

WIKIOMICS

Pioneering biologists are trying to use wiki-type web pages to manage and interpret data, reports **Mitch Waldrop**. But will the wider research community go along with the experiment?

Alexander Pico remembers just when the idea hit him. In January 2007, he and his boss, Bruce Conklin, were discussing how to push their software tool for visualizing intracellular signalling pathways to the next level of interactivity — when Pico blurted out, “What we really need is a wiki!”



Well, it was an original thought at the time, says Pico, a software engineer in Conklin's laboratory in the Gladstone Institute of Cardiovascular Disease at the University of California, San Francisco. In retrospect, it was one of those ideas that strikes everywhere at once. As soon as he and his colleagues started giving talks about ‘WikiPathways’, as they called their project, someone in the audience would invariably say, “Ah — we had the exact same idea.”

Scientist-edited interactive ‘wiki’-type websites have proliferated over the past year or so (see table), to the point where research-

ers have begun to joke about the new science of ‘wikiomics’. All the sites are modelled on the popular user-edited, online encyclopedia Wikipedia, and all aim to help biologists turn the data flooding into the large public gene and protein databases into useful knowledge.

The flood is going to rise even faster, says Amos Bairoch, executive director of the Swiss Institute for Bioinformatics in Geneva and creator of Swiss-Prot, a predecessor to the international protein sequence database UniProt: “As the price keeps going down, we’re reaching the point where every genome that can be sequenced, will be sequenced,” he says.

Ultimately, that could mean the genomes of most of Earth's 1.8 million named species, along with individual variants produced by projects such as the ‘1000 Genomes’ programme for humans. And there's all the rest of the quantifiable information about life on Earth — data on protein structure and function, biomolecular interactions, signalling

and metabolic pathways, and much more. The challenge is to make sense of the deluge.

Teams of scientist-annotators at the data repositories make valiant efforts to keep up, and bioinformatics programmers devise increasingly sophisticated annotation algorithms to help. Scientists write review articles and textbooks to make sense of it all. But it's still not enough.

Hence the proliferation of wikis, which have the potential to vastly multiply the number of annotators and bring in the most interested expertise: “The best people to do annotation are the researchers in the laboratories, the people who are producing this knowledge in the first place,” says bioinformatician Barend Mons at the University of Rotterdam in the Netherlands. Mons is one of the prime movers behind WikiProfessional Life Sciences, a site that links publications on a given topic and enables users to add their own annotations.

But will the bench scientists participate?

“This business of trying to capture data from the community has been around ever since there have been biological databases,” says Ewan Birney of the European Bioinformatics Institute in Hinxton, UK. And the efforts always seem to fizzle out. Founders enthusiastically put up a lot of information on the site, but the ‘community’ — either too busy or too secretive to cooperate — never materializes. So how do the wiki proponents know that this time around will be different?

They don’t. “This is an experiment,” says Pico, echoing just about everyone in the wiki movement. He is optimistic, however. This June he attended a workshop at the University of California, San Diego, on new communication channels in biology. “Many of the people had come to this from prior attempts,” he says, “and were very sober about the challenges.” From ensuring usability to ensuring users, these challenges go beyond the technical. As the developers of WikiPathways and several others have found, a truly cooperative web-based community requires a change in thinking — a shift in the way scientists work and in the way they get credit for that work.

Take but no give

Conklin’s original idea for software to help biologists visualize and draw pathways grew from his research exploring how hormones and their receptors direct tissue development. Pathway diagrams are flow-chart representations of the interactions between genes, proteins or metabolites involved in a particular cellular function, such as the response to an external signal. They enable researchers to interpret the biochemical functions of individual molecules in the broader cell-biological context. One protein might have a very limited function, marking another protein for destruction, for example. But seeing its place in a pathway gives a clue to the physiological significance of that tiny action and offers clues to the functions of similar-looking proteins.

Better still, says Conklin, pathways help make sense of DNA microarray data on gene expression. If administering a drug enhances the expression of a set of genes all involved in the same pathway, say one causing cell death, then that’s an important clue to what is going on.

So, back in 1999, Conklin’s lab began to develop software that would make it easier to visualize and modify cellular signalling pathways. Known as the Gene Map Annotator and Pathway Profiler (GenMAPP), it offered free, downloadable software that could turn a database of interactions into a pathway diagram, and also enabled the user to add a new entry to the database simply by

sketching in a new reaction. GenMAPP also offered the capability to match microarray gene-expression data against an extensive library of known pathways and identify the most likely matches.

To get the library started, says Conklin, “I went to Amazon.com and bought \$900 worth of textbooks. Then my students and I flipped through and redrew the pathways we found there by hand, making electronic versions.” They figured the tedium was worth it. Their library would grow fast, as soon as researchers who downloaded the drawing tool began uploading their own pathways.

But the team was overly optimistic. The GenMAPP drawing tool proved popular, and in the nine years since the launch, it’s been downloaded 17,000 times. But when it came to giving back to the library — the rate wasn’t so great. Only about 30% of the 557 pathways in the current GenMAPP library have come from outside the developers’ own labs.

There were some enthusiasts. The group run by Chris Evelo, head of the department of

bioinformatics at the University of Maastricht in the Netherlands, was such an active contributor that it became a formal collaborator on GenMAPP in 2003. “But it was frustratingly slow,” says Conklin. “We’d see publications with pathways created using our software, but half the time people wouldn’t submit them back to us.”

Make it easy

Two things broke the impasse, says Conklin. In 2005, the lab was approached by the developers of Cytoscape, an open-source software platform for very powerful, very high-end network analysis, much used in systems biology. They liked GenMAPP’s layout, with its easy-to-use sketching capability, which they wanted to incorporate into Cytoscape, where the pathway drawings were abstract and mathematically elegant, but hard for the uninitiated to understand.

Conklin and his group were happy to oblige. “Cytoscape turned out to be supported by a very robust open-source community, which we didn’t have,” says Conklin. “Here were people



A SELECTION OF WIKI-STYLE INFORMATION COOPERATIVES	
Name	Content
EcoliWiki http://ecoliwiki.net	The community-annotation component of EcoliHub, which integrates information from 19 websites relevant to <i>Escherichia coli</i> .
Gene Wiki http://en.wikipedia.org/wiki/Gene_Wiki	Not a site by itself, but an effort to create or update Wikipedia pages on some 9,000 human genes, using data and text from primary gene and protein databases
OpenWetWare http://openwetware.org	One of the oldest and largest scientist-edited sites, OpenWetWare has evolved into an active social network for biologists, hosting blogs, links to labs and special-interest groups and, of course, a long list of lab protocols
PDBWiki http://pdbwiki.org	A community-annotated knowledge base of biological molecular structures in the Protein Data Bank (PDB)
Proteopedia www.proteopedia.org	An interactive encyclopedia of three-dimensional structures of proteins, RNA, DNA and other molecules
Topsan www.topsan.org	The Open Protein Structure Annotation Network: a wiki devoted to protein three-dimensional structures, including some not yet deposited in the PDB
WikiGenes www.wikigenes.org	A specially designed wiki that tracks — and thus allows scientists to get credit for — every contribution that’s made
WikiPathways www.wikipathways.org	A wiki devoted to the community curation and enhancement of biological pathways
Wiki Professional Life Sciences www.wikiprofessional.org	An attempt to create a ‘Concept Web’ by linking publications on a given ‘concept’ and enabling user annotation

coming out of the walls, offering us all kinds of software solutions.” The GenMAPP team became active participants — Conklin now sits on the Cytoscape board — and soon decided to revamp their own drawing tool entirely; the next GenMAPP release, due out in 2009, will essentially be a slightly specialized version of Cytoscape.

That involvement led to the second innovation, says Pico. “The Cytoscape team was using a wiki to coordinate their work,” he says. “And that was my first experience with the idea.” So he decided to install a wiki in Conklin’s lab for internal use. “These were mostly wet-lab biologists, and what impressed me was that even the least technically inclined people in the group picked it right up,” says Pico. “Even biologists who would never add to a website would add to the wiki — it was easy and fun.”

Sketching the idea

So the next big idea was almost inevitable — a public wiki interface for GenMAPP to make it easier for researchers to contribute their new pathways.

As inspiration hit on that January day, Pico sketched out his idea. It would need an online version of the GenMAPP drawing tool, instead of a separate piece of software to download, and a one-click submission of a finished pathway to the library instead of a separate uploading process. When he e-mailed Evelo with the idea, two of Evelo’s graduate students, Martijn van Iersel and Thomas Kelder, replied. Surprise — they’d had the same idea.

Kelder and van Iersel in Maastricht and Pico and Kristina Hanspers at the University of San Francisco became the design group for WikiPathways. A top priority was to make the site very easy for bench scientists to use. Like most of the other wiki-inspired biosites, WikiPathways does this by using the open-source MediaWiki software that underlies Wikipedia. As EcoliWiki creator James Hu of Texas A&M University in College Station puts it, “We didn’t want to ask young scientists who were already editing Wikipedia to learn a new interface.”

The WikiPathways team did need tools not available on Wikipedia itself. “We completely gutted the MediaWiki text-editing functionality and replaced it with new applets that would represent pathway information graphically,” says Pico. The diagram is linked behind the scenes to a structured database of biochemical interactions, he says, but the goal is to make drawing the pathway on screen as easy as drawing it on a napkin. “And then once you’re done, it’s immediately available to you — or to the world. You can e-mail the link and do

collaborative editing with biologists globally, which is impossible with GenMAPP or any other tool that’s on your personal machine,” says Pico. “The wiki can put all of you on the same drawing board.”

A prototype WikiPathways was up and running by spring 2007. By autumn the team felt confident enough to promote the site more widely. And in January 2008 they got their first pathway contributed by a researcher they didn’t know directly. “I consider that the birthday,” says Pico.

By mid-summer 2008, WikiPathways had some 350 registered users, of whom 50 or so had made changes to at least one pathway. “It’s already more contributors than we’d gotten over the past nine years,” says Pico. And for several weeks after July 2008, when they pub-

lished a description of WikiPathways in the journal *PLoS Biology* (see A. R. Pico *et al.* *PLoS Biol.* 6, e184; 2008), the average of one new pathway contributed per month jumped to a new pathway every other day. The hope is that at some point soon, says Pico, “we’ll reach a tipping point, a critical mass, where people from areas of biology we know nothing about will start participating in the whole cycle of revision and correction while involving us less and less — and it will become self-sustaining”.

for Biotechnology Information’s Entrez Gene, together with links to data repositories and publications, and to Wikipedia’s rich resource of pages on diseases and physiology. Gene Wiki entries are already showing up on the first page of Google search results for particular genes, says Su. “And our hope is that some number of readers will actually stay to make an edit,” he says. “It could be as trivial as fixing a typo, or as substantive as summarizing a new paper in the literature. But it will start a positive feedback loop by making the page that much more useful.”

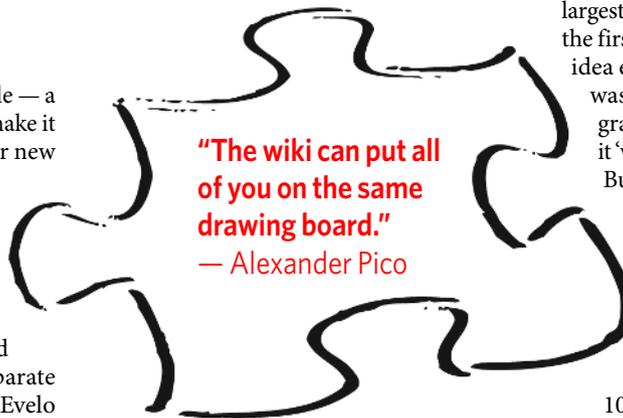
Building critical mass — that’s the real challenge in the wiki game, as everyone is acutely aware. It’s also a mysterious process that requires timing and luck just as much as skill. Wikipedia, for example, didn’t become the largest collaborative site on the planet by being the first. That honour goes to a programmers’ idea exchange called WikiWikiWeb, which was developed in 1995 by American programmer Ward Cunningham. (He named it ‘wiki’ after the Hawaiian word for ‘quick.’)

But Wikipedia, founded in 2001 by developers Jimmy Wales and Larry Sanger, was among the first to offer a free service — knowledge aggregation — that was useful to essentially any literate person on the planet. It proceeded to grow exponentially, to the point where it now claims more than

10 million articles in more than 250 languages — roughly a quarter of them in English. Wikipedia has also acquired the classic ‘long tail’ of contributors, with a comparative handful of people making lots of edits, and a multitude who make only a few.

The science wikis face a tougher challenge in building critical mass, if only because they’re aiming at a much smaller audience. One obvious strategy is to avoid fragmenting that audience. As Evelo points out, “biologists aren’t going to work on a dozen wikis to see which will survive”. They are going to want the various wikis to be interoperable and mutually supporting, so that the data they enter in one can be easily ported to another — or will even flow to all the appropriate sites automatically.

It should help that so many of the sites are based on the same MediaWiki software. That gives them the potential to act as one big open-source community, sharing code and improvements. And it’s not just potential, adds Pico: “We’ve been in close contact with Jim Hu and the Ecolihub folks about making our wikis interoperable.” Ecolihub is the ‘parent’ website of EcoliWiki, providing access to vast amounts of information on the bacterium *Escherichia coli*. Also critical to interoperability will be a standard language that can be understood by all



Critical mass

WikiPathways is a stand-alone site, but a few of the new bio-wiki sites are tapped into Wikipedia directly. Earlier this summer, a team led by Andrew Su at the Genomics Institute of the Novartis Research Foundation in La Jolla, California, launched a software ‘robot’ that systematically goes through Wikipedia creating or amending entries for every human gene that has been studied to any significant degree — some 9,000 in all. The result is Gene Wiki: a collection of Wikipedia pages in a standard format, populated with an integrated suite of information culled from the National Center

the databases. In the realm of pathways, says Pico, the closest to that right now is BioPAX, an XML-based standard for the exchange of pathway and interaction information. “We’re planning on converting our system to it.”

Interoperability is only part of the equation, however. Few scientists will contribute to these sites out of altruism. They need tangible incentives — starting with a real benefit to their day-to-day research.

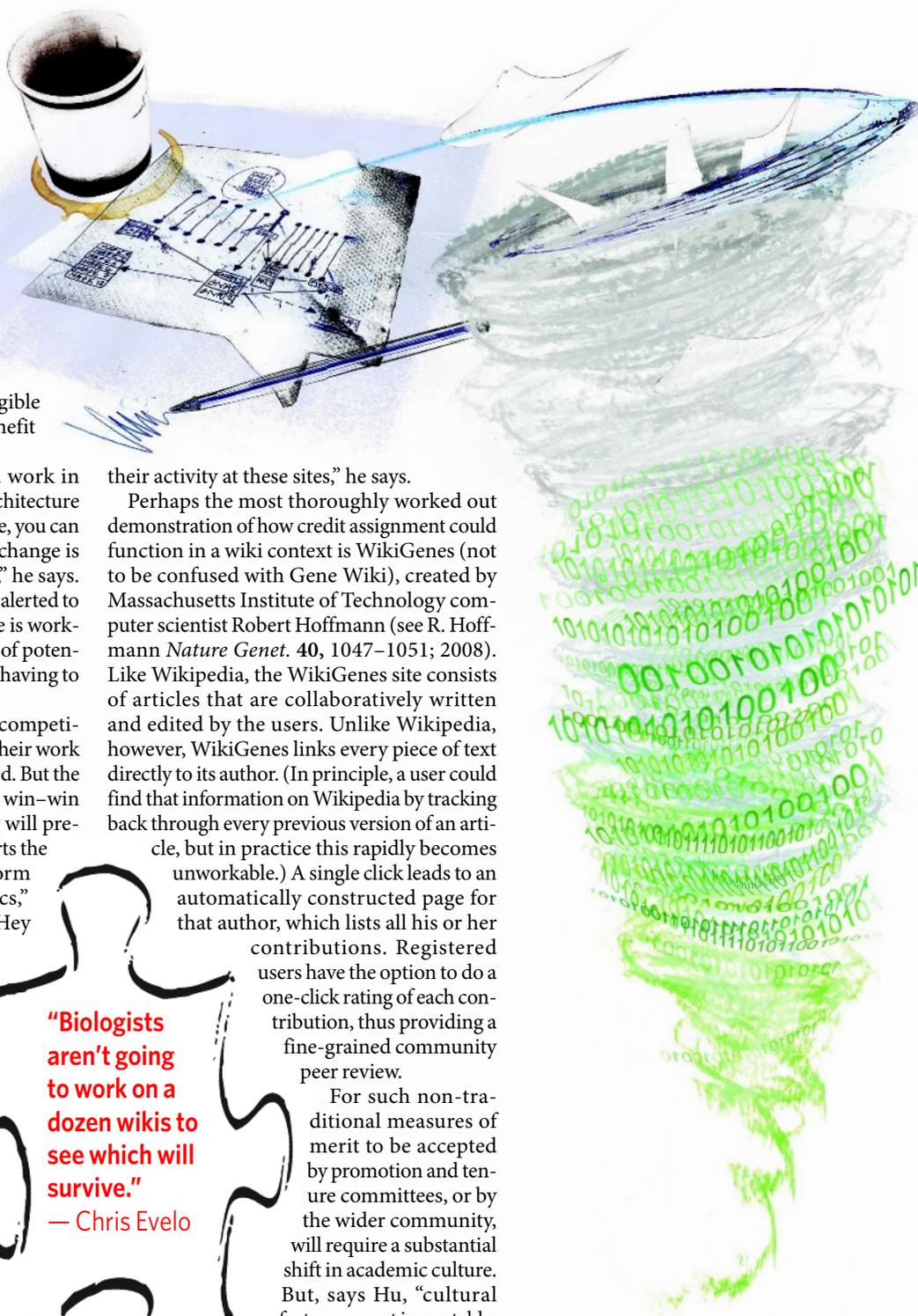
Giving them that is definitely a work in progress, says van Iersel. The wiki architecture offers some possibilities. “For example, you can sign up to be e-mailed whenever a change is made to a page you’re interested in,” he says. So a researcher could immediately be alerted to any new findings in an area he or she is working on, not to mention the existence of potential collaborators (or rivals), without having to wait for a paper to come out.

There will always be some hypercompetitive fields in which people will keep their work under wraps for fear of getting scooped. But the hope is that for most researchers, the win-win dynamics of real-time data sharing will prevail. “Community annotation supports the natural process in which people form intellectual networks around topics,” says Mons. “The system tells me, ‘Hey — if you’re interested in ABC, you’d better look at XYZ, as well.’ And that will become part of the workflow of a scientist’s life.”

Due credit

Academic culture being what it is, however, the wiki sites will have to crack the credit-assignment problem, and provide some way for scientists’ efforts there to be identified, recognized, cited and shown to funding agencies and tenure committees. Without a solution, says Hu, wiki-based community annotation will get nowhere.

“Everybody gets excited by the idea,” he says, “but then it always falls off the table, because it’s not one of the things that pays the rent.” Pico couldn’t agree more. At the San Diego workshop, “we had break-out sessions, lunches, lots of brainstorming trying to think of metrics for scientists to quantify



their activity at these sites,” he says.

Perhaps the most thoroughly worked out demonstration of how credit assignment could function in a wiki context is WikiGenes (not to be confused with Gene Wiki), created by Massachusetts Institute of Technology computer scientist Robert Hoffmann (see R. Hoffmann *Nature Genet.* **40**, 1047–1051; 2008). Like Wikipedia, the WikiGenes site consists of articles that are collaboratively written and edited by the users. Unlike Wikipedia, however, WikiGenes links every piece of text directly to its author. (In principle, a user could find that information on Wikipedia by tracking back through every previous version of an article, but in practice this rapidly becomes unworkable.) A single click leads to an automatically constructed page for that author, which lists all his or her contributions. Registered users have the option to do a one-click rating of each contribution, thus providing a fine-grained community peer review.

For such non-traditional measures of merit to be accepted by promotion and tenure committees, or by the wider community, will require a substantial shift in academic culture. But, says Hu, “cultural factors are not immutable.

If we can promote various incremental changes, then eventually this will take off.”

In the meantime, the wikis still have a lot of challenges to face — not least the need to prove to funding agencies that they are worthy of long-term support.

And that is why it is all very much still an experiment. “Community intelligence is a new concept for biology — and in broader society — and we certainly don’t claim to have the final

answer,” says Su. Still, the more mechanisms for harnessing community intelligence, the better: “The community will essentially vote on which model will be the most useful, and the beauty is that they will vote with their participation,” he says. “The only question is which model will resonate best.”

Mitch Waldrop is Editorials and Features editor for Nature.

See also Editorial, page 1.



“Biologists aren’t going to work on a dozen wikis to see which will survive.”
— Chris Evelo