

ICSTI Public Conference 2007 – Session 3

Practices and perspectives of research evaluation

Research evaluation: Evolving policies and practices for assessing impact

Jerry Sheehan

National Library of Medicine

The policy context for evaluation has been changing. There is growing recognition of the links between science, innovation, economic growth, health, etc. The funding for research has increased, but with some strings attached in the form of new models of governance of public research. Competition between universities and research laboratories has increased through research rankings, and there is increased emphasis on evaluation of institutions and their research output. National research funding has grown and average annual growth in real government R&D funding between 1995 and 2005 has varied between countries but has outpaced the inflation index by a few percent. However, in large countries which are big science producers, the growth has been somewhat constrained, which has caused quite a bit of angst in research institutions and the science system itself.

Some of these constraints in research funding are linked to the changes in governance of public research and the shift from basic research funding to a process of governance of the science system. Under the old model, the government finances basic research through funding institutions. Universities and government laboratories have greater autonomy in setting scientific objectives and the scientific community is expected to take care of the funds, allocate them appropriately to different fields, and identify those projects that have scientific merit and the universities and government research laboratories where to conduct quality research. The main mission is to conduct quality research and educate students. This type of model is somewhat limited when it comes to evaluation and provides few incentives to measure the quality and the impact of research.

In the new model, greater emphasis is placed on priority setting and some countries have formalized foresight and strategic planning processes. Greater attention is paid to the notion of return on investment because there have been significant changes in the way science is governed. Business and social groups play a greater role in government funding and the NIH is trying to find the gaps in its research portfolio, in collaboration with the scientific community, policy makers and the business community and also in consultation with advocacy and patient groups to help determine where the research dollar should be spent. There have been changes in the funding schemes to support research with an increase in competitive funding awards, especially in European countries and China, and an emphasis on collaborative activities targeting multidisciplinary research, and collaborative research bringing together investigators from universities and industries to move forward in a field oftentimes identified in the priority settings. There have been additional missions for research organizations. The idea that education and research are the main missions of universities and research organizations is still paramount but now a third mission

has been added and that is to contribute to science and society. At NIH, they talk of translational research or how to move research from the bench to the bedside. It has become a priority and there is funding for technology transfer and changes in intellectual property in universities and government organizations. Different types of infrastructures are being set up with emphasis on centers of excellence that will be the target of increased funding, and accountability criteria such as performance are being established and can affect subsequent funding.

With accountability, evaluation is playing a more important role and there is increased emphasis on all levels of evaluation such as ante-, in-process and post funding evaluation of projects, programs, institutions or national innovation systems. Some countries have formalized and institutionalized evaluation by establishing evaluation agencies and bureaus, and at NIH there is a Bureau for portfolio analysis and evaluation. Now as never before there are at national levels groups in charge of evaluation scientific activities in various countries. Evaluation results are being incorporated in policy making and are no longer systematically filed away to gather dust on a shelf. There are systematic evaluations such as the Dutch annual reviews of universities, evaluation-linked funding such as the UK Research Assessment Exercise and the DFG ranking of German universities.

Previous discussions covered Peer Review and the Impact Factor, which assumes that there is a research project, an output (article), a journal, and decisions regarding promotion, journal holdings, subscriptions, etc. Peer review decides whether an article is going to be published and the impact factor has an incidence on promotion, journal holdings, etc. However, there is a more diverse set of steps in the research process where evaluation can play a role. This not just a question of just one research project producing research outputs because, when you look at an institutional level, there are many research projects that produce outputs. It's not just articles and there is an increasing emphasis on technology development, innovation and patents. These research projects are often the result of institutional processes that are in place to design a research program and identify the areas that are going to be funded. The processes are put in place to choose among proposals and there are research grants midcourse and annual reports to give indication of progress. Also, besides the research outputs themselves, there can be outcomes such as greater collaboration or greater relevance to health or economic needs. The scientific, economic and societal impact of that research is much harder to quantify, but the assessment, the outcome and the impact of the output can then be factored in the funding and institutional processes.

There is clearly a role for different kinds of scientific and technical indicators, partly because there is more and more evaluation going on regularly in institutions. Indicators can reduce the burden of evaluation, especially large-scale evaluations, improve consistency and comparability across entities and countries, and generally increase transparency. Some of the measures used may be changing. There is direct interest in output and quality measures such as publications, patents and citation indices. Some process indicators can be developed for program management strategic planning (large group vs. individual or young investigator funding, fields with the greatest potential). Outcome and impact indicators can be matched against institutional objectives, but collaborations between the public and the private sectors are a problem, as it is sometimes difficult to see who works for whom and compiling indicators is very labor-intensive at this time. There has been a great increase of scientific articles over the year, half of which taking place in the Life Sciences. Patents are increasingly used as an indicator and there are difficulties with the naming of authors and institutions but institutions themselves are starting to collect this type of information. Patent citations can also be used to measure science–industry linkages and there has been a steady increase in the number of scientific publications from around the world cited in US-issued patents. There are also ways of measuring the impact of funding agencies on research. For example, studies show that the participation of government funding generates more innovative and

challenging research projects, increases the survival of collaborative projects beyond project completion, and the required funding agency reports improve R&D management.

When it comes to evaluation at NIH, two entities are in charge. The Office of Portfolio Analysis and Strategic Initiatives (OPASI)¹ was established in 2005, mainly to plan, conduct, coordinate and support program evaluation to inform strategic planning and assist in resource allocation, in other words to see where research money is or should be going. Most evaluations are based on bibliometric measures. The Office of Extramural Research develops funding policies for outside research and is interested in developing databases to track NIH grants and look at the success of investment decisions. Institutional and disciplinary archives such as PubMed Central,² as well as the recently established NIH Public Access Policy,³ are key elements for portfolio analysis and evaluation. By keeping a permanent central archive of NIH publications they provide electronic access to NIH-funded research publications for the community at large and advance science by creating an information resource that will make it easier for scientists to mine research publications and for NIH to manage its entire research investment.

Whether features such as linking PubMed to compound searches in PubChem, chemical structures in articles and 3-D views of chemical and protein structures have an impact on the advancement of science remains to be seen.

From ad hoc evaluation to monitoring systems

Stefan Hornbostel

DFG Institute for Research Information and Quality Assurance

Germany was a little late in establishing research evaluation and systematic quality assurance in the field of research. The process started more or less at the end of the 1980s with the publication in the *Spiegel* of the first German university ranking. In 1998, an international committee did an evaluation of the German Research Foundation (DFG – Deutsche Forschungsgemeinschaft), the major German funding agency, and the Max Planck Society (MPG) and came to the conclusion that “the German research system lacked a system for “continual monitoring” and that the DFG especially lacked a self-motivated quality assurance system”. The DFG was also criticized for the lack of output evaluation. In 2005, the Institute for Research Information and Quality Assurance (IFQ)⁴ was created in response to those criticisms. It is a young organization and is still mostly in the planning stage.

Germany might have started late, but currently there are flurries of evaluation activities, and Germany being a federal state, every activity is not carried out once but sixteen times using different methods and standards. Different types of actors and activities have to be evaluated at different levels, from the departmental level to the national level, and this places a tremendous burden on researchers and prevents the extended application of general measures. Also, there are some concerns in Germany about fragmenting research and increasing competition. The German way to cope with quality assurance was to externalize quality assurance, which means that block grants to universities were reduced and external funding from funding agencies increased. However, this shift did cause problems since, for example, more research projects were rejected than before (from 20 to 50% in ten years). This also caused problems to an agency

¹<http://opasi.nih.gov>.

²<http://www.pubmedcentral.nih.gov>.

³<http://publicaccess.nih.gov>.

⁴<http://www.forschungsinfo.de>.

like DFG since it has to justify the way it uses funding and this is one of the reasons why the IFQ was created. It is not funded as a state agency and it was a clear decision to put it in the self-government system of science. This is somewhat problematic since one might say that the DFG funds then “disfunds” the IFQ. The role of the IFQ is to evaluate the programs of the DFG and it is an independent research institute. It is not directly governed by the DFG but by an association regrouping different universities and the DFG. Its objectives are research and development activities to set up methods and indicators, and service activities such as monitoring, program evaluation and research information. The IFQ is using DFG databases on institutions, applicants, grants, reviews, etc. to avoid duplicating efforts whenever possible. The IFQ was completely DFG-funded during its formation stage (2005/2006), and is still heavily dependent on DFG funds today. The first task and/or goal of the IFQ is to provide a long-term monitoring and evaluation system of DFG funding activities to replace the situation of having to collate a number of different ad hoc evaluations coming from different actors. The idea is to build continuous database-driven monitoring systems to reduce the scientists’ workload. The next task is to provide evaluative information on the results of research programs funded and organized by the DFG. The third and fourth tasks are still in the future and concern the analysis of scientific trends in national and international research fields and the establishment of information tools and resources for the public.

Two projects will be used to illustrate the work currently carried out at the IFQ. The first project concerns the Funding monitor and the second, the ProFile online survey. The Funding monitor is a database with a web front end to DFG data covering some of the DFG programs but that will eventually cover all programs. It starts with all the information collected during the application process and already in databases. Then the institutional level data is collected such a staff size, number of PhD students, number of special programs, number of publications and patents, conference participation and collaborative structures. In the future, final project reports will also be collected. Right now, scientists collect these data all year long and at the end a report is generated to be sent to the DFG. The idea is to also use this database to produce data for internal project management, annual DFG reports, evaluation studies and public information. At this time, people are being asked the type of tools they would like to see in the database. This system can increase the acceptance rate not only with DFG but also with personal projects. It is planned to connect this database to repositories and other databases and eventually generate a scientist directory starting with the information in the database and containing information about the scientist.

The ProFile online survey project aims at collecting information on PhD students. The IFQ has very few information about PhD students in Germany. Today there is greater emphasis placed on the role of young investigators in research programs. Information could be collected through an online questionnaire which could span the predoctoral and 3–4 year postdoctoral career of young investigators. Once again, the starting information will be in the Funding Monitor. The Funding Monitor and the ProFile online survey will be connected and some universities are already participating in a pilot project with all their PhD students. Unfortunately, in Germany it is impossible to know how many students are registered in PhD programs and only the number of students having successfully defended their theses is known. The medical thesis is another problem. In Germany, contrary to the USA where you have two different degrees, you cannot tell whether a thesis was written to go into research or simply to satisfy the requirements before going into medical practice. To a lower extent, the situation is similar in chemistry and in biology, where it is not easy to assess the real quality of a thesis. The ProFile design is based on two basic modules with variants according to institutions and PhD careers.

The Emmy Noether Programme is a well-known, well-funded postdoc program in Germany in which young investigators have their own research groups. The goals are to promote outstanding young re-

searchers and to support them in their early stage research positions to prepare for a career path without having to go through “Habilitation”. It started in 1995 and was modified in 2005. It is mostly focused on Natural and Life Sciences because most applications come from these fields. There is an average 50% acceptance rate and around 20% of the funded researchers are women. The Programme evaluation uses a triangulated method combining online surveys, face to face interviews and bibliometric analysis using ISI’s Medical and Physics sections. The online survey includes questions on “soft” skills, such as research administration, communication, networking or leadership. The results show that acceptance rate is improved by the number of publications in the accepted group but that the results of the rejected group are quite good, which is an indication that there are excellent potential scientists in the rejected group. Bibliometric analysis shows almost no differences between the two groups. Paradoxically, although publications increase after the funding decisions, citation rates which were high before tend to drop down after the decisions.

Globally the evaluation results highlight that selection for funding is a complicated process.

Institutions, repositories and research assessment

Tim Brody

University of Southampton

The UK Research Assessment Exercise (RAE)

The UK’s Research Assessment Exercise (RAE)⁵ is held every 3–5 years, although recent exercises have had longer periods between them – the previous exercise being held in 2001 and the next to be completed in 2008. The 2001 RAE was responsible for roughly 5 billion UKP (10 billion USD) of public research funds given out to UK higher-education institutions (HEIs). The RAE provides ‘top-up’ funds that are given as grants to research institutions in addition to funds obtained through competitive tendering for research projects from UK and EU research councils. These two funding streams are collectively known as the ‘dual-support’ system, with the intention that research excellence is rewarded both on the quality of bids (project grants) and on post-hoc evaluation of the results (peer-review of the resulting research).

“Any higher education institution (HEI) in the UK that is eligible to receive research funding from one of these bodies is eligible to participate in the exercise.” (RAE 2001 website)

2008 marks the end of the current RAE process which will determine funding from 2009 onwards. HEIs can make submissions to any of 67 ‘units of assessment’ (subject areas), which are spread between 15 review panels (where each panel has a set of assessment criteria appropriate to the subjects being reviewed). Less research active HEIs are likely to make submissions to only a few subject areas. It’s the responsibility of the HEIs to determine which researchers they submit to which subject area – allowing some flexibility where a researcher’s work may (quite reasonably) overlap between RAE-defined subjects. Once the HEI submissions are complete (November 2007) they are farmed out to the review panels – made up of experts in the respective fields. The panels then review the submissions and give each submission a ‘quality profile’ (i.e. a score). In the RAE 2001 this score had been made into a discrete value of 1 (weakest), 2, 3a, 3b, 4, 5 or 5* (best), although the resulting funding given out wasn’t

⁵<http://www.rae.ac.uk>.

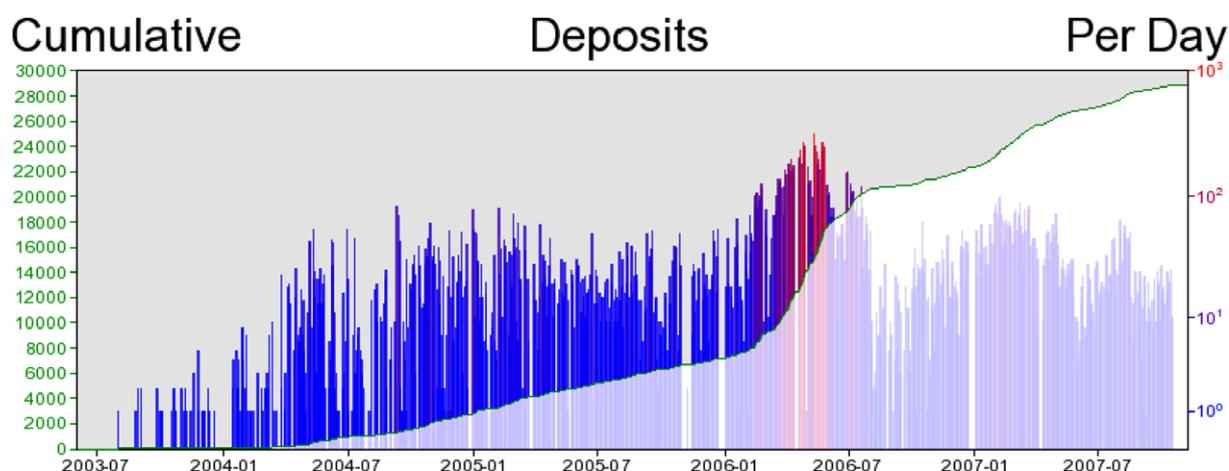


Fig. 1. Daily submissions to the Institutional Repository (line shows the cumulative total). Each bar is colour-coded to indicate the number of submissions on that day – ranging from dark-grey/blue* (few) to grey/red* (many). The bars are shown on a logarithmic scale (right axis).

proportional to the score given. The amount of funding allocated by the RAE 2008 and exactly how a profile will effect funding hasn't yet been determined – although, like previous RAEs, the results will be made public.

Southampton University and the RAE

The Institutional Repository at the University of Southampton is the primary tool being used to build Southampton's RAE 2008 submission. The submission is generated from University administrative systems and *eprint* records from the Repository. The administrative system provides information about who is submitted to which RAE units of assessment, who is available to be submitted and general University data required by the RAE – numbers of research students, project grant income etc. In addition to research outputs authors are required to submit their 'measures of esteem'. These measures are non-quantifiable but nevertheless provide evidence of research activity – editorships, awards (e.g. Nobel prizes), conference organisation, etc. The measures of esteem are captured as simple text and are the responsibility of individuals to contribute. The authors' research outputs are more complex because they have to be collected, organised and delivered to the RAE assessment panels. While some authors may have outputs such as shows or exhibitions the vast majority of research will be paper-based (i.e. PDFs or similar). The RAE allows up to 4 research outputs per research staff to be submitted – staff with fewer than 4 outputs are unlikely to be submitted due to not being 'research active' (hence potentially diluting the eventual RAE score . . .).

The Repository collects data in three ways: authors submit their measures of esteem through an on-line form, deposit their eprints/bibliographic records and lastly through the bulk import of records from school publication databases. Once the eprint records for an author are collected – by their own deposits or by bulk import – they can make a selection of their 4 research outputs from any that are/have been published within the RAE assessment period. Every selected eprint has to go through a validation/verification process that is a combination of automated tools and editorial checking. The automated

*The colors are visible in the on-line version of the article.

tool does things like checking all the fields required by the RAE for the type of document are filled out. A library-based editorial team corrects any errors found by the tool and verify the eprint is genuine. It's the responsibility of each school's RAE co-ordinator to put together their submissions (i.e. make sure the submission is as strong as possible). For the RAE 2008 every submitted published paper must be made available to the RAE reviewers. This can either be achieved by supplying a DOI (Digital Object Identifier) – which allows the reviewer access to the paper direct from the publisher – or by the institution supplying a PDF version of the paper. (For the RAE 2008 an agreement has been made with publishers to allow reviewers electronic access to all research papers for the purposes of accessing papers via DOI.) For papers without a DOI a PDF version has to be supplied on CD-ROM for distribution to the review panels. How the institution/author gets hold of a PDF version is up to them – for Southampton this has meant scanning a local copy (if available) or obtaining a version through inter-library loan. This scanning/collection activity has been undertaken by the editorial team – fortunately most papers now have DOIs.

A submission case study

The RAE 2008 has 15 panels reviewing 67 units of assessment (i.e. subjects). E.g. the School of Electronics and Computer Science at Southampton will submit two units: 'Electronics' and 'Computer Science and Informatics'. The review panels review all submitted papers and assign a quality score to each. In practise this could mean each panel member reviewing upwards of one thousand papers and with only one month to perform that reviewing in. Each author receives a score based on their research outputs and measures of esteem. The authors' scores are then aggregated to form (for this panel) 70% of the rating of the unit score. The remaining 30% is made up from a 'department-level' measure – based on how effective the panel thinks the department has done as a group.

Proposal for open access research metrics

“The government's firm presumption is that after the 2008 RAE the system for assessing research quality and allocating 'quality-related' (QR) funding from the DfES (Department for Education and Skills) will be mainly metrics-based.” (UK Treasury 2006)

The current RAE is cumbersome and expensive to administer – although exact figures are difficult to come by. There is not only the cost of running the RAE – costing the funding councils – but the cost to the HE sector as a whole of making submissions. Every researcher submitted to the RAE has to create his/her selection of outputs and supply a CV-like list of measures of esteem. Administrative staff select which researchers to submit, assign them to RAE units and collect University metrics. Library staff collate and check the researchers' selection of research outputs. In a bid to reduce this administrative burden it has been proposed to use 'metrics' instead of panel reviews – the hope being to automate the review process as much as possible. Replacing reviews with metrics begs three questions: what data can be collected, how can that data be collected and what analyses of those data should be made. To answer these questions we have put forward a proposal for 'Open Access Metrics'.

Evidence of research output

The Institutional Repository (IR) is ideally placed to collect research output – what the RAE is intended to assess. The IR is self-sustaining and increasingly part of the existing institutional practise – the institution wants to know what researchers are doing, as much as funding councils. Academics are also slowly starting to make use of IRs to maximise their individual research impact by making their

work more visible (IRs are freely accessible and the structured content leads to good indexing by Web crawlers). IRs collect research outputs from researchers, organise it and, depending on the IR policy, ensure it has good quality bibliographic metadata. The IR primarily provides Web access to its content but it also provides machine access through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), which enables services to harvest and index the content of the IR.

In addition to the content IRs are capturing usage data through their Web logs. While most IRs are not, currently, making use of that data it is there ready to be used as more quantifiable data in RAE metrics.

Building a metrics service

Citations are the most wide-spread tool for evaluating the *impact* of research outputs. The more citations a work has, the higher impact it is and hence can be used as an *indicator* of the importance/quality of the work. No current IR software is capable of counting the number of citations to its content, therefore the IR in isolation is limited as a tool for developing metrics. Indeed, a single IR's collection is not that interesting for citation counts anyway because it only captures a tiny part of the global scholarly environment. What is needed is an aggregation of IRs and other sources of citations to build a *citation index*. Such a service could harvest the content of many IRs, count the number of citations to each eprint, and provide citation metrics for each IR. In addition a metrics service could harvest the usage data from each IR to provide another dimension to the metrics data. Combining usage (how many times an eprint has been downloaded/read) with impact (how many times an eprint has been cited) could provide greater analytical power than citation data alone – particularly for very new eprints that haven't had time to acquire citations. There's nothing in this proposal for a metrics service that precludes commercial involvement or indeed commercial ownership. The 'value-add' of metrics isn't easy to implement – unlike the IR there's no clear body that would take responsibility for such a service. Existing tools like the ISI Web of Science citation index may be a way to quickly build metrics for IRs. Conversely there's a strong argument for making metrics as transparent as possible. Where institutions' funding is being determined having an open system, that is open to interrogation and verification, must be an important criteria for adoption.

Building an armoury of metrics

With metrics there is no reason to limit analyses to a pre-defined 'one size fits all' approach. Different fields and different types of research material have characteristics that aren't directly comparable. There is also a value judgement to be made by funding agencies for which criteria to use when determining 'quality'. We believe these questions are best answered by having multiple metrics that can be combined together to determine a quality score. The relative importance of different aspects of the data – citations and usage, longitudinal, comparative statistics – can be tweaked in the light of expert criteria to achieve the most favourable system. Until now only publication counts and citation impact (citations normalised by total papers) have been used in research evaluation. These can be augmented with usage data (downloads), influence-type adjustments (e.g. 'being cited by a highly cited author is better than being cited by a lowly cited author') and tools like the h-Index. If all of these factors could easily be interactively adjusted reviewers would have a powerful tool to pick apart how a researcher's work has developed and impacted on their community.

Validating the process

As stated earlier the goal of moving to a metrics-based RAE is to reduce the considerable administrative overhead of making and evaluating the submissions. Empirically this infers that a metrics system must at least be capable of replicating the results of existing RAE processes. This can be done by calculating the component of each input metric (citations, downloads, derivations etc.) that best fits the

results found by peer-review panels. Following this procedure also grounds the process of identifying metrics in evidence – rather than *a-priori* attempting to determine the ‘right’ metrics (an almost religious argument) we can start with the accepted, or at least tolerated, status quo. Once a metrics system is established and shown to replicate the existing process it would then be possible to tweak the system in response to the needs of funding agencies and any distortions in behaviour that may occur – it’s inevitable that researchers’ behaviour will change in response to a change in evaluation mechanism.

Metrics with Google Scholar

As part of the School of Electronics and Computer Science’s (ECS) RAE 2008 submission we undertook a small experiment using Google Scholar to compare ECS’ citation impact to its nearest 5 competitors. This was based on comparing the eprints for academic/research staff from each of the top-rated RAE 2001 departments. A list of staff was created for each institution and each authors’ name was queried with a simple key term for each institution (‘Southampton’ for Southampton, ‘Edinburgh’ for Edinburgh etc.). Each match from the Scholar search results were recorded with its title, the authors and total citation count. The data from Scholar was stored in a SQL database and then analysed using a combination of some simple scripts (necessary to calculate the h-Index) and Microsoft Excel.

From this small study the difficulty in ascertaining the research output of a school was evident – even just working out which researchers a school has could be troublesome. As a result the number of eprints found for each institution varied widely, from just 791 for Cambridge University or 3862 for Southampton. On a simple average Cambridge had the highest citation count, although that could be attributed primarily to a single, super-cited book. Southampton had the highest number of total citations, but the lowest average citations per eprint. Southampton’s low average could be attributable to the high uptake of the repository in ECS – one would expect authors to start by putting their most important work online first. This would seem to be supported when only the most highly cited 791 eprints are counted: Southampton jumps from bottom to top.

Conclusion

This study was a quick effort to produce some headline statistics for the RAE submission, however it shows how easy it is to quickly put together metrics based on open access content compared to collecting and reviewing vast numbers of research papers.

Ratings, reviews and repercussions: PLoS meets the web

Catriona MacCallum
Public Library of Science

Free access is not the same as open access. There is a huge growth of Open Access (OA) options, but OA means more than mere free and immediate access. It means depositing in public archives such as PubMed Central, but above all, it means unrestricted reuse with no permission needed. At PLoS we use the Creative Commons Attribution Licence, as many OA publishers do.

This presentation will cover the beginning and the future of the Public Library of Science (PLoS),⁶ its economic sustainability, its challenges and opportunities and its repercussions for the scholarly community in general.

PLoS was co-founded in 2001 by Nobel prize winner Harold Varmus, and by Patrick O. Brown and Michael B. Eisen. It started as an advocacy group but turned to publishing in 2003. PLoS publishing strategy is: establishing high quality journals and putting PLoS and OA on the map; building a more extensive OA publishing operation by giving a home to every paper; and to making the scientific literature more useful to scientists and the public. The Web 2.0 and its social interactions have brought many changes in what is relevant to scholarly communication. It is with this in mind that *PLoS ONE*⁷ was launched. It has an inclusive scope that aims to break down existing barriers within scholarly communication. It is not broken down into subject areas since this creates artificial barriers, so any emerging field immediately has a home for publication. It has no length or volume restrictions, is peer reviewed objectively by academic editors focusing on technical quality and its streamlined production speeds up acceptance to publication time with very little copy editing. Furthermore, communication about an article doesn't stop after it has been reviewed, since discussions and debates are encouraged through community comments and annotations. Ideally, the basic role of peer review is to select articles on the basis of their contribution to science, improve those articles, validate their findings by putting a stamp of authority on them, and provide status for the article rather than the journal. What PLoS wants to do is to extend the traditional peer review process by providing additional comments and it is attempting to do so with *PLoS ONE*.

PLoS ONE is taking advantage of the Web 2.0 functionalities to provide for comments and discussions that are not anonymous. The *PLoS ONE* site is a beta site where you can look either by what has been recently published or has been commented on. Comments, discussions threads and links to other sites can be added. In *PLoS ONE*, articles can also be rated on a 1–5 scale based on three attributes: insight provided by the article, reliability and style, and those attributes will be weighted. Ranking is still anonymous at this time. *PLoS ONE* provides the traditional review process of selection, validation and approval before publication, but with the ability to continue commenting and rating after publication. It is hoped that, by using these tools, greater focus will be put on the article rather than on the journal so that the article itself will be assessed and the attributes of the article brought out, and that, as time goes on, it can be linked to other articles. There are other experiments with peer review, so PLoS is by no means the first but, generally, the post publication process stops at one point. This is the case with *Atmospheric Chemistry and Physics* and the recently launched *Nature Precedings*, a preprint server for articles, conference proceedings, posters, etc., where the reviewing process is cut down, the filtering is weaker, articles are rated with a simple thumb-up thumb-down system, there is no facility for post publication comments and, presumably, some of these articles will never get published.

PLoS ONE submission started in August 2006 and publication in December, and so far, there have been 1387 submissions and 513 publications with a 75% acceptance rate. Peer review is based exclusively on technical rigour, there is not subjective judgment, so if the science is sound the article gets published. Some articles will naturally bubble up to the top but some small articles can also be important for collating data about different subject areas. Most journals tend not to publish articles in some areas because they do not get cited and have no impact factor but these articles are essential to scientific progress and *PLoS ONE* provides a venue for all papers whether they are high-end or low-end papers.

⁶<http://www.plos.org>.

⁷<http://www.plosone.org>.

PLoS ONE has a large editorial board (360 and growing) covering all areas of science. Acceptance to publication is about 19 days and, so far, commenting and discussion have been somewhat disappointing (600 post publication comments, 1 per article) but, with anonymous peer review so entrenched in the scientific community, scientists are still very reluctant to put their names on comments. However, this culture is changing with blogs and other Web 2.0 technology and *PLoS ONE* articles are getting noticed and picked up by scientific journalists, much more than was expected. *PLoS ONE* is a publishing and journal platform the features and functions of which are provided by TOPAZ,⁸ an independently-funded open software development project providing software to operate portals and gateways. TOPAZ deals with rich integrated information networks linking data with articles, authors and journals, and one aim of PLoS is greater integration of literature and data.

As far as current status is concerned, PLoS has a broad high quality publishing portfolio in all sciences. It offers a home for every paper wanting to be in open access. Community supports from scientists is building up and publishers are definitely more engaged with OA. If authors and publishers had to be relied on to change the landscape, it would not happen but fortunately funding agencies are getting involved and some of them like the Wellcome Trust are mandating that their articles be made freely available immediately. PLoS is trying to take advantage of technology and more and more people are willing to partner with PLoS. *PLoS ONE* is already sustainable and scalable, and covers all its overheads. Community journals are approaching sustainability. They run more along traditional lines with editorial staff and should be sustainable in 2–3 years. Flagships such as *PLoS Biology* and *PLoS Medicine* are much more expensive and publication charges do not cover the costs. There are editors involved, a large front section and like *Nature* and others, *PLoS Biology* and *PLoS Medicine* will have to resort to such things as advertising. Quarterly submissions are going up and some authors are transferring to *PLoS ONE*. As far as payment rates are concerned, they range from \$1,250 for *PLoS ONE* to \$2,500 for *PLoS Biology* and *PLoS Medicine*. Like all open OA journals, there are fee waivers but on average 81–90% of authors pay the PLoS journal fee.

One of the PLoS goals is to have journals hosted on a single platform with Web 2.0 tool applications so that all journals on TOPAZ will have same functionalities. New channels into contents should drive submission growth; build communities and increase traffic and usage. PLoS wants to take advantage of the technology to create, instead of journals, some subject specific portals where information can be collated for a particular established or emerging community, and connected with the data, the literature, conferences, blogs, debates, etc., that these communities need. PLoS aims to demonstrate the power of OA. OA is not going to work unless people see a good reason why they should be involved, and reuse and access are very convincing reasons.

Some challenges and opportunities remain. PLoS as a whole is not sustainable yet. The author-pay model is not enough and institutional memberships are needed. Scalable models for financial support of OA are needed and funding agencies, institutions and publishers must be encouraged to adopt OA policies. PLoS does not aim to publish all the scholarly literature; it just wants other journals to convert to OA. Many talk of the tyranny of the Impact Factor but OA provides opportunities for new metrics, and PLoS would like to see metrics that focus more on the article than on the journal. Some killer applications specific to OA are needed to demonstrate the power of reuse and *PLoS ONE* is a great way to demonstrate the new ways of scholarly communication but more are needed.

One of the repercussions of OA is that scholarly communication is no longer the sole remit of publishers and is coming out of their hands. Publishers and OA journals need to adapt to and be a reflection

⁸<http://www.topazproject.org>.

of the evolving Web. Even PLoS has had to respond to the changing environment. We have to adapt or else it is going to be a very simple process of natural selection. When it comes to “traditional journals” it is unclear what they are anymore and it is hard to say what they will be in five years.

Community peer review in Wiki environment: Semantic enrichment and the role of Wikis for professionals

Christine Chichester

KNEWCO Inc.

KNEWCO⁹ is a small company with collaborators in various countries and organizations (Erasmus MC, Leiden University Medical Center, Open Progress Foundation, etc.) which is attempting to create “the barcode of knowledge” by distilling down information and knowledge to a very simple form: a barcode, which we call “knowlet”, for every concept, that can be scanned and compared to other barcodes.

KNEWCO is attempting to do this first in biomedical research since Biology is the main area of expertise of the company’s staff, and they know its many challenges and difficulties. In Biology, there are huge amounts of data, with high throughput and textual data. These data are complex and there are many ways to look at them. They are found in distributed systems of databases, too often in different and incompatible data formats, and the context is multidisciplinary, which adds another difficulty to using biomedical research data. They are multilingual, a problem that not many people address since most of the time we deal with English, but there are huge amounts of data in China and South America that could be used for knowledge discovery, and, even though these data have been deposited in repositories, we cannot use them because of the language barrier. Ambiguity of terminology is another major difficulty in biomedical research. Right now, because of lack of adequate tools, the sharing of knowledge is hampered. However, wiki technology might well be the solution.

There is way too much to read, especially when we see that 1,000 new articles or abstracts are added each day in PubMed. This glut has started some trends and there is a shift from reading to consulting experts; from reading to Meta Analysis to digesting down the information for the scientists and save them time; from texts to facts since facts are repeated many times in a paper to establish the background, and it would save time if these facts could be extracted and put into a central database for central and community annotation, and this is what can be done with a wiki.

As mentioned earlier, ambiguity of terms and homonyms are two major problems in biomedical research. For example, the term PSA (which for us normally stands for ‘Prostate Specific Antigen’) can have at least ten meanings which make analysis of a text quite a difficult problem. The term cannot be disambiguated without its contexts. Using the context is what we call “first order semantic enrichment”, but a second order of semantic enrichment is needed and that is where the knowlet, or barcode of knowledge, comes in. A knowlet building block consists of a source concept, a target concept and the relations between the two. Some relations consist of what are called factual relations, which means that facts found in a database can be used to link the source concept with the target concept. The facts can be annotations found in a wiki for instance. Some relations are based on co-occurrence i.e. two concepts occurring together in one sentence or abstract. Other relations are: the Concept Profile Match or overlapping concepts, sequence similarity where a score can be given to a sequence to be included in a knowlet,

⁹<http://www.knewco.com>.

and co-expression with genes from expression databases. All these data can be captured in what we call the knowlet, and by putting the source and the target knowlets together you create the knowlet space.

The knowlet space is used to build an association matrix for large data sources. For example, knowlets can be built for groups of authors and for each paper of those authors, for organizations of objects (genes, diseases, drugs). Those knowlets are used to build a large association matrix. By taking all of Medline and disambiguating the names, 1 million author knowlets and 1 million gene, disease, and drug knowlets were created. In one study, proteins clusters showed proteins belonging to one functional group along with some unknown proteins. The clusters were used to help predict to which functional group these unknown proteins might belong and articles published later confirmed those predictions and proved that the knowlet system could be used for knowledge discovery. So, to arrive at a prediction, a protein is found in the literature, a concept profile is made and condensed into knowlets which are used to create an association matrix from which predictions can be derived to challenge experts to confirm or deny those predictions. All this can be done in a wiki. By building a wiki for proteins, the validated proteins could be included in databases such as Swiss-Prot¹⁰ and then feed back into the concept profile to improve the knowlet, therefore enhancing the system through continuous contributions.

This type of wiki is called *wikiProfessional*, it is not Wikipedia, it is a wiki for professional life scientists. The objectives at KNEWCO are first to concentrate on proteins, with *Wiki Proteins*,¹¹ then go on to *Wiki Medical* and *Wiki Physics*. It is really a professional system and real names will be needed to register for the wiki. *OmegaWiki*¹² is a related general initiative in which KNEWCO is involved, together with the Open Progress Foundation. It is a terminology system to create lists of synonyms and disambiguate homonyms. At term, the ambition is to have all terms and definitions in all languages. *OmegaWiki* is an ambitious project, it has not been advertised but it already has quite a following of people willing to contribute to improving this terminology. However, approval is required to register. *OpenWetWare*¹³ is another wiki where contributing biological laboratories can generate protocols. *Wiki Proteins* is the first prototype of a semantic wiki. *Wiki Authors* is to be developed where authors will have their own pages, which will help with disambiguation of names, and a consortium of people is already willing to contribute. Then, there is the Medical/Clinical wiki, the Phenotypes wiki and the Chemicals wiki. In all these cases, people from the whole community will be able to contribute and annotate concepts in a central place which will enable terminology growth and semantic linking between concepts.

Wiki technology is a million minds approach. Basically the defining characteristics of a wiki are that pages are easily created and updated, and the results of work are immediately visible. Jimmy Wales, co-founder of Wikipedia, said that his first project did not get off the ground because the results were not immediately visible. The annotations had to go through an editing process which slowed the whole process down, and he was adamant that this was the main problem with his first project. In his next Wikipedia project, he took off barriers and it was immediately a success. Therefore, it is critical for a wiki to succeed that there be no review, or minimal review, before modifications are accepted.

The authoritative sources of *OmegaWiki* will be databases such as UniProt,¹⁴ and KNEWCO has obtained permission from the NLM to include the UMLS (*Unified Medical Language System*), and UMLS contributors are quite excited to have the feedback they never had before. In *OmegaWiki*, data

¹⁰The Swiss-Prot database was incorporated into UniProt, see also: <http://www.ebi.ac.uk/swissprot>.

¹¹<http://www.omegawiki.org/Portal:Wikiproteins>.

¹²<http://www.omegawiki.org>.

¹³<http://openwetware.org>.

¹⁴<http://www.uniprot.org>.

from authoritative sources remains intact and, to be modified, it has to be copied into the community database. Terms in knowlets can be used to run PubMed searches, and citations, new definitions, and elements of disambiguation can be added to the wiki to substantiate an annotation.

Basically, a wiki is a place to validate and improve the knowlets so one can go from analysis of a concept to entering personal comments in the wiki. And from the knowlets, one can find appropriate reviewers based on their own knowlets. So, a knowlet of a concept can be matched to that of reviewers, and when a good match is found, one's annotations can be sent to that reviewer to get his or her feedback. For peer review, you can find scientists with a profile similar to yours, with an interest in the concept in question, or both.

At this website you can find a demo for the whole system: www.wikiprofessional.org.