



Illustration: John McKernan / Source: <http://whyperscience.org>

Research Evaluation in Informatics

This section was edited in cooperation with Informatics Europe. Guest editors: H el ene Kirchner (Inria) and Fabrizio Sebastiani (ISTI-CNR)

Research evaluation, as applied to individual researchers, university departments, or research centres, plays a crucial role in recognising and supporting research that can lead to advances in knowledge and benefits to society.

Evaluation can have a tremendously positive effect in improving research quality and productivity. At the same time, the effect of following wrong criteria or practices in research evaluation can have seriously negative long-term effects, potentially on several generations of researchers. These negative effects range from demotivating researchers who, despite performing good-quality research, get evaluated negatively, to wrongly promoting unworthy research endeavours at the expense of worthy ones.

To achieve the positive effects of research evaluation, its specific goals must be clearly formulated upfront, and it must be performed in a transparent way that is aligned with these goals. The evaluation should follow established principles and practical criteria, known and shared by evaluators, researchers, and the broader scientific community.

Research evaluation should indeed take into account the specificities of Informatics as a new science and technology, and its rapid and pervasive evolution. Informatics Europe recently published a report on this topic entitled "Informatics Research Evaluation". The report focuses on research evaluation performed to assess individual researchers, typically for promotion or hiring. The recommendations and conclusions are outlined in the first article of this section.

The Informatics Europe report also raises several questions on the publication culture and its evolution, on the importance of artifacts, and on impact assessment for this scientific domain. Several contributions in this special section are devoted to exploring these questions.

How to evaluate the quality and impact of publications?

Dino Mandrioli (Politecnico di Milano) addresses the conferences vs. journals controversy. He explains why focusing on conference publications, rather than on journal articles, may be harmful for research quality and evaluation in informatics, and he gives serious arguments in favour of journals. He argues that conferences should go back to their original, authentic goal, i.e., circulation and discussion of ideas; journals, in his view, then remain the best and natural (although not the only) medium for publication of research results, and research evaluations should treat them accordingly. The recent trend towards the tight coupling of conferences and journals, with conference papers appearing in a journal, or conferences incorporating journal-first papers into their program, could reconcile the divergent views.

Stefano Mizzaro (University of Udine) argues in favour of an alternative to traditional peer review that relies on crowd-sourcing, i.e., that exploits data and information from fellow researchers who read, spread, comment on, and cite papers. The idea is to compute a quality index of papers based on the post-publication comments they receive; in turn, this allows the computation of quality indexes for researchers as authors and researchers as readers, via algorithms in which these indexes mutually reinforce each other. Although it is unclear whether the quality of the scientific literature will be improved by crowd-sourcing peer review, proposals such as this warrant further study. The development of public archives and the concept of overlay publication might offer an opportunity to experiment with this idea.

How to evaluate software, artifacts and outreach?

Alain Girault and Laura Grigori (Inria) address the growing importance of software in academic research and, as a consequence, the necessity of taking software development activi-

ties into account when evaluating researchers. They support this idea by describing the software evaluation procedure at Inria that allows researchers to characterise their own software along different criteria, in the context of recruitment or team evaluation. Software evaluation goes together with the ability to reproduce experimental results described in a publication, and opens the way to a new publication model which, in addition to open access, also includes open data and open software, simultaneously released.

How to take into account open science criteria?

Laurent Romary (Inria) addresses the link between open access and research assessment and argues that, despite some fears, open access offers the potential of a wealth of information for the strategic management of research, typically for identifying experts or emerging topics in a given field. He points out challenges related to the possible fragmentation of the publication corpus and to the actual quality of the available information, but argues that if all the proper conditions are fulfilled for a trusted information corpus of scientific publications, we can actually foresee the basis for an open transparent process for research assessment. This opens a wide variety of usages that we have to invent and deal with in an ethical way. He suggests for instance a visibility index related to the dissemination effort of a researcher or a research institution. He concludes with a reminder that open access on research data and tools ultimately exist to help researchers themselves.

How to measure scholarly impact?

Giovanni Abramo (IASI-CNR) revisits the concept of scholarly impact, i.e., the footprint that a research paper leaves on scientific research performed after its publication. He remarks that, assuming that the amount of citations obtained by a certain paper is a good proxy of its scholarly impact, this impact can be measured only at the end of the “life cycle” of the paper, i.e., after it has ceased to influence current research (and thus ceased to be cited). This is too late for the practical needs of those who need to evaluate research (e.g., a researcher, for promotion or hiring); this means that there is a need of “early indicators of impact”, i.e., measures capable of predicting whether and how much a paper is going to be cited in the future. Giovanni Abramo recalls current efforts, essentially based on (a) using as features the early citation counts plus the impact factor of the journal in which the paper was published, and (b) using late citation counts as benchmarks for the accuracy of these predictors, but stresses in conclusion that more research is needed in this area.

Many more questions on research evaluation could be raised and good practices exchanged, but we should be aware that each of us, as soon as we are involved in any evaluation process, has the opportunity to encourage and promote research evaluation criteria which should be based primarily on quality and impact.

Please contact:

Hélène Kirchner, Inria, France
helene.kirchner@inria.fr

Fabrizio Sebastiani, ISTI-CNR, Italy
fabrizio.sebastiani@isti.cnr.it

Recommendations for Informatics Research Evaluation

In March 2018, Informatics Europe published a report [1] focusing on the main principles and criteria that should be followed when individual researchers in informatics (computer science) are evaluated for their research activity, addressing the specificities of this area. This subsumes evaluation of a specific piece of research and can often be generalised to university departments or research centres, since their research performance is largely determined by their individuals.

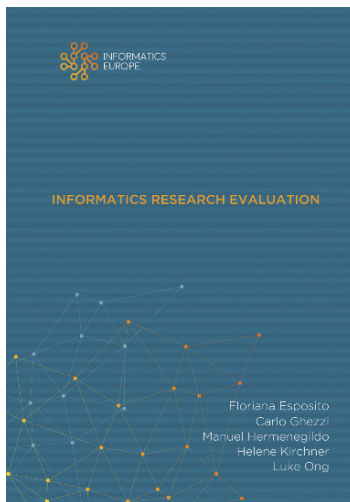
This report confirms the findings of the 2008 Informatics Europe report (Research Evaluation for Computer Science, Informatics Europe Report, Eds. Bertrand Meyer, Christine Choppy, Jan van Leeuwen and Jørgen Staunstrup) on the subject, and at the same time incorporates a number of new observations concerning the growing emphasis on collaborative, transparent, reproducible, and accessible research. The conclusions of the report are summarised in nine recommendations to people involved in evaluation committees and funding agencies:

1. Informatics is an original discipline combining mathematics, science, and engineering. Researcher evaluation must adapt to its specificity.
2. A distinctive feature of publication in informatics is the importance of highly selective conferences. Journals have complementary advantages but do not necessarily carry more prestige. Publication models that couple conferences and journals, where the papers of a conference are published directly in a journal, are a growing trend that may bridge the current gap between these two forms of publishing.
3. Open archives and overlay journals are recent innovations in the informatics publication culture that offer improved tracking in evaluation.
4. To assess impact, artifacts such as software can be as important as publications. The evaluation of such artifacts, which is now performed by many conferences (often in the form of software competitions), should be encouraged and accepted as a standard component of research assessment. Advances that lead to commercial exploitation or adoption by industry or standard bodies also represent an important indicator of impact.
5. Open science and its research evaluation practices are highly relevant to informatics. Informatics has played a key enabling role in the open science revolution and should remain at its forefront.
6. Numerical measurements (such as citation and publication counts) must never be used as the sole evaluation instrument. They must be filtered through human interpretation, specifically to avoid errors, and complemented by peer

review and assessment of outputs other than publications. In particular, numerical measurements must not be used to compare researchers across scientific disciplines, including across subfields of informatics.

7. The order in which a publication in informatics lists authors is generally not significant and differs across subfields. In the absence of specific indications, it should not serve as a factor in the evaluation of researchers.
8. In assessing publications and citations, the use of public archives should be favoured. When using ranking and benchmarking services provided by for-profit companies, the respect of open access criteria is mandatory. Journal-based or journal-biased ranking services are inadequate for most of informatics and must not be used.
9. Any evaluation, especially quantitative, must be based on clear, published criteria. Furthermore, assessment criteria must themselves undergo assessment and revision.

These recommendations are consistent with a recent statement of three national Academies (Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes) that also provides recommendations on evaluator selection, overload and training.



Links:

The Informatics Europe report is now available on its web site in its final form:
<http://www.informatics-europe.org/publications.html>.

The direct link to download is:
<http://www.informatics-europe.org/component/phocadownload/category/10-reports.html?download=76:research-evaluation-2018>

Reference:

[1] Floriana Esposito, Carlo Ghezzi, Manuel Hermenegildo, Hélène Kirchner and Luke Ong (eds): "Informatics Research Evaluation, Informatics Europe Report, 2018

On the Conferences vs. Journals Controversy

by Dino Mandrioli (Politecnico di Milano)

Conferences are harmful for research evaluation.

The document [1] addresses the critical issue of informatics research evaluation in a complete and thorough way, touching all the main facets of the problem. I substantially agree with most of the arguments and proposals presented therein. In this note, however, I discuss the only point of serious disagreement, i.e., the role assigned to conferences for the evaluation of research in computer science. I admit that it is a fact that in our community – probably the only scientific community that exhibits this specificity – conferences often overwhelm journals in the research assessment procedure, but I believe that this anomaly is seriously flawed, dangerous, and should be contrasted – not be accepted or, even worse, encouraged.

There are two main arguments to support preferring conferences over journals as the main publication medium:

1. Journal publishing takes too long
2. Some conferences are even more selective than many journals.

My main counterpoints to the above arguments are:

- 1.a) It is true that reviewing (nowadays not publishing!) journal papers often takes a long time, but for highly technical papers this is necessary. Research advances quickly but its evaluation needs its own time!
- 1.b) Some highly prestigious journals (e.g., JACM, TOPLAS) have adopted effective policies to dramatically reduce the average response time: typically, a preliminary quick review states whether the paper is readable and interesting enough to deserve a thorough further reviewing effort; if a paper passes this phase, the author knows that acceptance will depend mostly on technical soundness and can be more patient. See also the symmetric point 2.b) about conference reviewing.
- 1.c) From a more general, "social", point of view, this claim is a symptom of the typical "time to market" syndrome which often causes serious damage even in the industrial world: research needs serious evaluation before "going to the market" (remember the "cold fusion" phenomenon).
- 2.a) Top journals are highly selective too, sometimes even more than top conferences; furthermore, often people try to "submit anyway" to conferences even if the paper is not yet mature (e.g., due to the deadline); or they submit several papers maybe hoping for some "good luck" for some of them. Many authors are more conservative and careful when submitting to high-level journals.
- 2.b) Being highly selective does not guarantee high quality. The conference and journal reviewing processes are necessarily different: having many papers to review with hard deadlines, even for serious committee members, leads to superficial reading where generic interest, "sympathy for the topic" (or even for the author), read-

ability, and, in some cases, distributing the load to several sub-reviewers, overwhelms thorough and comparative evaluation. The typical conference bidding process tends to produce clusters of strongly related topics, thus producing the risk of self-referencing environments, just the opposite of the real goal of cross-circulating new ideas. Furthermore, in program committee (PC) meetings, whether in person or electronic, the prevailing opinion is often the one of the member who “speaks the loudest” (“over my dead body”).

The 2014 “NIPS experiment” has been mentioned a lot lately (see e.g., [2]), where the PC was split into two independent subcommittees which agreed only on the top and bottom evaluations and whose opinions varied substantially on about 60% of the papers in-between. This fact was interpreted in various ways; in some cases it was even used as an argument in favour of keeping top conferences highly selective: “If we pick up just the top 10%, we are sure that we reward just the best papers”. I have little doubt that the bottom 10% warrant rejection, but this approach risks rejecting many potentially important papers in the grey area – some of which are probably even more interesting than those unanimously accepted. The literature exhibits plenty of examples of pioneering papers that suffered from several rejections before the community appreciated their value. Important, novel contributions are often controversial and misunderstood; it is easier to gain general appreciation by means of results in fields that are already familiar to a wide audience. Of course, the same problem may occur even with journal publishing, but in this scenario it can be mitigated by a careful selection of reviewers, by the opportunity for rebuttal, and by ...more time and thought for the final decision.

- 2.c) The tight page limits compel authors to submit “ ϵ -papers”: small, maybe relevant, advances can be more easily published than well-developed, completed, long-term research. Technical details are omitted or confined to appendices, which are rarely checked.

In conclusion, I would like conferences to be relegated to their original, authentic goal, i.e., circulation and discussion of ideas. Journals remain the best, natural, certainly not only, medium for research evaluation.

References:

- [1] F. Esposito et al.: “Informatics Research Evaluation”, An Informatics Europe Report, Draft Version: 20-10-2017
- [2] M. Vardi, “Divination by Program Committee”, CACM, Vol. 60, N. 9, p. 7, 2017.

Please contact:

Dino Mandrioli, Politecnico di Milano, Italy
+39-02-2399-3522
dino.mandrioli@polimi.it
<http://home.deib.polimi.it/mandriol/SitoInglese/perswebsi-teengl.html>

Evaluation in Academic Publishing: Crowdsourcing Peer Review?

by Stefano Mizzaro (University of Udine)

In academic publishing, peer review is the practice adopted to evaluate the quality of papers before publication. Each submission undergoes a thorough review by some peers, who decide whether the paper is worthy of publication. The pre-publication filter by peer review is important to guarantee that scientific literature is of high quality. However, although peer review is a well known, widely adopted, and accepted practice, it is not immune to criticism, and its appropriateness has been challenged by many. One recent example is the paper by Tomkins et al. [1].

Some such opponents go beyond discussion and criticisms, and propose concrete alternatives. Such proposals essentially aim to exploit data and information that are available but are not currently used for evaluation purposes. Indeed, after publication, papers are (hopefully) read by fellow researchers, who usually form an opinion about their quality. However, all these opinions remain private, or are communicated among a few collaborators in an informal way, or perhaps manifest themselves (although in a rather implicit way) in citations. In extreme terms, we might say they are lost (at least, for the purposes of research evaluation). At the very least, the research community fails to leverage precious data and information that could help research evaluation.

Some alternative proposals, including Readersourcing [3] and TrueReview [2] make use of this data. Both rely on the opinions of readers, who are asked directly to judge the papers they have read. Reader assessments are then stored, in an open way, as numerical scores. This would then allow a numerical quality score to be computed for papers, authors, and readers too, by means of an algorithm capable of estimating the quality of the assessments. In other words (see the original publications for all the details), readers are invited to submit a score on the quality of the papers they have read; all the scores concerning a specific paper may then be aggregated (and weighted according to the quality of the readers) in order to compute an overall quality score for the paper; the scores of all papers by a certain author may in turn be aggregated in order to compute a quality index for researchers as authors; and the quality of the expressed assessments may be used to compute an overall quality index for researchers as readers.

These proposals, and similar ones, utilise crowdsourcing [3]: the activity of outsourcing, to large crowds of unknown people, tasks that are usually performed by a few experts. Although this approach might seem unfeasible – and it has had its share of criticism - there exist many successful examples of crowdsourcing. Wikipedia, for instance, is a free, good-quality crowdsourced encyclopaedia; marketplaces for crowdsourcing exist and are flourishing [L1], [L2]; and crowdsourcing platforms specialising in research and development activities are available [L3].

Indeed, crowdsourcing peer review would also address the issue of the scarcity of reviewers, which seems to be a growing problem, as discussed [3]. Although it is not clear whether crowdsourcing peer review will allow the quality of the scientific literature to remain high, this is an interesting research question that, in my opinion, is worth addressing.

Links:

[L1] www.mturk.com

[L2] www.crowdfunder.com

[L3] www.innocentive.com

References:

- [1] A. Tomkins, M. Zhang, W. D. Heavlin: “Reviewer bias in single- versus double-blind peer review”, *PNAS*, 114(48):12708–12713. November 28, 2017, <http://www.pnas.org/content/114/48/12708>.
- [2] L de Alfaro, M; Faella: “TrueReview: A Proposal for Post-Publication Peer Review” (white paper), technical report UCSC-SOE-16-13, 2016.
- [3] S. Mizzaro: “Readersourcing - A manifesto”, *Journal of the American Society for Information Science and Technology*, 63(8):1666-1672, 2012, Wiley.

Please contact:

Stefano Mizzaro, University of Udine, mizzaro@uniud.it

Taking Software Development Activities into Account when Evaluating Researchers

by Alain Girault and Laura Grigori (Inria)

Software is becoming increasingly important in academic research and consequently, software development activities should be taken into account when evaluating researchers (be it individually or teams of researchers). In 2011, a dedicated working group of the Inria Evaluation Committee addressed the issue of software evaluation and proposed criteria that allow researchers to characterise their own software. In the context of recruitment or team evaluation, this self-assessment exercise allows the evaluators to estimate the importance of software development in the activities of the authors, to measure the impact of the software within its target community, and to understand its current state and its future evolution.

In computer science and applied mathematics, software is crucial for several reasons:

- It can be used as a proof of concept to prototype an original idea.
- It can be distributed to an entire community, thus serving as a dissemination tool (quite often there can even be some competition between teams worldwide).
- It can be used as an experimentation platform (e.g., the Grid5000 management software).

- It can be delivered to industry and allow a faster transfer of academic results to industry.

Competitions between international teams address key challenges and aim at establishing new records with respect to a given criterion (e.g., size of the solved problems, speed or accuracy of the results). This is the case in several domains of computer science, including the model-checking, computer vision, high-performance computing, and the 3D meshing communities. In this context, it is mandatory to devise and implement new algorithms and to make the corresponding software publicly available. This allows the new algorithms to be tested by competitors.

A concrete proof of the growing importance of software is that several journals and conferences accept “software artifacts” to support a published paper, which are formally tested and evaluated. A recent trend focuses on the ability to reproduce the experimental section included in an article. Allowing this reproducibility is an important step forward towards reproducible research: given the same data and the same software, one has to derive the same results. Some journals already provide a “replication certificate” for the papers the results of which have been reproduced. We are convinced that this is an adequate step, and we support a publication model that, in addition to open access, also includes open data and open software, released simultaneously.

There are two different evaluation processes at Inria (in general, not only for software):

- The evaluation of the Inria teams that takes place every four years.
- The evaluation of the Inria researchers when they apply for various promotions or recruitments (in particular when they are hired as junior researchers and later when they apply for a promotion as senior researchers or higher ranks).

Taking software into account when evaluating a researcher or a research team at Inria has been common practice for over 15 years. Reflections on this topic started in the early 2000s within the Evaluation Committee of Inria, and a working group was created in 2011 under the guidance of Gérard Berry, which produced a document that allows the researchers and teams to evaluate their own software development [1]. The working group recommended the use of several qualitative and quantitative criteria to assess a software. This self-assessment process has been used at Inria for seven years and has been demonstrated to work satisfactorily. The criteria are then used by the evaluators to estimate the importance of software development in the activities of the authors of the software, to measure the impact of the software in a specific community, to understand its current state and its future evolution. More precisely, the criteria are:

1. *Audience*: ranging from personal usage to a massive usage.
2. *Software originality*: ranging from implementing known ideas to fully original ideas.
3. *Software maturity*: ranging from a software used for demonstration purposes only to a high assurance software certified by an evaluation agency (e.g., DO178 in civil avionics).
4. *Evolution and maintenance*: ranging from no particular plan to a fully organised support including, e.g., a user group or a forum where users can ask questions.

5. *Software distribution and licencing*: ranging from internal use within the authors' team to external packaging and distribution (e.g., as a Linux, Matlab, or R toolboxes or packages, just to take examples).

For each criterion, the software authors choose between the four or five available levels, spanning the extreme options listed above, the one that best characterises their software. Those levels are not used to rank the software, the higher levels are not necessarily the best. Indeed, the evaluators appreciate not only the size or the maturity of the software, but also the originality of the implemented ideas and algorithms.

In addition, the authors mention their own contribution along the different axes that matter for the production of software: design and architecture, coding and debugging, maintenance and support, and finally project management. For each of these criteria, an individual's contribution can range from occasional contribution to being (one of) the principal contributor(s). This second set of criteria is essential for highlighting the respective contribution of the different researchers to a particular software. As mentioned above, higher levels are not necessarily the best: to provide a concrete example, for the same software there can be the main architect who provided the core algorithms that form the basis of the software, and the main developer who coded the software and decided the architecture and the test and debug plan. Both roles are key to the success of the software.

In an additional succinct section, the researchers can add any information they consider relevant, such as the application targeted by the software, the user's community (research, education, industry), its impact, a short description of the differences with respect to other state-of-the-art software with same functionalities, and so on. This section should also include the programming language, the number of lines of the code, and a url from which the software can be downloaded. The url is very important so that the evaluators can download and test the software themselves. The criteria listed in this report are self-assessment criteria, so being able to download and test the software is essential to validate the self-assessment performed by the software authors. One trend that we are seeing at the moment is that source files are stored on repository systems like git, so the evaluators can monitor precisely the actual contribution of each author to a particular software.

In conclusion, we would like to emphasise that software development plays a crucial role at Inria, it definitely constitutes a key part of the activity of our research teams, both in the applied mathematics and computer science fields. As such, being able to evaluate the software produced by the teams and the researchers is of paramount importance. The criteria for self-assessment reported in [1] have been used for the past seven years and have proved to work very well.

Reference:

[1] Inria Evaluation Committee, "Criteria for Software Self-Assessment", 2011.

Please contact:

Alain Girault and Laura Grigori, Inria
alain.girault@inria.fr, laura.grigori@inria.fr

Open Access and Research Assessment: Why Would We Link the Two?

by Laurent Romary (Inria)

Despite some concerns within the scientific community, open access has the potential to offer a wealth of information for the strategic management of research.

Recent years have seen an increase in initiatives intended to foster open access, i.e. the free dissemination and re-use of scientific publications, at institutional, national and European levels. The European Union is taking a strong stance on the issue, making open access mandatory in its H2020 program, and funding the OpenAire initiative to coordinate publication repositories all over the continent. Several countries have also taken important legislative initiatives with the "Zweitveröffentlichung" principle in Germany and its equivalent in the recent "Loi pour une République Numérique" in France. Moreover, several institutions, such as Inria (FR), CentraleSupélec (FR) and the University of Liège (BE) have set up a "deposit mandate", whereby annual reporting is entirely based upon the scholarly publications that are made public in designated open publication repositories (e.g., HAL in France).

Whilst this clearly represents major progress for the fast and free dissemination of scientific knowledge, there are concerns within the scientific community that their activity and results may be scrutinised in an inappropriate way, and in particular that researchers' assessments may come to be based upon numerical impact measures computed from the material in publication repositories. Scientists are also uncertain with regards to the possible poorer quality of material put online at an early stage, as well as with the risk of being plagiarized, even if on the contrary open access is the best rampart against plagiarism. It is vital that we begin to balance these concerns by acquiring an understanding of the ways in which science and scientists may benefit from an ambitious open access policy, and that we communicate these benefits to the scientific community. For instance, a thorough coverage of the publication corpus of an institution, or of an entire country, could provide a means to analyse co-publication patterns, and in doing so could help identify collaboration strengths or weaknesses that could be fostered or improved. Text and data mining techniques can also provide mechanisms to identify expertise in a given research landscape or emerging topics within a given field. At the end of day, open scientific information could be used in a whole range of ways that we have to invent and deal with in an ethical way, in compliance with the DORA principles, for instance.

Right now we have even more pressing challenges, though, including: a) the possible fragmentation of the publication corpus and b) the actual quality of the available information. The first issue has to do with our capacity to have a coherent



Innovier pour retrouver de la croissance
Pour une République numérique

Mis à jour le 15 mai 2017

Contenu publié sous la présidence de François Hollande du 15 mai 2012 au 15 Mai 2017

La République du 21e siècle sera nécessairement numérique. Pour relever ce défi, la loi pour une République numérique, publiée au Journal officiel du 8 octobre, vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique.

Illustration 1: Loi pour une république numérique.

picture of the publication landscape and could be based upon the existence of a network of interoperable publication repositories where content, as well as access or download information, can be used, re-used and above all mined under open licences. Relying on private third parties (under the “gold open access model”) carries with it the risk that we may be left with diminished re-use possibilities for our own publications, which may end up spread across various publishers’ servers.

The second issue is even more central when considering publication information as the basis for the assessment and design of scientific policies. We need to ensure that the documentation associated with publication, comprising author identification, affiliations and precise publication information, is recorded and curated in a way that guarantees a trusted background for further analyses. For instance, it is important to maintain proper (open) authorities for research institutions upon which affiliation information can be based. In this domain we need to determine how much we can rely on third party initiatives such as ORCID where publishers may have too strong a voice.

If all the conditions are met for a trusted information corpus related to scientific publication, this could conceivably form the basis for an open, transparent process for research assessment exercises. In fact, this is exactly what the French assessment institution Hcéres is proposing, with their announcement in early 2018 that future assessment campaigns would rely entirely on the French national publication repository HAL to disseminate a) reference publications, as selected by the research organisations under assessment, as the sole source of information for reviewers and b) the self-assessment reports themselves, so that this information is made available to the wider research community and the public.

In conclusion, if we are serious about the issue of open access to publications - and probably to research data too - we could imagine setting up a visibility index that would relate to the proportion of openly available and re-usable

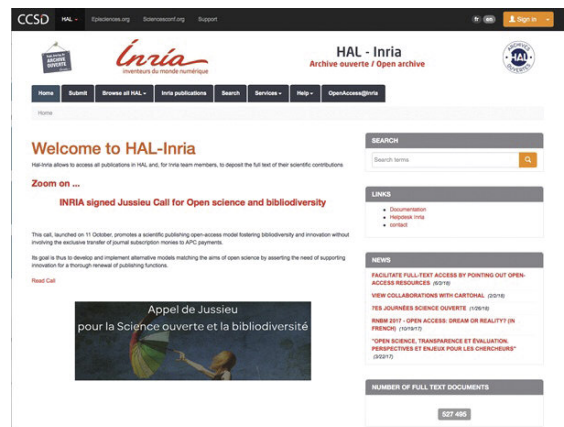


Illustration 2: The HAL-Inria publication portal.

content in a public repository in relation to the actual research output of an individual, team laboratory or institution. While not necessarily useful for “assessing” scientific outputs, such an index could encourage scholars to feel concerned with the wider dissemination of their own research.

At the end of the day, if open access policies are to be used for any form of assessment, its goal should be to assist the researchers themselves, providing them with appropriate data and tools to reflect on their own research trends, practices and, possibly, impact [1] [2]. This implies, in addition to the factors that we have touched on above, that openness also applies to research data, expressed in formats compliant with international standards, and tools, developed and distributed according to open source principles.

References:

- [1] S. Harnad: “Open access scientometrics and the UK Research Assessment Exercise”, *Scientometrics*, 79(1), 147-156, 2008.
- [2] L. Romary, C. Armbruster: “Beyond institutional repositories”, *International Journal of Digital Library Systems*, 1 (1), pp.44-61, 2010. <hal-00399881>

Please contact:

Laurent Romary, Inria, France
 laurent.romary@inria.fr

Revisiting the Scientometric Conceptualisation of Impact and its Measurement

by Giovanni Abramo (IASI-CNR and The University of Waikato)

“Impact”, jointly with “citation”, is probably the most frequent term in scientometric literature. It recurs so often that its definition is more and more taken for granted, hardly ever spelled out. Impact is often defined by means of the indicators employed to measure it, rather than as a concept in itself. The obsessive pursuit of measurement resembles a case of gold fever. One apparent aspect of this is to “first measure, and then ask whether it counts”. The aim of this short note is to revisit the conceptualization of impact, and relevant measurement indicator, which could serve in opening a discussion.

To better appreciate the role of scientometrics in assessing impact, we need to observe its position within the overall research value chain. Governments and private organizations invest in research to accomplish scientific and technical advancement, which has the potential of contributing to socio-economic progress. Taking for granted that the measurement of “social” impact is beyond the scope of scientometrics, what kind of impact is measured by scientometrics? To answer this question we have to consider scientific activity as “information processing”: the science system consumes, transforms, produces, and exchanges “information”. Scientometrics studies and analyzes, through the reference lists, the exchanges of information encoded in published papers. Scientists collect and analyze prior knowledge encoded in verbal forms and add value to it for producing new knowledge, nearly always encoded in papers made accessible to other scientists. For the production of new knowledge “to have an impact” it has to be used by other scientists. We can state that evaluative scientometrics is aimed at measuring/predicting the impact of research on future scientific advancement. We can conclude that scientometrics concerns “scientific impact” or “scholarly impact”.

But how to measure it? We frequently hear that the proper basis of research evaluation would be for experts to review the work of their colleagues. In fact, reviewers cannot assess impact, at the most being able to predict it, by examining the characteristics of the publication under evaluation.

In scientometrics citation is the natural indicator of impact, as it certifies the use of the cited publication towards the scientific advancement encoded in the citing publication. This position derives from the Mertonian or normative theory [1], according to which scientists cite papers to recognize their influence, being aware that exceptions (uncitedness, undercitation, and overcitation) occur. Although a reviewer might

judge certain characteristics of a publication, he/she cannot certify its “use”. Also, no other indicator certifies it. Journal impact indicators reflect the distribution of citations of all hosted publications, not the individual ones. Also altmetrics (i.e. Mendeley reader counts) do not certify the proper use of a publication. But this does not mean that journal impact and/or altmetrics cannot be useful in assessing the scholarly impact of a publication.

Ideally, when the life of a publication is over, i.e. when it is no longer cited, the accrued citations reflect the publication’s impact. Unfortunately policy-makers and managers, hoping to make informed decisions, cannot wait the many years (or decades) for citation life-cycles to end, and so to conduct research assessments.

The very challenge is: how to predict future impact? In literature the answer is not univocal, but what should not be in discussion is that late citation counts (as proxy of long-term impact) serve as the benchmark for determining the best indicator (and its predictive power), in function of the citation time window. Initially, scholars faced the problem of how long the citation time window should be for the early citations to be considered as an accurate proxy of later citations [2]. They have also investigated the possibility of increasing the accuracy of impact prediction by combining early citation counts with journal impact indicators [3].

The identification of alternative indicators or combinations, which could vary in type and weight across disciplines, as a function of the citation time window, and early-citations accrued, should be conducted by comparing their prediction of impact to the benchmark of late citation counts. This is the actual challenge for our community.

The multitude of assumptions, conventions, limitations, and caveats of evaluative citation analysis still apply; we just hope to have stimulated a proper discussion for a definitive convergence on the meaning and measurement of a fundamental concept of the scientometric science. Differently, every assessment exercise becomes metaphysical.

References:

- [1] R. K. Merton: “Priorities in scientific discovery”, in R. K. Merton (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 286–324), Chicago: University of Chicago Press, 1973.
- [2] G. Abramo, T. Cicero, C.A. D’Angelo: “Assessing the varying level of impact measurement accuracy as a function of the citation window length”, *Journal of Informetrics*, 5(4), 659-667, 2011.
- [3] G. Abramo, C.A. D’Angelo, G. Felici: “Predicting long-term publication impact through a combination of early citations and journal impact factor”, working paper, 2017.

Please contact:

Giovanni Abramo, IASI-CNR Italy and The University of Waikato, New Zealand
+39 06 72597362, giovanni.abramo@uniroma2.it