DOI:10.1145/1810891.1810911

Computing research ages more slowly than research in other scientific disciplines, supporting the call for parity in funding.

BY DAG I.K. SJØBERG

Confronting the Myth of Rapid Obsolescence in Computing Research

COMPUTING TECHNOLOGIES ARE changing everyone's social, political, economic, and cultural worlds.[12] Meanwhile, scientists commonly believe that research in computing is advancing more quickly and just as quickly becoming obsolete more quickly than research in other scientific disciplines. A notable indicator is how quickly it stops being cited in the literature. Common measures of this phenomenon are called "cited half-life," "citing half-life," and the Price Index (see the sidebar "Definitions and Measures of Obsolescence"). These measures show that research in computing does not cease being cited any more quickly than research in other disciplines, indicating (contrary to popular belief) that research in computing does not become obsolete more quickly than research in other disciplines. The extent to which this is the case is important for several reasons:

Demand for innovation. Though computing has made great strides, society continues to demand more complex, reliable, robust, usable hardware and software systems. Advances in computing technology needed to meet it depend on long-term funding of fundamental research.[11] However, it can be difficult to convince funding bodies to support long-term fundamental research programs in computing. One reason may be the already quick pace of development of computing applications, perhaps suggesting that the research is not as difficult as in other disciplines and that progress can be made with less funding than other disciplines. Hence, as has been reported in the context of U.S. National Science Foundation research-funding policy, when competing for research money, computer scientists argue that society has a compelling need for the results of their research, as well as CS as a basic research discipline to maintain its standing within the scientific community.[19] Competing for funding with researchers from other sciences in a university setting, CS researchers must counter the argument that research funding in computing will not be prioritized because everything useful is already being done both faster and better by the IT industry anyway.

» key insights

- With respect to aging in the research literature, CS is in the middle of the scientific disciplines.
- The research challenges in computing are as fundamental and long-lasting as those in other disciplines.
- Publication delay is not a major problem within CS.



ILLUSTRATION BY GARY NEILL

Aging research. Though relevant, the CS literature may still be considered obsolete and thereby ignored due to its age. As a researcher and journal editor, I find that reviewers frequently mention "old references," and, as a supervisor, I find Ph.D. students are often reluctant to read older literature.

Publication delay. Researchers in computing sometimes claim the relatively long lag between submission and publication of a journal article renders the research outdated before publication, arguing for submitting their manuscripts to conferences rather than to journals.

Library ROI. Due to the ever-increasing volume of research literature, libraries must make cost-effective decisions, identifying the core journals within a discipline, canceling their subscriptions to less-accessed journals, and archiving less-accessed material to save shelf space. To maximize return on their investment, libraries must collect statistics on the use of their materials.[9] Research literature on computing being accessed less often or quickly becoming obsolete may affect decisions about the archiving and retention of computing journals.

Results

Table 1 reflects CS within the various disciplines with respect to average aggregated cited, citing half-lives, and Price Index. This result is in striking contrast to the only other work I found on obsolescence of the computing literature—Cunningham and Bocock[3]—which found a citing half-life of four years (I found 7.5), concluding that their study supported "...a commonly held belief about computer science, that it is a rapidly changing field with a relatively high obsolescence rate for its documents. This hypothesis is confirmed for the field of computer operating systems and network management..." They also reported a half-life of five years for the field of "Information Systems." The main reason for the discrepancy between their results and mine is likely that they based their analysis on a small sample—only two journals (one that no longer exists) and four issues of the proceedings of one conference, the International Conference on Information Systems. By contrast, *ISI Journal Citation Report* (JCR) provided me with values for 382 computing journals.

The extent to which the cited and citing half-life measures are equivalent or complementary has been covered in the literature.[17] For individual

Definitions and Measures of Obsolescence

All disciplines include foundational research that is relevant to ongoing research but that does not need to be (and is not) cited; an example is Newton's *Principia Mathematica*. However, it is generally reasonable to assume that researchers cite other research in their publications if they consider it relevant. Consequently, the median age of the references to, or in, published articles within a field is an indicator of how quickly the literature becomes obsolete. Inspired by the same term in nuclear physics, this indicator, called "half-life," has been used for a long time in bibliometric research.[18]

There are basically two ways—retrospective studies and prospective studies[2]—to weigh the obsolescence of an article, a journal, or the body of the literature of a (sub)field. In retrospective studies, one proceeds backward from a particular date. The JCR[8] provides two retrospective half-life measures:

Cited half-life. The cited half-life of a journal for a particular year is the number of years (counting backward from and including that year) accounting for 50% of the citations received from the sample of journals under consideration. Cited half-life shows how quickly articles published in a given journal, taken as a set, cease to be cited. To illustrate, the red arrows in the figure here indicate articles published in 2007 in various journals that cite articles in *Communications* independent of year; assume in this example only one article per journal. One must then go back to 1998 to include 50% of the citations to *Communications*, giving a cited half-life of 10 years. In reality, JCR listed 8,969 citations from articles published in 2007 in 624 journals or other sources to articles published in *Communications*, but 63% of the citations were to articles published in 1997 or earlier; that is, the cited half-life was greater than 10 years. (For half-lives >10, JCR reports only the text ">10.") The definition of cited half-life can be modified to cover subject categories or research fields by considering citations to articles in a set of journals representing the category or field. This aggregate cited half-life is an indicator of the turnover rate of the body of work on a subject or in a field.

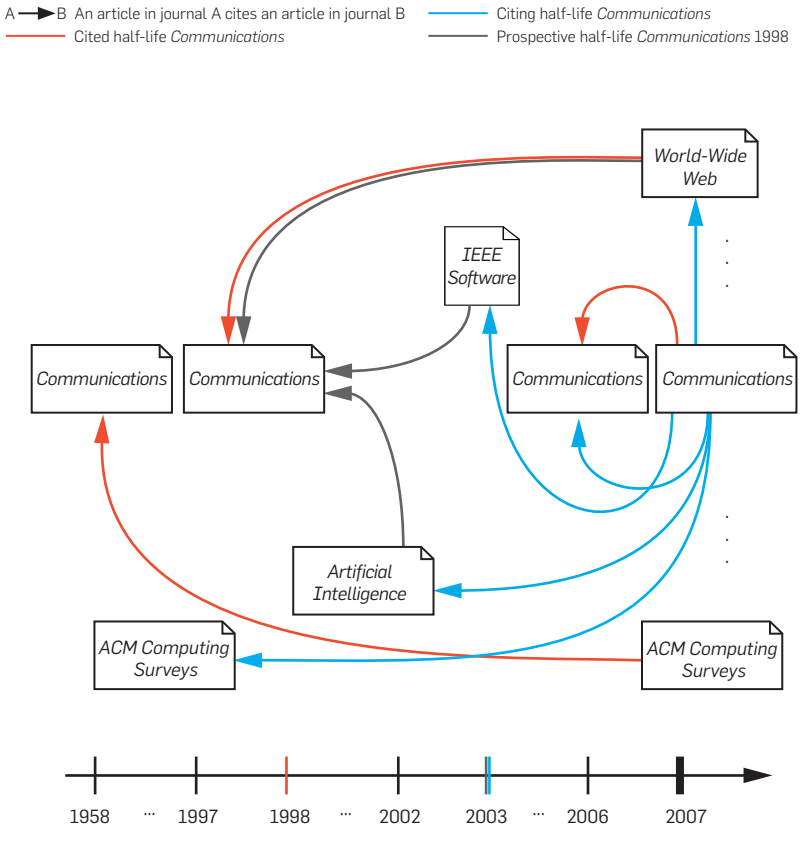
Citing half-life. The citing half-life for a particular year is the median age of all articles cited in a given sample of articles. In the figure, the blue arrows indicate citations in 2007 *Communications* articles to five articles in various journals in different years. The median year of publication of these five articles is 2003; the citing half-life is five years. In reality, the citing half-life for *Communications* in 2007, as reported by JCR, was 5.5 years, calculated on the basis of 1,607 citations to articles in 155 journals or other sources. (JCR reported half-life values in decimals because it used an interpolation formula in the calculation.) The citing half-life shows how quickly a journal ceases citing articles from itself or from other sources. This definition can be modified to cover subject categories or research fields.

Related to the measure of citing half-life is the Price Index, defined as the proportion of articles cited by a publication that is no more than five years older than the publication doing the citing.[4] The index is an outcome of Derek J. de Solla Price's work at the University of Malaya in Singapore, Cambridge University, and Yale University in the 1950s and 1960s on the growth of knowledge in science and the "research front of recent papers."[5] A large index value indicates a discipline characterized by quick growth and an active research front.

In prospective studies, one investigates the history of citations that have been made to a particular article or set of articles after publication over a given time period, typically 10 to 15 years.[6] Half-life is defined as the time period over which half the citations to the (set of) articles were made; for example, if we were to calculate the prospective half-life of *Communications* for 1998 in a 10-year window, we would have to determine the number of citations to the 1998 *Communications* articles from 1998 to 2007. The figure outlines how articles published in the journals *Artificial Intelligence* in 2002, *IEEE Software* in 2003, and *World Wide Web* in 2007 cited articles in *Communications* in 1998 (dark gray arrows). If the citations were from one article in each of the three journals, the prospective half-life would be six years, and the median citation would be in *IEEE Software*.

The advantage of prospective half-lives is that researchers are able to track the use of individual articles. However, calculating prospective half-lives is challenging and not provided by JCR. Consequently, I don't report prospective half-lives here but include the definition for the sake of completeness.

Citation half-lives.



journals, the values may be quite different (such as the respective values >10 and 5.5 for *Communications*). Nevertheless, as in Table 1, there is strong correlation among the three measures of obsolescence at the level of overall disciplines: $r_{\text{cited-citing}} = 0.90$, $r_{\text{cited-Price Index}} = -0.83$, $r_{\text{citing-Price Index}} = -0.89$). Literature with a long lifetime has high cited and citing half-life values but low Price Index, and vice versa, giving negative correlations between cited/citing half-live and Price Index.

Looking at the subdisciplines within computing, one finds the citation lifespan of the literature is shortest within Information Systems and longest within Theory & Methods (see Table 2). The variations among the subdisciplines of the various disciplines are generally small; on average, the cited half-life stddev = 0.8. The variation among the journals within a discipline is much greater; on average, the cited half-life stddev = 2.2. For computing journals, stddev = 2.0, with the extremes varying from cited half-life = 1.7 and citing half-life = 3.7 to >10 for both half-life measures.

Based on the assumption that everything is changing quickly throughout society, it is easy to believe that the scientific literature is becoming obsolete more quickly than it used to. However, a comprehensive study shows that the median citation age (citing half-life) of scientific publications has increased steadily since the 1970s.[10] One likely reason for this increase is the availability of online bibliographic databases and the Internet, making it easier to access older references. A 2008 study reported, "The Internet appears to have lengthened the average life of academic citations by six to eight months."[1] Another reason may be the significant increase in the number of references per article.[10] Having space for more references allows for increasing the time period for included references.

The reported study[10] focused on medical fields, natural sciences, and engineering. To study the evolution of the aging distribution of the computing literature compared to all other disciplines, I investigated cited half-lives from 2003 to 2007 (see the sidebar "How the Study Was Done"); JCR did not provide such information earlier than 2003. I found the cited half-life of computing literature increased from 7.1 years in 2003 to 7.4 years in 2007 (4.7%), the fifth highest increase among the 22 disciplines. Geosciences was tops, with an increase from 8.2 years to 8.8 years (7.7%). The disciplines with the most decreasing cited half-life were Environment/Ecology and Engineering, with declines of 2.4% and 1.8%, respectively. The average increase among all disciplines was 0.1 year (1.9%). Hence, there seems to be a trend that the age of useful computing literature is increasing, not decreasing relative to other disciplines.

The increasing interest in research related to environment and ecology may have contributed to less old work being cited in more recent issues of the related journals. Moreover, if my study is replicated in, say, five years, we may observe different trends; for example, the financial crisis at the time of this writing (2009) may contribute to more

Table 1. Half-lives and Price Index for all scientific disciplines.

Discipline	Mean Half-Life		Cited Half-Life		Citing Half-Life		Price Index	
	Years	*Rank*	Years	*Rank*	Years	*Rank*	%	*Rank*
Immunology	5.9	*1*	5.8	*1*	6.1	*1*	41.9	1
Molecular Biology and Genetics	6.5	*2*	6.2	*2*	6.7	*2*	29.3	13
Space sciences	6.6	*3*	6.3	*3*	6.8	*3*	40.1	2
Pharmacology	6.6	*3*	6.3	*3*	6.9	*4*	36.2	4
Biology and Biochemistry	6.7	*5*	6.5	*7*	6.9	*4*	35.3	6
Microbiology	6.8	*6*	6.3	*3*	7.2	*7*	35.3	5
Clinical Medicine	6.8	*6*	6.7	*8*	7.0	*6*	36.5	3
Chemistry	7.0	*8*	6.4	*6*	7.6	*9*	33.7	7
Neuroscience & Behavior	7.3	*9*	6.9	*11*	7.6	*9*	33.3	9
Physics	7.3	*9*	6.8	*9*	7.7	*11*	33.6	8
Multidisciplinary	7.3	*9*	6.8	*9*	7.8	*12*	33.1	10
Computer Science	**7.5**	***12***	**7.4**	***14***	**7.5**	***8***	**31.7**	***11***
Engineering	7.7	*13*	7.2	*13*	8.3	*13*	29.7	12
Environment/Ecology	7.9	*14*	7.4	*14*	8.4	*15*	26.5	17
Materials Science	7.9	*14*	7.1	*12*	8.8	*17*	28.7	14
Agricultural Sciences	8.1	*16*	7.5	*16*	8.7	*16*	25.5	19
Social Sciences, General	8.2	*17*	8.1	*17*	8.3	*13*	27.9	15
Plant & Animal Sciences	8.8	*18*	8.4	*18*	9.2	*18*	26.4	18
Psychiatry/Psychology	9.1	*19*	9.0	*20*	9.2	*18*	25.0	20
Geosciences	9.2	*20*	8.8	*19*	9.5	*21*	27.1	16
Economics and Business	9.6	*21*	9.9	*22*	9.2	*18*	24.8	21
Mathematics	9.7	*22*	9.5	*21*	9.8	*22*	23.9	22
Mean	**7.7**		**7.3**		**8.0**		**31.2**	

Table 2. Half-lives and Price Index for computing.

Subdiscipline	Mean Half-Life		Cited Half-Life		Citing Half-Life		Price Index	
	Years	*Rank*	Years	*Rank*	Years	*Rank*	%	*Rank*
Information Systems	6.8	*1*	6.7	*2*	6.8	*1*	35.1	1
Interdisciplinary Applications	7.0	*2*	6.2	*1*	7.8	*4*	31.1	4
Artificial Intelligence	7.6	*3*	7.2	*3*	8.0	*5*	28.2	6
Software Engineering	7.7	*4*	8.1	*5*	7.2	*3*	33.8	3
Hardware & Architecture	7.8	*5*	8.8	*7*	6.8	*2*	35.1	1
Cybernetics	7.9	*6*	7.2	*3*	8.5	*7*	26.1	7
Theory & Methods	8.3	7	8.6	6	8.0	5	29.7	5

research being done in economics and business, with more recent work being cited, or shorter half-lives.

Journals vs. conferences. I found an average of 5.9 years for the citing half-life of the 307 conference and workshop proceedings available in the ACM Digital Library. Their citing half-lives are shorter than for computing journals (7.5 years). The two main explanations for why conferences have shorter half-lives are shorter publication delay and fewer references per article.

Publication delay means the cited references grow older due to the publication process per se; that is, the references were younger when the article was submitted than when the article was published. A list of publication delays for computing journals, conferences, and other venues shows a clear tendency for journals to have longer delays than conferences (http://citeseer.ist.psu.edu/pubdelay.html). The average publication delay of journals common to the CiteSeer list and JCR was 20 months. The average publication delay of the conferences common to the CiteSeer list and the ACM Digital Library was eight months. About one-third of the JCR journals and one-quarter of the ACM Digital Library conferences were included. It is unlikely these samples were biased with respect to publication delay. Hence, we can infer that the average difference in publication delay between computing journals and conferences is approximately one year, even though the increasing use of Web-based support tools in the review process of many journals may have contributed to slightly shorter publication delays today than when the list was assembled in 2003.

The 11,719 articles in the ACM conferences (as of 2008) include, on average, 16.1 references, while the 36,004 articles in the JCR computing journals include, on average, 27.1 (26.2 if review articles are excluded); that is, journals include 70% more references than conferences. Journal articles are also generally longer than conference articles; thus, more space is available for related work. Consequently, the citing half-lives of journals may be higher than the citing half-lives of conference proceedings due in part to journals citing more references.

One should take care criticizing or ignoring literature just because it is "old"; other criteria must be used to judge quality and relevance.

When calculating the half-lives of the conference proceedings, I excluded references to URLs because their year of "publication" was rarely indicated in their citations; moreover, for those ULR references with the year indicated, it's likely that the content of the actual Web site has changed, meaning we cannot necessarily use the indicated year to calculate the age of the content of a given Web site (unlike printed publications). However, another study[16] investigated how long URLs are accessible by inspecting the URLs referenced in articles in *IEEE Computer* and *Communications* from 1995 to 1999, reporting, "A noteworthy parallel can be observed between the four years we calculated as the half-life of referenced URLs and five years given as the median citation age for computer science." One may reasonably question the extent to which one is able to compare the accessibility of URLs with the inclusion of references in articles.

Nevertheless, the claim that the half-life in CS is five years is from four issues of the *Proceedings of the International Conference on Information Systems*.[3] Due to difference between journals and conferences, it would be more correct to compare the four-year half-life of URLs with the citing half-life of 7.5 years in Table 1, as both figures result from analyzing journals. In this case, referenced articles would have a useful life approximately twice as long as the URLs. However, given that I found large variations in the citing half-lives between journals and conferences with respect to printed publications, one may find large variations in the half-lives of referenced URLs as well. Therefore, one should analyze much larger samples than only two journals to make a general statement.

Conclusion

My investigation found that the aging of the computing literature is not atypical compared with other scientific research disciplines, indicating that the research front in computing does not move more quickly than its counterpart in other disciplines. It is also a sign that computing is an established research discipline with long-lasting challenges and complex research problems taking years to solve. For example, developing software systems that

are reliable, efficient, user-friendly, and maintainable has been, and probably always will be, a grand challenge in computing research. Moreover, it typically takes 10 to 20 years for a technology to mature from being a good research idea to being widely used in practice.[14,15] This fundamental aspect of computing, combined with the importance of software in modern society, means there is no reason funding for computing research should not be at a level comparable to that found in other scientific disciplines, including physics and clinical medicine.

These results have further consequences. First, half of the citations in the computing literature are more than seven years old. Publications older than seven years may be viewed as old but still considered relevant by the authors citing them. Therefore, one should take care criticizing or ignoring literature just because it is "old"; other criteria must be used to judge quality and relevance.

The relatively long cited half-life of computing literature also indicates that the time lag between submitting a paper to a journal and it being published in that journal should not be a major concern; such work is rarely obsolete before publication. In any case, the delay may be significantly shorter in the future, as an increasing number of journals publish their articles online shortly after accepting them for publication.

My results also indicate that computing journals are not more likely to have their subscriptions cancelled or stored for a shorter time than journals of other scientific disciplines. There are significant variations, so decisions regarding particular journals must be based on more detailed information about the journals.

Here, I've discussed obsolescence at a coarse level (disciplines and subdisciplines). It would be interesting to study obsolescence within categories of computing topics and research. For example, how does obsolescence vary between research that aims to solve (minor) practical problems and research that aims to develop comprehensive theories? However, this would require substantial effort, given there is no database that easily provides relevant data similar to what JCR provided for the study I've reported here.

How the Study Was Done

I used Thomson's *JCR Science Edition* (6,417 journals in 172 categories) and *Social Sciences Edition* (1,865 journals 55 in categories) for 2007, with most journals in the (natural) sciences covered in the selection. The coverage in the social sciences was less comprehensive.[13] To comprehensively compare the overall computing discipline with other scientific disciplines, I first aggregated the JCR journal categories into 22 disciplines, per *ScienceWatch* (http://sciencewatch.com/about/met/fielddef/) and discarded eight of the 227 JCR categories because I could not fit them into the scheme of the aggregated disciplines.

JCR provided citation data at the level of both journals and categories but did not provide half-lives for new journals or journals cited fewer than 100 times. Among the 382 CS journals, 8% and 2%, respectively, lacked cited and citing half-lives. In the calculations of the aggregated results by discipline (see Table 1), I weighted the categories with respect to their number of journals.

For half-lives >10, JCR used only the value 10 in the calculation of aggregated half-lives. In the aggregation of half-life values from the JCR categories into the disciplines, I used the same approximation. A half-life ">10" was reported for individual categories in nine of the 22 disciplines; on average, 25% of the categories had the value ">10." Note that even if these nine disciplines were registered with exact values, it would not affect the CS position relative to the other disciplines in Table 1.

JCR focused on journals for citation data. However, though a study[7] reported that conference proceedings were less cited in the computing literature than books and journals, conferences play an important role in computing research. I therefore investigated the proceedings in the ACM Digital Library (from conferences and workshops in 2007) to make the data comparable with the data from the JCR 2007 edition. I included all scientific papers with at least one reference where the publication year was given; I thus excluded 2.7% of the papers on this ground. A script crawled the Web sites and extracted the references of each article in 307 proceedings. I then analyzed the output using a regular expression to identify the year of publication, enabling me to calculate the citing half-life. The 0.9% of the references lacking a clear year of publication required manual inspection.

Acknowledgments

I thank Chris Wright for help clarifying basic concepts and stimulating comments; Gilles Brassard and anonymous referees for valuable comments; and Alexander Ottesen, Birgitte Refsland, and Bjørnar Snoksrud for help collecting the reported data. C

References

1. Barnett, G.A. and Fink, E.L. Impact of the Internet and scholar age distribution on academic citation age. *Journal of the American Society for Information Science and Technology 59*, 4 (Feb. 2008), 526–534.
2. Burrell, Q. Stochastic modelling of the first-citation distribution. *Scientometrics 52*, 1 (Sept. 2001), 3–12.
3. Cunningham. S.J. and Bocock, D. Obsolescence of computing literature. *Scientometrics 34*, 2 (Oct. 1995), 255–262.
4. De Solla Price, D.J. Citation measures of hard science, soft science, technology and nonscience. In *Communication Among Scientists and Engineers*, C.E. Nelson and D.K. Pollack, Eds. D.C. Heath and Company, Lexington, MA, 1970, 3–22.
5. De Solla Price, D.J. Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science 149*, 3683 (July 1965), 510–515.
6. Glänzel, W. Towards a model for diachronous and synchronous citation analyses. *Scientometrics 60*, 3 (Dec. 2004), 511–522.
7. Goodrum, A.A., McCain, K.W., Lawrence, S., and Giles, C.L. Scholarly publishing in the Internet age: A citation analysis of computer science literature. *Information Processing and Management 37*, 5 (Sept. 2001), 661–675.
8. ISI Web of Knowledge. *Journal Citation Reports on the Web 4.2*. The Thomson Corporation, 2008; http://www.isiknowledge.com/JCR
9. Ladwig, J.P. and Sommese, A.J. Using cited half-life to adjust download statistics. *College & Research Libraries 66*, 6 (Nov. 2005), 527–542.
10. Larivière, V., Archambault, É., and Gingras, Y. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology 59*, 2 (Jan. 2008), 288–296.
11. Lazowska, E.D. and Patterson, D.A. An endless frontier postponed. *Science 308*, 5723 (May 2005), 757.
12. Misa, Y.J. Understanding 'how computing has changed the world.' *IEEE Annals of the History of Computing 29*, 4 (Oct.–Dec. 2007), 52–63.
13. Moed, H.F. *Citation Analysis in Research Evaluation*. Springer, Dordrecht, The Netherlands, 2005.
14. Osterweil, L.J., Ghezzi, C., Kramer, J., and Wolf, A.L. Determining the impact of software engineering research on practice. *IEEE Computer 41*, 3 (Mar. 2008), 39–49.
15. Redwine Jr., S.T. and Riddle, W.E. Software technology maturation. In *Proceedings of the Eighth International Conference on Software Engineering* (London, Aug. 28–30). IEEE Computer Society Press, Los Alamitos, CA, 1985, 189–200.
16. Spinellis, D. The decay and failures of Web references. *Commun. ACM 46*, 1 (Jan. 2003), 71–77.
17. Stinson, R. and Lancaster, F.W. Synchronous versus diachronous methods in the measurement of obsolescence by citation studies. *Journal of Information Science 13*, 2 (Apr. 1987), 65–74.
18. Száva-Kováts, E. Unfounded attribution of the 'half-life' index-number of literature obsolescence to Burton and Kebler: A literature science study. *Journal of the American Society for Information Science and Technology 53*, 13 (Nov. 2002), 1098–1105.
19. Weingarten, F. Government funding and computing research priorities. *ACM Computing Surveys 27*, 1 (Mar. 1995), 49–54.

Dag I.K. Sjøberg (dagsj@ifi.uio.no) is a professor of software engineering in the Department of Informatics at the University of Oslo, Norway.

© 2010 ACM 0001-0782/10/0900 $10.00