

# This revolution will be digitized: online tools for radical collaboration

Chris Patil<sup>1,\*</sup> and Vivian Siegel<sup>2,\*</sup>

**What if everyone in the world were in your lab – a ‘hive mind’ of sorts, but composed of countless creative intellects rather than mindless worker ants, and one in which resources, reagents and effort could be shared, along with ideas, in a manner not dictated by institutional and geographical constraints?**

## The cool fantasy: radical sharing in a global ‘hive mind’

Most scientific work is performed in the context of a laboratory group or small-scale collaboration – islands of cooperation floating in an ocean of competition. For any given project, it is nearly certain that the participants represent a tiny sampling of the available expertise on the subject. Groups of colleagues assemble themselves largely by happy accident – past acquaintance, geographical proximity, institutional affiliation – and pursue their ambitions in relative isolation, punctuated by episodes of formalized sharing: publications in peer-reviewed journals or presentations at conferences.

Few would argue that the typical size of a research group has been optimized for the rapid progress of science as a whole. We go to the bench with the army we have, rather than the army we might want.

But let us entertain the thought that the ideal size of the collaborative unit might be much larger than the average research group of today, and that we lived in a world in which scientific efforts were organized around this principle. How might evolving information technologies allow science to progress more rapidly? In such a world, we might choose to organize scientific efforts differently: not according to physical proximity in labs or departments, but rather by aptitude, expertise and availability. Rather than thinking of projects as the virtual property of small groups, we would simply broadcast ideas (or data) until they reached the right person(s) to take the next step.

Suppose that your unique combination of training and expertise leads you to ask a novel question that you are not currently able to address. You advertise your idea to the world, seeking others who might be able to help. You find that Miranda has an idle machine, built for another purpose, that could be modified *just so* to help answer your question, if only she had a few samples from an appropriate patient. Hugo, busy with clinical responsibilities, has no time, but has a freezer full of biopsy tissues from such patients. Steve has the time and inclination to modify Miranda’s machine and to write the scripts to drive the analysis. Polly watches the whole process to make sure that the study has sufficient statistical power. Correspondence among the interested parties could be recorded in a publicly available forum, along with data and analysis as they emerge – allowing the entire scientific world to look on and to offer advice on the framing of the question, the design of the machine, the processing of the samples and the interpretation of the results.

In other words, what if you could think a thought at the world and have the world think back? What if everyone in the world were in your lab – a ‘hive mind’ of sorts, but composed of countless creative intellects rather than mindless worker ants, and one in which resources, reagents and effort could be shared, along with ideas, in a manner not dictated by institutional and geographical constraints?

What if, in the process, you could do actual scientific research? Granted, it would be research for which no one person (or group) could take credit, but research all the same. Progress might even occur more rapidly than it does in our world, where new knowledge is shared in the form of highly refined distillates of years of work.

<sup>1</sup>Chris Patil is at the Buck Institute for Age Research, Novato, California, USA

<sup>2</sup>Vivian Siegel is at the Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

\*Authors for correspondence (e-mail: cpatil@buckinstitute.org; dmmeditor@biologists.com)

### The reality check

Achieving this vision would require a series of significant changes in outlook. Rather than conceiving of communication as an operation between individuals or within small groups, we need to envision it as an operation between individuals and the world: you contribute your idea within the informational ecosystem, and there it stands, for others to develop or elaborate as they see fit, making their own contributions to the system in the process – like Wikipedia, but for scientific concepts.

Beyond raising concerns about the philosophy of communication, our utopian fantasy ignores important aspects of human nature. In any real world, finding collaborators would require a great deal more than shooting questions into the void and cocking an ear for the echo. In particular, in order to find a colleague with exactly the right complement of skills, interest and dependability, we need not only openness but trust. Within a laboratory group (at least, in a functional one), trust is part and parcel of lab citizenship; we and our colleagues voluntarily suspend our competitive urges in order to create a cooperative (and mutually beneficial) environment. In the wider world, however, the presumption is reversed: we tend to be cagey and suspicious in our interactions with other scientists. When we step outside the laboratory door, we transform from Musketeers ('All for one...!') to Mulder and Scully ('Trust no one!').

But, assuming that we 'Want to Believe,' how do we establish trust among strangers? Often, an introduction or referral by a mutually trusted third party is required. Conventional means for making such introductions, however, are slow and haphazard, sometimes requiring plane tickets, face-to-face meetings and late nights at conference-hotel bars. Our world of radical sharing will require tools that allow robust networking at low resource cost.

Another clash between utopia and human nature occurs at the level of publicly sharing preliminary data. In particular, during the period of transition between the status quo and the glorious future, openness may be provably irrational from a game-theoretical standpoint. If I share my data but my competitors do not, I've laid all of my cards out on the table, whereas others play theirs close to the vest – a bad bet under any circumstances. At best, my openness allows my adversaries to strategize; at worst, it allows them to steal my ideas. Perhaps the term 'stealing' is too harsh: in the words of our estimable thesis advisor, Peter Walter, 'you can't unthink a thought.' Once an idea is in the field, can anyone be blamed for reacting to it in a way that is personally optimal? We already live with this moral conundrum every time we agree to review papers and need to balance the expectation of confidentiality with our own desire to shape our own future plans on the basis of the best and most current information. Radical sharing will require ways for individuals to protect themselves from the occasionally deleterious consequences of rational self-interest.

Perhaps most importantly from a practical perspective: information doesn't share itself. From establishing an open record of preliminary discussions to freely disseminating experimental results, each step in the process requires an infrastructure. A framework, composed of software and web tools, is necessary in order to empower individual scientists to share information without each of them having to write the enabling code from scratch.

### The tool chest

A great many of the logical components of the sharing circuitry already exist, under the umbrella of 'Web 2.0' – a movement in web development that seeks to enable users to share information, collaborate and post self-generated content online.

Several efforts are already underway to create Facebook-like connections between potential colleagues who haven't yet met. These networks (e.g. SciLink, Epernicus, Academia.edu and the Nature Network) are driven primarily by prior associations,

demonstrated expertise and current interests, and these could easily be adapted to establish the circles of trust required for meaningful collaboration.

In addition to trusting each other, we must also trust the repositories where scientists deposit openly accessible descriptions of their experimental designs and results. In the most prominent of these, OpenWetWare, each notebook entry is associated with a specific time stamp and history of editorial changes to the content. To the extent that the community has confidence in these records, the risk of outright thievery is diminished: if I can demonstrate that I did an experiment first, it will be difficult for an unsavory competitor to claim precedence.

After new knowledge is published, it can be shared through social bookmarking systems such as de.licio.us and Connotea, which allow scientists with similar interests to bring relevant new information to each others' attention – serving a filtering function that we discussed in a previous article (Patil and Siegel, 2009) and also accelerating the spread of novel ideas. Post-publication tagging and annotation, along with reader commentaries that are permanently linked to articles (e.g. the 'post-publication peer review' at *PLoS ONE*<sup>3</sup>, which takes the 'Letter to the Editor' concept to the next level), have the potential to turn publications from static 'snapshots' into dynamic entities that change over time.

Meanwhile, microblogging and forum tools – both of the threaded (organized by topic; e.g. FriendFeed) and unthreaded (e.g. Twitter) varieties – can provide a means to create a public, persistent record of the generation and development of scientific ideas. Heretofore, this ephemeral process has been difficult to capture, but now we could potentially observe the evolution of concepts at very high resolution, one inspiration at a time.

### **They have built it; why have we not come?**

Very well then: the requisite components are in place. Where is our hive mind? Is their potential being ignored, or are there barriers to adoption? There are several reasons to expect a priori that scientists might resist using new tools.

The principle among these is time – specifically, we have none to spare. Therefore, we won't start using a new communication widget unless its value has already been demonstrated. Unfortunately, perniciously, many Web 2.0 tools don't accrue value unless people spend time with them. Social bookmarking is not very social if you're the only one bookmarking; forums are lonely when you're the only one there. Systems that take advantage of the wisdom of crowds only work if there is a crowd – and why would the crowd assemble if there is not yet any wisdom?

Social networking tools also suffer from a variant of the 'no one will go there until everyone goes there' problem – the 'me too' dilution factor. Just as in the social/job space (Facebook, LinkedIn, MySpace, Bebo), there are myriad networks to choose from and many are too similar to distinguish. To a new user with limited time, it's not obvious whether to try and join multiple networks, arbitrarily choose one, or wait for a clear winner to emerge.

There is more to fear than merely wasting time. Some of the tools of open science are genuinely scary, particularly when we consider adopting them before our peers. For example, although traditional peer review is safely anonymous, we might avoid commenting publicly on a published paper in order to escape the ire of colleagues. On another front, we might resist opening our notebooks out of fear of being scooped, tipping our hand to a competitor, or being caught in a logical error. Despite assurances to the contrary, these fears cannot be entirely assuaged by the idea of time-stamped entries (mentioned above), in part because that safeguard requires every stakeholder in the status quo to buy in before it has teeth. Would you rather be the smiling scientist holding up her *Nature* paper or the one weeping while pointing furiously to the time stamp in an online notebook?

There are other forces at work as well: ignorance of the tools available, which may or may not be a failure of marketing on the part of the tool builders; generational differences in comfort with computer-based communication; and simple inertia. Taking these all into account, it does seem unlikely that anyone, other than science bloggers and the denizens of networking forums, would be prone to explore and adopt the tools of the interactive web.

### **E pur si muove! (And yet it moves!)**

And yet, despite these considerable obstacles, there are signs of life emerging in the open-science universe.

The simplest form of catalysis – the facilitation of meaningful networking that could provide a basis for shared effort – is well underway. On her blog *'I Was Lost But Now I Live Here'*, bioinformaticist, open-science advocate and self-proclaimed introvert Shirley Wu described how new tools help her meet other scientists in the 'sandbox for grownups'<sup>4</sup>: 'social networking is not new, of course, but until last year I hadn't fully appreciated the ability of the web to bring people together as real functional communities de novo. ...But then I found FriendFeed and with that I discovered that there are tons of clever, interesting people out there who get excited about the same things as I do.' These communities, initially formed from electrons and spare time, can sometimes achieve physical form: 'in no time, I felt like I was part of a real community, and the best part is that it really is, well *real*. As in...who's going to be at XYZ conference? And a dozen or so people chime in, meet at the conference for the first time, and end up publishing a paper about it.'

Similar tools not only bring people together but also allow them to answer questions more efficiently. In public forums like FriendFeed, a sort of low-level distributed intelligence is emerging based on a simple premise: if I have a conceptual question, there are probably several people in the world who have the expertise and perspective to answer it more quickly and easily than I could, and who might even be willing to spare the time to help a stranger. So I ask my question, someone answers, and it's all over in a few minutes.

In fact, this distributed intelligence played a significant role in the writing of this very article. Our investigation began with us posting on FriendFeed and receiving responses from a number of forum participants<sup>5</sup>. One of the respondents, Drexel University chemist Jean-Claude Bradley, pointed out that this sort of interaction can provide not only colleagues for one-off projects but also help to establish and maintain long-term collaborations. Bradley himself serves as an example; he uses a combination of microblogging, social bookmarking and other tools to coordinate a distributed effort to collect solubility information for a wide range of chemical compounds<sup>6</sup>. He also maintains an online spreadsheet<sup>7</sup> of collaborative projects that have been initiated or conducted using the interactive web.

After a project is finished, interactive tools may also be useful to disseminate results. Partial proof of principle has been provided by the attendees of the International Society for Computational Biology (ISMB) 2008 conference, who used microblogging to cover the meeting far more comprehensively than would have been possible using the traditional, centrally organized 'meeting review' approach (Saunders et al., 2009).

Taken together, the early adopters of Web 2.0 in open-science applications are inching toward the creation of a movement. BioGang<sup>8</sup>, 'an informal, distributed collection of geeky life scientists', was inspired by a blog post<sup>9</sup> and continues to pursue its mission ('to try and think of cool problems and ideas that can be solved collaboratively') using the communication methods of the social web. There, the emphasis is less on long-term models than on the concept of 'bursty work', i.e. using networks to define, and then solve, small-scale problems of interest in a distributed way, assembling relevant

expertise on the fly and maintaining specific ‘collaborations’ for only as long as it takes to move the next step forward.

Those who are actively incorporating the interactive web into their science have grappled with the question of trust that we’ve raised throughout this piece. In some cases, their answers reveal that they are operating within a novel paradigm: openness doesn’t require trust; rather, true openness makes trust irrelevant. In an online interview, Jean-Claude Bradley addressed the issue of establishing trust between colleagues by ‘un-asking’ the question: ‘Within the context of scientific collaborations, one of the points that I am very keen on is that there should be no trust – we should use proof as much as possible to evaluate. And you can only do that if you have access to all the data.’ At the same time, he acknowledges that it’s best if we can reassure ourselves with the *potential* to confirm everything. ‘It does not mean that it is practical to look up every detail, but being able to when something is not adding up. ...As you get to work with people, you can develop a form of trust over time where you can estimate how careful someone is and how they work. But anyone can make a mistake.’

But what about trust of those outside collaborative groups, however distributed and informal those groups might be? Are the adherents of Web 2.0 science concerned about competitors stealing their data? Again, Bradley’s answer subverts the dominant paradigm, taking a leaf from a free spirit of an earlier era (Hoffman, 1976): ‘I had no concerns because I chose an area (malaria, solubility) where I would be happy if someone used my results. IP [intellectual property] is not an issue so that simplifies things greatly.’ In his view, even rather malicious behavior would yield useful data, providing a useful opportunity to think about and test the new system. ‘I keep waiting for someone to plagiarize or misrepresent what they have done because I think it would make a great case study and something to blog about. But that hasn’t happened yet.’

### Boom or bubble?

So, the new tools are being adopted by some, but are not yet predominant. Do the early adopters represent the beginnings of an exponential expansion in novel online collaborations or merely a flash in the pan? Amazon.com or pets.com? In the early days, exponential growth is notoriously hard to distinguish from a flatline.

In order to spread more widely, the technologies that enable radical sharing – and indeed, the entire philosophy of an open, distributed science – must attract new adherents, and in the process overcome considerable barriers to adoption. Ultimately, in order to demonstrate their worth to an occasionally skeptical (and invariably busy) scientific community, the new tools must provide an affirmative answer to one question: *can they help us do better science?*

*In the interest of full disclosure: C.P. is a member of the PLoS ONE editorial board. This article is freely accessible online from the date of publication.*

### REFERENCES

- Hoffman, A. (1976). *Steal this book*. New York: Amereon Ltd.  
 Patil, C. and Siegel, V. (2009). Drinking from the firehose of scientific publishing. *Dis. Model. Mech.* **2**, 100-102.  
 Saunders, N., Beltrão, P., Jensen, L., Jurczak, D., Krause, R., Kuhn, M. and Wu, S. (2009). Microblogging the ISMB: a new approach to conference reporting. *PLoS Comput. Biol.* **5**, e1000263.

### Internet resources

- <sup>3</sup>[http://scienceblogs.com/clock/2008/08/postpublication\\_peerreview\\_in.php](http://scienceblogs.com/clock/2008/08/postpublication_peerreview_in.php)  
<sup>4</sup><http://shirleywho.wordpress.com/2009/03/17/when-the-net-works-for-networking/>  
<sup>5</sup><http://friendfeed.com/e/6792b5b3-8c7e-4454-b210-19c44e725fdd/Has-anyone-used-Web-2-0-tools-like-FriendFeed/>  
<sup>6</sup><http://friendfeed.com/jcbradley>  
<sup>7</sup>[https://spreadsheets.google.com/ccc?key=plwwufp30hfqwGp0\\_E3iKNQ&hl=en](https://spreadsheets.google.com/ccc?key=plwwufp30hfqwGp0_E3iKNQ&hl=en)  
<sup>8</sup><http://biogang.openwetware.org/>  
<sup>9</sup><http://mndoci.com/blog/2008/05/30/a-global-informatics-collaboratory/>