

## The Bourne Laboratory - Scholarly Communication

*Last Update May 24, 2013*

Over the past few years our laboratory has become engaged in the broad question *can we improve the way science is communicated and comprehended?* The following is a brief description of some of our motivations and contributions to date.

Our work began in 2005 with the co-founding and Founding Editor in Chief role of the journal *PLoS Computational Biology* [1]. Originally our intent was to provide a venue for an expanding area of scientific research where the focus was on biological outcomes from computation, a focus that at the time was under represented by existing journals. Founding an open access rather than closed access publication seemed the right approach to provide the broadest access to our corpus. The implications of open access soon became startlingly apparent to us for reasons other than the obvious ones of free access and copyright retention. We now had XML marked up documents that could be used computationally in new and exciting ways, beyond the well-trodden text mining approaches. This is one opportunity afforded by open access that we have been attempting to utilize since then.

Being responsible for a major open access journal and heavily involved in providing a major free and publically available biological database (The RCSB Protein Data Bank, used by over 300,000 scientists and students per month) led us to address the question *is a biological database different than a biological journal?* The short answer is they are converging; the difference is how the relative contributions are perceived [2]. We then set about bringing down that divide through a project we termed BioLit. BioLit provided a version of PubMed Central (PMC) that was semantically enriched. Enriched through the addition of ontological terms from the OBO Foundry and access to relevant database content immediately from the journal article page [3]. Viewed from the database, in our case the PDB, BioLit content (appropriate components of the literature) are returned to the database user in a context sensitive way [4]. Other database providers are now using BioLit in the same way.

It became obvious that semantic enrichment, as described above, is best added by the author at the time the paper is being drafted, rather than generated automatically from the published text. In a collaboration with Microsoft we developed an open source plug-in to Word that provides that service [5]. What remains is the incentive for an author to use it and hence a publisher to support such semantic enrichment and provide incentives for its addition.

Scholarly communication has, in our view, been very slow in using the full power of the Internet. Publishers certainly have adopted the Internet as a distribution medium, but not fully as a tool to foster understanding. Some aspects of science are simply better conceived through video and sound than they are through the printed word and still figures, for example a complex biological experimental protocol. What we have tried to do is integrate both together in the form of a mashup. SciVee Inc. ([www.scivee.tv](http://www.scivee.tv)) is our effort to explore this combined form of communication through pubcasts (open access papers integrated with video or podcats), postercasts (posters integrated with video or podcasts) and so on [6]. About 50,000 unique users per month currently use SciVee. While SciVee Inc. was started with NSF (SGER) and NIH (SBIR) seed money, persistence is an issue and we

have a business model that provides free access to the generic corpus, but provides a fee-based service on top of that content. We believe this an example of sustainability so important to new models of scholarly communication.

We continue to write perspectives on where we think scholarly communication should be headed. For example, in the need for a unique author identifier to tag each virtual component of an individual's scholarly output [7], how open access has yet to find a killer application [8], what a publisher of the future should offer [9], and how science should be rewarded [10].

We have dubbed our work here as contributing to "Beyond the PDF." It is based on the following observed shortcomings:

1. The ability for a laboratory to manage research in the eScience era is hampered by a lack of good tools.
2. The publishing process and the research upon which that publication are based are currently disjoint entities; there is no need for this to be so.
3. As already indicated, publishing today falls short of what is possible to make the work as accessible as possible to the largest number of people.
4. Rewards for scholarship are distorted and do not reflect one's true value; in an eScience era we can be more quantitative.

Our continuing goal is to work with the community to address these shortcomings. *To crowd source the 21<sup>st</sup> century digital printing press and to openly disseminate science.* We held a workshop on this topic in Jan. 2011 at UC San Diego with over 150 attendees [11]. This was followed by a second workshop in Dagstuhl, Germany, which led to the formation of FORCE11 [12] a grass roots effort to bring like-minded people together to address these problems. Associated with FORCE11 is the FORCE11 Manifesto, which is a living document describing the scope, stakeholders and goals of the group [13]. Meetings, special interest group formation etc. continue amongst this energetic group.

[1] P.E. Bourne, S.E. Brenner, M.B. Eisen 2005 PLoS Computational Biology: A New Community Journal. *PLoS Comp Biol* 1(1): e4.

[2] P.E. Bourne 2005 In the Future will a Biological Database Really be Different from a Biological Journal? *PLoS Comp Biol* 1(3) e34.

[3] J.L.Fink, S. Kushch, P. Williams, P.E.Bourne 2008 BioLit: Integrating Biological Literature with Databases *NAR* 36(S2) W385-389.

[4] A. Prlic, M.A. Martinez, B.T. Yulich, D. Dimitropoulos, B. Beran, P.W. Rose, P.E. Bourne, J.L. Fink 2010 Integration of Open Access Literature into the RCSB Protein Data Bank Using BioLit. *BMC Bioinformatics* 11:220.

[5] J.L. Fink, P. Fericola, R. Chandran, S. Parastatidis, A. Wade, O. Naim, G.B. Quinn & P.E. Bourne 2010 Word Add-in for Ontology Recognition: Semantic Enrichment of Scientific Literature. *BMC Bioinformatics* 11:103.

[6] J.L. Fink and P.E.Bourne 2007 Reinventing Scholarly Communication for the Electronic Age. *CT Watch*, 3(3) 26-31.

[7] P.E. Bourne and J.L. Fink 2008 I am Not a Scientist I am a Number *PLoS Comp Biol* 4(12):e1000247.

[8] P.E. Bourne, J.L.Fink, M.Gerstein 2008 Open Access: Taking Full Advantage of the Content *PLoS Comp Biol* 4(3) e1000037.

[9] P.E. Bourne 2010 What Do I Want from the Publisher of the Future? *PLoS Comp Biol* 6(5): e1000787.

[10] P.E. Bourne 2010 Ten Simple Rules for Getting Ahead as a Computational Biologist in Academia *PLoS Comp Biol*

[11] <https://sites.google.com/site/beyondthepdf/>

[12] <http://www.force11.org>

[13] [http://www.force11.org/white\\_paper](http://www.force11.org/white_paper)

### Recognition

- 2010 Jim Gray eScience Award
- 2009 Benjamin Franklin Award

### Related Contributions

- Scientific Advisory Board Microsoft Research
- Scientific Advisory Board PubMed Central

### Press Releases

- Joint Tool Development with Microsoft and Science Commons - [O'Reilly March 11, 2009](#)
- A YouTube for Scientists - [Xconomy | San Diego November 24, 2008](#)
- Video sites lets scientists show off experiments [msnbc December 2, 2007](#)
- Science Gets its YouTube on with SciVee - [ars\\_technica September 10, 2007](#)

### Keynotes

- See [http://www.sdsc.edu/pb/pages/invited\\_lectures.php](http://www.sdsc.edu/pb/pages/invited_lectures.php)