community. In this work, we investigate social networking models as an approach to scientific knowledge sharing, and present an implementation called genSpace, which is built as an extension to the geWorkbench platform for computational biologists. Last, we discuss the approach from the viewpoint of social software engineering.*

## 1 Introduction

Collaborative work requires more than just sharing data (notes, files, documents, *etc.*) or tools (shared whiteboard, *etc.*). It requires *knowledge sharing* [14]. This is especially important for scientific researchers, be it domain knowledge (which analysis is appropriate for a given data set) or functional knowledge of a particular tool (which configuration will yield the best results).

A somewhat static approach to knowledge sharing is to have experts write online documentation or have a group of users create and maintain a knowledge base, *e.g.* a wiki. However, these sources of knowledge can quickly and easily become out of date and stale if they are not adequately and frequently maintained, and other users may not always trust the source of the information. Furthermore, different

interpretations and variations of the knowledge may not be properly reflected.

Often times, particularly in scientific research, the tools that are used by researchers indirectly encapsulate this important knowledge, which could then be extracted based on observations of how someone actually uses the tool. For example, if a large number of users employ a particular tool to analyze a particular type of data, or configure an analysis tool in a particular way, it may be safe to assume that that represents knowledge that should be shared.

Our approach, therefore, is to automatically build organizational memory and enable knowledge sharing by monitoring what users do with a particular tool or set of tools, through the addition of activity and usage monitoring facilities to standalone applications. Assuming that a majority of users correctly use the tool(s), the system can then make this gathered knowledge available to novice users who are just getting started, or to intermediate users who need to become aware of advanced features of the tool(s); optionally, if a small subset of experts has the ability to annotate and correct the accumulated knowledge, that enhances its quality even further.

In order to make this knowledge available to end users, we expose it via social networking models, which have proven intuitive and easy to use. In this paper, we consider features of popular social networking applications and websites, and apply them to scientific collaboration. We also introduce an implementation of our approach, called *genSpace*, built on a tool used by researchers in computational biology. Last, we discuss the social software engineering implications of such an approach.

## 2 Background and approach

We are currently working with researchers at Columbia University's Center for Computational Biology & Bioinformatics (C2B2) and its MAGNet (Multiscale Analysis of Genomic and Cellular Networks) Center to explore new ways in which researchers in computational biology and bioinformatics can collaborate to exchange data, analyze results,

and share knowledge. Prior online “collaboratories” for bioinformatics scientists include applications for *tool sharing* (e.g., BioCoRE [3] from UIUC) or for *data sharing* (e.g., Biological Systems Collaboratory (BSC) [8] from Pacific Northwest National Laboratory). These make it possible for researchers in this field to assist and understand each other without being physically colocated.

According to the computational biologists we spoke with, one drawback of the tool-centric and data-centric approaches to scientific collaboration is that they rely on the collaborators themselves to actively contribute to the community. We seek to overcome this obstacle by introducing an approach to collaboration that is based on automating the process of *knowledge sharing*. In particular, our goal is to make it easier for bioinformatics researchers to find answers to their own questions, based on the passive gathering of the collective knowledge of other users. We do not seek to replace the tool sharing and data sharing collaborative facilities, but rather enhance them by creating a repository of user knowledge that is transparently collected and can be mined to answer such questions as, What tools/datasets should I use to investigate this problem? Who do I know who also uses this tool/dataset? Which tools and datasets work well together? Where does this tool/dataset fit in a typical workflow? When did I previously use this tool/dataset? How can I get help (from an expert who is online right now)?

The approach to knowledge sharing described here is based on a “social networking” metaphor for collaborative work, taking a lead from such popular websites as Facebook, Netflix, and Amazon. Social networking facilitates collaborative filtering, by allowing users to get answers to questions like, What movies would I like? Who also likes this book? Which food and wine go together? Where does this song fit in a playlist? When was this restaurant last reviewed? How can I get help about this MP3 player? These social networking questions are analogous to those that come up in scientific collaboration, and have been identified by Carroll *et al.* as ways of creating communities of practice as a facet of activity awareness [6].

Our implementation of this approach is called *genSpace* (a combination of “genomics” and “MySpace”), which is built as an extension to the geWorkbench [4] platform. This proof-of-concept includes such features as an expert finder, suggestions on analytical tools to use, and automatically-generated workflows based on past usage amongst the members of a social network or the entire community.

### 3 Implementation of genSpace

Many researchers in the field of computational biology use the geWorkbench tool [4], a Java-based open-source platform for integrated genomics. Using a component architecture, it allows independently developed plug-ins to

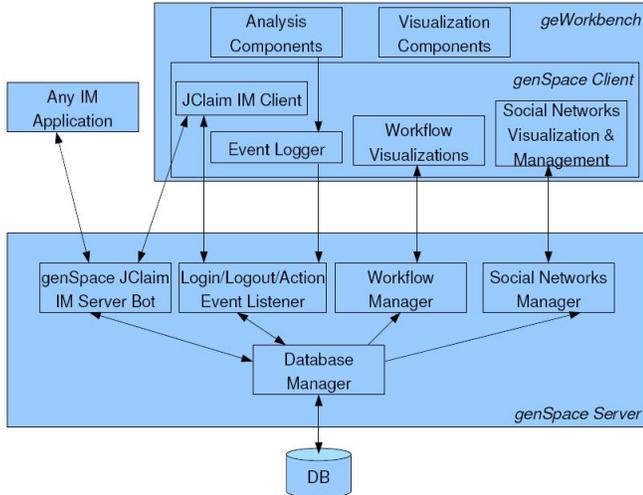
be configured into complex bioinformatic applications. At present there are more than 30 available plug-ins supporting the visualization and analysis of gene expression and sequence data; however, geWorkbench is purely a stand-alone application and does not have any collaboration capabilities. We have developed genSpace as a set of plug-in components for geWorkbench, conceptually orthogonal to the others, and have also developed a central genSpace Server that stores the collected knowledge, performs data mining, and provides facilities for user interaction.

#### 3.1 Summary of features

In order to provide collaboration facilities for geWorkbench, genSpace offers features including:

- **Social Networking:** allows users to associate with each other and share knowledge
- **Collaborative Workflow Composition:** past history of analysis tool usage is used to identify and visualize commonly-occurring sequences/workflows
- **Instant Messaging (IM)-based User Interface:** communication with the system may be done via IM clients such as Yahoo! Messenger, Windows Live Messenger, or Google Talk
- **Peer Suggestions:** suggests other genSpace users who work with similar analysis tools
- **Tool Suggestions:** suggests analysis tools that may be useful, based on what tools were previously used
- **Expert Finder:** identifies genSpace members who appear to be experts in using geWorkbench, a particular analysis tool, or a set of tools

Note that all of the features listed here are likely to be applicable in many applications and domains, particularly for collaboration in other scientific research fields. That is, although many of the genSpace features grew out of requirements from our work with MAGNet and C2B2, the social networking metaphors used to enable collaboration in the domain of computational biology can also be applied to other domains, and the same benefits would follow. Furthermore, although genSpace users are using a specific set of software tools, the same approach could in particular be applied to the software developers who are developing such tools, be it in computational biology or other domains. It is this approach to collaboration and knowledge sharing, based on social networking metaphors, that is the main contribution of this work.



**Figure 1. genSpace architecture**

### 3.2 Architecture

genSpace is implemented in a classic three-tier architecture, as shown in Figure 1. It consists primarily of the genSpace Client, which is a set of Java plugin components in the geWorkbench application framework, a separate centralized genSpace Server (also implemented in Java), and a backend database.

In order to capture the data needed to enable the collaborative facilities, geWorkbench is instrumented to monitor and record usage information, particularly the use of analysis tools such as Hierarchical Clustering [12], MEDUSA [25], *etc.* The records include the user’s name, the host-name of the system, the name of the analysis tool, the name of the dataset, and a timestamp. The usage logs are sent as XML files to the genSpace Server at pre-determined time intervals. Of course, geWorkbench users have the option of disabling this logging, or having their actions logged anonymously, in which case data mining and collaboration can still be done, even without identifying the source of the data. Once the logs are sent to the genSpace Server, the data is stored in a relational database, which is then later mined for event patterns.

### 3.3 Privacy concerns

Further investigation of the privacy and security aspects of genSpace will be required to ameliorate some concerns expressed by a few members of the geWorkbench user community. To date we have so far taken only initial measures to allow for user control of privacy (*e.g.* as mentioned above, users can choose anonymous logging or disable it entirely), but users may want to change the data that has already been collected, for example to anonymize data that previously

was associated with them. Moreover, users may not want to be suggested as an expert or as a peer, or they may want to restrict the use of that data to people in their own networks. While genSpace does currently allow for some opt-in/opt-out and limitation features, users may not want their own usage history included in the data mining that is used to generate such suggestions. One possible approach is to use the overlapping social networks to drive data privacy rules, for instance to ensure that results are published only to friends in the social networks, and not inadvertently passed on to friends of friends. We intend to address these concerns in future work.

Security of the data will also need to be investigated, for instance to transmit and store the log files in encrypted form (though they will need to be unencrypted at the server for data mining). It may also be possible to use privacy-preserving data formats to aid validation of analysis results (did someone else independently get the same result I did?) in the collaboration framework while retaining confidentiality and, when warranted, anonymity; this has previously been investigated in our lab for separate purposes (collaborative intrusion detection), as described in [27].

## 4 Features

This section describes the main features to date of the genSpace system. Note that all of these concepts could also be applied to other applications in other domains, and are not necessarily tied only to computational biology.

### 4.1 Social Networking

Social networks form the foundation of genSpace. genSpace users can perform operations such as creating and joining networks, listing all the networks to which they belong, listing all public networks in the system, and maintaining lists of friends. Users can also find out who else is currently logged into genSpace, and lookup additional public information on other genSpace users (such as their email address) based on their login name. genSpace also allows users to find out which tools are most popular (by frequency of use) within a particular network or across all genSpace users.

Also, genSpace provides visualization of the social networks to which the user belongs, using an interface based on the HyperGraph [18] tool, shown in Figure 2. This allows users to see the members of their networks, and see which other networks those users belong to. Nodes in the hyperbolic graph can be moved in the visualization to make it easier to read; future work in this area may include the ability to highlight user IDs to show annotations, like their email address (if publicly available) or tools for which they are considered an expert.

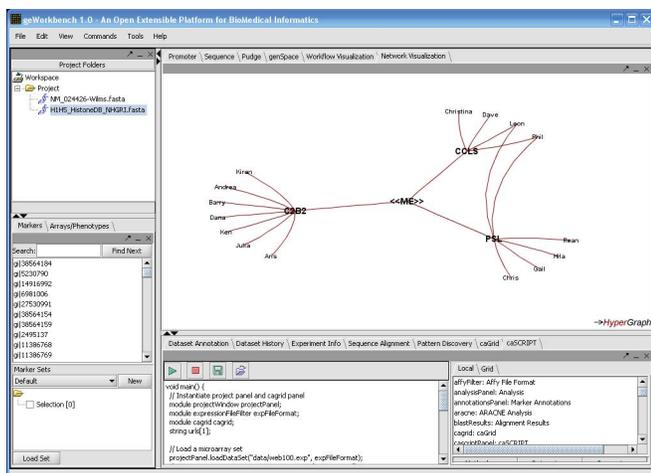


Figure 2. Visualization of social networks

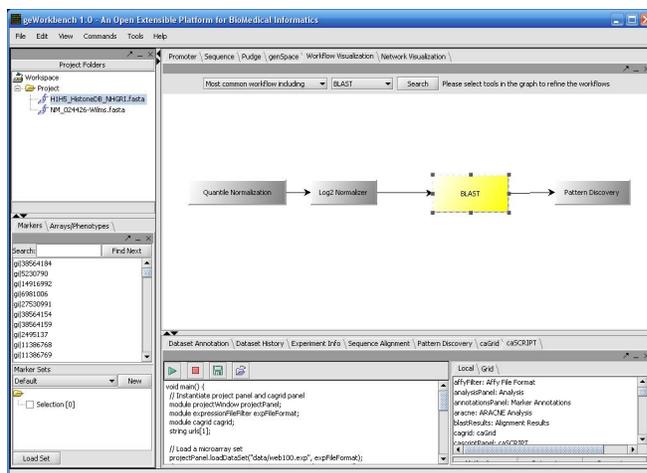


Figure 3. Visualization of single workflow

## 4.2 Collaborative workflow composition

A user of geWorkbench may be faced with the challenge of selecting a number of analysis and visualization tools for a particular data set, but also may not be certain of the proper order in which to execute them, or even which tools can be used together, regardless of order. We define a “workflow” as a sequence of execution of analysis tools in geWorkbench; genSpace supports the “collaborative composition” of workflows by investigating what other workflows have been performed in the past by other users and then suggesting one or more workflows based on previous activity. Note that whereas a “collaborative workflow” typically refers to a workflow whose constituent parts are executed by various sources working in collaboration, the collaborative composition of workflows refers to using organizational memory and history of collaboration to create a workflow, whose execution can then be performed by a single user (or conceivably multiple users, though this is not currently supported in geWorkbench).

The genSpace user requests a workflow suggestion by specifying an analysis tool that he or she wants to use, and then querying (via a UI component in geWorkbench) either for (1) the most common workflow starting with that tool, (2) the most common workflow including that tool, or (3) all workflows including that tool. genSpace then searches through all records of workflows and finds the results, which are weighted so that more recent workflows are given higher priority than workflows from further in the past (this is done using an exponential time-decay function; see [11] for more details); noise is eliminated by filtering out workflows that appear in only a small percentage of the results. Users also have the option of searching only workflows conducted by other users in their social networks.

The suggested workflows are displayed in geWorkbench

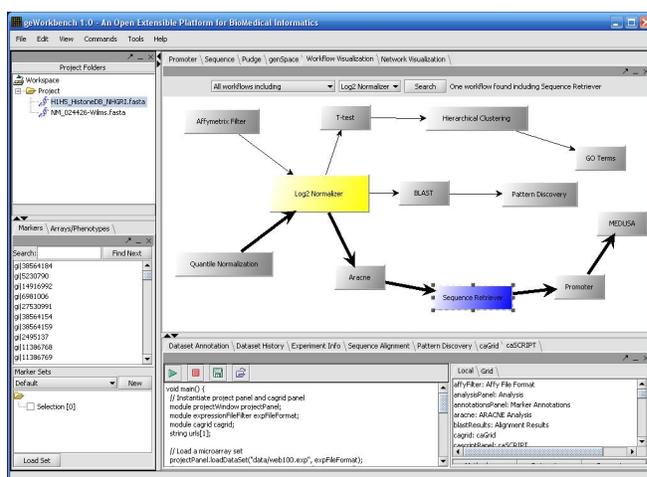


Figure 4. Workflow relationships

using a directed graph built using the jGraph [22] API. For a single workflow (most common starting with or including the selected tool), this is just shown as a sequence of tools, as in Figure 3.

When the user chooses to see all workflows including a given analysis tool, the results are displayed in a workflow relationship graph, as shown in Figure 4. The specified tool is highlighted in yellow and is shown as a larger box than the other tools. Tools that tend to be towards the beginning of workflows are initially placed towards the left of the screen, and those that tend to be towards the end are placed towards the right, so that on average, workflows typically go from left to right. The vertical placement of the boxes is random, however, but the display allows the user to move the boxes for easier readability.

The display also allows the user to select and highlight other analysis tools (in addition to the one initially selected)

so as to refine the graph and highlight only those workflows involving the selected tools. The number of corresponding workflows is indicated in the upper right of the display, and workflow paths are indicated with bold arrows. If a user knows that he or she needs to use a set of tools in a workflow, this allows for a way of expanding the search criteria so that only a small number of workflows are suggested.

We are currently looking into ways to enhance this important feature of genSpace. We have investigated the possibility of saving the search results and expressing the suggested workflow(s) as scripts for automation in geWorkbench. We also are considering ways to enhance the visualization, such as the use of weighted edges to indicate which workflows are more common, and providing additional information about the tools shown within the visualization, like links to online help and documentation, or advice on which genSpace users are the corresponding experts.

### 4.3 IM-based user interface

Almost all of the user interaction with genSpace can be performed via an Instant Messaging (IM)-based user interface, as previously explored in [7]. This design decision was made to keep the system in line with the social networking metaphor: as most social networks include some sort of real-time chatting feature, we thought it would make sense to allow the user to “chat” with the genSpace system, using the same interface that one uses to chat with human friends and collaborators.

As shown in Figure 1, the genSpace Server includes a component built on the JClaim [21] API. This API allows for the development of an IM Server Bot that can be used over such popular IM networks as Yahoo! Messenger, Windows Live Messenger, or Google Talk. When users are logged into geWorkbench, they can communicate with the genSpace system through an integrated JClaim chat client (which supports multiple chat networks). Moreover, the IM Server Bot can be accessed regardless of the IM Client that is being used, so a user can use one of the native clients or a multi-network client such as Trillian to chat with genSpace, even when they are not using geWorkbench.

The genSpace Server does not support fully free-text input commands, but the complete listing of all valid commands is available from the genSpace Server by typing “help”. Because of the use of a text-based interface, though, the genSpace Server performs some processing on the typed input, in case of any typing mistakes. The domain of valid input commands is limited so it is possible to accurately guess the user’s intention even if the commands are not spelled correctly; we currently use custom-built code that employs such techniques as comparing edit distance and considering permutations. Thus, if a user types “my ntwrks”, the genSpace Server can still detect that the user

meant to type “my networks”, which is the command for finding all the networks to which he or she belongs.

Of course, some users will not be comfortable with a command-line interface or may not want to use a chat client; we are currently implementing a separate geWorkbench plugin that will allow users to perform all of the social networking actions through a GUI component.

The JClaim API on which the genSpace IM Server Bot is based only supports a stateless request-response model of interaction for IM chatting, but we have modified it to allow the server to push messages to the clients, and to allow for stateful conversations; we intend to explore this further in future work. We are currently investigating ways to proactively send users suggestions based on current activity, or even to send a warning if there appears to be an anomaly in the way the user is working with geWorkbench. However, a feature like this must be wary of the possibility of a “Clippy Effect” [16], in that unwanted or unwarranted messages may prove to be more annoying than helpful.

### 4.4 Peer suggestions

An important part of the social networking metaphor is to allow genSpace users to find other members of the community who are like them in some way. Although genSpace does not currently go so far as to suggest potential collaborators, as in [24], which requires much more domain knowledge, genSpace does allow users to find others who have similar operational profiles in terms of patterns of analysis tool usage. This “friend finder” or peer suggestion feature works by looking at all of the analysis tools that the user has worked with, then looking at all users (either in the same social networks or in all of genSpace) who have used those same tools, and identifying the one who has the greatest number of tools in common (recent usage is weighted more so as to increase relevance). By doing so, genSpace allows users to locate and add these as “friends”, and then possibly collaborate with - or at least get help from - them as they use similar tool sets.

Of course, this feature raises some questions around privacy concerns, as discussed above, as users may not want to be considered for peer suggestion or may not want other users to know which analysis tools they have used in the past. These issues will need to be investigated further, though users currently have the ability to control how data is collected about them and to opt out of being suggested as a peer.

### 4.5 Tool suggestions

In addition to suggesting peers, genSpace also can suggest individual analysis tools to users. This is particularly useful in numerous situations, for instance if a new analysis

tool is made available, if the genSpace community finds that a tool is appropriate for a certain datatype and starts using it more, or in the cases where a user simply is unaware of a tool's existence.

The tool suggestion feature operates in a similar manner to the peer suggestion. Based on the analysis tools that the user has previously employed, genSpace finds all previous workflows including those tools, sorts the tools based on their frequency of use, and returns a ranked list of suggestions. Users can specify that certain tools *not* be included in the list of suggestions, for instance if they already know that they are not appropriate. As with the other genSpace features, the tools' frequency of use is weighted so that more recent invocations are more likely to appear than ones from further in the past. Additionally, users can also limit the scope of the search, if desired.

#### 4.6 Expert finder

Many novice geWorkbench users will need help understanding the various analysis tools, or even getting started with geWorkbench itself. Of course, online documentation or training would be ideal for getting a user up to speed, though as noted above, online documentation can quickly become stale if not properly maintained. In order to augment any online documentation (which may not even exist), genSpace enables users to find experts on a particular analysis tool, based on other users' past history. Currently it uses a simple model of finding the user who has most often worked with the analysis tool in question, using a weighting such that more recent activity is considered more relevant than activity far in the past; users can also specify the scope of users in which to search for experts. When a user searches for an expert, genSpace indicates whether the expert is currently online (logged into genSpace) and provides the expert's login ID, so that he or she can be contacted over an IM chat network. If that user is not online, genSpace also provides the ID of the user with the most expertise who is currently logged in.

A variant of this feature can also be used to identify geWorkbench "power users", irrespective of the individual analysis tools. This simply looks for genSpace users who have been working with geWorkbench the most, regardless of which tools they have been using. These experts would be the ideal candidates to help out users who are new to geWorkbench in general.

Users who are labeled as experts or power users may not want genSpace to treat them as such. Therefore, they currently have the ability to opt out and not be considered when searching, as described above. Future work in this area could also include user ratings of experts or searching for experts by other criteria, such as combinations of tools (workflows) or specific data types.

## 5 Software engineering considerations

As this is ongoing work, we continually consider the different challenges faced in the engineering of such "social software", particularly in the case of retrofitting collaboration capabilities onto an existing application. As previously identified in [33], one of the challenges is that the single-user application (in our case, geWorkbench) must expose an API for capturing users' activities and associated data, otherwise potentially major modifications might be necessary to give the collaboration engine access to that information. We benefited from geWorkbench's component-based architecture and its publish/subscribe model for transmitting data between components; this meant that we required the geWorkbench developers to make only a minor change (approximately 145 lines of code), which then allowed us create a simple plugin to capture, record, and centralize information on users' activities. Note that this plugin, from an architectural perspective, was no different from any of the other plugins supported by geWorkbench.

We were also able to easily integrate IM capability and social network and workflow visualization into the application similarly because of geWorkbench's plugin architecture for user interface components; the addition of these features required no changes to geWorkbench whatsoever.

We therefore recommend that those who plan on enabling knowledge sharing capabilities in standalone applications use a publish/subscribe model for internal events within the application, and a component-based architecture for visual components to facilitate the easy creation of new user interface features. When this is not possible (for example, in existing applications without an exposed API), a multi-layer approach such as the one in [33] may be best for separating application functionality from collaboration functionality, though it will require more extensive changes to the application itself.

## 6 Related work

Tools like BSC [8] and BioCoRE [3] are similar in spirit to genSpace in that they provide collaborative features for bioinformatics researchers, but they do not provide the social networking capabilities, nor do they build on past usage history and organizational memory or provide recommendations based on those. This is where genSpace primarily differentiates itself from those tools.

Important components of genSpace are related to previous work in social networking and computer supported cooperative work, such as using IM bots as a user interface [7], shared workflows and scripted processes [19], expert finder utilities [15] [23] [31] [34], and a limited chat-based tool for collaboration [28]. The KnoSoS project [10] also seeks to apply social networking concepts to knowledge sharing,

by investigating how to create group boundaries and track content, but they do not specifically address scientific collaboration, and do not consider the modification of existing applications, as we do with genSpace. Additionally, the use of past history to generate workflows has been shown to be effective in performing procedural tasks in online environments [19]. These important works provide the foundation upon which parts of genSpace are built.

Others have conducted experiments on the effectiveness of many principles that genSpace espouses, such as ways of motivating contributions to the online community [2], the psychology of collaborative technical help and help-giving [13] [30], and using organizational memory for collaborative help [1]; all of these are areas that could be investigated with genSpace in the future. Still others have looked at mechanisms of and rationale for “knowledge sharing”, including but certainly not limited to ARPA’s Knowledge Sharing Effort [26], research into knowledge sharing in virtual communities [9] [17], and investigation of ways to visualize the merging of users’ knowledge [20]. Carroll *et al.* in particular have looked at extending the idea of knowledge sharing (what they term “knowledge in common” [6]) and investigating concept sharing, and eventually activity sharing and activity awareness techniques [5]. We are building on this work by looking at the challenges of applying it to collaboration in scientific research, taking into account the specific activities in that domain, and hope that our work contributes to this overall effort to explore the possibilities of knowledge sharing.

## 7 Future work

At the time of this writing, we have not yet been able to perform any empirical studies to demonstrate the usefulness of genSpace. However, the client-side event logging feature will be included in the next release of geWorkbench, and other aspects will be rolled out incrementally as we acquire more data that can be used to drive these features. After that future release, we hope to conduct user studies to gain a better understand of how computational biologists can benefit from such a knowledge sharing infrastructure.

A future prospective challenge in our implementation is that of concept drift [32]. Any recommendations constructed from what users have “done” in the past would likely lose significance and accuracy when users move on to “something new”. Also, as users’ interests change and familiarity with the tools grows, they may want to migrate to other social networks; thus, it may be suitable to create dynamic, ad hoc networks and then automatically add users to them.

Additionally, genSpace may also later include more in-depth exploration of the social networking metaphor, such as the use of news feeds (possibly using RSS or Atom)

or watch lists to get instant notification of the activities of other genSpace users. Also, aside from real-time chat, a bulletin board or email-like system could also be included, or the system could make recommendations on a sequence of tools to use (workflows), rather than just individual tools. Other collaboration features not necessarily related to knowledge sharing could include data sharing and tagging/annotating, distributed workflow management, and sharing of computational load [29].

Most importantly, future work should consider the specific challenges of scientific collaboration with respect to the scientific process: gathering observable, empirical and measurable evidence through experimentation, and then formulating and testing a hypothesis in a reproducible manner. And, beyond that, to consider the effect that such challenges have on those of us in the social software engineering community.

## 8 Conclusion

We have presented an approach to scientific collaborative work that goes beyond data- and tool-sharing and uses social networking metaphors to enable knowledge sharing. We have also presented an implementation called genSpace, which is a system built on the geWorkbench tool for computational biology. We believe that other domains and other applications can benefit from this approach, particularly as social networking becomes more pervasive in everyday life, and knowledge sharing becomes more critical to successful scientific collaboration and other endeavors.

## 9 Acknowledgments

The authors would like to thank Aris Floratos, Kiran Keshav, Ken Ross, and Diana Chang for their guidance and assistance. Murphy, Sheth, and Kaiser are members of the Programming Systems Lab, funded in part by NSF CNS-0717544, CNS-0627473, CNS-0426623 and EIA-0202063, and NIH 1 U54 CA121852-01A1.

## References

- [1] M. Ackerman and D. McDonald. Answer Garden 2: merging organizational memory with collaborative help. In *Proc. of the 1996 ACM conference on computer supported cooperative work (CSCW)*, pages 97–105, 1996.
- [2] G. Beenen, K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. Kraut. Leveraging single-user applications for multi-user collaboration: the cword approach. In *Proc. of the 2004 ACM conference on computer supported cooperative work (CSCW)*, pages 212–221, 2004.
- [3] M. Bhandarkar, G. Budescu, W. F. Humphrey, J. A. Izaguirre, S. Izrailev, L. V. Kal, D. Kosztin, F. Molnar, J. C.

- Phillips, and K. Schulten. BioCoRE: A collaboratory for structural biology. In *Proc. of the SCS International Conference on Web-Based Modeling and Simulation*, pages 242 – 251, 1999.
- [4] A. Califano, A. Floratos, M. Kustagi, and J. Watkinson. geWorkbench: An Open-Source Platform for Integrated Genomics. <http://www.geworkbench.org/>.
- [5] J. M. Carroll, D. C. Neale, P. L. Isenhour, M. B. Rothson, and D. S. McCrickard. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies*, 58(5):605–632, May 2003.
- [6] J. M. Carroll, M. B. Rosson, G. Convertino, and C. H. Gano. Awareness and teamwork in computer-supported collaborations. *Interacting with Computers*, 18(1):21–46, January 2006.
- [7] S. Chan, B. Hill, and S. Yardi. Instant messaging bots: accountability and peripheral participation for textual user interfaces. In *Proc. of the 2005 international ACM SIGGROUP conference on supporting group work*, pages 113 – 115, 2005.
- [8] G. Chin Jr. and C. S. Lansing. Capturing and supporting contexts for scientific data sharing via the Biological Sciences Collaboratory. In *Proc. of the 2004 ACM conference on computer supported cooperative work (CSCW)*, pages 409 – 418, 2004.
- [9] C.-M. Chiu, M.-H. Hsu, and E. T. G. Wang. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3):1872–1888, December 2006.
- [10] T. Coenen, D. Kenis, C. Van Damme, and E. Matthys. Knowledge sharing over social networking systems: Architecture, usage patterns and their application. *LNCS 4277: On the Move to Meaningful Internet Systems 2006*, pages 189–198, 2006.
- [11] E. Cohen and M. Strauss. Maintaining time-decaying stream aggregates. In *Proc. of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems (PODS)*, pages 223–233, 2003.
- [12] F. Corpet. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 16(22):10881–10890, 1988.
- [13] A. Crabtree, J. O’Neill, P. Tolmie, S. Castellani, T. Colombino, and A. Grasso. The practical indispensability of articulation work to immediate and remote help-giving. In *Proc. of the 2006 20th anniversary conference on computer supported cooperative work (CSCW)*, pages 219–228, 2006.
- [14] T. Davenport and L. Prusak. *Working Knowledge*. Harvard Business School Press, 1998.
- [15] K. Ehrlich, C.-Y. Lin, and V. Griffiths-Fisher. Searching for experts in the enterprise: combining text and social network analysis. In *Proc. of the 2007 international ACM SIGGROUP conference on supporting group work*, pages 117–126, 2007.
- [16] R. G. P. Galluccio. Humanizing CALL: The use of pedagogical agents as language tutors. New England Regional Association for Language Learning Technology, Oct. 2006.
- [17] M.-H. Hsu, T. L. Ju, C.-H. Yen, and C.-M. Chang. Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International Journal of Human-Computer Studies*, 65(2):153–169, February 2007.
- [18] HyperGraph. <http://hypergraph.sourceforge.net/>.
- [19] IBM CoScripter. <http://services.alphaworks.ibm.com/coscripter>.
- [20] S. Iwasaki, Y. Hirakawa, H. Mase, E. Tokunaga, and T. Nakajima. Towards computer-supported face-to-face knowledge sharing. In *Proc. of CHI’06 Conference on Human Factors in Computing Systems*, pages 911–916, 2006.
- [21] JClaim. Java Compliant Logging and Auditing Instant Messenger. <http://www.itbsllc.com/jclaim/>.
- [22] jGraph. <http://www.jgraph.com/>.
- [23] D. McDonald. Evaluating expertise recommendations. In *Proc. of the 2001 international ACM SIGGROUP conference on supporting group work*, pages 214–223, 2001.
- [24] D. McDonald. Recommending collaboration with social networks: a comparative evaluation. In *Proc. of the SIGCHI conference on human factors in computing systems*, pages 593–600, 2003.
- [25] M. Middendorf, A. Kundaje, M. Shah, Y. Freund, C. H. Wiggins, and C. Leslie. Motif discovery through predictive modeling of gene regulation. *Research in Computational Molecular Biology*, pages 538–552, 2005.
- [26] Neches, R. and Gunning, D. ARPA knowledge sharing effort. <http://ksl.stanford.edu/knowledge-sharing/>.
- [27] J. Parekh. *Privacy-Preserving Distributed Event Corroboration*. PhD thesis, Columbia University, Mar 2007.
- [28] A. Ribak, M. Jacovi, and V. Soroka. “Ask before you search”: peer support and community building with ReachOut. In *Proc. of the 2002 ACM conference on computer supported cooperative work (CSCW)*, pages 126–135, 2002.
- [29] S. Sheth and G. Kaiser. CPU torrent - enabling CPU cycle sharing to reduce user wait time and service provider computation requirements. Technical Report cucs-030-08, Dept. of Computer Science, Columbia University, 2008.
- [30] M. Twidale and K. Ruhleder. Where am I and who am I?: issues in collaborative technical help. In *Proc. of the 2004 ACM conference on computer supported cooperative work (CSCW)*, pages 378–387, 2004.
- [31] A. Vivacqua and H. Lieberman. Agents to assist in finding help. In *Proc. of the SIGCHI conference on human factors in computing systems*, pages 65–72, 2000.
- [32] G. Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101, 1996.
- [33] S. Xia, D. Sun, C. Sun, D. Chen, and H. Shen. Leveraging single-user applications for multi-user collaboration: the coWord approach. In *Proc. of the 2004 ACM conference on computer supported cooperative work (CSCW)*, pages 162–171, 2004.
- [34] J. Zhang and M. Ackerman. Searching for expertise in social networks: a simulation of potential strategies. In *Proc. of the 2005 international ACM SIGGROUP conference on supporting group work*, pages 71–80, 2005.