



Carlo Ghezzi

DEIB-Politecnico di Milano

Being a researcher

spreading research results: publication and beyond

Outline

- Research as an open global process
- Different kinds of diffusion
- External validation via peer review
- Publication
 - Goals and process
 - Different kinds of publications
 - Journal vs conferences
- Finding your way through publication
 - Understanding the culture of a research community
 - Publication strategy
- Research diffusion beyond publication

Research is an open global process

- No barriers, truly international
- Results MUST be diffused to other researchers and more generally to society
- Diffusion requires special effort

Diffusion

- Written, through scientific publications
- Oral, through presentations
- Implemented in research prototypes that are made available in some controlled fashion
- Packaged, both written and oral, through targeted dissemination efforts, which may include educational or reach-out purposes

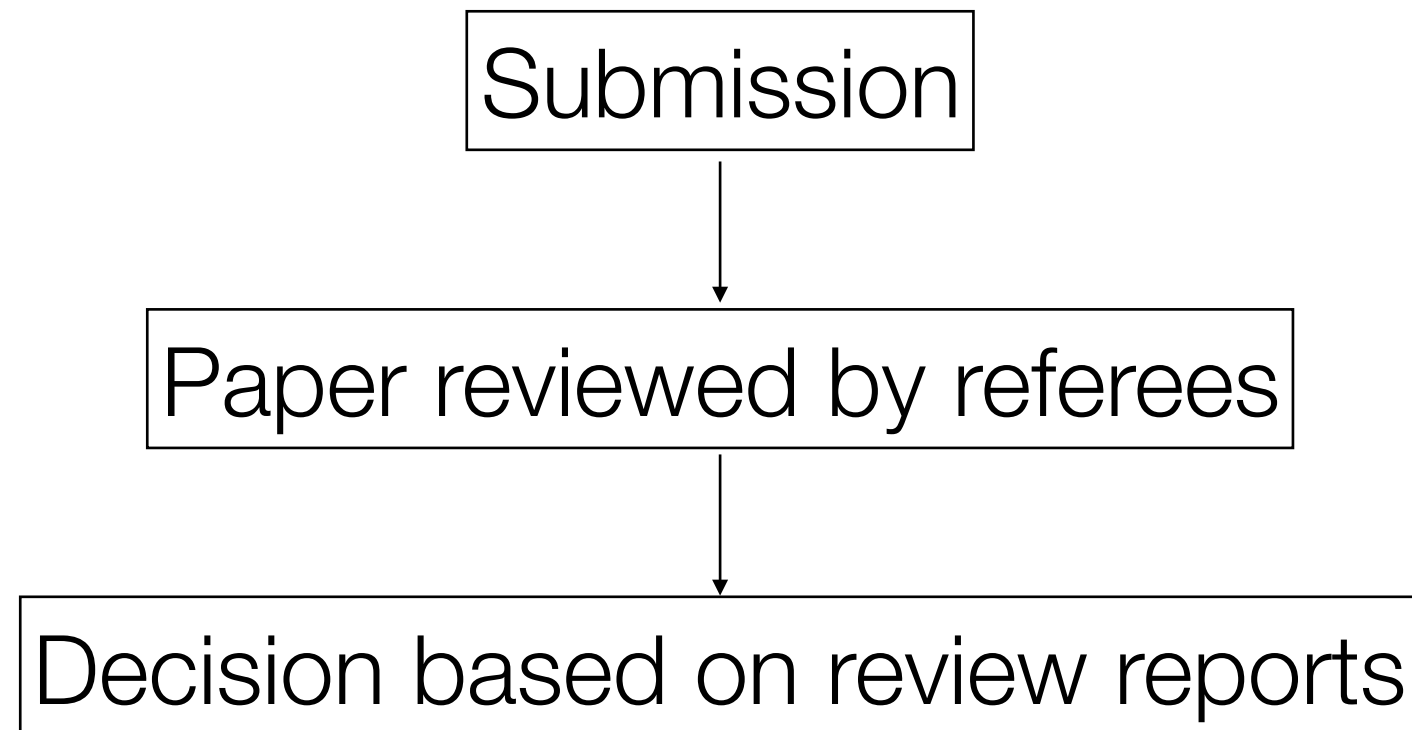
Diffusion and validation

- They go hand-in-hand
- External validation as a pre-requisite for a kind of "certified diffusion"
 - peer review for scientific papers
- Certification aims at assessing validity of the work and the way it is presented

Peer review — a preview

- Peer review is fundamental to advance science: it both certifies (filters) and helps improve quality
- Peers are recognized experts who voluntarily accept to provide reviews
- Reviews are used to decide whether a paper reaches certain quality standards and should be accepted for publication
- Reviews may suggest that a paper should be revised (and perhaps re-reviewed) prior to publication
- They may also recommend rejection

Peer-review-based publication process



variants exist, also depending on conference vs journal,
different conference processes (rebuttal, different committees, blindness)
talk more later about evaluation

Publication

- Main kind of diffusion
- Typically, a scientific paper
- But also (less common in ICST) a book
 - often resulting from a systematic unification of a series of papers
- A successful research must lead to papers, but first you must do the research!
- Your goal is to do good research, not to publish papers
- Papers follow from good research

Kinds of papers

- Working papers
- Workshop papers
- Conference papers
- Journal papers

Working papers

- Written in intermediate stage, primarily to get feedback and establish ownership
- May specify the draft status, prevent further distribution, citation
- May be put on archive like arXiv (repository of electronic preprints —e-prints— approved for publication after moderation)
 - *see later discussion on Open Access*

Workshop papers

- Workshops are specialized events for specific research subfields, often non permanent and focusing on hot topics
- They try to attract a cohesive audience and aim at interaction among participants
- Submitted papers may be formally reviewed and may be published in workshop proceedings, thus becoming formal publications

Conference papers

- Tradition about conferences varies according to areas: in some areas conference papers have little or no value
- Conferences play a primary role in ICST
- Consolidated conferences series in most ICST areas: they are held regularly, typically annually
 - e.g., ICSE exists continuously since 1975, AAAI exists since 1980
- Some conferences cover an entire area, others are specific of subareas
 - e.g., ICSE and ISSTA

Journal papers

- In most areas, they are the main and more respected kinds of publication
- In ICST often mainly used for archival reasons
- Each area has its main journals, and subareas have more specialized journals

Journal vs conferences in CS

- Acceptance rate of top conferences (20% or less) often lower than top journals
- Top conferences can be at least as prestigious (or more prestigious) than top journals
- Accepted length for conference papers shorter (many journals have no page limit)
- Processing dates fixed and known for conferences (journals have longer and less predictable turnaround time)
 - more timely and hot material
- Conferences for timeliness and feedback, journals for comprehensive and consolidated results
- Mixed forms emerging

Emerging new forms blurring the boundary

- VLDB has integrated the journal continuous submission and reviewing style with the conference. Papers reviewed within 2 months and published in the Proceedings of the VLDB. For presentation at the conference in any year, a paper must be accepted by a specified date in May of that year.
- ICSE has begun to form partnerships with prestigious journals to incorporate journal-first papers into the ICSE program. Through this initiative, authors of journal-first papers accepted in the partnering journals will be invited to present their work at ICSE, allowing the authors to speak directly to the community and offering the ICSE attendees a richer set of presentations.

Understanding the publications world

Essentials about the publication world

- Publications have identifiers helping in finding information on an article or publication using a coded information
 - much alike SSN in USA or Tax Identification Numbers (TIN) in Europe (codice fiscale in Italy)
- International Standard Book Number (ISBN) is for books (13 digits)
- Digital Object Identifier (DOI) is used mainly in scientific journals. Each article in each of the thousands of journals has its own unique DOI
- Similar concept for researchers: ORCID provides a persistent digital identifier that distinguishes you from every other researcher

Professional societies

- International (and national) professional societies exist, which mainly group researchers working in various areas
- Main international ICST societies are ACM and IEEE
 - ACM: <https://www.acm.org/about-acm>
 - IEEE: <https://www.ieee.org/>
- Societies support the diffusion of research through journals and conferences (the sponsor them)
- They also help make the professions visible to society, e.g., through their prestigious awards
- Scientists run the scientific side of societies

Journals

- May be published by professional societies
 - in ICST, mainly ACM, IEEE, but also national societies, like BCS
- May be published by commercial publishers
 - like Springer, Wiley, Elsevier
- Some universities have their own publishing organization

Side remark about the traditional business strategy

- Research largely (mainly) funded by public money (including grants and researcher salaries)
- Researchers are not paid for diffusing (publishing) research results: their return is in terms of knowledge, recognition, and prestige
- Copyright transferred to publisher
- Researchers sustain the process by providing voluntary unpaid support to the reviewing and editorial processes
- Publisher's costs: advertising, editorial process, printing and distribution, electronic archiving
- Research institutions/individual researchers must pay a subscription fee to access published material (an institution does that also for its own financed research)
- Journal subscription fees are substantial. They may include access to digital libraries
- Research institutions often face with cost reduction problems...

Copyright

- Traditionally, copyright is transferred to the journal publisher. Publishers claimed this was necessary to protect author's rights and coordinate permissions for reprints or other use
- Over time, authors found this unsatisfactory and used their influence to evolve towards a license to publish (edit, print, distribute commercially, while authors retain other rights themselves)
- Today, even if they retain the copyright to an article, most journals allow certain rights to authors

Copyright policy example: ACM

Copyright © 2016 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page in print or the first screen in digital media. **Copyrights for components of this work owned by others than ACM must be honored.** Abstracting with credit is permitted.

To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Send written requests for republication to ACM Publications, Copyright & Permissions at the address above or fax +1 (212) 869-0481 or email permissions@acm.org.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Copyright policy example: ACM

Permanent Rights held by original Owners/Authors

.....

Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, (3) any repository legally mandated by an agency funding the research on which the Work is based, and (4) any non-commercial repository or aggregation that does not duplicate ACM tables of contents, i.e., whose patterns of links do not substantially duplicate an ACM-copyrighted volume or issue. Non-commercial repositories are here understood as repositories owned by non-profit organizations that do not charge a fee for accessing deposited articles and that do not sell advertising or otherwise profit from serving articles.

Access to knowledge

- An increasing number of researchers believe that traditional publishing restrict access to knowledge
 - knowledge locked behind technical, legal, financial barriers
 - at the same time, technology today supports immediate knowledge sharing through many channels and allows everybody to access and create new knowledge
 - traditional practice mainly generates profit for commercial publishers
- This generated **open access**, and also forced publishers to evolve their business roles

What makes work open access?

- Available in digital form
- Available online
- Free of charge
- Free of **most** copyright and licensing restrictions

Brief history of OA

- Until 1990's, scientific papers were published on subscription journals paid by individual researchers or research institutions. OA challenges the model, arguing that scientific knowledge is a public good which should be made freely available to anyone, anywhere
- By 1990's the web made it possible to post and share papers online. This eliminates printing costs, makes distribution instantaneous, unlimited, no borders, potentially free
- Paul Ginsparg created AsXiv for particle physicists to allow them to make their preprints freely acceptable
- Starting from 2000, an increasing range of peer reviewed OS journals, primarily in life sciences
 - well known examples are PLOS <https://www.plos.org/publications>, Frontiers <https://www.frontiersin.org/>
- Hybrid models pushed primarily by traditional publishers: Green OA, Gold OA

Two main kinds of OA

- Self-archiving, **green open access**: delivered by repositories
 - author archives the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication
- Open access publishing, **gold open access**: delivered by journals
 - an article is immediately published in open access mode. Payment of publication costs is shifted away from subscribing readers
 - most common business model is based on one-off payments by authors' side (aka Article Processing Charges—APCs)
 - in other cases, the costs of open access publishing are covered by subsidies or other funding models.

Gold vs Green: the process

- To make new articles gold OA, authors simply submit their manuscripts to OA journals, as they would to conventional journals
- To make articles green OA, authors simply deposit their manuscripts in an OA repository
 - they may be papers submitted and accepted by a traditional journal

OA and rights

- Gold
 - OA journals obtain the rights or permissions they need directly from the rights-holders
- Green
 - Repositories ask depositors to obtain the needed rights or permissions on their own

Example: ACM and Green OA

All ACM published authors retain the right to post the pre-submitted (also known as "pre-prints"), submitted, accepted, and peer-reviewed versions of their work in any and all of the following sites:

Author's Homepage

Author's Institutional Repository

Any Repository legally mandated by the agency or funder funding the research on which the work is based

Any Non-Commercial Repository or Aggregation that does not duplicate ACM tables of contents. Non-Commercial Repositories are defined as Repositories owned by non-profit organizations that do not charge a fee to access deposited articles and that do not sell advertising or otherwise profit from serving scholarly articles

For the avoidance of doubt, an example of a site ACM authors may post all versions of their work to, with the exception of the final published "Version of Record", is ArXiv. ACM does request authors, who post to ArXiv or other permitted sites, to also post the published version's Digital Object Identifier (DOI) alongside the pre-published version on these sites, so that easy access may be facilitated to the published "Version of Record" upon publication in the ACM Digital Library.

Examples of sites ACM authors may not post their work to are ResearchGate, Academia.edu, Mendeley, or Sci-Hub, as these sites are all either commercial or in some instances utilize predatory practices that violate copyright, which negatively impacts both ACM and ACM authors.

Example: ACM and Gold OA

Unlike Hybrid Open Access journals, Gold Open Access journals are completely open via the ACM Digital Library with all articles requiring either a paid Article Processing Charge or a Financial Waiver, issued by ACM and based on certain criteria defined by ACM.

Open-ness

- Is a principle that applies to other products than papers
 - Open Data
 - Open Software
- See EU directive on Open Access and Data Management (http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)

Further distinction: libre vs free

- Every kind of OA removes price barriers (**gratis**)
- Many different additional permission barriers may exist
 - **libre** indicates the case where at least some copyright restrictions are removed
 - most libre OA is gold OA, but not vice-versa
 - open licenses from Creative Commons (CC) are the best-known and most widely used (<https://creativecommons.org/>)
 - “some-rights-reserved” copyrights rather than “all-rights-reserved” copyrights
 - CC-Zero if you assign your work to the public domain
 - CC-BY allows any use, provided the user attributes the work to the original author

Back to research publications (and the publication obsession of researchers)

Papers matter (for the good and bad)

- Papers are the traditional outputs of research
- Unavoidable tendency to assess productivity in terms of delivered output
- And (worse) to consider quantity over quality of delivered output
- And (worse) to only consider papers as outputs

Publish
or
perish

From the CRA—Computing Research Association

The evaluation of computer science and engineering faculty for promotion and tenure has generally followed the dictate “publish or perish,” where “publish” has had its standard academic meaning of “publish in archival journals”. Relying on journal publications as the sole demonstration of scholarly achievement, especially counting such publications to determine whether they exceed a prescribed threshold, ignores significant evidence of accomplishment in computer science and engineering. For example, conference publication is preferred in the field, and computational artifacts — software, chips, etc. — are a tangible means of conveying ideas and insight. Obligating faculty to be evaluated by this traditional standard handicaps their careers, and indirectly harms the field.

Once more: understand the publication culture

- Different research communities have different cultures, which must be taken into account when referring to publications
- In particular, the issue conferences vs journals
 - ICST (and CS in particular) considers conferences and journals alike
 - both envisage a thorough "peer review"
 - "top" conferences as prestigious as "top" journals

Publication quality

- Publications may differ a lot in terms of quality, namely
 - degree of novelty
 - degree of relevance
 - rigor and completeness of treatment
- Your publications will not always have the same level of quality!

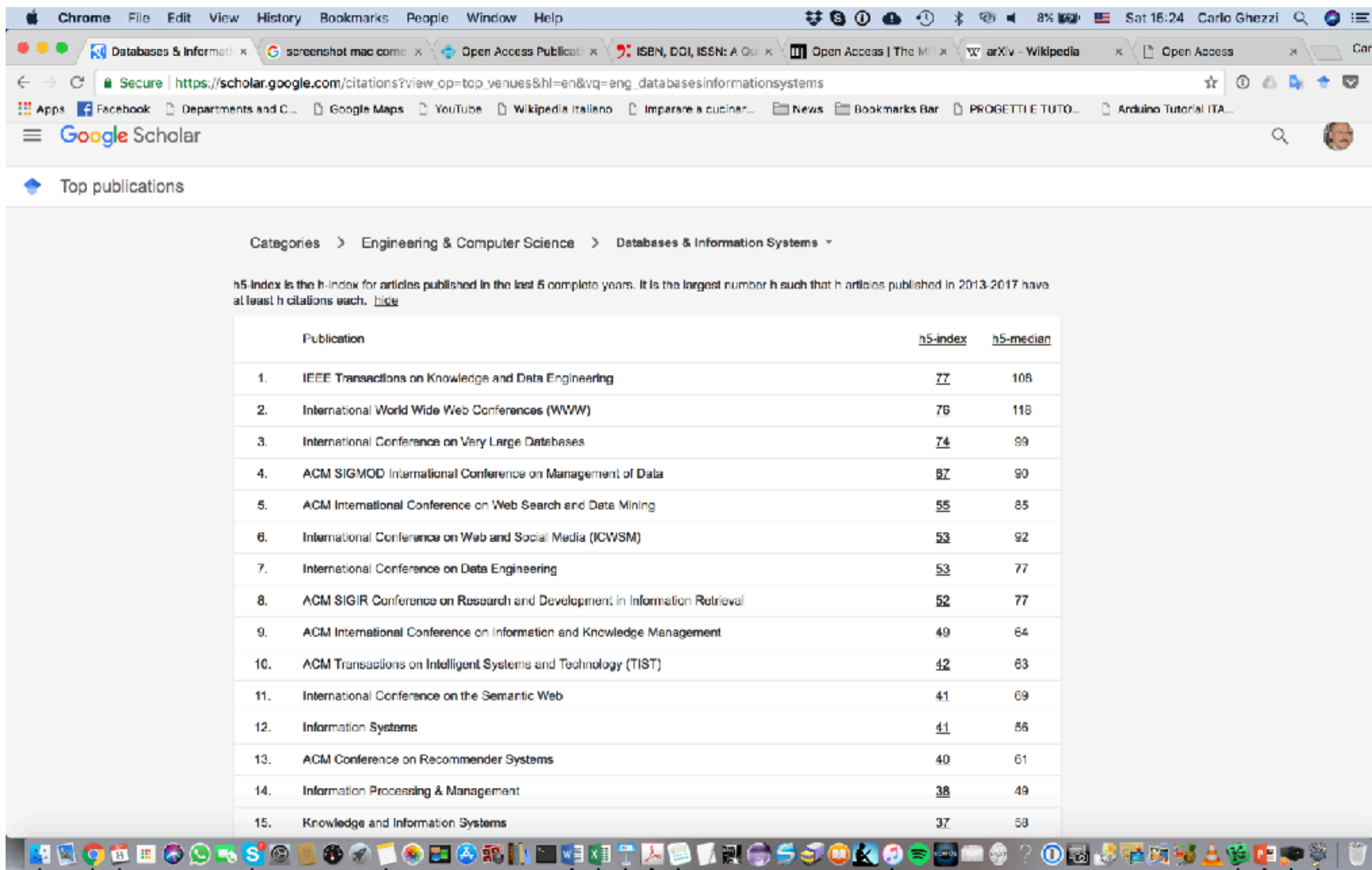
Quality of the venue

- Not all venues (conferences, journals) have the same quality (i.e., prestige, severity of the review process)
- You normally choose the venue depending on the context and on your self-assessment of the quality of your paper
 - try the top for top results

How do I find out about top venues?

- Info is part of the culture of a research area
- Ask your supervisor, top scientists in the area, you'll probably have similar answers
- There are also several ranking systems (more on rankings later)

Example from Google Scholar



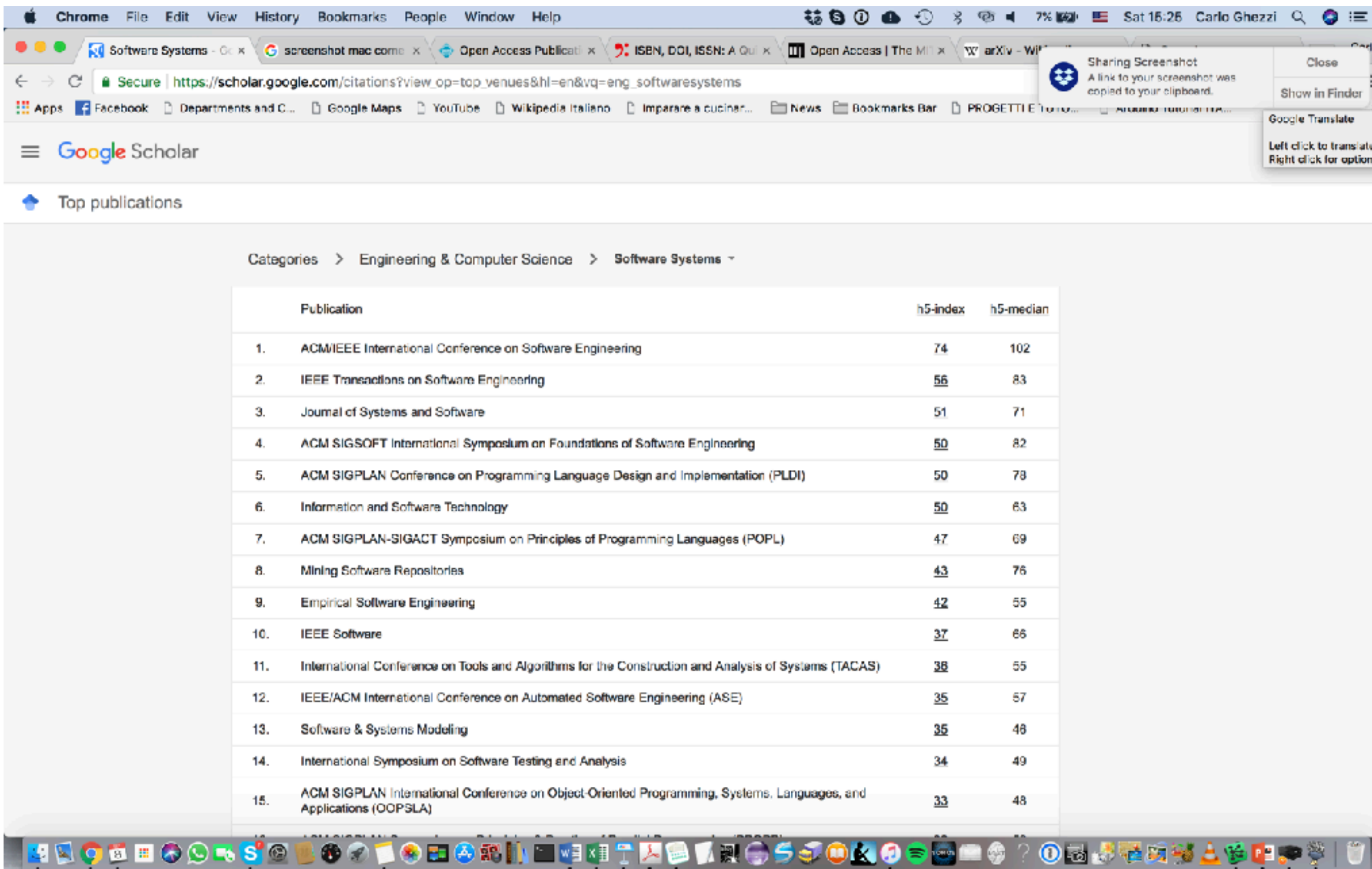
The screenshot shows a Google Scholar page titled "Top publications" for the category "Databases & Information Systems". The page lists 15 top venues with their h5-index and h5-median values. The browser's address bar shows the URL: https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_databasesinformationsystems. The browser's taskbar at the bottom shows various application icons.

Categories > Engineering & Computer Science > Databases & Information Systems

h5-index is the h-index for articles published in the last 5 complete years. It is the largest number h such that h articles published in 2013-2017 have at least h citations each. [hide](#)

	Publication	<u>h5-index</u>	<u>h5-median</u>
1.	IEEE Transactions on Knowledge and Data Engineering	<u>77</u>	108
2.	International World Wide Web Conferences (WWW)	<u>76</u>	118
3.	International Conference on Very Large Databases	<u>74</u>	99
4.	ACM SIGMOD International Conference on Management of Data	<u>67</u>	90
5.	ACM International Conference on Web Search and Data Mining	<u>55</u>	85
6.	International Conference on Web and Social Media (ICWSM)	<u>53</u>	92
7.	International Conference on Data Engineering	<u>53</u>	77
8.	ACM SIGIR Conference on Research and Development in Information Retrieval	<u>52</u>	77
9.	ACM International Conference on Information and Knowledge Management	<u>49</u>	64
10.	ACM Transactions on Intelligent Systems and Technology (TIST)	<u>42</u>	63
11.	International Conference on the Semantic Web	<u>41</u>	69
12.	Information Systems	<u>41</u>	56
13.	ACM Conference on Recommender Systems	<u>40</u>	61
14.	Information Processing & Management	<u>38</u>	49
15.	Knowledge and Information Systems	<u>37</u>	58

Example from Google Scholar



The screenshot shows a Google Scholar page for the category 'Software Systems'. The browser is Chrome, and the URL is https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems. The page displays a list of 15 top publications with their h5-index and h5-median values. A 'Sharing Screenshot' notification is visible in the top right corner, and a 'Google Translate' bar is at the bottom.

Categories > Engineering & Computer Science > Software Systems

	Publication	h5-index	h5-median
1.	ACM/IEEE International Conference on Software Engineering	74	102
2.	IEEE Transactions on Software Engineering	56	83
3.	Journal of Systems and Software	51	71
4.	ACM SIGSOFT International Symposium on Foundations of Software Engineering	50	82
5.	ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)	50	78
6.	Information and Software Technology	50	63
7.	ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)	47	69
8.	Mining Software Repositories	43	76
9.	Empirical Software Engineering	42	55
10.	IEEE Software	37	66
11.	International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)	36	55
12.	IEEE/ACM International Conference on Automated Software Engineering (ASE)	35	57
13.	Software & Systems Modeling	35	48
14.	International Symposium on Software Testing and Analysis	34	49
15.	ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA)	33	48

Possible strategy: combine connect and publish

- In progress, preliminary results may be submitted to a workshop (often associated with a main conference)
 - you may also produce a preliminary draft that you send out for comments/feedback to inner circles
- Significant results may be submitted to a conference
- Mature collection of results may be submitted to a journal
- Feasibility of the strategy also depends on available funding

Other kinds of research outputs: artifacts

- Several research areas (including ICST) develop artifacts of various kinds as part of the research
 - artifacts may be prototype implementations acting as proof-of-concept
- It is vital, however, that others may pick them up and use, possibly keeping them alive over time.
Building on top of others' work is vital to advance research
- personal study: TOSEM papers 2001-2006, 60% refer to a tool, only 20% passed installation test!

Bottom line: repeatability

- Reproducibility
 - which the act of creating a fresh system from first principles to duplicate an existing result under different experimental conditions.
 - may be an expensive undertaking
- Repeatability
 - the act of checking the claims made in the paper, usually, but not only, by re-running a bundled software artifact

S. Krishnamurthi, J. Vitek The Real Software Crisis: Repeatability as a Core Value, CACM March 2015

The case for software artifacts

- Researchers cannot be expected to develop industrial-quality software, difference between research prototypes and production software
- Software artifact should not be measured against some absolute notion of quality, but rather how the artifact stacks up against the expectations set by the paper
- Also, not all papers need artifacts: they may contain valuable theoretical results or profound observations that do not lend themselves to artifacts

Experience

- Several ACM SIGPLAN conferences (OOPSLA, PLDI, and POPL) and closely related conferences (SAS, ECOOP, and ESEC/FSE) have begun experiments to run artifact evaluation processes
- personal experience in 2011, the ESEC/FSE conference had 14 artifact submissions (for 34 accepted papers) and 7 of those met or exceeded expectations
 - artifact evaluation only voluntary, and for accepted papers

Should artifacts be published?

- Many good reasons for making them publicly available, possibly through a maintained public repository
- Some arguments against: *opinions?*
 - The artifact may have been produced in a company and may therefore be regarded as proprietary
 - The data used in the paper's experiments may be proprietary or have high privacy needs
 - The artifact may depend on expensive or proprietary platforms that are difficult or impossible for anyone but authors to access
 - By making the tools public, it becomes easy for others to continue that line of research, which reduces the payoff for the original researcher

Summary

- We discussed the different kinds of research products
- We discussed how products may be diffused
- In particular, we focused on publication process
- We have examined how other artifacts can be evaluated and diffused