

*Learned Publishing* (2004)17, 3–6

---

# Evolving scholarly communication

---

**B**ackground  
Scientists and funding bodies are actively questioning the role of the article and of publishers in scholarly communication. This view is driven by lack of adequate funding to cope with the growing output of research and by a vision of scientists that the Web can and should offer seamless searching and access to research when and where they need it.

However, the output of research is not just the published paper, it also includes the data that have been gathered. These data can and should be free wherever possible, and indeed publicly accessible databases have been available for some time in certain disciplines (e.g. GenBank, Ensembl, Swiss-Prot, TrEMBL, Protein Data Bank, Gene Expression Omnibus and ArrayExpress). Most journal articles now offer links to at least some of these. There is also a project under way in Europe to help provide linking and cross-searching of databases. E-BioSci (<http://www.e-biosci.org/>) is a next-generation scientific information platform that will interlink genomic and other factual data with the life sciences research literature. The platform aims to offer scientists new forms of navigation through databases and the published literature.

As a science publisher, Nature Publishing Group was therefore interested in the opportunity to work with database developers to create a new form of publication that combines primary scientific data with reviews and reports in new and innovative ways. The Alliance for Cellular Signaling (AfCS) is a consortium of eight US-based cell biology and bioinformatics laboratories that is engaged in a 10-year project initiated in 2000 to understand as fully as possible the internal signalling mechanisms of cells. This question is at the heart of much modern biology, from immunology and neuroscience to cancer research and drug discovery. The

**Jayne Marks and Timo Hannay**  
*Nature Publishing Group*

© Jayne Marks and Timo Hannay 2004

*ABSTRACT: Nature Publishing Group and the Alliance for Cellular Signaling formed a partnership in 2001 to develop a comprehensive website for researchers in the field of cellular signalling. This website aimed to evolve the traditional form of publication (an article with supplementary information and links) into something that made better use of web and database technologies, and provided more value to scientists. This article gives an overview of the project, some of the thinking behind it, and how it has developed.*



Jayne Marks



Timo Hannay

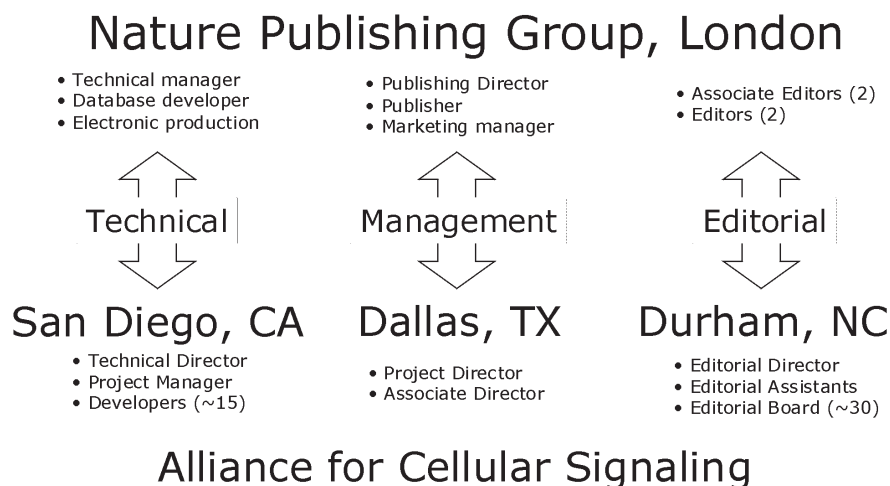


Figure 1

ultimate aim of the AfCS is to be able to construct computer simulations that can accurately predict the response of these cells to novel stimuli. This holds out the hope of one day enabling *in silico* research, which could accelerate the pace of scientific progress and drug discovery by orders of magnitude.

#### The project proposal

The plan was to create a comprehensive site for anyone interested in cellular signalling. This would gather together primary research data sets, review information summarized from the broader scientific literature, and news updates on the latest research findings. It would also include tools to view and manipulate the data, as well as features such as email alerts. The AfCS would contribute data from its experimental laboratories and extended network of experts while NPG would contribute publishing expertise. The website and associated databases would be hosted at AfCS facilities in the San Diego Supercomputer Center.

One of the most important aspects of this project to our partners in the AfCS was that access to the site must be cost-free for all users. One of the first questions to address, therefore, was how to pay for the resources that NPG was committing to the project, which included editorial, technical and publishing staff. In fact, these costs are now

covered by sponsorship income from pharmaceutical and biotechnology companies, and from *ad hoc* advertising on the email alerts and certain sections of the website.

It has been more important to NPG to test this new approach to publishing than to make a profit, but it has so far proved possible to cover the costs of the site.

#### Project set up

Figure 1 illustrates the main areas of collaboration between NPG in London and various AfCS centres in the US. Most NPG staff work on the project part-time, but it is nevertheless a substantial team. The AfCS as a whole is much larger than indicated because there is little direct contact between NPG team members and the laboratory scientists who represent the majority of AfCS staff.

#### The signaling gateway

The website itself (located at <http://www.signaling-gateway.org/>) comprises three main sections: Signaling Update, the Data Center and the Molecule Pages.

Signaling Update is a weekly news and information service covering developments in cellular signaling research. The content is written, edited and commissioned by NPG editorial staff working in London. This type of service is not uncommon at professional

*it has so far proved possible to cover the costs of the site*

publishers' sites and, indeed, NPG runs a series of similar services in other areas of research. But we do not know of another example in which it forms part of an academically hosted and curated scientific database. Feedback from users, together with analysis of usage statistics, suggests that its existence greatly improves the usefulness of the site and encourages regular return visits even when researchers do not have immediate need to access the databases. This service is accompanied by a weekly email alert, which currently goes to over 65,000 unique active email addresses, as well as an XML news feed in the popular RSS format.

The Data Center is a repository for all the primary experimental data collected by AfCS laboratories. It currently contains about half a terabyte of information gathered from immune system B cells (antibody-producing cells) and cardiomyocytes (heart muscle cells). The experiments conducted are large-scale investigations that would be beyond the capacity of any individual laboratory. Data sets gathered so far include analyses of the cells' internal responses to various external stimuli, as well as investigations of the interactions between different proteins in the cell. The data are published online as soon as they become available and AfCS staff have absolutely no priority over them. In this way the entire research community is encouraged to participate in their analysis and interpretation. This unusually open approach to scientific research was inspired to a large extent by the publicly funded genome sequencing initiatives. There are currently few parallels elsewhere in biomedical research but we expect to see them emerge in coming years.

The Molecule Pages are perhaps the most innovative section of the site and have also provided us with the greatest challenges. They currently consist of about 3,500 database records, each one concentrating on a particular protein that plays a significant role in cellular signaling. We expect the list to grow to about 5,000 within the next two years. The data themselves come in two types, automated and author-entered. Automated data (such as amino acid sequence, biochemical properties, protein family infor-

mation and structural data) are fed in automatically from other online databases and updated monthly. These are supplemented by information gathered from the scientific literature, entered by carefully selected expert authors and updated annually. This author-entered information currently takes the form of short, abstract-like 'Mini Molecule Pages', now available for more than 800 molecules. However, starting in December 2003, these will be replaced by much more detailed and highly structured author-entered information that will include details of all known states and interactions of a protein as well as their functions and locations within the cell, all backed up with references to the traditional literature. In this way, we expect the Molecule Pages to become an extremely comprehensive source of information that can be searched in highly specific ways. Once enough data have been gathered, it should also be possible to use the Molecule Pages to discover new cellular signaling networks that are currently opaque to us simply because of the high degree of specialization of individual researchers, who are therefore unable to see the signaling network 'wood' for the protein 'trees'.

In addition, this richer form of author-entered information will be subjected to an anonymous peer-review process managed by NPG editors, and the resulting publications will be formally citable using Digital Object Identifiers (DOIs). The aim is thus to create a new type of publication that makes the most of the potential offered by Web and database technologies while at the same time retaining the most valuable characteristics of traditional papers.

### Lessons learned

So far this site has been successful from the point of view of Web hits. With 200,000–250,000 page downloads per month and 50,000 registered users, the site is well used.

However, any project with the scope and ambition of the Signaling Gateway inevitably creates challenges. Indeed, it was largely to better understand and overcome these difficulties that NPG first embarked on this work.

*the entire  
research  
community is  
encouraged to  
participate in  
analysis and  
interpretation*

*other data  
sets and  
communities of  
interest may  
benefit from  
this approach*

Perhaps the most obvious challenge is the basic logistical problem of co-ordinating dozens of collaborators separated by geography and time zone, and working in a wide variety of disciplines – from publishing to information technology and molecular biology. There is no single easy solution to this problem, but NPG's decision to send two of our London-based staff to work alongside AfCS developers in US for six months proved, in hindsight, to be one of enormous value. Even after they returned, it enabled a degree and quality of communication that would otherwise have been impossible to achieve.

Another challenge was one that comes up in developing almost any biological database: data schemas (i.e. the framework within which the information is captured) have to be rigid and tightly defined. By comparison, biological data tend to be extremely varied and rather loosely defined. In the end, compromises were inevitable since it is not always possible to define everything as rigidly as a computer scientist would like, nor is it possible to store in an inflexible database every nuance that a biologist would like to convey. An associated challenge was one of creating user interfaces into a richly functional database that nevertheless remained intuitive to bench biologists with little knowledge of information technology. This remains a constant challenge as we develop the site, and is likely to be so for the foreseeable future.

Finally, there is the challenge of working with partners from a different realm of society. Academic research groups and commercial companies have different priorities, time horizons and attitudes to risk. Of course,

this is exactly why both types of organization need to exist, but it can be a source of difficulty when they collaborate. For example, what might represent an acceptable delay or risk to one might not to the other. We believe that in this particular project a strong mutual desire to appreciate the other's point of view – as well a strong mutual respect that has developed during the course of our close collaboration – has all but eliminated such potential problems. But this may be the exception rather than the rule.

### Conclusions

NPG became involved in this project to use its core competencies to enhance services for scientists and to test the theory that a new type of information service could be developed using Web technology and publishing expertise. To this extent the venture can be said to have been a success, although some of the more ambitious plans to present data are still to come to fruition.

This is a project that was not expected to be profitable but has been able to cover its costs. The main benefit for NPG has come from an enhanced profile among the cell-signalling community. But in addition to this we have been able to test how easy or indeed difficult it is to present and enhance scientific data in new ways. It is likely that there are other data sets and communities of interest, which may benefit from this approach.

**Jayne Marks and Timo Hannay**

*Nature Publishing Group  
The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK*