

Open Archives and the Invisible College: driving forces towards a new publication model

Nicolas Holzschuch

► **To cite this version:**

Nicolas Holzschuch. Open Archives and the Invisible College: driving forces towards a new publication model. Kurt Mehlhorn and Moshe Y. Vardi. Dagstuhl Perspectives Workshop: Publication Culture in Computing Research, Nov 2012, Dagstuhl, Germany. 2012, <10.4230/DagRep.2.11.20>. <hal-00749806v3>

HAL Id: hal-00749806

<https://hal.inria.fr/hal-00749806v3>

Submitted on 13 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archives and the Invisible College: driving forces towards a new publication model

Nicolas Holzschuch¹
INRIA

Abstract: Publication methods have been and are constantly evolving in Computer Science. My position, expressed in this paper, is that in this current landscape, there is room for a new publication model, combining OpenArchives for immediate access with editorial peer-reviewing. A key interest of this proposal is that peer-reviewing happens after the results have been released. I start by reviewing recent changes in publication methods in Computer Science and some of their consequences, drawing practical examples from Computer Graphics. I then review existing forces that are contributing to changes in publication methods, and how these forces push towards a new publication model. I describe this model, and review practical requirements to make it work.

Publications in Computer Science: a changing landscape

Computer Science has long been an outlier among scientific disciplines. Research in our domain is published preferably in conferences and workshops, while in other disciplines the usual benchmark is papers published in journals. This specificity is partly related to a positive feedback loop: we submit papers in conferences because they have a strong reputation in the field, conferences select from a larger pool of papers, improving the overall quality of papers published, thus increasing the reputation of the conference. Papers published in selective conferences have a larger impact². The fast reviewing cycle in conferences is also attractive for the authors, especially at an early stage of their scientific career.

The classical landscape has been evolving in recent years, with cooperation and exchanges between conferences and journals. This cooperation can take multiple forms³:

- The conference proceedings are published as a regular issue of a journal (for example the *Siggraph* proceedings as issue 3 of *ACM Transactions on Graphics*, since 2002). Most major conferences in the field are now following this model.
- Selected papers from the conference are published in a journal, after a second review cycle and addition of extra material. This model is preferred for less selective conferences.
- Papers from a journal can be presented at the conference, either in a separate session or merged with regular conference papers (such as *ACM Transactions on Graphics* papers at *Siggraph* since 2010). Most major journals are offering this service.

These changes have brought Computer Science (or at least, Computer Graphics) more in line with the other disciplines; they also had a positive impact on the journals: the total review cycle—the amount of time between the day the paper is submitted for review and the day it is published—is now less than a year for all major journals in Computer Graphics, with a first decision in less than 4 months.

¹ INRIA, France.

² Jilin Chen and Joseph A. Konstan. 2010. [Conference paper selectivity and impact](#). *Commun. ACM*, vol. 53, n° 6 (June 2010), 79-83. DOI=10.1145/1743546.1743569

³ I am taking examples from the field of Computer Graphics, mainly because I happen to know it better, but also because the convergence between journals and conferences started earlier.

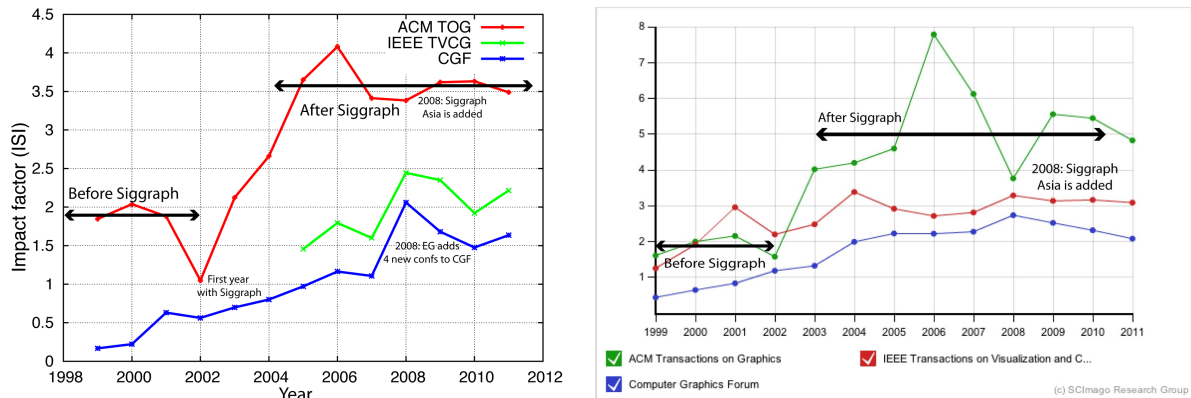


Figure 1: Evolution over time of the Impact Factor for several reviews in Computer Graphics. Merging a major conference (*Siggraph*) with a journal (*ACM Transactions on Graphics*) increased the impact factor of the journal. **Left:** 2-years IF measured with ISI Web of Knowledge. **Right:** number of citations per document over 2 years, measured with ScimagoJR (using Scopus). Adding *Siggraph Asia* in 2008 doubled the number of citable papers.

An interesting side effect is that the impact factor of the journals has been significantly increased by these changes. This could be an experimental confirmation of the study by Chen and Konstan. Figure 1 shows the impact factor of three major journals in Computer Graphics as a function of time. The first year after *Siggraph* proceedings were added to *ACM Transactions on Graphics*, the impact factor of the journal decreased because the number of citable papers had been multiplied by 6. After a short transition, the impact factor was multiplied by 2.

A less positive side effect of this convergence between conferences and journals is that journals are more likely to reject papers, not because they do not have the sufficient scientific quality, but because guiding and nurturing them until they are worthy of publication would take too long. For at least one major journal, if three reviewers say “Accept with major revisions”, the editor has to *reject* the paper.

Forces driving towards further changes

Several forces can be seen as driving current and future changes in the publication culture. Some are specific to Computer Science, while others are more generic.

Open Access mandates:

Any publications by researchers that are employees of the United States government is in the public domain⁴. This policy has had little practical impact because the federal government employs few researchers, but it has been the inspiration for other Open Access mandates in recent years. Research funded by the NIH⁵ must be placed in an open archive within 12 months of publication. Major research universities (including, but not limited to, the MIT⁶ and Harvard⁷) require that work authored by their employees must be placed in an Open Archive. The European Union is considering a similar requirement on research projects it finances⁸.

Several reasons explain this strong push for Open Access. Some of them are ethical: these research have been funded by the public, and the public should be able to see the results. Others are practical: the subscription

⁴ US Code, Title 17, Chapter 1, § 105: Subject matter of copyright: United States Government works.

⁵ National Institutes of Health Public Access. 2008. <http://publicaccess.nih.gov/>

⁶ MIT Faculty Open Access Policy. 2009.

<http://libraries.mit.edu/sites/scholarly/mit-open-access/open-access-at-mit/mit-open-access-policy/>

⁷ Harvard University. 2010. *Open Access Policies*, <http://osc.hul.harvard.edu/policies>

⁸ Neelie Kroes. 2012. *New Commission measures to open up science in Europe*.

<http://blogs.ec.europa.eu/neelie-kroes/open-science/>

costs for major publishers are high, reaching a point where even major research institutions can not afford them. Open Access mandates are seen as a way to drive these subscription costs down.

The role of the industry:

Another specificity of Computer Science is the strong connection between our research and the industry. Companies are working with researchers, re-using results from scientific papers, sometimes funding research groups. Major companies have their own research centers. This strong connection is visible in the industrial exhibits at major conferences. This relationship between research and industry is not a one-way street: the industry also does research and publish; 50% of technical papers at *Siggraph* 2011 had at least one author working for a private company (the same proportion holds for *Sigcomm* 2011 and other conferences). Some of these technical papers from the industry are seminal, long-term research papers.

This strong connection is a driving force behind many important research problems. It is also a force pushing for fast dissemination of results. Private companies want fast results and quick publications. In that respect, they might be more impatient than academic researchers: one of the most important recent contributions in Global Illumination came from the industry and was only published as a technical report⁹, yet it had a major impact on both research and industry. The author chose the technical report format for faster dissemination and discussion.

Individuals and Social Networks:

Most publishers in Computer Science have opted for a Green Open-Access model: authors can place the final version of the manuscript, post-reviewing, on a personal web site or institutional repository. Researchers have embraced this model because they can propagate their results faster: final versions of papers are published as soon as authors receive the “final accept”, several months before the paper is available on the publisher digital library (usually on the first day of the conference). This quick publication the web has empowered individuals, and changed the way we do research: papers are downloaded, discussed on social networks and sometimes replicated months before the actual presentation at the conference. The discussion on social networks¹⁰ is a little like the discussion you can have after presentations at a conference, but more lively: people are less shy, don't hesitate to engage the author, ask questions, or to point out that this paper is useful (or not) for their industrial problems. This discussion looks a little like an *Invisible College*, a group of people working together even though they are not physically in the same room.

This immediate publication on web pages has also filled a need in the community. There is always a delay between the time when notice of acceptance is sent to authors and the time when the conference program is made public, with the full list of accepted papers. Researchers often complain that they would like to see the full list of accepted papers earlier. Now, armed with a classical web search engine, anyone can assemble a list of papers accepted at a conference, within hours of the decision. It might be a partial list, but that is better than no list. More importantly, researchers can share the list of papers they have found, and this list will likely include links towards the authors version of the papers. <http://kesen.realtimerendering.com/> provides this list of papers, for all major conferences in Computer Graphics since 2000; these pages have become an invaluable tool for all researchers. In 2009, ACM tried to take them down arguing about copyright issues, but quickly reversed its position after a very strong show of support from the community¹¹.

A new publication model

⁹ Per H. Christensen. 2008. Point-Based Approximate Color Bleeding. Pixar Technical Memo #08-01. Cited 24 times according to Google Scholar, implemented by Pixar, ILM and Dreamworks Animation as early as 2008.

¹⁰ In Computer Graphics, researchers and industry developers are using mainly Twitter and blogs.

¹¹ Including support by the *Siggraph Asia* 2009 technical papers chair. For more on this, please see, e.g.: <http://www.realtimerendering.com/blog/ke-sen-huangs-paper-pages-are-down-will-soon-go-back-up/>.

I argue that the convergence between journals and conferences is unlikely to slow down or be reversed, because it answers the needs of journals, conferences and researchers: it increases the number of high quality papers published in journals, it gives conferences a better visibility outside of Computer Science and it helps researchers in their job applications and tenure reviews. Yet it leaves an empty niche in the publication landscape for Computer Science. Because of their short review cycle, conferences in Computer Science tend to accept for publication only papers that are close to perfection, both in their scientific content and presentation. Shortening the review cycle for journals has brought them in line with conferences in that respect: a paper with great scientific content but poorly written is now likely to be rejected at major journals, because improving the writing would take too long. The same applies to ideas that are too original or too innovative. Author can of course revise and resubmit their papers, either at the same venue or at a different venue. But this increases the delay before the idea is finally available for the public. Many shareholders, including those from the industry, are pushing for faster dissemination of ideas.

To fill this niche, I suggest to revert the order of publication: first, the author makes his paper available by placing it in an Open Archive. Once it is there, it becomes open for both *comments* (by the community) and *reviewing* (by journals):

- *Comments* by the community allow a discussion around the paper and the ideas or algorithms it presents. It provides several levels of information: opinion about the paper or its usefulness, replicability. It is also a good place to ask about details of the algorithm or implementation details. It reproduces the discussion in e-mails, social networks and conferences, but the questions and answers can be made visible to all.
- *Reviewing* by journals gives a level of scientific certification to the paper. Journal editors assign reviewers to a paper, and based on their reviews decides to accept the paper in this journal. This acceptance is both a stamp of quality for the paper (it appears on the paper web page), and a way to access the paper from the journal portal.

Based on readers comments or reviews, the author can revise his paper (either minor or major revisions). Reviewers for a journal can be anonymous or not (single blind reviewing). New ideas will be available faster, but we maintain the scientific validation through peer-reviewing. This system has been proposed already by several people, in different domains. It is both feasible and desirable at this point in time, but we must meet several conditions:

- We need an open-archive system:
 - that can manage the discussions around a paper, both public and private.
 - that can handle all kind of supplemental materials (pictures, sample code, videos, models) both in the paper and the comments.
 - robust enough to handle the load, and nimble enough to implement the required changes.
 - shared by researchers from different research domains.
- We need the researchers to trust the system:
 - from the scientific point of view, this means high-quality scientific reviewing and editing.
 - from a practical point of view, the system must have independence, durability and stability.
 - to avoid negative incentives, I suggest financial independence from the number of accepted papers (unlike in the Gold Open-Access model).

These requirements point towards a central repository such as HAL¹². This repository would be managed by a consortium and have financial independence through backup from public shareholders, such as states and research institutes. INRIA is ready to support this new publication model.

¹² Laurent Romary and Chris Armbruster. 2010. [Beyond institutional repositories](#). *International Journal of Digital Library Systems*, vol. 1, n° 1, p. 44-61. DOI: 10.4018/jdls.2010102703.