What's in a name?

An international author identification system could allow scientists to receive credit for all their scientific contributions and would solve the problem of identity in a world of limited surnames

In *Romeo and Juliet*, William Shakespeare (1546–1616) asked: “What’s in a name? That which we call a rose / By any other name would smell as sweet” (2.2.1–2). But, unlike Shakespeare’s characters, whose names were a burden to them, names for scientists are extremely important and are attached to discoveries, publications, careers and even fame. In her poem *Sacred Emily*, American poet Gertrude Stein (1874–1946) wrote, “Rose is a rose is a rose is a rose”, which is reminiscent of the trouble that scientists face. However, unlike two roses, two scientists with identical names are extremely different. Indeed, this growing problem of how to unambiguously identify members of an ever-growing international community has triggered a serious debate.

...mistaken identity has resulted in the wrong person being invited to work on a project [...] or to undertake the peer review of an article

Matthew Falagas, Director of the Alfa Institute of Biomedical Sciences (Athens, Greece) and Adjunct Associate Professor in the Department of Medicine at Tufts University School of Medicine (Boston, MA, USA) described the problem. “Every country has its own common names: Smith and Jones in English-speaking countries; here in Greece we have Papapapoulous; in China, Li is very common,” he said. “When you enter the last name, or even with the initial of the first name, you may produce thousands of papers; you cannot understand who produced what. The result is that you have difficulties finding the best collaborator or the best person to ask to do peer review.” Falagas, who serves on the editorial board of the scientific journal *PLoS One* (San Francisco, CA, USA), said that he has heard of cases in which mistaken identity has resulted in the wrong person being invited

to work on a project, appear on a television programme or to undertake the peer review of an article.

Names are also important with regard to the funding of research. Kyle Brown, who founded ResearchCrossroads (San Mateo, CA, USA; www.researchcrossroads.org), explained that the company started as a database to help funding agencies and researchers look up who was getting money from government and private foundations, including the National Institutes of Health (Bethesda, MD, USA), the National Science Foundation (NSF; Arlington, VA, USA) and the Community Research and Development Information Service run by the European Union (CORDIS; Brussels, Belgium). Brown discovered the confusion inherent in similar names almost immediately as researchers contacted him to try to correct their information posted on the ResearchCrossroads website. “We struggle with this constantly,” he said. “I get e-mails every day or two saying: ‘My profile is combined with this other guy’s profile. Can you help us untangle them?’ And we’ll do that.” But he added that this is often a tough assignment as scientists can change jobs and even names, and do not use their middle initials consistently.

The same problem applies to publishing, in which it is not always clear whether the authors of different articles are the same person or not. “I’d love a user of Nature.com to be able to click on an author’s name and to be able to see a list of everything that we publish by them,” commented Timor Hannay, Publishing Director of Nature.com (London, UK). “And that kind of thing, which seems really trivial, should be very straightforward, but actually isn’t because we don’t have identifiers associated with them [...] We’ve got a world in which scientists have assigned numbers to all kinds of things: to genes, to species, to stars, to molecules, to the articles they write. The one thing they left out was themselves,”

he said, adding, "it does create real problems for us as publishers wanting to provide certain services."

...younger scientists [...] could make a name for themselves in the scientific community even before they publish their first papers

Yet, Hannay also has an idea of how to solve this problem: "A global author ID does bring you the same benefits that you already have from [a] unique article ID, and you can locate an article very quickly and easily online if you know what its DOI [digital object identifier] is." However, despite the elegance and simplicity of the idea, introducing a unique author ID for scientists is anything but a simple measure, as Peter Binfield, Managing Editor of *PLoS One* commented: "It's not a simple problem at all. I have worked at a few publishing companies and nobody is able to make unique author IDs happen within their own databases. [Authors] may submit with a different middle initial, or [they may have] moved institutions. Even our internal databases get incredibly messy and cluttered. It's hard to keep anything consistent."

But, the idea itself is appealing and, as Brown noted: "What is needed is a serial number or social security number for scientists." To this end, several publishers, scientists and other interested parties have begun to make proposals on this, and the related problem of assigning and recording credit for scientific contributions. Ernst Hafen, Professor at the Institute for Molecular Systems Biology at the ETH Zürich in Switzerland, and Manuel Peitsch, Professor of Bioinformatics at the University of Basel in Switzerland, have already circulated a draft white paper in which they call for the creation of a unique author identification (UAID), which would be used to identify the contributions that scientists make in both their papers and database entries, as well as in reviews and discussions (Hafen & Peitsch, 2008). However, in contrast to Wikipedia (www.wikipedia.org), where anyone can sign up as an author and/or editor, participation in the UAID scheme would be limited to the scientific community and the UAID would be assigned on graduating with a MSc or higher degree.

"This should on one hand reduce the risk of abuse," Hafen and Peitsch wrote in their draft White Paper, "and on the other increase the value of the academic degree since it would be the license to actively participate in the scientific community and contribute to the advancement of science. The universities could award the licenses as part of their academic degree [...] This would achieve a rapid distribution of the UAID and also increase the value of the degree. Furthermore, it is a logical step from the alignment of academic degrees [through the Bologna Declaration] in Europe and other countries including the United States" (E Hafen & MC Peitsch, unpublished data).

Their proposal was discussed in February 2008 at a meeting held by the European Molecular Biology Organization (EMBO; Heidelberg, Germany), which is exploring ways to make progress on the issues surrounding author IDs. The meeting included representatives from funding agencies, research institutions and publishers, as well as journal editors. Thomas Lemberger, Editor of *Molecular Systems Biology*, attended the meeting and commented that the Hafen–Peitsch proposal "addresses many issues that people are discussing. It has great merit. It takes these issues that everybody is talking about and tries to move the right people to take some concrete steps."

Hafen is clear that the UAID would change the way in which a scientist's contributions are rated by moving beyond the current measures of assessment. "These are the only qualifications that we have now," he said, in reference to the currently limited scope of looking at what articles a scientist has published in which journals, the impact factor of those journals and the number of citations each article has received so far. The UAID would help to credit scientists for all their other contributions such as peer review, and would benefit younger scientists as they would be able to earn credit for commenting on articles and could make a name for themselves in the scientific community even before they publish their first papers.

"The problem," Hafen said, "is that [...] if you have more experience, then you spend 40–50% of your time reviewing other people's work and giving feedback that is valuable to that work. But, there is no way that you get credit for this because it is usually done anonymously and there's

no way to trace what you have done [...] So, I want to change this. I think it's absolutely necessary to change this, to give a more balanced view of the contributions of a scientist, especially with the Web 2.0 capabilities that you have now."

He continued to explain that such systems are already in place. "If you go to Amazon [www.amazon.com], you look to see the book reviews, then you can rate the book reviews. Were they helpful or not helpful? You can also write your own reviews and so on; so there are all these possibilities out there," Hafen said. "The only thing, the only feedback or measuring system that you have in the scientific publishing world at the moment is the impact factor of the journal the paper is published in, or the citations the paper attracts, and it usually takes four to five years before you can actually measure whether the paper had a significant impact in the field or not."

"[S]cientific activity has many aspects and people do not only write papers; they also serve as reviewers, they organize conferences, they teach, they may write a blog that is very influential."

Lemberger also sees the value in such a system: "The strength of the idea from Hafen and many other people is that by providing this author ID, you give a technical means to follow all the aspects of scientific activity," he said. "[S]cientific activity has many aspects and people do not only write papers; they also serve as reviewers, they organize conferences, they teach, they may write a blog that is very influential. All these side activities are completely ignored. An author ID would give us a way to track them."

Hafen and Peitsch are not alone in their attempts to address the issue, and Binfield pointed out that several organizations are already holding discussions and trying to tackle the problem of authors with similar names and crediting contributions not published in a journal. "I think it's quite a big problem in the sense [that] if we could solve it there would be a great advance," he said. "The fact that we don't have [an author ID] doesn't stop us doing things, but I think that if we had one, it would allow all sorts of new stuff to happen," such

as the Web 2.0 ideas mentioned by Hafen and Peitsch in their draft White Paper.

One of the businesses already involved in trying to find a solution is Thomson Reuters (New York, NY, USA), who have created ResearcherID (www.researcherid.com), which they describe as a “global, multi-disciplinary scholarly research community”. Under the ResearcherID system, each researcher is assigned a unique identifier, but Binfield commented that, “[t]he problem with [ResearcherID] is that it’s proprietary; it’s locked up behind a company that is trying to make money out of this. What you really need is an open data source that everyone uses and everyone deposits into [...] Even within Nature there is Nature Network [network.nature.com],” he added, “which has places where authors can build their own profile. The problem is [that] you don’t want a system where authors want to set up their profiles in six different places. You need a single one, which is the key.”

Those who are thinking about unique author IDs agree that it will require a collaboration of scientific organizations, governments, agencies and publishers for the system to work. David Lipman, Director of the US National Library of Medicine’s National Center for Biotechnology Information (NLM and NCBI; Bethesda, MD, USA), who is in charge of NCBI’s GenBank, PubMed and PubMed Central platforms—digital archives

of sequencing data, biomedical literature and full-text scholarly articles—said that the Hafen–Peitsch proposal is among several aimed at improving scientific communications and collaboration. “When you think about it,” he said, “if you compare how we’re communicating in science compared to how we could be communicating and how people are communicating in many other areas, there seems to be a real gap in how we use the technology [...] The real challenge is how do we move from where we are to something which everybody hopes would work better. Not to say that everything doesn’t work well; it’s just that there are other kinds of opportunities.”

Yet, although some consider establishing an author ID system to be a simple matter, Lipman does not see it that way. “A lot of people don’t understand how challenging it would be to really have a universal ID system for scientists because people change their names, or they change their institution and you’d have to constantly maintain it,” he said. However, he added that he is starting to see a system emerge organically, such as the collaboration between the NLM, the Wellcome Trust (London, UK), the Howard Hughes Medical Institute (HHMI; Chevy Chase, MD, USA) and other institutions, who are establishing compatible authentication systems. “If universities and funding agencies see [an

author ID system] as an advantage, it is possible [that their own systems] could morph into that,” he said. “Ironically, for some of the reasons that make it more challenging, [an ID system] is also desirable for things that scientists don’t even think about; not high-level ways of improving scientific communications, but low-level, concrete administrative things that have to happen on [a] daily basis.”

“...if you compare how we’re communicating in science compared to how we could be [...] there seems to be a real gap in how we use the technology.”

Hannay agreed that cooperation between publishers, funding agencies and others would be necessary. “I understand people’s hesitancy in adopting a system that would appear, on the face of it, to be owned, managed and run by a single commercial entity and [wouldn’t be] open and federated and fully participatory for other parties that want to get involved,” he said. “Any system that is closed and proprietary and owned and run by one organization I suspect is doomed to fail, and in many ways, I hope it fails because I don’t think such a system would be particularly healthy for the information ecosystem.” In

fact, he added, it is likely that many ID systems will emerge; in which case, it would be desirable to make them interoperable.

Despite the cautious optimism and thought going into how to create an author ID system, there are lingering issues. As Hannay noted, some scientists are put off by the idea of a system that could track all their contributions, not only the articles they publish. "I've heard scientists complain that they get rated for everything they do: teaching, publishing and so on. They have metrics coming out of their ears and if you apply even more metrics then some of them could find that oppressive and intrusive. It might even inhibit their contributions, which of course would be counterproductive," he said. "There can be unintended consequences. You give people IDs because you want to give them credit for what they do and actually they turn around and say: 'I don't want this thing.' Scientists are human beings too, and I think we have to take into account how they feel about what we're doing in addition to the apparent objective benefits."

However, although some might worry about an Orwellian 'Big Brother' invading the scientific world, Falagas commented that, "Personally, I don't mind if all persons around the globe have to have a universal number, but I can also see how this may be abused very easily by people who would like to do so. Still, for scientists, I see more advantages than disadvantages."

Lemberger also feels that invasion of privacy is unlikely to be the real issue. "Rather," he said, "the concern is about anonymity; that everything is getting diluted in the mass of things that get done by anonymous hands. The author ID is a response to this, to bring these people and their contributions into the light [...] I don't see this [ID system] moving out of the professional/scientific sphere and endangering personal privacy. The problem is too much anonymity rather than not enough."

In the end, whether an author ID system is a universal database or a connected and compatible network of databases, it has to serve the needs of the scientific community. There is a careful balance to be struck between giving credit where credit is due and knowing everything about everyone. Where that balance lies will be up to the community and those who collaborate to make such a system a reality.

Howard Wolinsky

doi:10.1038/embor.2008.217