

Journal Relevance

Richard Snodgrass

rts@cs.arizona.edu

As I write this, I am thinking back to the SIGMOD and PODS conferences, held at FCRC in San Diego just a few weeks ago. What again struck me at these conferences is the merit of all of the sessions, whether research paper, industrial, or tutorial. Our field is blessed with several first-rate conferences and a wide range of specialized workshops through which much quality research can be communicated.

I'm also struck with the wealth of material in the SIGMOD *DiSC* DVD included with the June issue of *SIGMOD Record*, which I received upon my return. DiSC2001 includes 14 conferences and a total of 1158 PDF files; DiSC2002 includes an incredible 26 conferences and 1630 PDF files. My appreciation goes out to editors Isabel Cruz and Aidong Zhang and their editorial boards for assembling this valuable resource, and for continuing to expand it so effectively.

In this internet age, where one can access anything on the web anywhere in the world in just a few seconds or communicate in real-time with instant messaging on their computer or cell phone, a common concern is whether those dinosaurs of scientific publishing, the technical journals, are still relevant. This raises the related question of, "precisely, what is relevance in the context of a publication?" In this column, I come at this question from four related metrics, and use *ACM Transactions on Database Systems* as a particular example, in part because I have access to more statistical information on this journal. I compare this journal with two of the prevalent conferences, SIGMOD and PODS. For many of the statistics, I'll use the just completed year, 2002.

1 Turnaround Time

Turnaround time is the interval between the submission, usually electronic, of a manuscript or a revision and the sending of the editorial decision, now almost entirely by electronic mail. A publication with a short turnaround time may be considered more relevant than one with a longer turnaround time.

Conferences advertise their turnaround times in advance, by announcing the submission deadline and the notification date. SIGMOD'02 requested papers on November 2, 2001 and notified authors on January 22, 2002, a turnaround time of 11.5 weeks¹. It achieved this fast turnaround time by requiring all submissions to be synchronized and by assembling in advance a program committee. PODS'02 requested papers on November 9, 2001 and notified authors on January 29, 2002, a turnaround time also of 11.5 weeks.

TODS allows submissions at any time, and assembles a unique set of reviewers for each manuscript. Such flexibility necessarily increases the turnaround time, but as Figure 1 shows, not by much².

The turnaround time has been slowly decreasing over the past two years. This figure shows four sets of data. The bottom line is the *average turnaround time*, a moving average of the turnaround time for papers submitted in the indicated month. To smooth monthly variations, the moving average includes all of the submissions for the previous year. Each data point represents dozens of papers. The value for February 2003, 13.7 weeks, is the average turnaround time for all of the papers submitted between (inclusive) March 2002 and February 2003.

The next line up is the average turnaround time for external reviews only, a moving average of the turnaround time for papers submitted in the indicated month. This includes only submissions that went out to external reviewers and specifically excludes desk accepts and rejects. The value for February 2003, 17.1 weeks, is the average turnaround time for external reviews of all the papers submitted during the year up through February 2003.

¹For computational simplicity, months are considered to have thirty days.

²This figure is from <http://www.acm.org/tods/TurnaroundTime.html> where it is updated monthly.

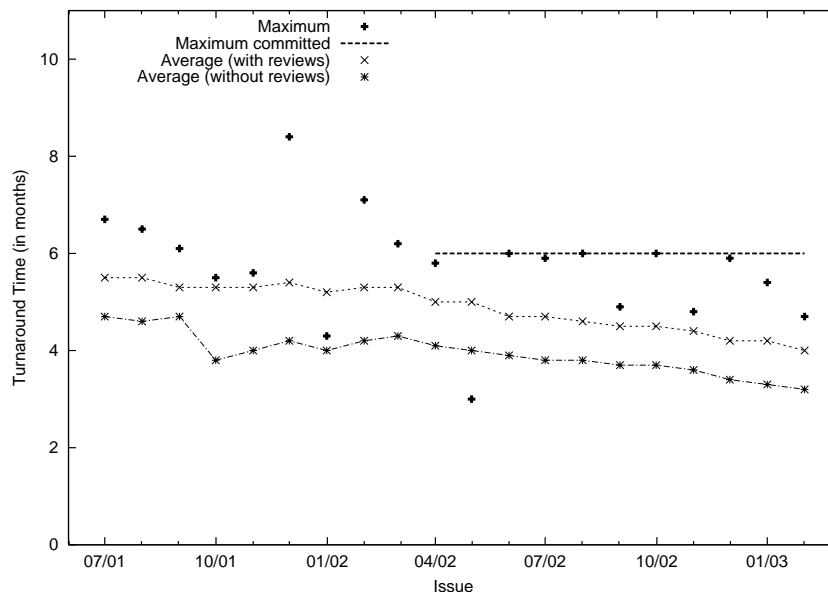


Figure 1: ACM *TODS* Turnaround Time

The straight line is the *committed maximum turnaround time*, the boundary that the Editorial Board has committed to not exceed, for any submission. The Editorial Board recently established a formal policy stating its commitment to providing an editorial decision within 6 months. I don't know of another journal that has been willing to publicly announce such a commitment and then provide statistics indicating compliance.

The points, one per month, denote the maximum or peak turnaround time for submissions in the indicated month. Each point represents a single, unusually slow paper submitted during the indicated month. For all the papers submitted in February 2003, the longest turnaround time was 4.7 months. For the year this commitment has been in place, no paper has required more than six months.

In terms of turnaround time, *TODS* at 13.7 weeks is approaching conferences (as exemplified by SIGMOD and PODS at 11.5 weeks), and *TODS* follows conferences in publicly stating a deadline for notification.

2 End-to-End Time

End-to-end time is the interval between the original submission of a manuscript and the appearance in print of (generally a revision of) that manuscript. (An alternate definition uses the time the paper appears electronically, though that time is much harder to determine for papers in the past, and so is not reported here.) A publication with a shorter end-to-end time might be considered more relevant.

One difference between conference and journals is that the former is one-shot, up or down, whereas journals permit a dialog between the reviewers and the authors to obtain a publishable paper. Many papers rejected by a conference appear in a subsequent conference, though I am not aware of any study of how many reviewing cycles are endured by the average conference paper. (The same holds for some journal articles, which are rejected from one or more journals before finally being accepted.)

For SIGMOD'02, submissions were due November 2, 2001 and the paper appeared in print the first day of the conference, June 3, 2003, for an end-to-end time of seven months. For PODS'02, the submission date was a week later, so the end-to-end time was 6.8 months.

For journals, the end-to-end time comprises the turnaround time for each cycle, the time for the author(s) to prepare zero, one, or several revisions, the time the paper sits in the queue waiting for a slot in an issue,



Figure 2: ACM *TODS* End-to-End Time

and the time for the publisher to copy edit, typeset, proof and print the paper. Figure 2 shows the data for *TODS*, calculated from the submission date as indicated on the last page of the article in the journal and from the cover month of the issue³. This data does not take into account that some issues over the past few years were printed late, nor does it include data for the first volume, as papers in that volume do not have a “submitted on” date.

The end-to-end time started at 11 months in 1977, grew to almost 3.5 years in 1991 (can you imagine waiting over three years for your paper to wind its way through the reviewing and production process?!), then fell in spurts, to 17 months in 2003 and 16 months for the first two issues of 2003. The 2002 and 2003 values are below all but those of the first three years of the journal, over twenty years ago.

3 Reference Age

There is an interesting CiteSeer page⁴ with statistics on *publication delay*, which “is estimated using the average age of the most recent citation in articles at the time of publication.” 621 conferences and journals are included. PODS comes in about half-way down, at position 291 at 9.13 months, SIGMOD is right there at position 294 at 9.26 months, and *TODS* is way down at the bottom, at position 597 at 27.00 months (to four significant digits!). There is a big difference here between conferences and journals.

Or is there? A recent *CACM* article⁵ entitled “Of course it’s true; I saw it on the Internet!: critical thinking in the Internet era,” by Leah Graham and Panagiotis Takis Metaxas, studied use of the Internet by students for research. The paper concludes, “Clearly, students consider the Internet a primary source of information. The results presented here suggest many students have difficulty recognizing trustworthy sources, though perhaps the underlying problem is a lack of understanding of the Internet as an unmonitored source of information.”

³Thanks to Tessa Chalberg and Eric and Melanie Brucks for collecting these statistics. This figure is also on <http://www.acm.org/tods/TurnaroundTime.html>.

⁴<http://citeseer.nj.nec.com/pubdelay.html>

⁵Volume 46, Issue 5, May 2003, pp. 71–75, <http://doi.acm.org/10.1145/769800.769804>

Many researchers use CiteSeer to locate useful articles. Some even quote statistics from CiteSeer⁶. But where does CiteSeer get its data? Quoting from the CiteSeer publication delay page, “Generated from documents in the CiteSeer database. This list is automatically generated and may contain errors. Only venues with at least 15 articles are shown.” The CiteSeer database contains only publicly accessible papers. Most journals, including *TODS*, are not publicly accessible. So what CiteSeer does is find papers on the web (preprints, previous versions, even technical reports) with the same title and authors. What is problematic is that it uses those papers on the web, which are not from the journals, to make statements about the journals themselves. At the very least, CiteSeer should state up front this critical proviso.

Ah, but SIGMOD and PODS papers are both freely available through the ACM DL. So shouldn't those statistics be accurate? Unfortunately, no. The ACM DL provides a query interface, so its content, even if freely accessible, cannot be retrieved by web crawlers, and thus is out of reach of CiteSeer. I checked, and CiteSeer doesn't know about most of my SIGMOD and PODS papers. So these statistics are over an unreported number of articles, related in some undetermined way to the actual papers in the conferences and journals.

Unlike CiteSeer, I have access to all of the SIGMOD, PODS and *TODS* papers. So I manually computed these statistics for the year 2002, which was a pain. The *reference age*, the average interval from the citation of the most recent published paper to the print publication date, was 8.2 months for SIGMOD, 10 months for PODS, and 13 months for *TODS*. (As the turnaround time for *TODS* is falling, so is the average reference age, which is right at 10 months for the first half of 2003.) So the CiteSeer statistics are misleading, in that, at least for 2002 and 2003, there is little difference between *TODS* and top conferences concerning reference age. (From Figure 2 we see that end-to-end time has decreased dramatically in the last decade. My guess is that the reference age follows this trend, and so the *average* reference age for *TODS*, computed over the last 25 years, is probably much higher than that for the last few years.)

The National Research Council published a study about ten years ago⁷ with an appendix entitled “Comparing Journal and Conference Publication.” In addition to time to publication⁸, this study also examined two other metrics: “fraction of references to papers less than two years old,” with conferences having a higher fraction, and “median age of reference,” with conferences having a lower age. It is not clear whether these metrics are useful indicators of journal relevance. While it is certainly important to reference the most recent work, it is also important to reference related work done in the past, even long in the past. In this aspect, I feel that CiteSeer got it right, that the average age of the most recent citation is a more appropriate metric.

4 Selectivity

Finally, the NRC study mentioned selectivity, commenting that “Although the journal reviewing process may be more thorough, the prestigious conferences are highly selective.” They reported acceptance rates of 18–23%; database conferences are in that range: SIGMOD'02 had an acceptance rate of 17.5% and PODS'02 of 22.0%. Continuing, “An informal survey of editors of other major ACM and Institute of Electrical and Electronics Engineers (IEEE) journals found acceptance rates that vary from 25 to 35 percent. The private journals tend to be in the same range, although at the higher end.” *TODS* is at the very bottom of that range, more similar to conferences than to private journals in terms of selectivity.

⁶I confess: I did so in a September 2001 *TODS* editorial (<http://doi.acm.org/10.1145/502030.505049>): “*TODS* fares similarly well in an summary of estimated impact from the Research Index database (<http://citeseer.nj.rec.com/impact.html>), which ranked journals according to their average citation rate. *TODS* was judged the database journal with the most impact, appearing in the top four percent of the 800-odd journals and conferences analyzed.”

⁷“Academic Careers for Experimental Computer Scientists and Engineers,” Computer Science and Telecommunications Board, National Research Council, 1994.

⁸This is our end-to-end time. They measured *ACM TOPLAS* to have a time to publication of 32.4 months and *ACM TOCS* of 21.3 months. At the time the study came out, 1994, *TODS* had an end-to-end time of 30.5 months, almost double what it is now. It would be interesting to see if the end-to-end times of *TOPLAS* and *TOCS* have also halved in the intervening ten years.

5 Editorial Board

In summary, *TODS* is within just a few weeks of conferences (specifically SIGMOD and PODS) in terms of turnaround time (13.7 versus 11.5 weeks), about twice that of conferences for end-to-end time (17 versus 7 months, reflecting multiple review cycles), and similar to top conferences in reference age and selectivity.

When an editorial decision is slow in coming, editors handling papers often blame the unresponsiveness of reviewers and of authors in revising papers. What the editors don't want you to know is that the turnaround time is *entirely* under the control of the editor. I can say this because the *TODS* Associate Editors handle papers by the same authors as papers handled by other database journals and use the same pool of reviewers as the other database journals. Additionally, the end-to-end time is mostly under the control of the editor, the reference age for a paper mostly follows its end-to-end time, and so that too is indirectly under the control of the editor, and selectivity is solely under the control of the editor. Those that claim otherwise are trying to pass the buck.

So I want to thank the following people, who comprise the *TODS* Editorial Board, for their hard work in ensuring that, as just discussed, *TODS* is just as relevant as the top conferences, by several metrics.

Rakesh Agrawal, IBM Almaden Research Center
Peter Buneman, University of Edinburgh
Michael J. Carey, BEA Systems, Inc.
Surajit Chaudhuri, Microsoft Research
Michael Franklin, Univ. of California at Berkeley
Christian S. Jensen, Aalborg University

David Lomet, Microsoft Research
Alberto Mendelzon, University of Toronto
Z. Meral Özsoyoğlu, Case Western Reserve
Gerhard Weikum, Max-Planck Institute
Marianne Winslett, University of Illinois

These 11 people are providing a truly valuable service to readers, to authors, and to reviewers.

6 Surveys

A year ago I announced in this space⁹ that *TODS* is now accepting focused survey articles. I'm pleased to report that this month's issue has the first such survey, "Are quorums an alternative for data replication?" by Ricardo Jiménez-Peris, Marta Patiño-Martínez, Gustavo Alonso and Bettina Kemme. This survey is somewhat unusual in that it uses extensive experimentation to compare existing quorum-based data replication protocols with each other and with the conventional read-one/write-all-available approach. Their conclusion may surprise you. This article, and other upcoming articles, were on the web¹⁰ several months before they appeared in print. Check them out!

⁹<http://www.acm.org/sigmod/record/issues/0209/ricksnodgrass.pdf>

¹⁰<http://www.acm.org/tods/Upcoming.html>