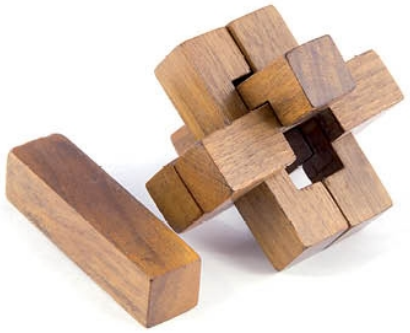


BANFF, CANADA September 7–11, 2014

The International School on Research Impact Assessment

Using Bibliometrics to Assess Research Impact

Éric Archambault
CEO at Science-Metrix

September 9, 2014

Hosted by: Alberta Innovates Health Solutions

In partnership with: Agència de Qualitat i Avaluació Sanitàries de Catalunya
AGENCY FOR HEALTH QUALITY AND ASSESSMENT OF CATALONIA

RAND EUROPE

Session overview

- What is bibliometrics?
- How to perform bibliometric analyses
- Limits of bibliometrics
- When and when not to use bibliometric approaches

Learning outcomes

- Learn about the main types of indicators
- Learn what questions to ask vendors of bibliometric expertise/information products
- Identify the purpose of using bibliometrics
- Identify when this method is appropriate to use for an assessment

Definitions

- Scientometrics – measure of science
- Technometrics – measure of technology
- Bibliometrics – measure of bibliographic records such as records on books, papers, patents
- Metadata – information contained in bibliographic records

Metadata in bibliographic records

Bibliographic information

- Counts of papers by year (trends)
- Delineation of scientific fields/subfields
- Counts of papers by researcher
- Counts of papers by institution, province, region and country
- Citations counts, i.e., number of times paper appears in references of other papers to measure scientific impact

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Biological Conservation 118 2004 583–592
BIOLOGICAL CONSERVATION
www.elsevier.com/locate/biocon

Comparison of Coleoptera assemblages from a recently burned and unburned black spruce forests of northeastern North America

Michel Saint-Germain [a,*], Pierre Drapeau [a], Christian Hébert [b]

[a] Groupe de recherche en écologie forestière interuniversitaire, Département des sciences biologiques, Université du Québec à Montréal, CP 8888, succ., Centre-ville, Montréal, Que., Canada H3C 3P8
[b] Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Center, 1055 rue du PEPS, CP 3800, Sainte-Foy, Que., Canada G1V 4C7

Received 2 April 2003; received in revised form 15 September 2003; accepted 14 October 2003

Abstract

Several insect groups have adapted to fire cycles in boreal forests, and can efficiently use new habitats created by fire. Our study aimed at producing a first characterization of post-fire Coleoptera assemblages of black spruce forests of eastern North America. For two years, we sampled Coleoptera using flight-interception traps in burned stands of contrasting age and structure in a 5097-ha wildfire and in neighbouring unburned mature stands. More than 40 species were exclusively captured in burned stands. Time elapsed since fire and proximity of unburned forests were the most significant parameters affecting Coleoptera assemblages. Stand age and structure had limited effects on assemblage structure; the Scolytid *Polygraphus rufipennis* Kirby was the only common species to clearly favor older stands. Fire-associated Coleoptera assemblages found in our study area were clearly distinct from those found in similar unburned stands; we should thus be conservative in our management approach concerning recently burned stands.

Keywords: Boreal forest; Forest fires; Habitat selection; Fire-associated Coleoptera; Salvage logging

References

Anhlund, H., Lindhe, A., 1992. Endangered wood-living insects in coniferous forests– some thoughts from studies of forest-fire sites, ourcrops and clearing in the province of Sörmland, Sweden. Enromologisk Tidskrift 113, 13–23 (in Swedish).

Bergeron, Y., Archambault, S., 1993. Decreasing frequency of forest fires in the southern boreal zone of Quebec and its relation to global warming since the end of the "Little Ice Age". The Holocene 3, 255–259.

A continuum from macro to micro

- World level – overall trend in research, overall trend in specific fields (e.g., material sciences)
- Country level – how do countries compare, how is a country doing (e.g., in biotechnology)
- Regions – states, provinces, counties
- Organisations – universities, government, health, companies
- Individuals

Choice of database

- Database should comprise a comprehensive set of high-quality relevant metadata
  - Authors, authors' address, journal name, paper title, abstract, keyword, year, issue, number, pages, references
- Database constitution must be documented and transparent
- Local access to metadata vs. web access

Choice of database *(continued)*

- In practice, very few databases present all these features
  - Web of Science by Thomson Reuters
  - Scopus by Elsevier
- Bespoke use
  - Medline
  - Inspec
  - Etc.

Field/topic delineation

- Having a notion of fields (i.e., subject areas for research) is essential
- One use is for analytical purpose – what is happening in “green energy”
- The other is normalization – how do I compare researchers in maths to those in radiology
- Several types of field and/or delineation
  - Journal-based classification
  - Article-based clustering
  - Bespoke queries
  -

Bibliometric indicators

- **Publication count** – number of publication (full or fractional) from entity in given time period
  - Beware of differences between research fields: mathematicians publish papers less frequently than researchers in biomedical research
  - Do not compare researchers/institutions from different fields of research based on raw number of publications – apple vs. orange

Bibliometric indicators

- **Specialization Index (SI)** – indicates the relative intensity of an entity in a given research field relative to the world

$$SI = \frac{\left(\frac{x_s}{x_t}\right)}{\left(\frac{N_s}{N_t}\right)}$$

$\frac{x_s}{x_t}$ → % of [local] papers in field

$\frac{N_s}{N_t}$ → % of [world] papers in field

Bibliometric indicators *(continued)*

- **Growth Ratio and Growth Index (GR, GI)**
  - **GR** compares the *output* of an entity over two time periods to assess growth in output
  - **GI** compares the *growth* of an entity (i.e., its GR) with the growth of the world
  - Percentage of growth can be obtained by regression/geometric mean

Bibliometric indicators *(continued)*

- **Growth Ratio and Growth Index (GR, GI)**
  - Growth Ratio more convenient with noisy data
  - Use GI for more robust comparisons between research fields

$$GR = \frac{\chi_b}{\chi_a}$$ → Ratio (or percentage) change in output between recent period (b) compared to prior period (a)

$$GI = \frac{GR_E}{GR_W}$$ → Ratio (or percentage) change in output between recent period (b) compared to prior period (a)

Bibliometric indicators *(continued)*

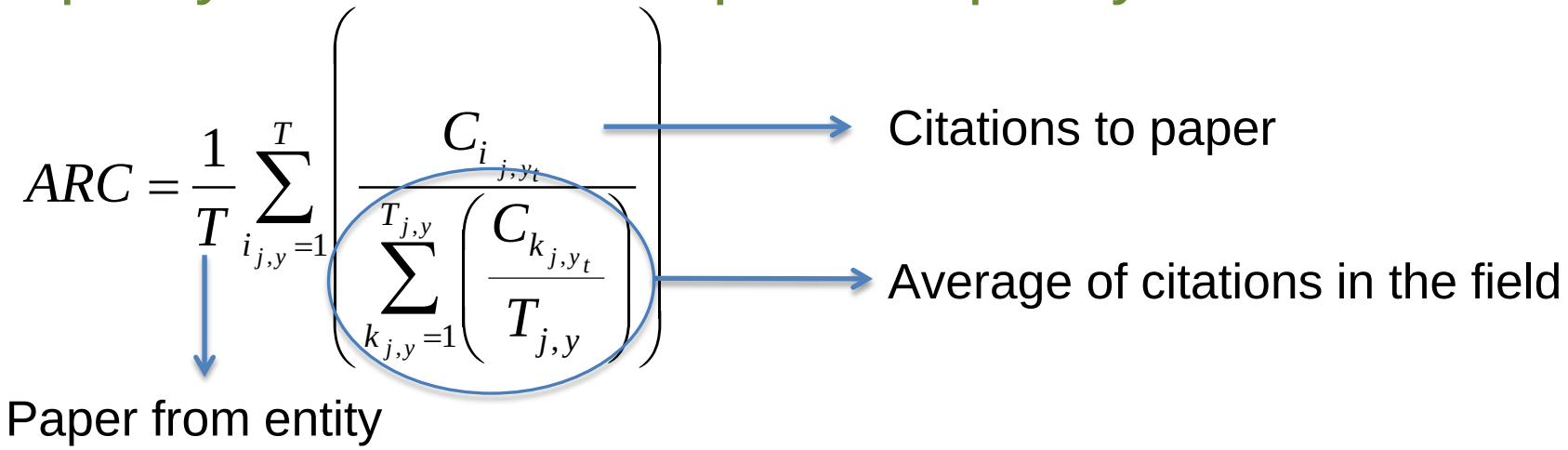
- **Citations/Average citations** – number /average number of times outputs are cited
  - Beware of differences between research fields: mathematicians publish & cite less frequently than researchers in biomedical research, thus reducing the size of the citation network
  - Do not compare researchers/institutions from different fields of research based on raw number of citations – apple vs. orange

Bibliometric indicators *(continued)*

- **Average of relative citations (ARC)** – scientific impact based on average number of citations of an entity relative to average number of citations received by world papers, normalized by field to account for differences across fields
  - Prefer this measure to number/average number of citations as it reflects differences between research fields

Bibliometric indicators *(continued)*

- **Average of relative citations (ARC)** *(cont'd)*
  - Allows comparisons across fields of research
  - Commonly used in bibliometric analyses as a proxy for research impact or quality



Bibliometric indicators *(continued)*

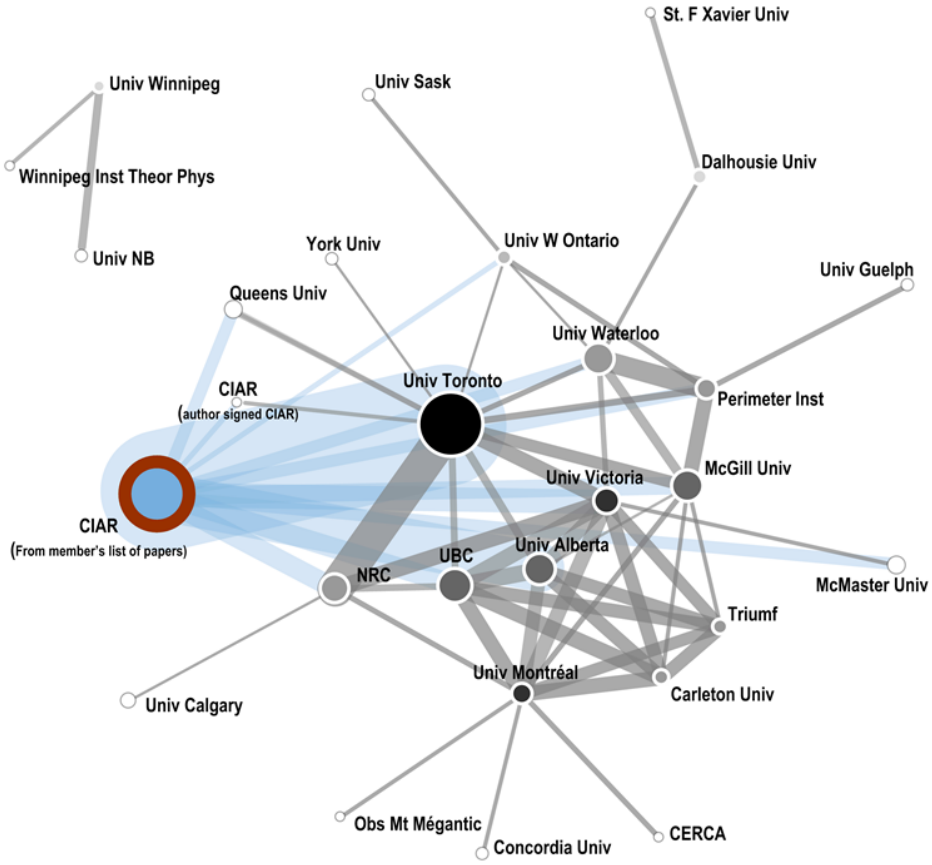
- **Impact factor/average of impact factor** – measure of scientific journal citedness based on Thomson Reuters Journal Citation Report
  - Beware of differences between research fields: mathematicians publish & cite less frequently than researchers in biomedical research, thus altering the value of impact factors between fields
  - Do not compare researchers/institutions from different fields of research based on IF

Bibliometric indicators *(continued)*

- **Average of relative impact factors (ARIF)** – modified impact factor that corrects for a number of deficiencies in the IF, and normalized for difference between fields
  - Prefer this measure to number/average number of IF as it reflects differences between research fields
  - Requires access to whole database to compute

Bibliometric indicators *(continued)*

- **Collaboration** – number/percentage of co-authored papers between entities (e.g., country, universities)
  - Provides basic data for **Social Network Analysis**



Bibliometric indicators *(continued)*

- Data should be presented in a user-friendly manner

Scientific Performance of Leading Countries in Subatomic Sciences, (1996-2010)

Number of papers, specialization index (SI), average of relative citations (ARC), average of relative impact factors (ARIF)

Country	Papers		SI	ARC	ARIF	Composite Indicator
	Score	Trend*				
United States	137,847		1.08	1.40	1.15	57.1
Germany	59,155		1.88	1.34	1.14	40.9
Switzerland	21,746		3.13	1.50	1.18	37.1
United Kingdom	36,521		1.09	1.50	1.18	36.2
Italy	37,585		2.22	1.18	1.08	35.0
France	34,833		1.50	1.32	1.12	34.5
Russia	34,075		3.17	0.99	0.93	33.0
Netherlands	11,417		1.18	1.67	1.24	32.9
Canada	**16,752**		**0.95**	**1.55**	**1.23**	**32.6**
Japan	47,409		1.40	0.96	0.96	32.4
Israel	6,063		1.45	1.51	1.26	31.3
Spain	16,619		1.27	1.41	1.15	31.1
Finland	5,310		1.54	1.49	1.20	30.5
Sweden	9,206		1.35	1.46	1.16	30.3
Denmark	5,050		1.37	1.46	1.19	29.7
Belgium	8,869		1.68	1.30	1.12	29.1
Austria	5,305		1.44	1.35	1.18	29.0
Poland	12,336		1.98	1.22	1.04	28.8
Greece	4,848		1.55	1.29	1.18	28.6
Rep. of Korea	11,030		1.12	1.24	1.12	28.0
Czech Republic	4,772		1.65	1.24	1.01	26.4
Australia	8,052		0.69	1.24	1.11	26.3
China	31,854		0.62	0.86	0.91	26.0
India	15,131		1.28	1.01	1.00	25.9
Brazil	10,097		1.30	0.96	1.04	25.0
World	**428,848**		**1.00**	**1.00**	**1.00**	-

Source: Computed by Science-Metrix using the Scopus Database (Elsevier).

Note: *The scale is not the same across countries.

Limits of bibliometrics

- All indicators are proxies
  - They represent a construct that reflects reality—not reality itself
  - Never lose sight that knowledge production and diffusion, which are presented in a highly reduced form in bibliometric measurement, are complex phenomena
  - Publications not indexed in databases are not included in analyses (e.g., books, certain types of papers, briefing papers, white papers)

Limits of bibliometrics *(continued)*

- Although best practices exist, there is also:
  - Lack of standardisation of indicators' names
    - Terms and concepts are also used rather loosely (production and productivity, quality, performance, demand)
  - Lack of standardisation of methods
  - Variations between databases produces different results
  - Results are indicative rather than definitive

Limits of bibliometrics *(continued)*

- Development/use of bibliometrics in web-service
  - Instantly gave access to thousands of bibliometric users
  - “Black-boxing” of bibliometric statistics and blind faith
  - Growth of “subliminal bibliometrics” in peer-review
- There can be perverse incentives associated with bibliometric measures
  - E.g., altered publication behaviours

Limits of bibliometrics *(continued)*

- Still vast problems of data “ambiguity”
  - Very difficult to determine how many papers published by Smith, Tremblay, Wang, Lee and Kumar
  - Accuracy is a direct result of care placed in data cleaning and dataset preparation

Limits of bibliometrics *(continued)*

- Data "ambiguity" illustration

ISI Web of Knowledge℠

Essential Science Indicators℠

WELCOME | HELP | RETURN TO MENU | IN-CITES

SCIENTIST RANKINGS IN MATERIALS SCIENCE

Display items with at least: 0 Citation(s)

Sorted by: Papers SORT AGAIN

1 - 20 (of 4284) [1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10] Page 1 of 215

	View		Scientist	Papers	Citations	Citations Per Paper
1			WANG, Y	2,081	20,544	9.87
2			ZHANG, Y	1,976	17,095	8.65
3			LIU, Y	1,964	13,204	6.72
4			WANG, J	1,891	15,493	8.19
5			LI, J	1,701	14,215	8.36
6			LI, Y	1,628	14,360	8.82
7			ZHANG, J	1,605	11,770	7.33
8			ZHANG, L	1,441	9,317	6.47
9			WANG, L	1,423	10,597	7.45
10			WANG, H	1,420	10,558	7.44
11			LEE, JH	1,293	10,448	8.08
12			KIM, JH	1,286	8,724	6.78
13			ZHANG, H	1,200	10,694	8.91
14			WANG, X	1,193	10,063	8.44
15			LI, L	1,150	8,961	7.79

When to/not to use bibliometrics

- More robust use is in the presence of the law of large numbers
- Smaller datasets more prone to show exceptional as opposed to regular behaviour
- Use in proper context or avoid absolute indicators as much possible
  - OK to compare Western countries in the NSE, not OK to compare a medical to a sociology faculty

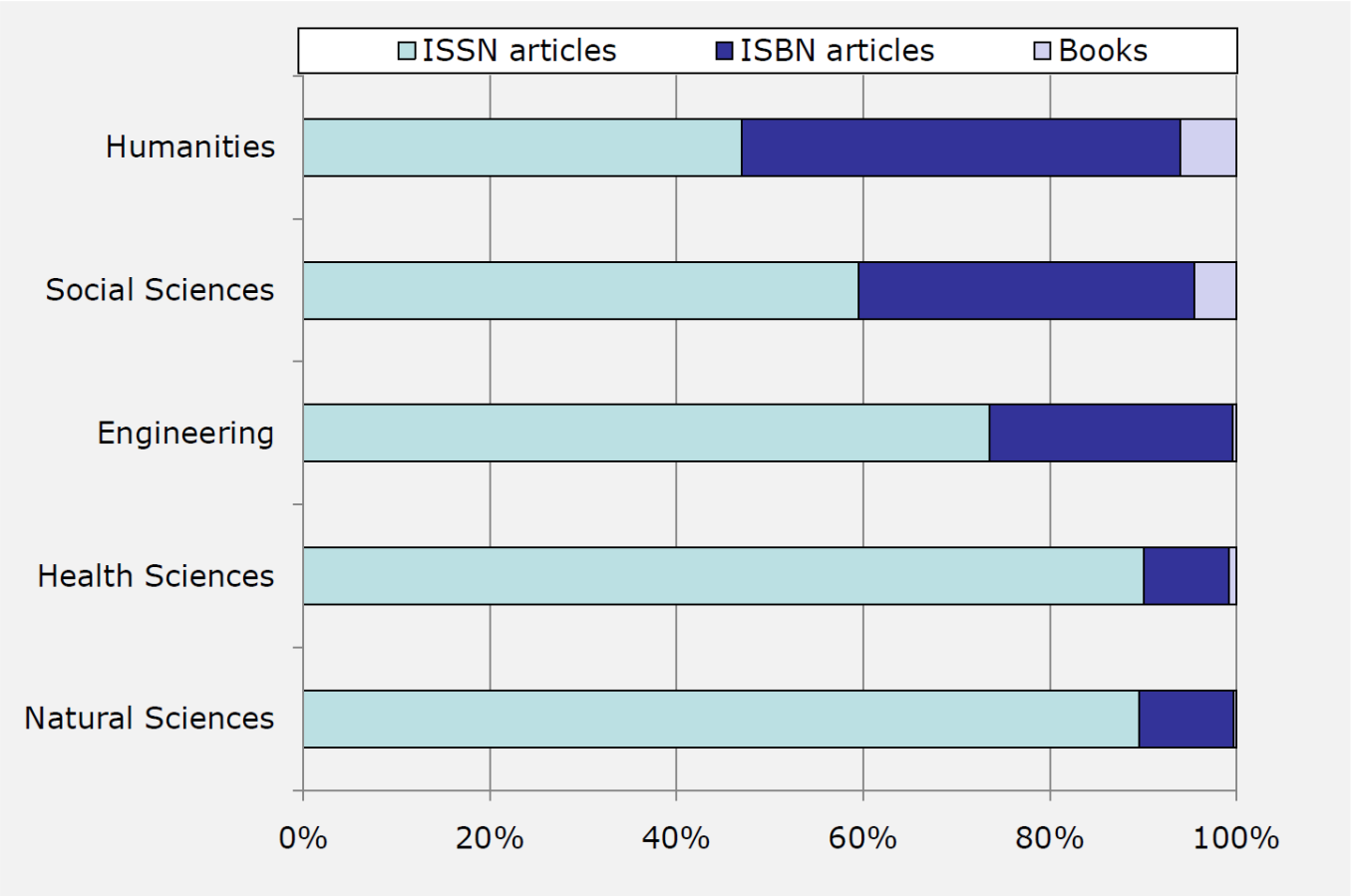
When to/not to use bibliometrics

- Very good in the natural and health sciences, acceptable in engineering (proceedings ideally included)
- Careful use in social sciences, language, and local biases
- Extreme caution when used in the humanities and the arts – books and other forms of knowledge diffusion and expression count

When to/not to use bibliometrics

Publication types in five major fields.

Based on 33.000 (fractionalized) publications from Norway's HE sector 2005-2008

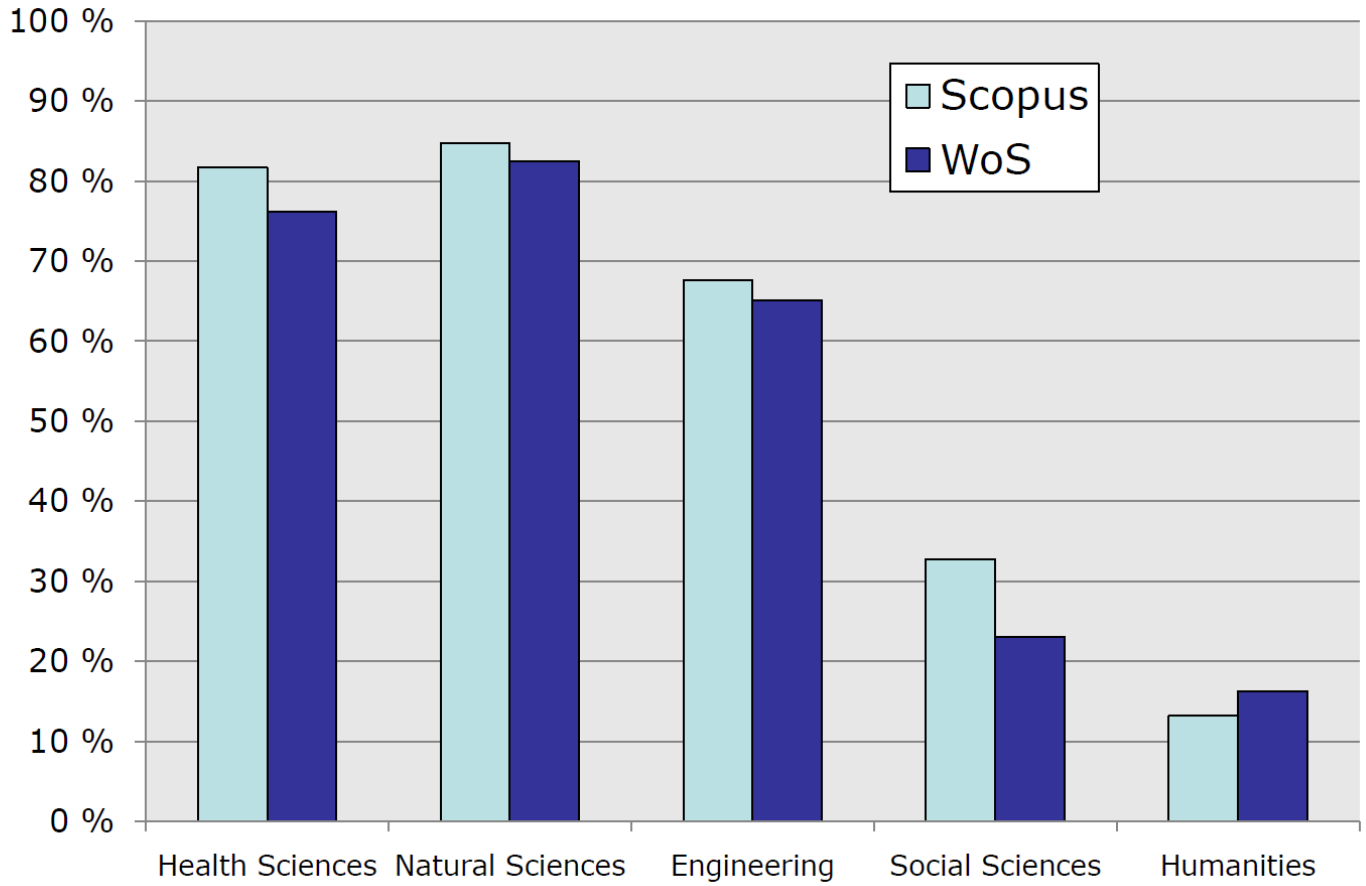


Source: Gunnar Sivertsen. 2009. A Bibliometric Funding Model based on a National Research Information System. ISSI 2009 - 12th ISSI Conference.

When to/not to use bibliometrics

Coverage of all scientific publications in Norway's HE sector

Based on 33.000 (fractionalized) publications from 2005-2008



Source: Gunnar Sivertsen. 2009. A Bibliometric Funding Model based on a National Research Information System. ISSI 2009 - 12th ISSI Conference.

Learning activity

- Individual work, followed by table discussion
- Answer the questions in the learning activity handout
- You will need to use the calculators on your table or on your smartphones
- 15 minutes

Key messages

- Be careful with non-normalized indicators – papers, citations, impact factor, H-Index
  - Having access to numbers does not mean they are adequate for the job
- Be critical of database coverage
  - Databases used in bibliometrics were designed for bibliographic search
  - Databases are used for convenience, not because they are ideal

Key messages *(continued)*

- Use with great care outside the natural and health sciences
- Bibliometrics is a complex science and technological undertaking
- If you're not comfortable fixing your car's fuel injection system, you shouldn't be uncomfortable asking experts to diagnose your research engine

Recommended readings

- Archambault, É. (2002). Methods for using patents in cross-country comparisons. *Scientometrics*, *54*(1), 15–30.
- Archambault, É. et al. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329–342.
- Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics*, *79*(3), 635–649.
- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, *60*(7), 1320–1326.

Recommended readings *(continued)*

- Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): past, present, and future. *Scientometrics*, *52*(3), 365-377.
- Bordons, M., Fernández, M. T., & Gomez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, *53*(2), 195-206.
- Campbell, D., et al. (2010). Bibliometrics as a performance measurement tool for research evaluation: The case of research funded by the National Cancer Institute of Canada. *American Journal of Evaluation*, *31*(1), 66–83.
- de Solla Price, D.J. (1963) *Little Science Big Science*. New York: Columbia U. Press. 119 pages.
- Glänzel, W., & Schoepflin, U. (1994). Little scientometrics, big scientometrics... and beyond?. *Scientometrics*, *30*(2), 375-384.

Recommended readings *(continued)*

- Katz, J. S., & Hicks, D. (1997). Desktop scientometrics. *Scientometrics*, *38*(1), 141-153.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington, D.C: Computer Horizons. Cherry Hill, N.J.; Computer Horizons, Inc. NTIS-PB252339/AS.
- Van Raan, A. F. (1997). Scientometrics: State-of-the-art. *Scientometrics*, *38*(1), 205-218.
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences?. *Scientometrics*, *62*(1), 117-131.

Thank you!

Éric Archambault

Science-Metrix

eric.archambault@science-metrix.com

Science-Metrix