Alzforum and SWAN: the present and future of scientific web communities

Tim Clark and June Kinoshita

Received (in revised form): 27th March 2007

Abstract

Scientists drove the early development of the World Wide Web, primarily as a means for rapid communication, document sharing and data access. They have been far slower to adopt the web as a medium for building research communities. Yet, web-based communities hold great potential for accelerating the pace of scientific research. In this article, we will describe the 10-year experience of the Alzheimer Research Forum (‘Alzforum’), a unique example of a thriving scientific web community, and explain the features that contributed to its success. We will then outline the SWAN (Semantic Web Applications in Neuromedicine) project, in which Alzforum curators are collaborating with informatics researchers to develop novel approaches that will enable communities to share richly contextualized information about scientific data, claims and hypotheses.

Keywords: Neuroinformatics; knowledge management; Alzheimer disease; semantic web; e-Science

INTRODUCTION

A research library today is a forlorn place. The reading rooms are often vacant, and the only reason to prowl the stacks is to exhume journal articles published before the digital era. To access articles and data in active fields of research, scientists today go to the web, where most full-text articles are but a few search queries away. Scientists have rapidly adopted and indeed driven the development of digital information technology, building many vitally important databases and tools to support their research needs.

Practically speaking, biomedical research is now impossible for the person without a computer and a network connection.

Furthermore, there is a steady increase in both formal and informal interactions between biomedical researchers, computer scientists and software engineers. Research institutes and consortia are beginning to evolve for the specific purpose of enabling joint work by natural scientists and computer scientists on important problems. One of the problems being tackled in such collaborations is, in fact, adapting the structure of the World Wide Web, its software protocols and associated metadata, to better enable scientific discovery and communications.

Human computer interaction theorists such as Nardi [1] stress the importance of intensional networks of collaborators within evolving web-mediated communities of practice. These phenomena in science seem to be convergent with web-mediated distance collaboration trends in industry, and have led, with mixed success, to attempts to create special scientific collaboration environments or collaboratories.

Bos et al. [2], classifying 212 collaboratories observed in their study, notes seven types of collaboratories, with three principal barriers to their successful construction.

(i) Scientific knowledge is difficult to aggregate, i.e. the transition from information to knowledge is still difficult to negotiate.
(ii) Scientists work independently and informally—they do not like to be told how to work, or what tools to use.
(iii) Scientific work across institutions is complicated by various legal and organizational barriers, principally having to do with intellectual property.
One particularly successful example of a web-based science collaboratory was unfortunately overlooked by the Bos et al. [2] study—the Alzheimer Research Forum (Alzforum, www.alzforum.org), founded in 1996 and today having over 4000 registered members of whom the majority are professional Alzheimer Disease (AD) researchers. We believe Alzforum’s very strong specialist focus, neutrality, and close integration with its constituency, have been enormous factors promoting its success. At the same time, perhaps these factors also have caused it to be overlooked by researchers outside the immediate AD research community. We sincerely hope that the present article helps to remedy this situation, as many valuable lessons can be learned from Alzforum.

In our experience, successfully overcoming Bos’s first barrier requires a sensitive consideration of exactly how scientific information is transformed into knowledge. We need to be equally clear on how discourse regarding this knowledge occurs, and how hypothesis evolves into theory through testing and discourse. Because the Alzforum captures a significant amount of scientific discourse around Alzheimer disease hypotheses, the SWAN (Semantic Web Applications in Neuromedicine) project was conceived with the aim of integrating hypothesis management into the design of this web community.

SWAN [3,4] builds on Alzforum’s successes and strong social network, to construct a semantically structured network of hypotheses, claims, dialogue, publications and digital repositories. Later in the discussion we will describe SWAN’s methods and technology, which we believe point the way to a general semantic framework for specialist web communities in biomedicine.

While digital community technologies such as blogs, social networking, chat rooms and virtual worlds are permeating contemporary culture and enabling the formation of communities around almost every conceivable theme, very few web communities are by and for working scientists. Yet, we do not believe biologists are to blame for slowness in embracing the concept of specialist digital communities. We believe it is the informaticians who have failed the biologists—to date—by not approaching the problem from a sufficiently holistic, information–ecosystem perspective, including an active social component; and by maintaining an overarching emphasis on technology at the expense of other factors.

The lack of scientific web communities represents a significant missed opportunity. Much of the creative activity in science occurs through informal discourse as scientists interpret and critique new findings, puzzle over discrepancies and integrate data to formulate new ideas and hypotheses. The lack of open web resources to promote and capture such informal discourse may be impeding the progress. An experiment might stall for the lack of a key ingredient, or a researcher might waste months traveling down a blind alley, unaware that someone in a different field has solved a missing piece of the puzzle. Web communities could promote greater transparency and access to this type of valuable information.

An approach to biological web communities which appropriately combines both social and technological knowledge infrastructure will allow researchers to share information with very little added effort, and returns high value. We hope, through the examples of Alzforum and SWAN, to illustrate one instance in which the balance has been appropriate, resulting in substantial benefit to this area of biomedicine.

A NEW NICHE FOR SCIENTIFIC DISCOURSE

The Alzheimer Research Forum Web site (www.alzforum.org) is a rare instance of an online community that is widely used by its target scientists for rapid, informal discourse [5]. Known popularly as ‘Alzforum’ by AD researchers around the world, the Web site was founded in 1996 as an independent, not-for-profit resource dedicated to Alzheimer disease and related disorders. Alzforum has since become a global online research community, with ~30–50 percent of all active Alzheimer researchers visiting it regularly.

The idea for the Alzheimer Research Forum Web site arose in late 1995, a time when AD research was exploding with new findings driven by the landmark discoveries a few years earlier of familial Alzheimer genes and the association between late-onset AD and the epsilon4 allele of apolipoprotein E. The field was ripe with professional rivalries and competing hypotheses. It seemed that AD research might benefit from more open dialogue and better management of the streams of data pouring out of laboratories. The project attracted support from an anonymous foundation, and in July 1996, the Web
site made its debut at the International Conference on Alzheimer’s Disease and Related Disorders in Osaka, Japan.

During the planning phase, the team considered whether the Web site should be hosted at an academic institution, but surveys indicated that individual and institutional rivalries could undermine the goal of creating an open community resource. The Alzforum was therefore established as an independent, not-for-profit entity.

Without an institutional imprimatur, the Alzforum faced the challenge of establishing its credibility within the scientific community. The Editor invited prominent leaders with diverse scientific backgrounds to join the scientific advisory board. The composition of the board was intended to signal that the site is dedicated to high scientific standards; that it is neutral with regard to hypothesis or dogma; and that perspectives from multiple disciplines are welcome.

At its launch, the Alzforum featured a ‘Papers of the Week’ list of peer-reviewed publications in the AD field, virtual seminars (slide and audio) and a Milestone Papers list of seminal publications dating back to Alois Alzheimer’s original article describing the first documented case of the disease that now bears his name. Although the initial Alzforum offerings seem modest in retrospect, the feedback was positive. By the end of its first year, the site had 1200 registered members.

The Alzforum upgraded in 2000 to a dynamic, data-driven Web site. Currently, the site contains more than 40 000 literature citations, 1400 research news articles, 4000 comments, 16 000 antibodies, 200 research models, 400 genes from published association studies of late-onset AD, all known mutations causing familial Alzheimer disease and frontotemporal dementia with parkinsonism (FTDP-17), all drugs in Phase 2 and 3 clinical trials, and a wealth of community resources such as databases for grants, conferences and jobs. The site receives over one million visits per year. A Google Scholar search finds 131 000 references to Alzforum. More than 4000 people have registered as members of Alzforum, representing a significant portion of AD researchers worldwide.

Feedback from scientists has been strongly positive. ‘[Alzforum] is the local newspaper for Alzheimer research’, writes John Hardy, Director of the Laboratory of Neurogenetics at the National Institutes of Aging. ‘I visit it one to two times a week just to see what’s going on...to check up on recent papers,...to see who’s hiring people and so on. I read people’s comments on papers, and I go from there to PubMed for anything I’ve missed. I think pretty much everyone in the field uses it in the same way’ (J. Hardy, personal communication) [6].

ALZFORUM’S KEYS TO SUCCESS

From the start, the creators of the Alzforum assumed that simply building a Web site would never be sufficient to ensure its success as an online community. The site would have to be compelling and highly accessible, and beyond that, it would have to be proactive, constantly soliciting the participation of scientists.

A guiding principle of the homepage is that the site would be ‘the daily tabloid for AD research’, one that AD researchers would designate as their personal homepage. Member registration is optional and free. To keep readers coming back regularly, the homepage is dynamic, useful and entertaining. Almost daily, readers will find something of interest: the latest news, a live discussion, ‘image of the week’, opinion polls, conference reports and commentaries by prominent scientists. Comments posted on the Alzforum become topics of conversation out in the nonvirtual world. The engine driving this activity is a professional team of highly informed science journalists, who provide expert reporting and know how to solicit provocative and timely commentaries from leading thinkers.

Many scientists have remarked on how effective the Alzforum has been in nurturing productive discussion of their ideas and findings. For example, in 2005 Vincent Marchesi, a cell biologist at Yale University, published an alternative interpretation of the amyloid hypothesis that might ordinarily have been quietly ignored by most AD researchers. Instead, when the paper was featured by Alzforum, 17 scientists posted lengthy, detailed and productive commentaries. ‘The postings on the Alzforum site regarding my PNAS paper have been incredibly rewarding for me, and I suspect, for many of the others that participated’, wrote Marchesi. ‘I don’t see how so many candid exchanges could have taken place any other way’ (V. Marchesi, personal communication) [7].

As a highly visible presence in the AD community, Alzforum is ideally positioned to develop and
host community data repositories. These repositories contain publicly available data only, in part to avoid the security issues that attend the distribution of raw, unpublished data, but also because there are significant opportunities to add value to public data, simply by providing a system to curate and organize it around the community’s interests.

The Papers of the Week database, for example, is a subset of PubMed citations selected for relevance to AD, related disorders, key genes, relevant developments from broad areas of basic research, and advances in methodology. Although PubMed is more comprehensive, many AD researchers prefer to browse Papers of the Week because it is tailored to their interests and eliminates the need to carry out multiple searches. The Alzforum further enriches the Papers of the Week by linking citations to news stories, commentaries and related articles. High-impact articles are designated as ‘ARF Recommended’ papers and ‘Milestones’. For AD scientists, this enriched content provides context that is missing from PubMed.

Alzforum also designs and develops databases for key findings and reagents. These data are publicly available through publications and disparate data stores, and individual scientists expend much time and effort to keep current. In many cases, there is little incentive for an individual to take on a curatorial task on behalf of the community, and Alzforum has stepped into the breach to develop databases of antibodies, animal models, drugs in clinical trials and the like.

Interestingly, Alzforum has also developed collaborative databases that serve both individual and community needs. The AlzGene is a comprehensive database of all published genetic association studies for late-onset AD. AlzGene was conceived and curated by Lars Bertram and his colleagues [6] at Massachusetts General Hospital, and designed, developed and hosted by Alzforum. The database can be browsed by chromosome or searched by gene, polymorphism, protein, keyword or author. Each gene is summarized in a table listing details of all published studies, and a meta-analysis of the findings can be calculated with a single click. The AlzGene was launched publicly in 2004 and is updated almost daily, as soon as a new study appears in PubMed. Bertram and his colleagues [6] identified the genes for which meta-analysis showed the strongest positive association with late-onset AD, and published their findings in Nature Genetics [8] more than 2 years after first providing open access to the database. This example shows that it is possible for a scientist to develop a valuable community resource without sacrificing professional advancement.

In summary, Alzforum’s 10-year experience has demonstrated some keys to creating a successful scientific Web community. These are neutrality, inclusiveness, trust, high quality, timeliness, proactive solicitation of community participation and value. Attaining these goals requires leadership, professional staffing and sustained financial support. The funding required to maintain Alzforum is not insignificant, comparable in magnitude to the funding that would be needed to produce the experimental results for a journal publication. Seen in this light, however, one could argue that Alzforum is a high value-added investment for the Alzheimer field.

SWAN—SEMANTIC WEB APPLICATIONS IN NEUROMEDICINE

Because of its proven success as a web scientific community, the Alzforum offers an opportunity to engage the community in a collective knowledge curation effort. This concept evolved into SWAN and is being developed through a collaboration between Alzforum and informaticians at Harvard University, Massachusetts General Hospital and IBM.

SWAN aims to develop a practical, common, semantically structured, web-compatible framework for scientific discourse using Semantic Web technology [7–9], applied to significant problems in Alzheimer Disease research.

The initial concept for SWAN was proposed in a talk at the W3C Semantic Web in Life Sciences workshop, October 2004 [10]. SWAN has since been developed through a pilot application [3, 4] (Figure 1) and is currently in the development stage of its first production-quality application.

In many formal models of knowledge acquisition in science, research proceeds in a cycle—from hypothesis development; through experiment and data collection; to interpretation and drawing of conclusions; to communicating results to other scientists; to assimilating, criticizing and synthesizing the communications of colleagues (Figure 2). This outline of course is too schematic, but thinking in these terms helps us understand a certain point.
These practice–theory–practice cycles are socially interconnected in an extremely rich and complex way in what has been termed the ‘knowledge ecosystem’ of science (Figure 3).

Science cannot function without these ecosystem-level interconnections, and today very, very many of the connections are made digitally, using the medium of the web. In the scientific knowledge ecosystem, as elsewhere, it is the social activity (practice) of the participants, and not their thoughts, which is primary, and which drives the system at a fundamental level. Therefore, we began...
our analysis for SWAN with the social and practical aspects of the ecosystem’s actual functioning, and set our goal to improve it.

Rather than attempting to construct a logically coherent model of the known facts about AD, we begin instead by attempting to understand how the technology-mediated activity of scientific experimentation and discourse in AD research functions in order to improve it. This naturalistic approach emphasizes social factors, that is, what people actually do in conducting and communicating science. It places developers of the mediating technology inside the virtual community they support. A diagram, based on the approach of Suh et al. [11] illustrating this approach for both Alzforum and SWAN appears in Figure 4.

Theoretically this approach derives from work in knowledge management [12, 13] and activity-theory approaches to human–computer interaction [14]. Practically it is based on many experiences in constructing information systems to support rapidly evolving science, in which social factors and the social frame of the system were seen to strongly interact with the technology and content, critically influencing its ultimate success [15].

One of the key drivers behind SWAN was the realization that the ecosystem around Alzforum suffers from serious information loss at the critical point shown boxed in Figure 3. When digital resources such as scientific publications and their metadata are searched and retrieved on a Web site, the Web site presents an information architecture or model around these resources, which is richly connected but whose semantics are only implicit and not machine interpretable.

Alzforum, for all its content and community activity, still is little more than collections of documents and data with links.

When Alzforum—or any science Web site’s—resources or metadata (e.g. bibliographic references) are transferred to a researcher’s local computer, the web of semantics—the context—around the resource may only be transferred by hand transcription, and most likely this will not occur at all. This information is not embedded in a machine-interpretable knowledge model, and so the transfer

Figure 4: Web-mediated collaboration models of (A) Alzforum and (B) SWAN.
of context around the content occurs with a far lower efficiency, completeness and fidelity than the transfer of content alone, which by itself is machine-processable.

What does occur then is that the researcher constructs her/his own implicit and idiosyncratic information model around the resource, through a combination of bibliographies, personal directory structures, notes, tags and other annotation, and of course human memory. The last is—as we in the neurodegeneration field know quite well—a particularly fragile form for preserving context. And yet knowledge can only be transferred adequately as content plus context. This is because context records the ‘warrant for belief’ [16, 17] that distinguishes knowledge from mere opinion, and also provides the anchor for its validation, which is the social goal of the ecosystem.

When a person reads a paper or follows a link, he or she fills in the contextual blanks, such as ‘this paper challenges hypothesis X’, or ‘so-and-so draws the opposite conclusion from this data’. With SWAN, we will provide scientists with a tool to embed their documents, data and other digital materials in a knowledge model, and then to share the entire model with other scientists and communities, who can then build upon it.

The SWAN knowledge model [18] is defined in a formal ontology declared in OWL [19] and stored as RDF triples [20] in the Boca triple store [21].

SWAN incorporates a large part of the biomedical research life cycle in its ontological model, including support for personal data organization, hypothesis generation and digital pre-publication collaboration. Community, laboratory and personal digital resources may all be organized, interconnected and shared using SWAN’s common semantic framework. Figure 5 provides an example of the application of SWAN’s scientific discourse ontology to modeling the context of digital resources. We plan to extend this ontology to cover the most common forms of

![Figure 5: SWAN semantic relationships (example).](image)
Individuals will use a version called ‘MySWAN’ as a personal tool to find and organize information, to extend their knowledge, motivate discoveries and to form and test hypotheses. At the community level, the same software and the same ontological framework can be used to organize and curate the research of a laboratory or of an entire research community (such as the Alzforum). Therefore, elements of the personal SWAN can be shared with the community at a low incremental effort in curation. What’s more, community SWAN contents may be shared back with individuals and re-used in new contexts.

SWAN team members participate actively in the World Wide Web Consortium, W3C [22], through its Health Care and Life Science Task Force [23]. This has led to numerous fruitful collaborations [24–26] and valuable discussions [27]. An emerging theme has been the potential to use self-annotated discourse as a bridging ontology connecting the many specialized research subdomains contributing to Alzheimer research.

The advantage of this approach is that the bridging ontology will automatically track the knowledge as it emerges, and is not required to make ‘value judgments’ about the proper bridging level concepts. The bridging level, in fact, becomes concrete as speech acts, which are documented only as to what is said, and its logical and/or evidentiary relationship to other statements. The bridging level is a constructive (or $+$-constructive) ontology, which links dynamically to many separate nonconstructive (or $-$-constructive) domain ontologies and folksonomies [28].

In general, there has been a strong interest in this work from the semantic web community, particularly those working in eScience. The mutual interest of neuroinformaticians who wish to use semantic web technologies to increase the productivity and pace of science; and semantic web researchers to prove their technologies in this domain [29]; enlarges the knowledge ecosystem of neurodegenerative disorders into a knowledge-and-technology ecosystem of these disorders.

The SWAN team is currently developing semantic content to seed the community curation effort. An expert neuroscience curator is now converting the ‘AD Hypotheses’ section of the Alzforum Web site into SWAN’s schematized knowledge network. The curator breaks down each hypothesis into its component assertions and supporting evidence. Assertions themselves can be associated with supporting or refuting assertions, each with its own author and supporting evidence. This process provides a level of granularity that will enable the reader to examine the individual assertions comprising the hypothesis narrative, and to drill down to supporting evidence and their associated assertions, alternate hypotheses, news and other content. Quality assurance will be provided by an editorial board composed of leading experts in AD—and supplemented by the comments and support of the Alzforum member community at large.

On the front end, each hypothesis will be presented in an attractive, table-of-contents–like format that displays an abstract, links to full text articles and news, lists of assertions and comments, a knowledge matrix chart, and a link to the ‘SWAN view’ where new comments, findings and assertions can be added by readers.

The editorial team’s effort is made tractable because of the work of the community members. The community members, unlike those in a process such as Wikipedia [30], are principally concerned with advancing their own research program. SWAN documents all sides of the discourse, tracks the provenance of concepts and assures that content is not overwritten by competing interests. The incremental effort required to share knowledge from the team to the community is intended to be relatively small, and in many ways can be seen as an enhancement of the standard publication process for scientific literature.

The SWAN content will develop through the type of partnership that already exists between Alzforum editors and the AD community, with editors laying the groundwork and inviting community members to contribute. We have high hopes that, with the social infrastructure provided by the Alzforum, SWAN will open up exciting new possibilities for community-driven scientific knowledge curation and creation.

**Key Point**
- This article provides a case study of a successful online scientific community that integrates information from multiple disciplines in the context of understanding and treating a major disease, and shows the technology and social frame for its future evolution.
Acknowledgements

Heartfelt thanks to the Alzheimer Research Forum team comprising Elaine Aliabrandi, Tom Fagan, Donald Hatfield, Hakon Heimer, Sandra Kizley, Colin Kneip, Paula Noyes, Nico Stanculescu, Gabrielle Strobel and Elizabeth Wu and the SWAN project team comprising Paolo Ciccarese, Yong Gao, Marco Ocana belonging to Massachusetts General Hospital Neurology Department, Boston, MA, Ben Szekely of IBM Advanced Internet Technology Group, Cambridge, MA, Gwen Wong and Elizabeth Wu of the Alzheimer Research Forum team, Boston, MA.

We are grateful to the Ellison Medical Foundation for its support of the SWAN project, and to an anonymous foundation for its unstinting support of Alzforum and SWAN.

Many thanks as well to Steve Cayzer and Andy Seabome (Hewlett-Packard Laboratories) and Sean Martin (IBM Advanced Internet Technology Group) for many valuable technical discussions.

Lastly, we wish to thank Bradley Hyman (Massachusetts General Hospital and Harvard Medical School, Department of Neurology) and Carole Goble (University of Manchester Computer Science Department) for their continuing moral support and helpful advice.

References

21. Szekely B. Features of an enterprise-ready triple store, presentation at the RDF, Ontologies and Meta-Data Workshop, June 7–9, 2006, Edinburgh, UK.