Bibliometrics

A new service of the university library to support academic research

Carolin Ahnert and
Martin Bauschmann

What is bibliometrics? – short definition

„Bibliometrics is the statistical analysis of bibliographic data, commonly focusing on citation analysis of research outputs and publications, i.e. how many times research outputs and publications are being cited" (University of Leeds, 2014)

➡ Quantitative analysis and visualisation of scientific research output based on publication and citation data

What is bibliometrics? – a general classification

	descriptive bibliometrics	evaluative bibliometrics
cognition interests	• Identification of relevant research topics or thematic trends • Identification of key actors • Exploration of cooperation patterns and communication structures	• Evaluation of research performance (groups of researchers, institutes, universities, countries) • Assessment of publication venues (especially journals)
examined constructs	• Interdisciplinarity • Internationality • Topical cluster • Research fronts/ knowledge bases	• Productivity → visibility → impact → quality?
methods/ means	• Social network analysis: co-author, co-citation, co-word-networks etc. • Science Mapping	• Indicators: number of articles, citation rate, h-Index, Impact Factor, Altmetrics • Score-Cards, Rankings

Own visualisation

Why evaluative bibliometrics?

"Every enterprise and almost every organisation or corporation is confronted with the task to monitor and evaluate the performance [...] of its teams, or of the whole unit."
(Wagner-Döbler, 2003)

➡ With an everyday growing research output, evaluative bibliometrics is a useful addition to qualitative evaluation such as peer-review

Classification of bibliometric indicators for evaluative purposes

Productivity indicators

Measures for the publication output of a researcher or a group of researchers within a certain time frame

- Number of publications
- Average number of publications per researcher
- Normalised number of publications (document type, number of co-authors, etc.)

Impact indicators

Measure for the impact of publications of a researcher, a group of researchers or a journal within a certain time frame

- Number of citations
- Citation rate: average number of citations per publication
- h-index/variations (productivity partly included)
- Journal impact factor(s)
- Normalised indicators (according to subject, number of co-autors etc.)

Impact indicators

h-index – determination

If publication number and related citation numbers are known, it is easy to determine the h-index manually. Publications are sorted descending according to the number of citations – then the last publication where the publication rank is not higher than the number of citations determines the h-index

Publication 1	24 citations
Publication 2	20 citations
Publication 3	15 citations
Publication 4	10 citations
Publication 5	8 citations
Publication 6	6 citations
Publication 7	5 citations
Publication 8	3 citations

h = 6
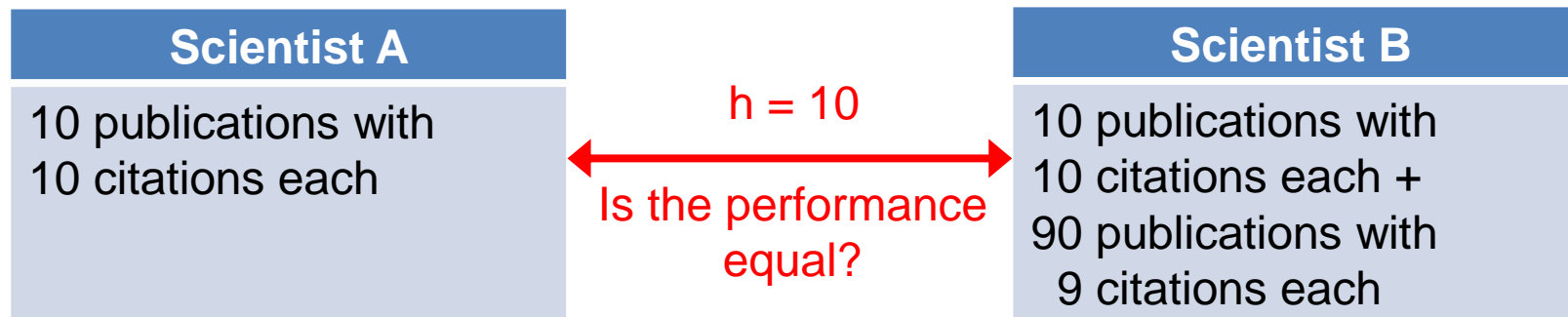
Own depiction

Response indicators

h-index – pros

- combines the measure of visibility (number of articles) and the impact (citations) of research output

- productivity is rewarded, but too many publications aren't

- a few much-cited publications have no influence on the h-index if the number of articles stays the same, only a broad peak is being rewarded

A growing number of studies show, that there is a correlation between h-index and other forms of evaluation, such as the acceptance of applications for research scholarships (Bornmann & Daniel, 2005), the granting of external funds (Lovegrove & Johnson, 2008) and the evaluation of research groups by peers (van Raan, 2006).
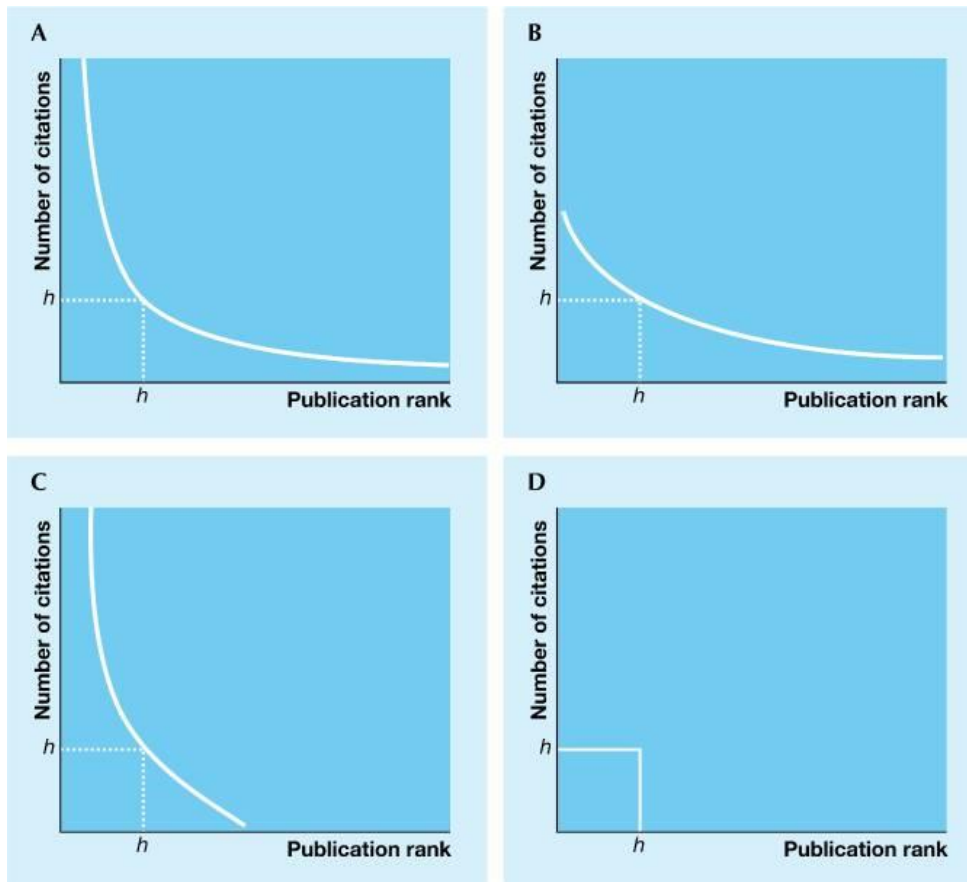
Response indicators

H-Index – cons

- number of publications and citations are equated arbitrarily which makes interpretation difficult

- variations in perception of scientific research or productivity (less citations or no publications in years) are not influencing the h-index

- age of scientists and therefore length of an academic career have a huge impact on h-index

- source determines the scale (Scopus, Web of Science, Google Scholar)

- it is not clear how the citations are distributed among the publications just by examining the h-index:

Scientist A	Scientist B
10 publications with 10 citations each	10 publications with 10 citations each + 90 publications with 9 citations each

h = 10

Is the performance equal?

Response indicators

H-Index – cons 2

Darstellung aus Bornmann & Daniel (2014): The state of h index research. Is the h index the ideal way to measure research performance?

A – a few highly cited works which are not represented and do not change h-index

scientist with many publications but few citations (B) and few publications but highly cited (C) = same h-index

Only if each publication (h) would get the same amount of citations (h) then the h-index would depict the reality (D). Such constant performances are very unlikely, if not impossible to find.

Journal Impact Factor

Definition

- developed by Eugene Garfield at the Institut for Scientific Information (ISI)
- indicator from Journal Citation Report (Web of Science) which depicts the average citation frequency of articles in a certain journal
- journal set: Science Citation Index and Social Science Citation Index
- eliminates problems which we encounter when dealing with absolute numbers of citations (gaps between articles)
- tool to determine the relevance and popularity of a journal
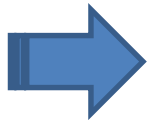
➡ JIF of a journal = average amount of citations in one year (i.e. 2017) to publications in this journal from two years (i.e. 2015 and 2016)

Journal Impact Factor

calculation

JIF = A/B ;

Example: JIF 2015 for IEEE
Transactions
on Industrial Electronics

citations from 2015 to articles from
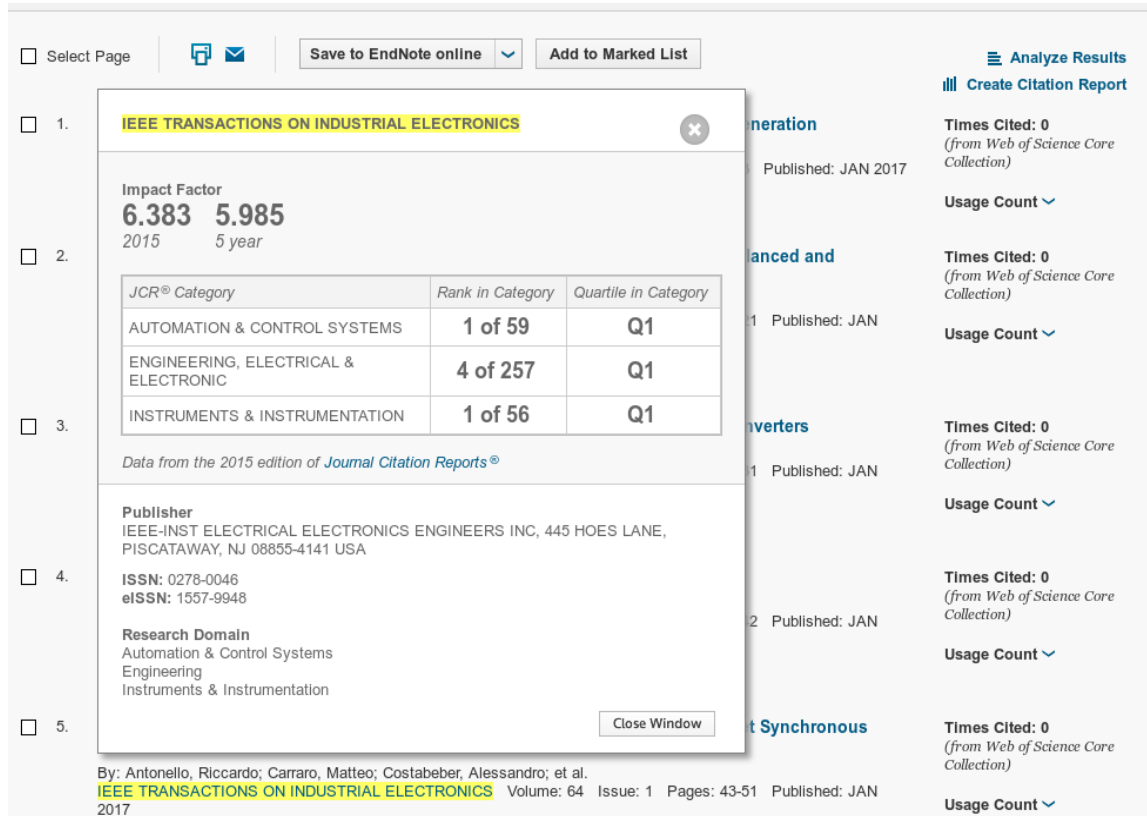2014 and 2013
→ **3.864 + 4.095 = 7.959**
divided by
amount of articles 2014 and 2013
→ **553 + 694 = 1.247**

7.959 / 1.247 = 6,383

Thus: an average article from
IEEE Transactions on Industrial Electronics
is cited about six times in within the first two years after publication

Source: Web of Science
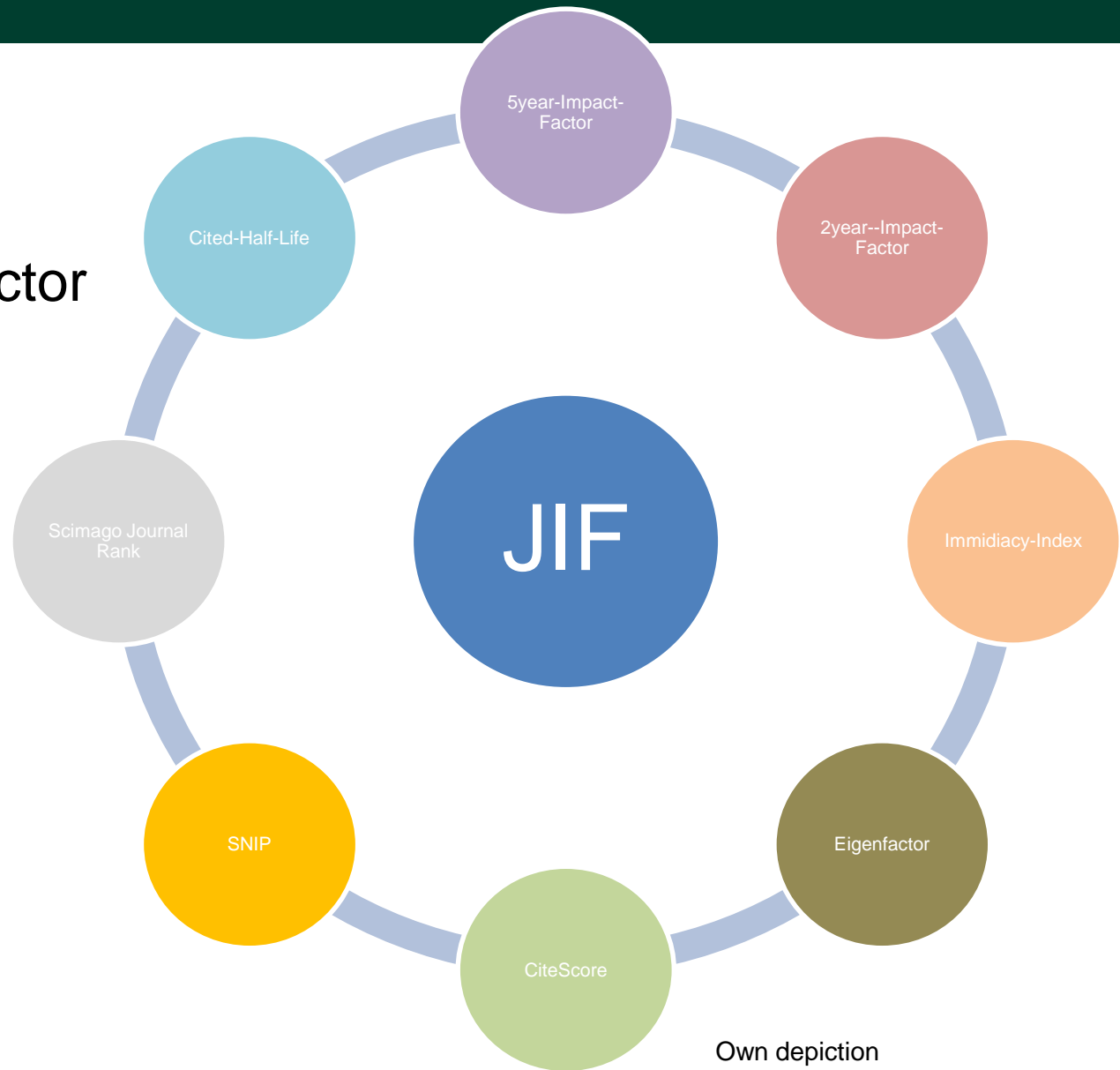
Journal Impact Factor

criticism

- highly depending on the subject area (journals in general fields do have higher impact factors than in more specific subjects)

- depending on type of publication (journal with short communication = higher impact factor bc published faster, read faster, cited faster)

- fluctuation each year, especially with less lengthy journals, up to 40%

Do not confuse with Article Impact Factor! Citations are not distributed equally (few papers highly cited, many papers rarely or not cited at all); average number of citations (JIF) is unrepresentative of the majority of articles in a journal

Journal Impact Factor

Variations

JIF

5year-Impact-Factor

2year--Impact-Factor

Cited-Half-Life

Immidiacy-Index

Scimago Journal Rank

Eigenfactor

SNIP

CiteScore

Own depiction

Criticism of evaluative bibliometrics

Methodological problems
- quality is a multi-dimensional construction and can not be measured with a single indicator
- do impact (citations) and relevance/value (assessment by peers) really correlate?
- distortions: subject dependant publication and citation culture, co-authors and honorary authorship, „Matthäus-Effect" (success breeds success) – new articles from well-known authors are cited more often, citation agreements, papers not being cited at all,…
- error and gaps in data

Dangers
- evaluation on the basis of single indicators or indicators that are misapplied: i.e. Impact Factor used to evaluate single papers or scientists
- scientists more and more comply with evaluation criteria → decreasing intrinsic motivation
- research aside of mainstream is not valued
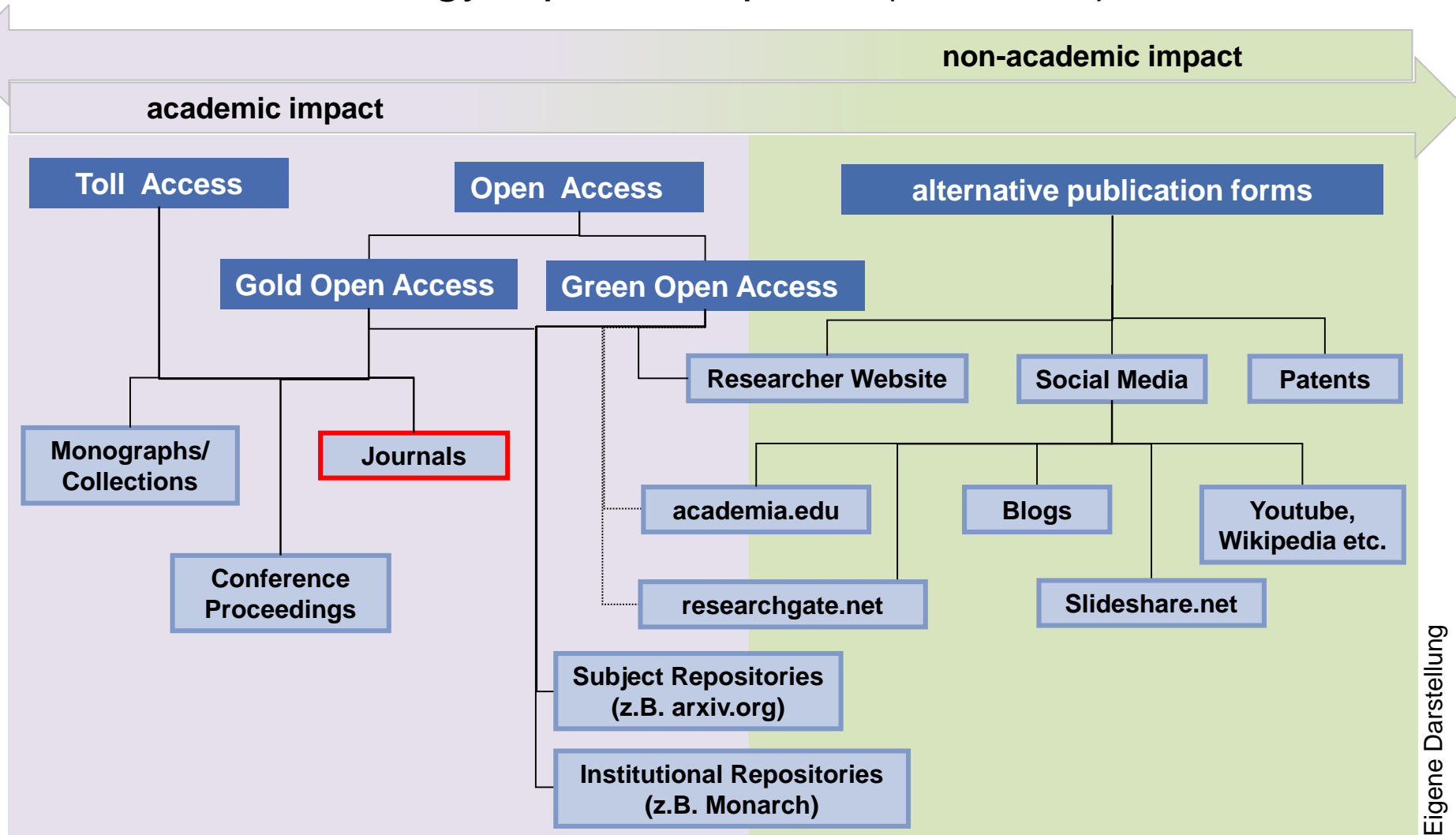- displacement of motivations, performance in research more important than teaching efforts

Why is it worth for scientists to care about publication and citation metrics?

- indicator, if only one among many others, for a researcher's contribution to the scientific discourse

- the increasing relevance of bibliometrics in research evaluation, performance oriented allocation of funds, tenure decisions as well as third-party funding is a reality

- understand who publishes what with whom and who cites whom in your subject areas:
    - identify potential collaboration partners
    - gain inspiration and new perspectives for the advancement of your own research focus and spectrum

- bibliometrics as a supplement to your repertoire of methods revealing clusters and catalysts of co-publication and co-citation within or between disciplines, institutions or countries; the evolution and demise of thematic trends (sleeping beauties and obsolence of publications); the knowledge transfer between science and technology etc.

How to improve visibility and impact of your research?

Publication Strategy – possible paths (selection)

non-academic impact

academic impact

Toll Access

Open Access

alternative publication forms

Gold Open Access

Green Open Access

Researcher Website

Social Media

Patents

Monographs/ Collections

Journals

academia.edu

Blogs

Youtube, Wikipedia etc.

Conference Proceedings

researchgate.net

Slideshare.net

Subject Repositories (z.B. arxiv.org)

Institutional Repositories (z.B. Monarch)

Eigene Darstellung

Publication Strategy – selection criteria for publication venues

- focus and scope (description in „About" section, published articles)

- target group („About" section)

- (inter)national orientation (language of articles)

- reputation (Editorial Board, publisher, published authors)

- bibliometrics: Journal Impact Factor (2y and 5y), *SCImago* Journal Rank (SJR), Citescore etc.

- peer review (procedure: single-blind or double-blind, number of reviewers; rejection rate)

- promptness of publication (journals: between 1 week and 6 months; monographs: 1 month with University Press of TU Chemnitz, up to 8 months with commercial publishers)

- archiving and long-term availabilty

- charges: article processing charges (open access), color or page charges, subscription prices (toll access)

- copyright (transfer of copyright, self-archiving rights)

- media type: electronic, print or hybrid

- editing services (copyediting, proofreading, translation)

Publication Strategy – recommendations

Publish open access

Cost-free access to and liberal usage of open access publications (gold or green) usually solicit a well documented citation advantage over subscription based publications.

(Sparc Europe (ed.) (2015). The Open Access Citation Advantage: List of studies until 2015. http://sparceurope.org/oaca_table/)

Consider writing reviews

Reviews are cited up to twice as much as research articles.

(Abt, H. A. & Garfield, E (2002). Is the relationship between numbers of references and paper length the same for all sciences? *Journal of the American Society for Information Science and Technology,* 53, 1106-1112)

Deposit research data

Independently of the impact factor, author country of origin und date of publication, making data to research articles publicly available is associated with a 69% increase in citations.

(Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE, 2*(3), 308. http://dx.doi.org/10.1371/journal.pone.0000308)

Find suitable data repositories here: http://www.re3data.org

Communication/ Networking – possibilities

- participate in conferences and present your work to peers

- create a personal researcher websites with up-to-date list of publications

- use web 2.0 technology (science blogs, podcast, videos) to widely distribute your research findings and connect with researcher within and outside your discipline

- deliberate selection and use of academic networks to upload papers, uncover new papers, contact collaborators, e.g.:

  - academia.edu
    - 37 m registered users, 12 m uploaded documents (source: Website)
    - massively populated by humanists and social scientists (Ortega, J. L. (2015). Disciplinary differences in the use of academic social networking sites. Online Information Review 39(4), 520-536. DOI: 10.1108/OIR-03-2015-0093)

  - researchgate.net
    - 9 m registered users, 100 m uploaded publications (source: Website)
    - predominently used by natural scientists, biologists and engineers (Ortega 2015)

  - mendeley.com
    - 4 m registered users (source: howtoweb.com)
    - balanced user community, with largest group from natural sciences (Ortega 2015)

Communication/ Networking – citation advantage of academic networks

Homepage Academia.edu (17.08.2015)
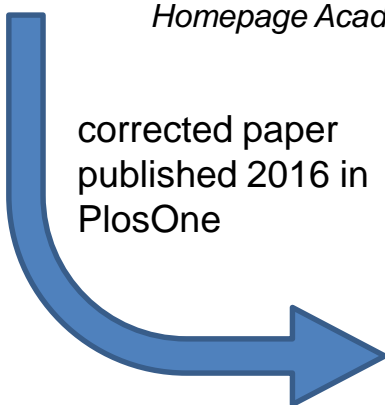
- Though critizised for methodological errors (z.B. Blog-Post on Phil Davies), a citation advantage of papers uploaded to academia.edu is supported by the study.
- This advantage can most certainly be attributed to an increased visibility among audiences beyond the usual target group (scientiest within the own discipline, region) and in the non-academic realm.

corrected paper published 2016 in PlosOne

Age	On-Academia	Online	Pred. Cites	95% Conf. Int.	95% Pred. Int.
3	N	N	4.78	(4.7, 4.86)	(0.05, 30.81)
		Y	6.06	(5.94, 6.19)	(0.28, 37.88)
	Y	N	7.24	(6.85, 7.64)	(0.5, 44.34)
		Y	8.54	(8.2, 8.9)	(0.73, 51.52)
5	N	N	7.85	(7.66, 8.03)	(0.61, 47.7)
		Y	9.82	(9.56, 10.08)	(0.96, 58.53)
	Y	N	13.31	(12.25, 14.42)	(1.6, 77.68)
		Y	15.57	(14.48, 16.72)	(2.01, 90.15)

doi:10.1371/journal.pone.0148257.t015

Niyazov, Y. et al. (2016). Open Access Meets Discoverability: Citations to Articles Posted to Academia.edu. PloS One, 11(2), e0148257. http://doi.org/10.1371/journal.pone.0148257

Cooperation – empirical findings

- Co-authorship advantage

    Co-authored articles are significantly more cited than articles by a single author.

    (Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. Science, 316(5827), 1036-1039. http://dx.doi.org/10.1126/science.1136099)

- Impact and interdisciplinary – a charged relationship

    Although the correlation between interdisciplinarity and impact is not straightforward, research shows that the variety (number) of disciplines a publication draws upon is positively correlated with citation impact, when the distance between the disciplines (subdisciplines) and the evenness of the distribution of different disciplinary influences are low.

    (Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLoS ONE*, *10*(8), e0135095. http://doi.org/10.1371/journal.pone.0135095)

- Perks of internationality

    Publications with an international co-authorship are cited up to four times as much as publications with authors originating from a single country.

    (Jones, K., & Evans, K. (2013). Good Practices for Improving Citations to your Published Work. University of BATH)

Identity Management – problem of author disambiguation

Authorship confusion has various reasons:

- homonymy: common names (e.g. in Korea three family names represent 50 % of the population: Kim, Park, Lee)

- different spelling of names in different languages

- misspelling of names

- name changes (e.g.: marriage)

- alternating mentions/ommissions of first names (initials or written out) and middle names

Identity Management – How to avoid disambiguation problems?

- If possible, use a consistent name structure throughout your career (abbreviation of first name, mention/ommission of middle name, spelling).

- Use consistent affiliation statement (university, faculty, department, research group).

- Use consistent denotations for recurring terms in your publications titles (e.g. research topics, methods).

- Create a researcher profile.

Identity Management – How to avoid disambiguation problems?

Profiling tools

ResearcherID (via Web of Science)

All logos on this slide are registered trademarks of the respective organisations and corporations

Identity Management – How to avoid disambiguation problems?

ORCID - characteristics

- well established, non-proprietory, international system for author identification
- founded in October 2012 as a not-for-profit initiative by academic institutions, professional bodies, funding agencies (e.g. CERN) and publishers (e.g. Elsevier, Nature Publishing Group, Springer)
- researchers can register (for free) on the ORCID website, where they can create a profile identified by the ORCID iD and a complete online record of their research and publications
- open data that is freely available via the web page, data feeds and an API
- authors can choose to make their profile public, private, or limited access only
- example of an ORCID iD: iD orcid.org/0000-0002-4301-0466
- solution for author name ambiguity problem:
  - creating and maintaining a registry of persistent unique researcher identifiers
  - embedding ORCID iDs in key workflows, such as manuscript subscription, grant applications and patent applications
  - enabling the integration of ORCID iDs with different identifier schemes and research systems (e.g. literature and citation databases, current research information systems)
- information and tutorials on registration and synchronisation of ORCID data: http://libguides.lb.polyu.edu.hk/orcid/ORCIDintro

Identity Management – How to avoid disambiguation problems?

ORCID – benefits

- cross-community structure, the open and international nature of the organization, data, interface and code, and the guarantee of permanence

- **researchers** avoid the hassle of entering biographical and bibliographic data in multiple systems

- links scholarly works and profiles across multiple platforms and systems (e.g. Scopus Author ID and ResearcherID)

- improves discoverability of scholarly output

- facilitates networking with other researcher

- increases overall research visibility and impact

- **Institutions and funding bodies** benefit from easier, more reliable analysis of scholarly output

- **Publishers** enjoy a simplified publication workflow

Identity Management – How to avoid disambiguation problems?

Recommended practice for managing researcher profiles

1

Clean Scopus Author ID:
1. run an Author Search
2. select the author profiles that include your work
3. click *Request to merge authors*
4. Follow the instructions

Scopus Author ID

3

Synchronise Scopus and ORCID profile:
- either from Scopus or ORCID
- detailed description: http://libguides.lb.polyu.edu.hk/orcid/orcidscopus

2

Register with ORCID

ORCID iD

5

Synchronise ResearcherID and ORCID iD:
- detailed description: http://libguides.lb.polyu.edu.hk/orcid/orcidresearcherid

7

Check for and delete duplicates

6

Import/Export with BibTex

4

Register Researcher ID and add publications

Researcher ID

Google Scholar Citations

Bibliometric Services of the University Library Chemnitz

Service portfolio and goals
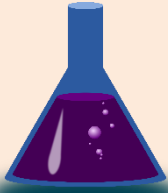
- answering specific bibliometric enquiries (z.B. „What is and how do I find my h-index?")
- detailed analyses, reports and presentations in explorative and evaluative bibliometrics: e.g. co-authorship and citation networks
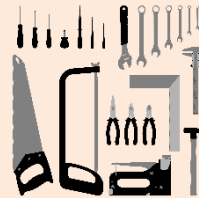
online manuals und tutorials on bibliometric software, methods and tools, for:
- data research (WoS, Scopus, Google Scholar)
- data munching and cleaning (e.g. OpenRefine, web scraping tools, Python)
- data analysis (e.g. Pajek, BibExcel)
- Data visualisation (e.g. Gephi, VosViewer, Tableau)
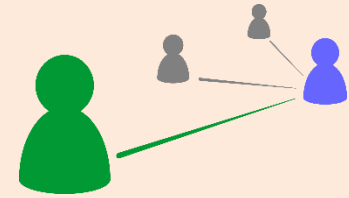
- consultation (personal or online)
- lectures and workshops
- knowledge exchange
- triggering and moderating a university-wide discussion about wisdom and folly of bibliometric practices in the academic environment (especially in evaluations)

Acquire and document up-to-date bibliometric know-how

Bibliometric Services of the University Library Chemnitz

Target Groups

Young scientists
- improve the visibility and impact of research publications
- selection of suitable venues for publication and potential partners for collaboration (co-authorship analyses)
- collection of publication data for applications for third-party funding
- bibliometric research on structures and clusters of scientific cooperation and communication as well as on thematic publication trends within the own field of research

Executives (university, faculties, institutes) und administrators of third-party funding
- performance analyses of the organisation, a structural unit or researchers
- network analyses to explore interdisciplinary and international cooperations
- detection and examination of thematical structures and trends (co-citation, co-publication and co-word networks)

Public relations and outreach
- data driven communication of research output
- interactive web-based visualisation of networks in order to incarnate interdisciplinary and international networks of the researchers

Library
- substantiate (de)aqcuisition decisions with publication, citation and usage data

Bibliometric Services of the University Library Chemnitz

Web apperance

Main contact: bibliometrie@bibliothek.tu-chemnitz.de

Contact persons

Carolin Ahnert
Phone: +49 371 531-31989
Room: Straße der Nationen 62, Böttcher Bau, 220B
E-Mail: carolin.ahnert@bibliothek.tu-chemnitz.de

Martin Bauschmann
Phone:+49 371 531-31135
Room: Straße der Nationen 62, Böttcher Bau, 137
E-Mail: martin.bauschmann@bibliothek.tu-chemnitz.de