CWTS
Meaningful metrics

Issues related to the application of bibiometrics in the research assessment of social sciences, humanities and law

Dr. Thed van Leeuwen

Research seminar, Dortmund, 14th January 2015,

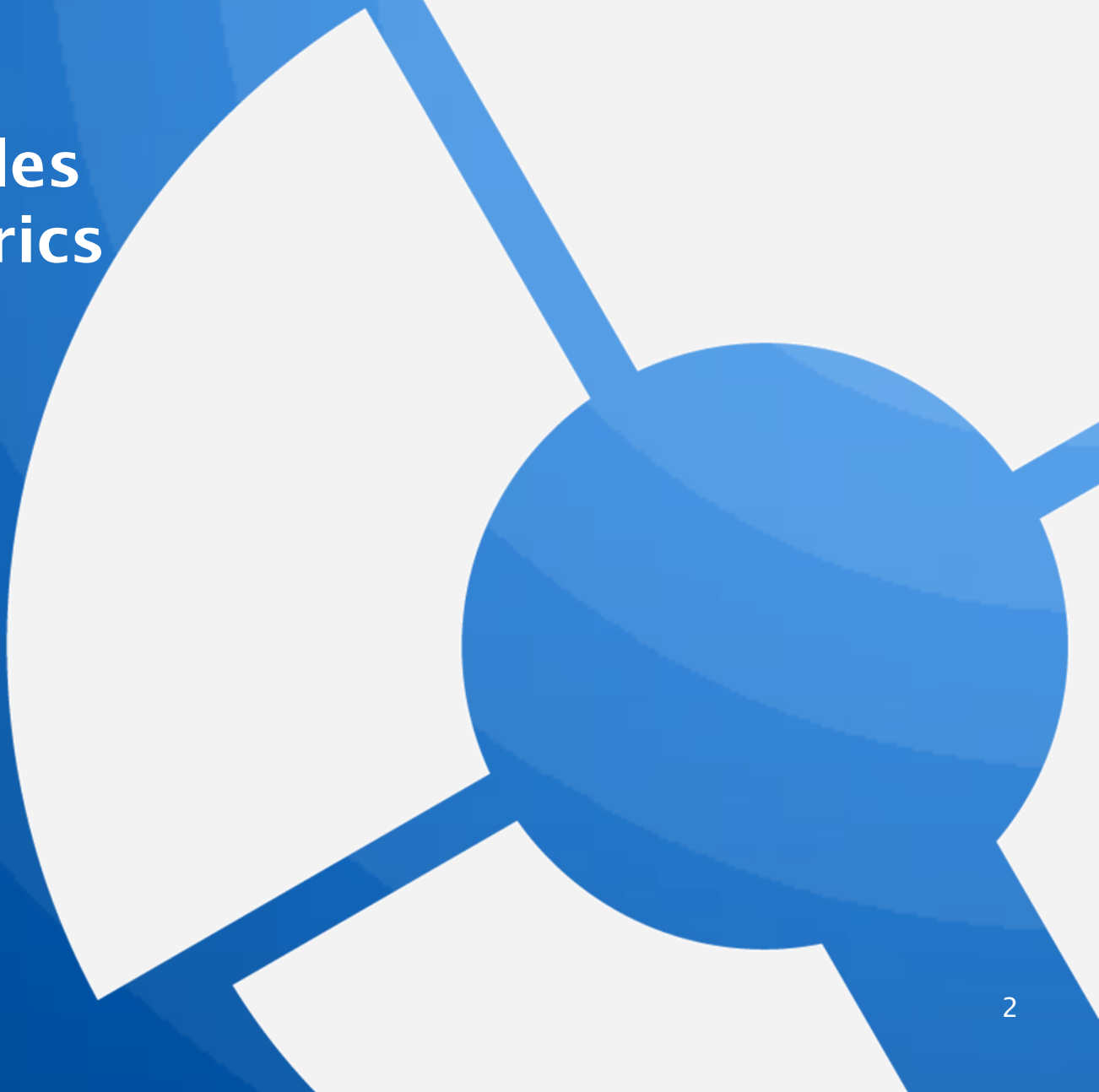
Universiteit Leiden

Outline

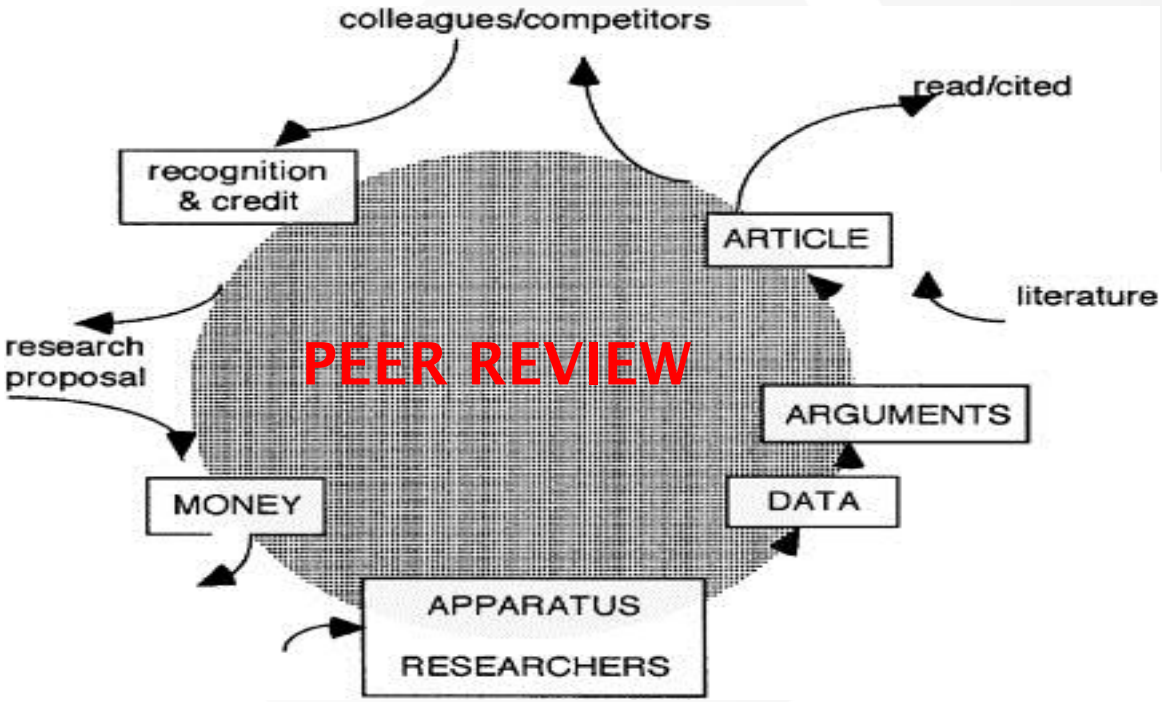
- Academic credibility cycle and bibliometrics
- A short historical perspective on evaluation of science and technology
- Coverage in bibliometric studies
- Specific issues related to bibliometrics and the SSH & Law

CWTS

Academic credibility cycles and bibliometrics



‘Classical’ image of the Credibility cycle



Credibility cycle (adapted from Latour and Woolgar (1979) & Rip (1990)

CWTS

Rise of performance indicators & bibliometrics

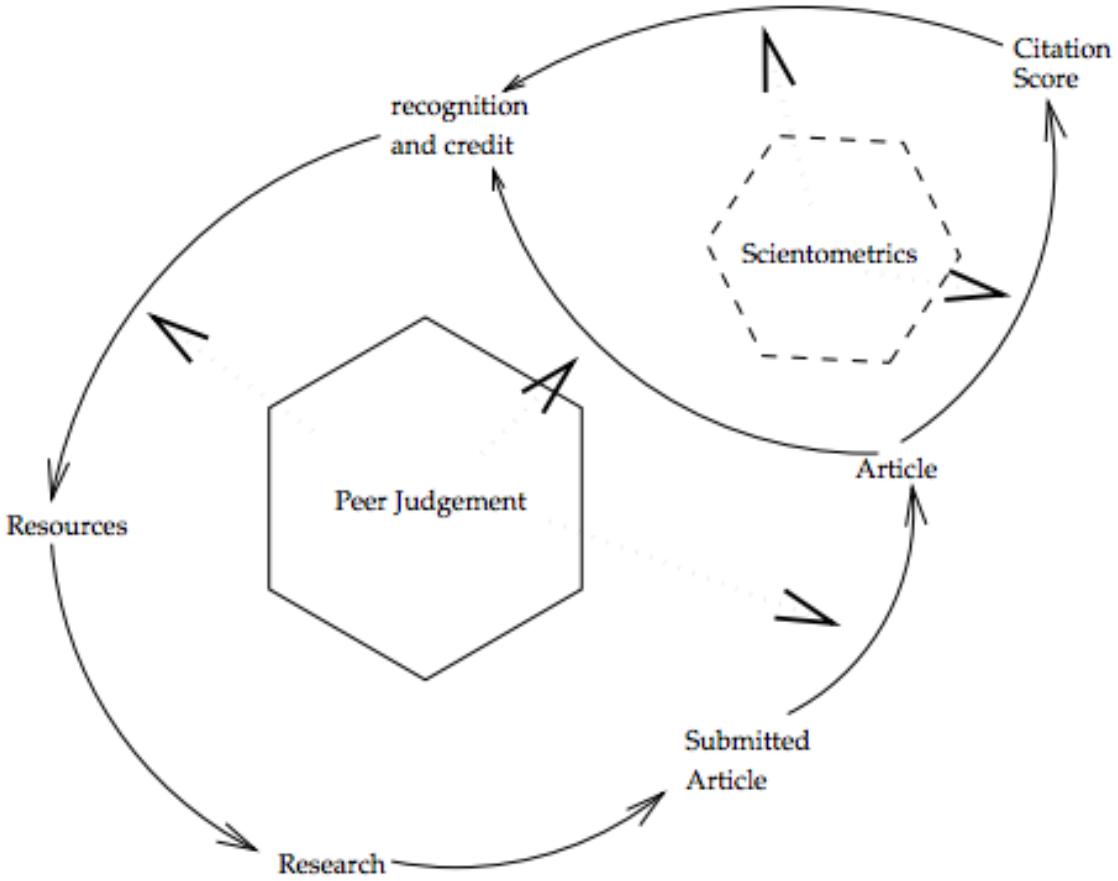
Increasing need for formalised measures:

- 'Push' from science policy (from 1970s onwards)
- Independent of peer review
- New Public Management / Neo-liberalism (from 1980s onwards)

Growing pressure on the research community:

- Researchers part of international community
  - Peer review
- ... but also part of local institutions
  - Specific management practices (yearly appraisals, external evaluations)
- Institute managers not alsways part of international expert community
- From the 1990's: tighter forms of management
  - Distance

Extended credibility cycle



An historical perspective on evaluation of science and technology

Macro developments on a global scale

- First national indicator reports on S & T in U.S. 1970s
  - *Science Indicators 1972*
  - *Science Indicators 1974*
- Europe followed some 20 years later
  - Dutch Observatory of S & T Indicators (NOWT, 1992)
  - French Observatory of S & T (OST, 1993)
- After that, the European Union followed
  - First European Report on S & T Indicators (1994)

CWTS

National developments in Europe

- 1st European research assessment systems in UK and Finland
  - *Research Assessment Exercises (RAE) in UK from 1986 onwards, nowadays labeled as Research Excellence Framework (REF)*
  - *Finland started in 1986, implementation in early 1990s*
- Other European countries followed these initiatives
  - In the Netherlands, the national assessment cycles started in 1993/4
  - Austria was relatively late with implementing such a system (2002)

CWTS

Lessons learned from national developments

- UK
  - *RAE/REF is a heavy burden for the university system*
  - *Re-orientation on peer review*
  - *Awareness of effects system (i.e. concentration of funding)*
  - *…as well as political prioritizing of STM disciplines*
- Netherlands
  - *A broadly accepted quality assurance system*
  - *Criteria seem useful for evaluation institutes and groups*
  - *Continuity is important, but changes should be adopted more easily*
  - *System possibly more complex*
  - *No direct link funding and evaluation*

CWTS

Money makes the world go ‘round …

- Important element in design, implementation & application of any assessment procedure:
  → link between review outcomes and funding schedules

- Any system (partially) based on metrics invites ‘*playing the numbers game*’, and may lead to ‘*impact engineering*’

- This may lead to perversions of the review system
  - –Focus on ‘best publications’ caused UK system to benefit natural sciences, life sciences, and biomedicine
  - –Focus on one bibliometric indicator had similar consequences
  - –Focus on publications in Web of Science led to decrease of the national scientific position of Australia (*Butler, Nature, 2002)*

CWTS

Coverage in bibliometric studies

Introduction

- The use of evaluative bibliometrics can only become meaningful when used in a the right context.
- Publication culture of the unit(s) under assessment are shaping that context.
- As such, any bibliometric study should start with an assessment of the adequacy of metrics in that particular context.
- Therefore, CWTS has developed methods to assess that fit of metrics in a certain context.

CWTS

How to define adequate coverage ?

- In order to determine whether metrics applied in an assessment context are meaningful, one needs to know what is represented through the metrics.

- We distinguish two types of coverage:
  - ***Internal*** (from inside the perspective of the WoS)
  - ***External*** (from the perspective of a total output set)

CWTS

Assessing the adequacy of WoS for bibliometrics: The *Internal* coverage method

- Look at publications in WoS across fields,
- Use the references given by the authors of the publications,
- Analyze the communication channels referred to,
- Usage of WoS journals as share of the total number of references is an indication of the relevance for the authors involved,
- Thereby constituting a basis for the usage of bibliometrics as evaluation tool !

CWTS

Assessing the adequacy of WoS for bibliometrics: The *External* coverage method

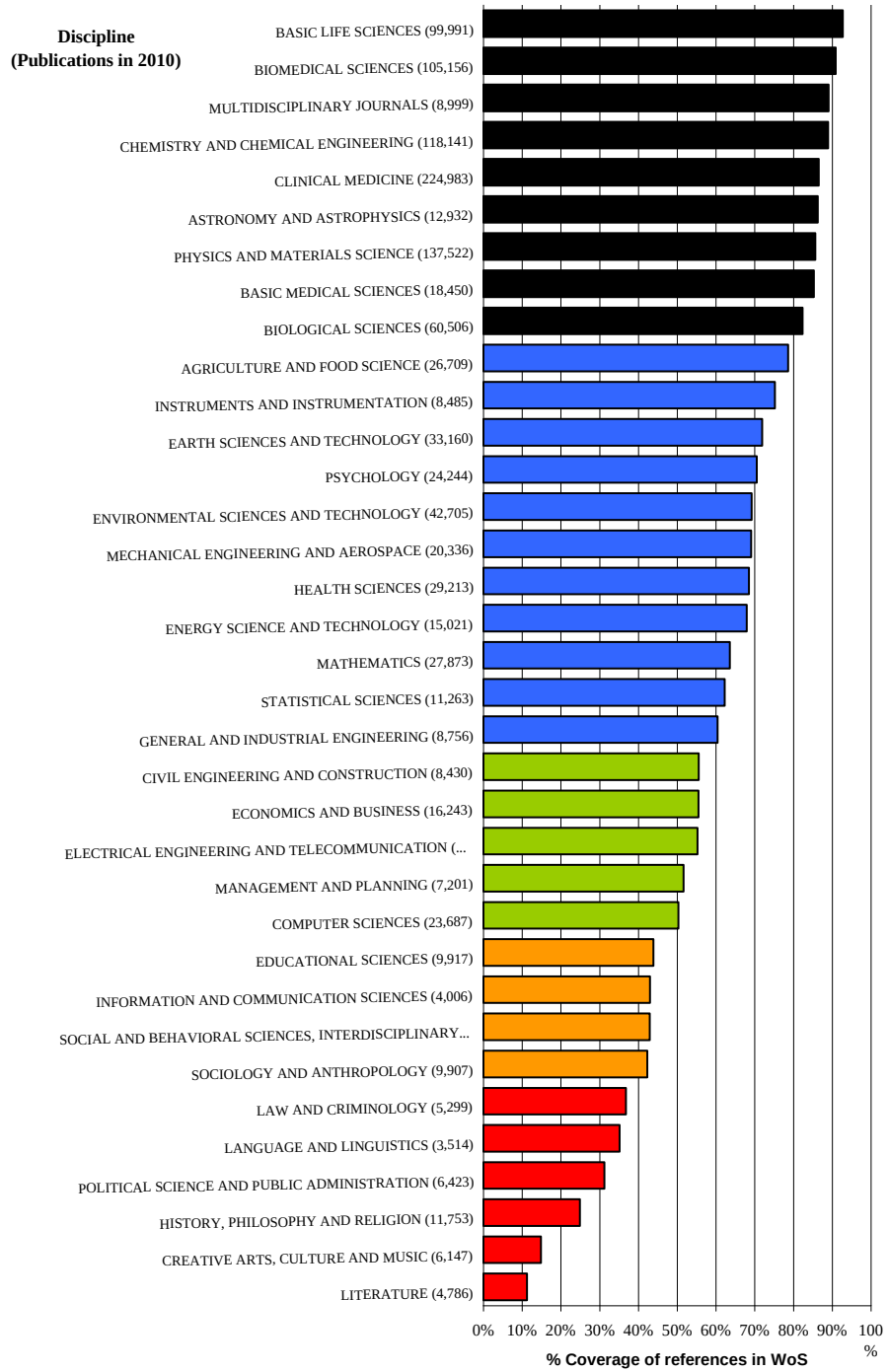
- Use the list of publications of an organization, subject of a bibliometric analysis
- Match the submitted list with the WoS
- Degrees of covered scientific outlets indicate the relevance of WoS journals
- Thereby constituting a basis for the usage of bibliometrics as an evaluation tool !

CWTS

AU	Moed, HF; Garfield, E.	in WOS
TI	In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter	
SO	SCIENTOMETRICS 60 (3): 295-303, 2004	Y
RF	ABT HA, J AM SOC INF SCI T, v 53, p 1106, 2004	Y
	GAR TION INDEXING, 1979 (BOOK!)	N
	GA S INFORMATION S, v 8, p 403, 1985	N
	GILBERT GN, SOC STUDIES SCI, v 7, p 113, 1977	Y
	MERTO	Y
	RO	Y
	ZUC	Y

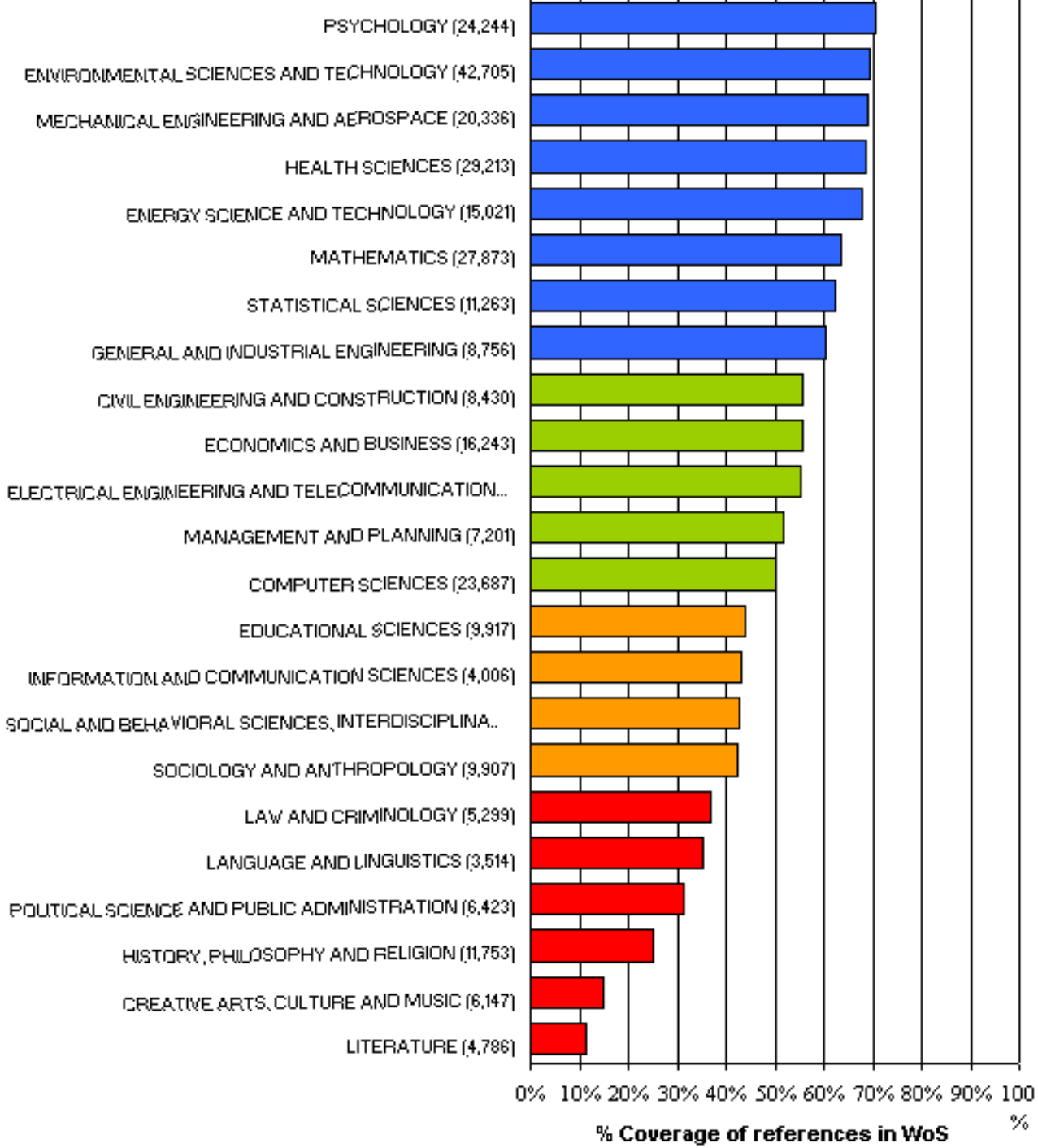
Not in WoS

WoS Coverage
= 5/7 = 71%

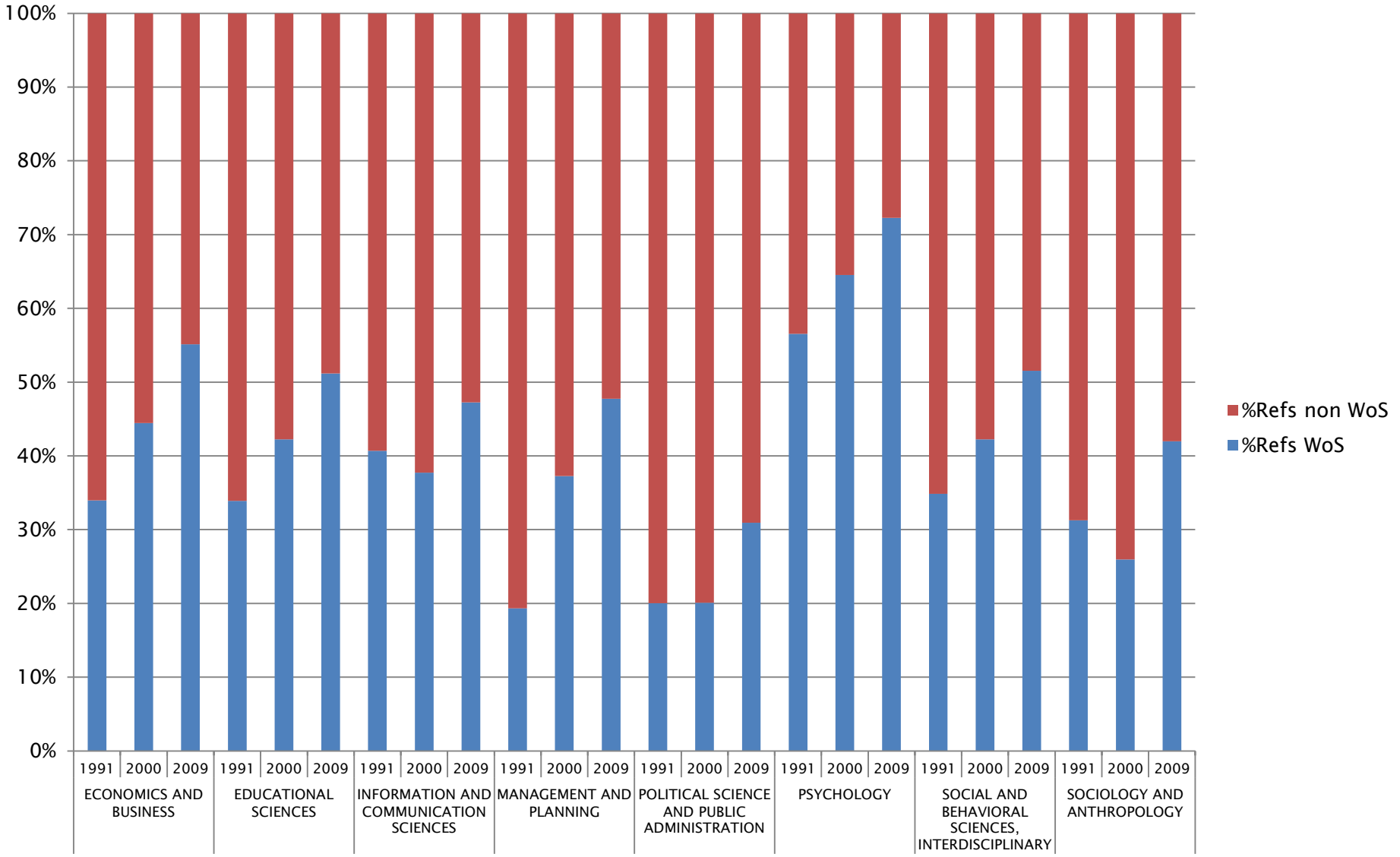


WoS Coverage in 2010 across disciplines

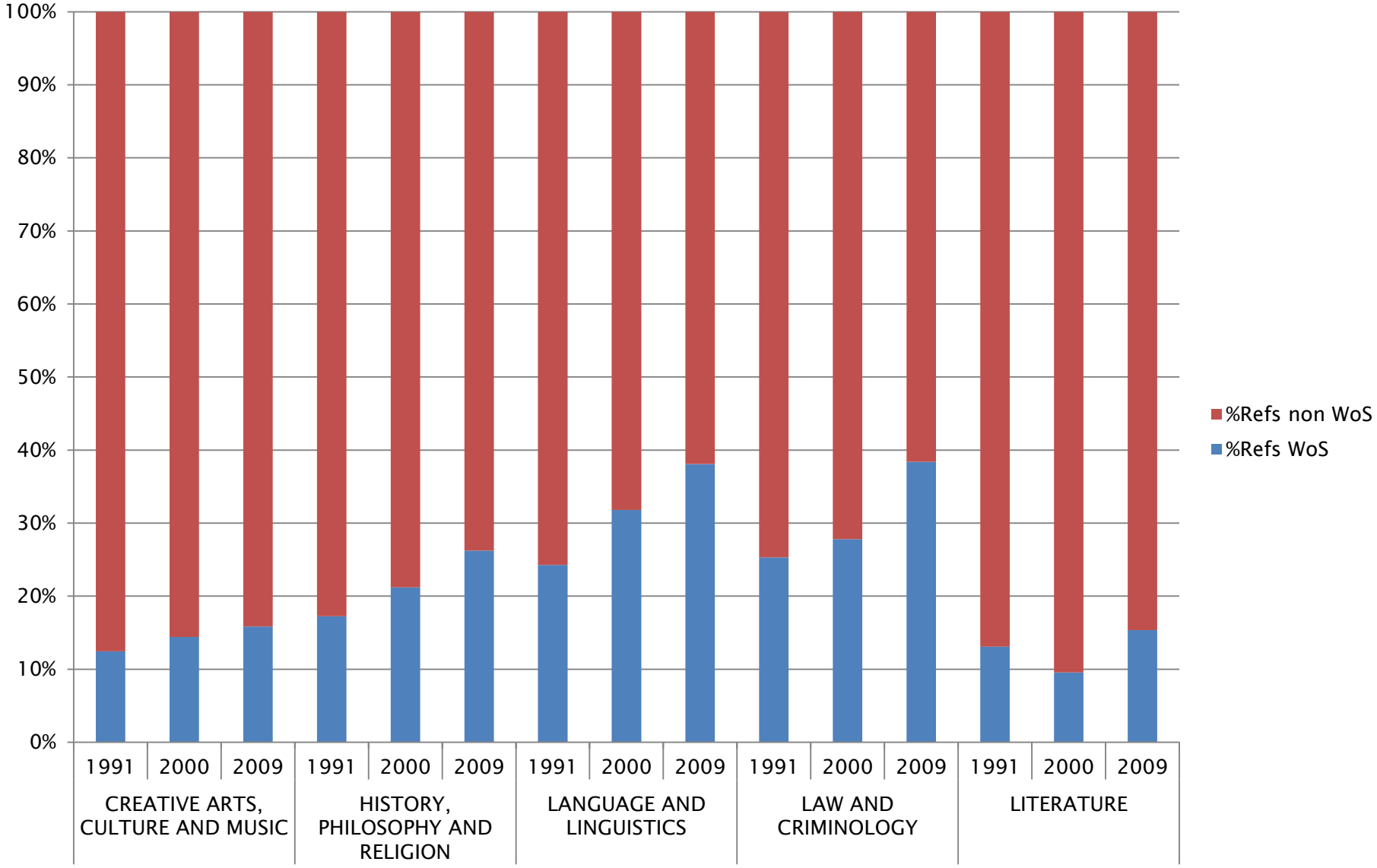
- **Black=Excellent coverage** (>80%)
- **Blue= Good coverage** (between 60-80%)
- **Green= Moderate coverage** (but above 50%)
- **Orange= Moderate coverage** (below 50%, but above 40%)
- **Red= Poor coverage** (highly problematic, below 40%)


PSYCHOLOGY (24,244)
ENVIRONMENTAL SCIENCES AND TECHNOLOGY (42,705)
MECHANICAL ENGINEERING AND AEROSPACE (20,336)
HEALTH SCIENCES (29,213)
ENERGY SCIENCE AND TECHNOLOGY (15,021)
MATHEMATICS (27,873)
STATISTICAL SCIENCES (11,263)
GENERAL AND INDUSTRIAL ENGINEERING (8,756)
CIVIL ENGINEERING AND CONSTRUCTION (8,430)
ECONOMICS AND BUSINESS (16,243)
ELECTRICAL ENGINEERING AND TELECOMMUNICATION...
MANAGEMENT AND PLANNING (7,201)
COMPUTER SCIENCES (23,687)
EDUCATIONAL SCIENCES (9,917)
INFORMATION AND COMMUNICATION SCIENCES (4,006)
SOCIAL AND BEHAVIORAL SCIENCES, INTERDISCIPLINA..
SOCIOLOGY AND ANTHROPOLOGY (9,907)
LAW AND CRIMINOLOGY (5,299)
LANGUAGE AND LINGUISTICS (3,514)
POLITICAL SCIENCE AND PUBLIC ADMINISTRATION (6,423)
HISTORY, PHILOSOPHY AND RELIGION (11,753)
CREATIVE ARTS, CULTURE AND MUSIC (6,147)
LITERATURE (4,786)
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100 %
% Coverage of references in WoS

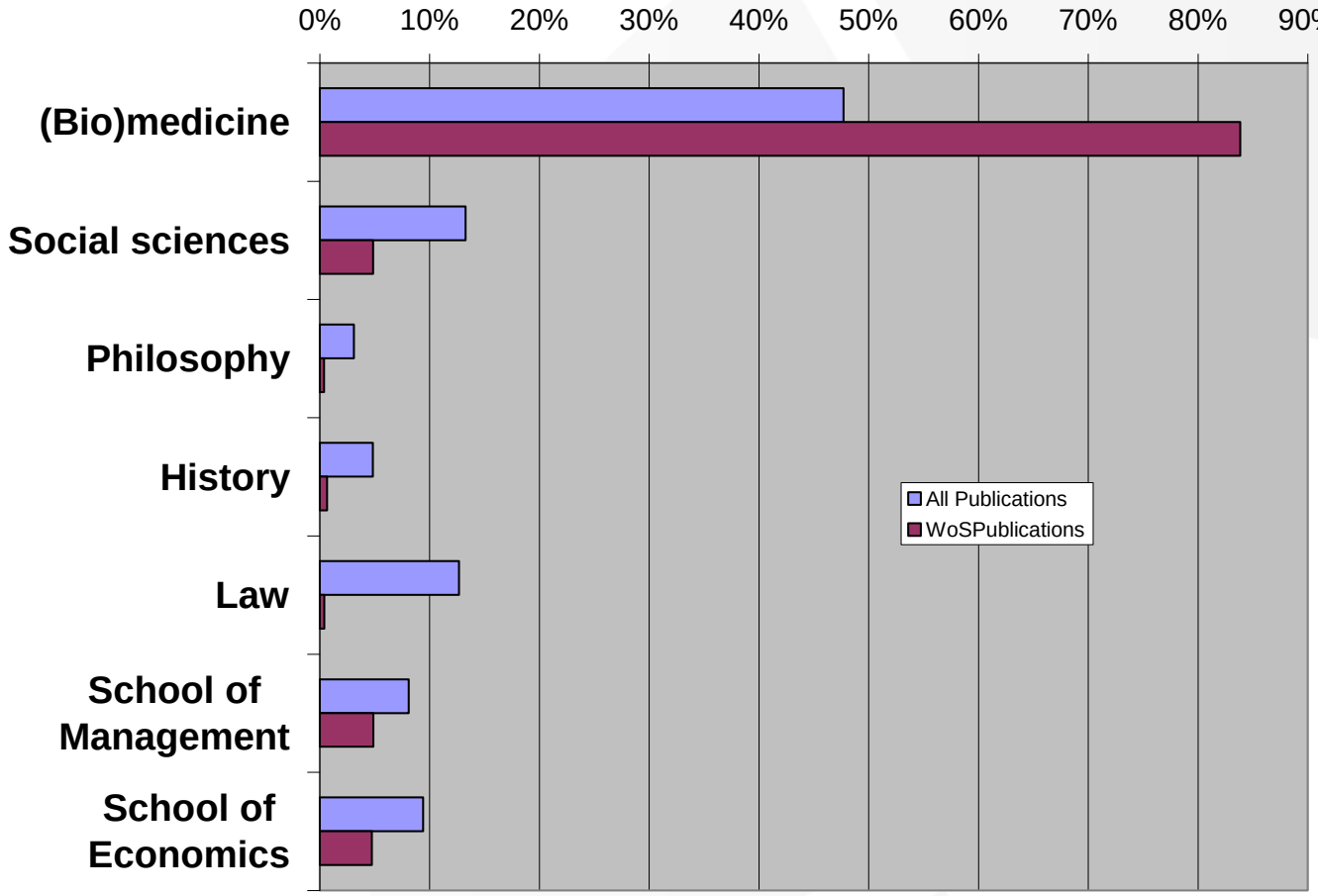
Social & behavioral sciences



Humanities & Law

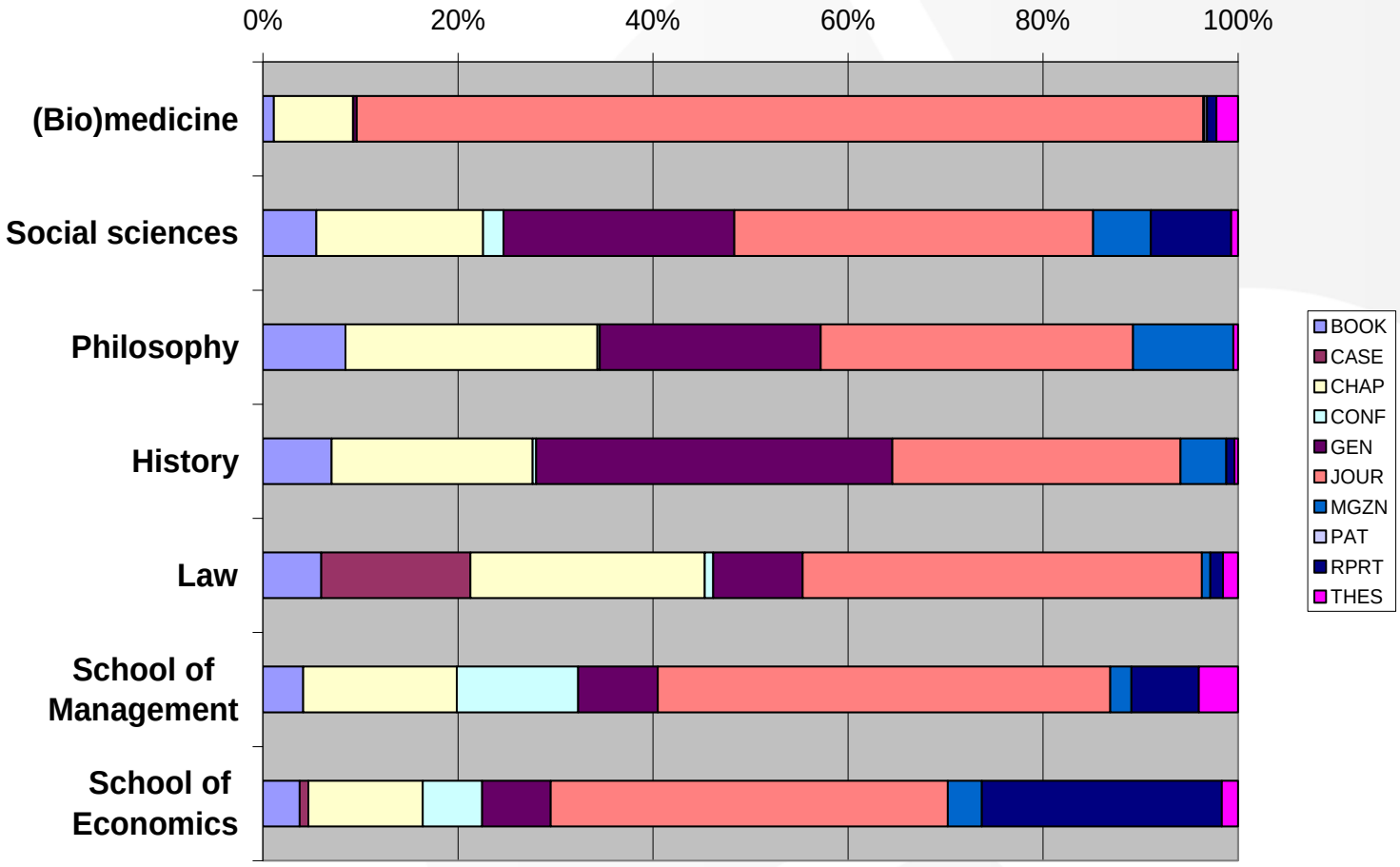


Difference between the internal registration system & representation WoS



- Dominance university hospital in WoS realm extremely visible
- Law and Humanities ‘disappear’ in WoS realm

Composition of the output of the university in METIS:
The external coverage of a university



- The category General is in some cases voluminous
- All units do have journal publications !

Specific issues related to bibliometrics and the SSH and Law

The language issue …

- English is the major language for communicating research findings (Garfield, 1990).
- Even in Medicine, publishing in other languages than in English will influence the impact scores, …
- … as even on the level of a whole country, some 20% difference in scientific impact can be observed (van Leeuwen et al, 2001).

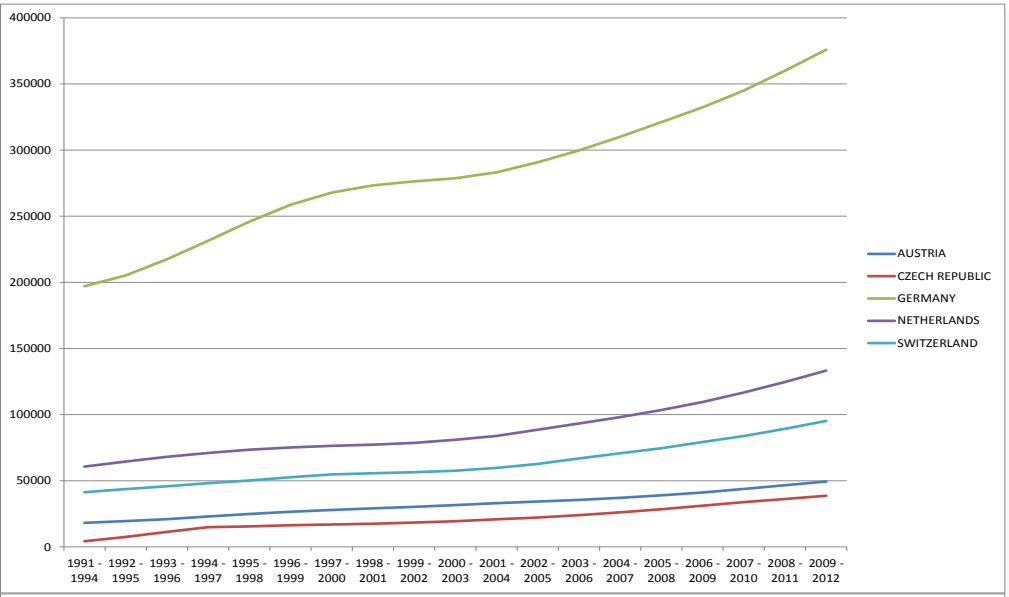
CWTS

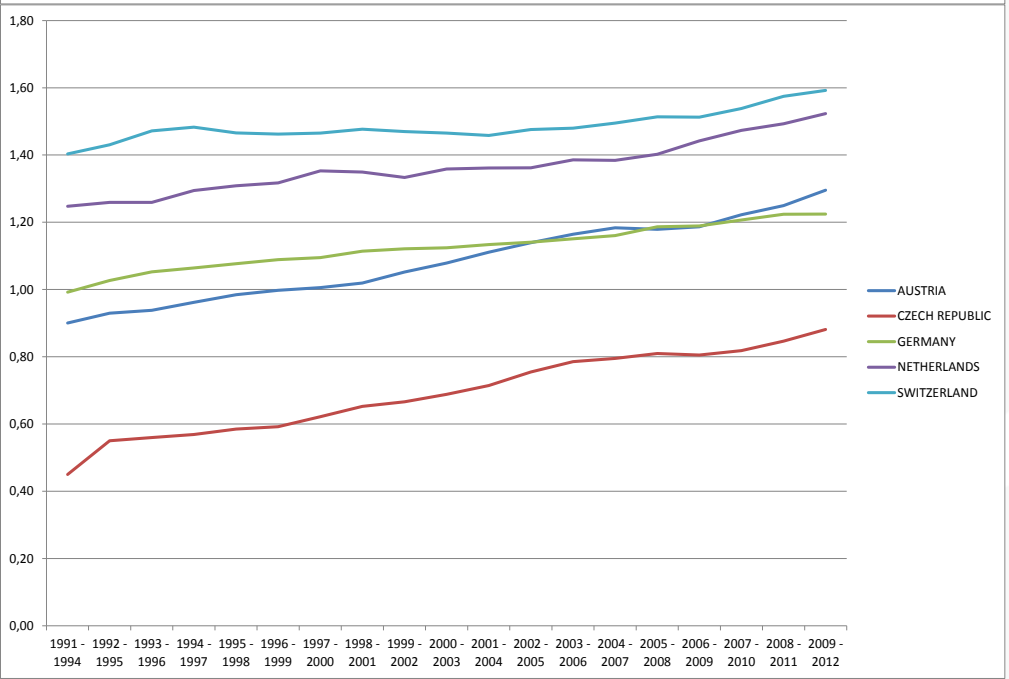
Introduction

- Follow up on the Regensburg lecture (2012) and the publication in Bibliometrie – Forschung & Praxis (2013)
- Focus on language issues in WoS output, distinguish between English and non English language output.
- Selected output for Germany, Austria, Switzerland, the Czech Republic, and the Netherlands.
- Further focus on scientific disciplines in SSH and Law.

CWTS

The international comparison

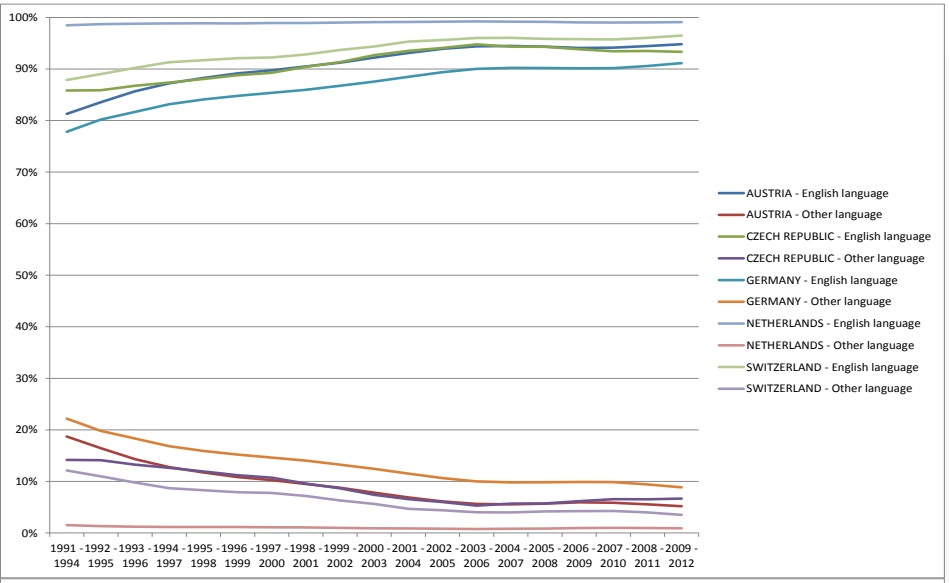


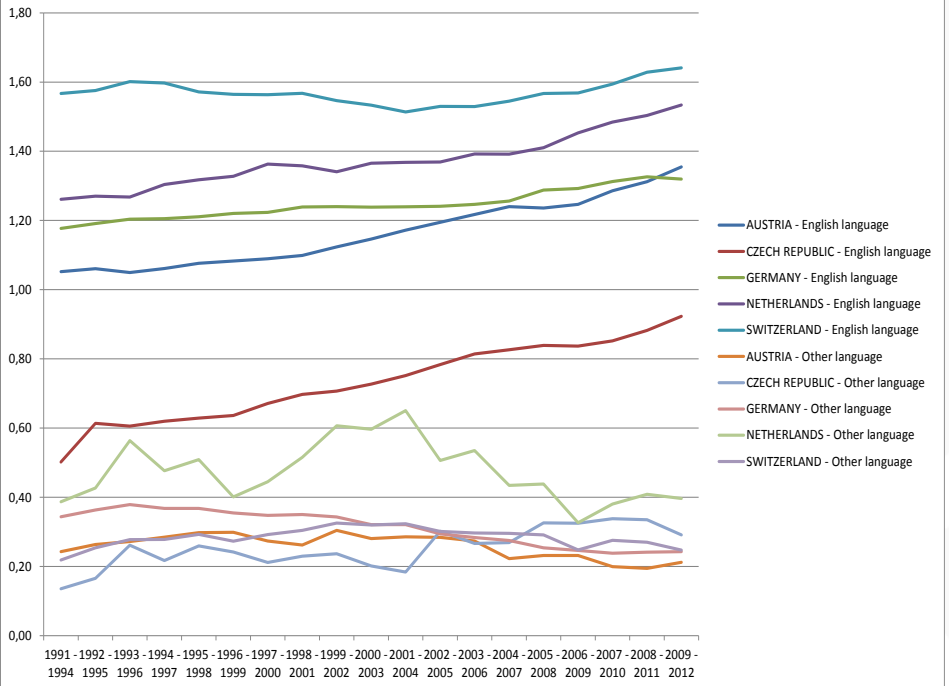


Overall output and impact

- Output wise, we observe a strong increase of output, particularly for Germany.
- Overall impact is increasing as well !

The international comparison

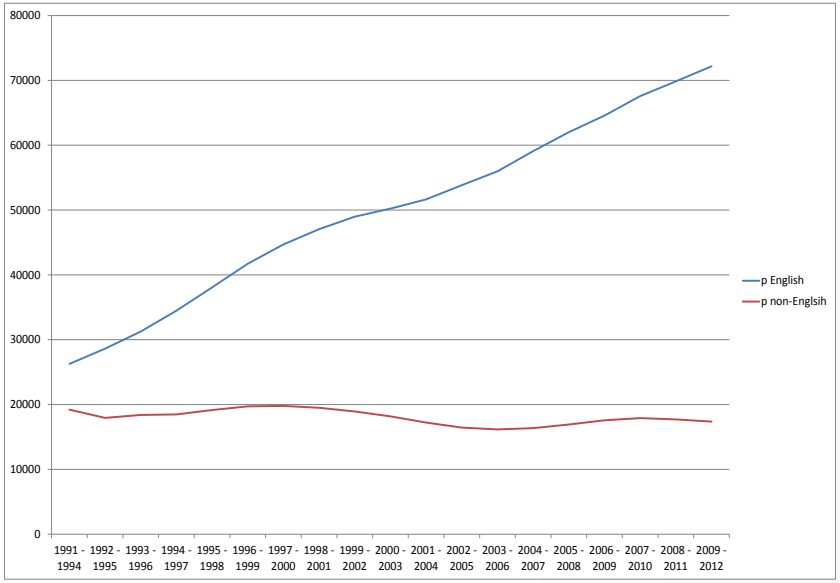


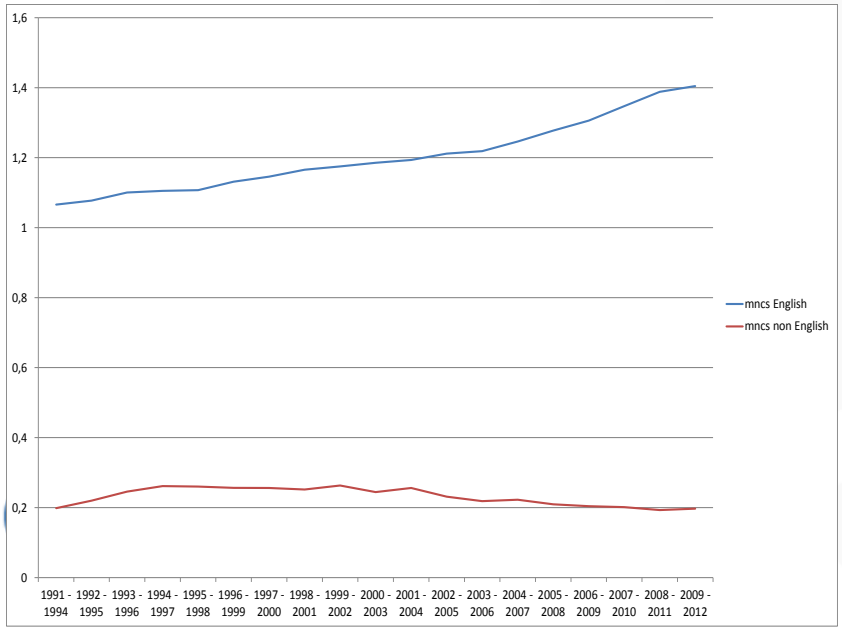


Output and impact by language

- Output wise, we observe a strong relative divergence, with a stronger focus on English.
- Impact of English-language part is increasing, and higher as compared to overall impact !

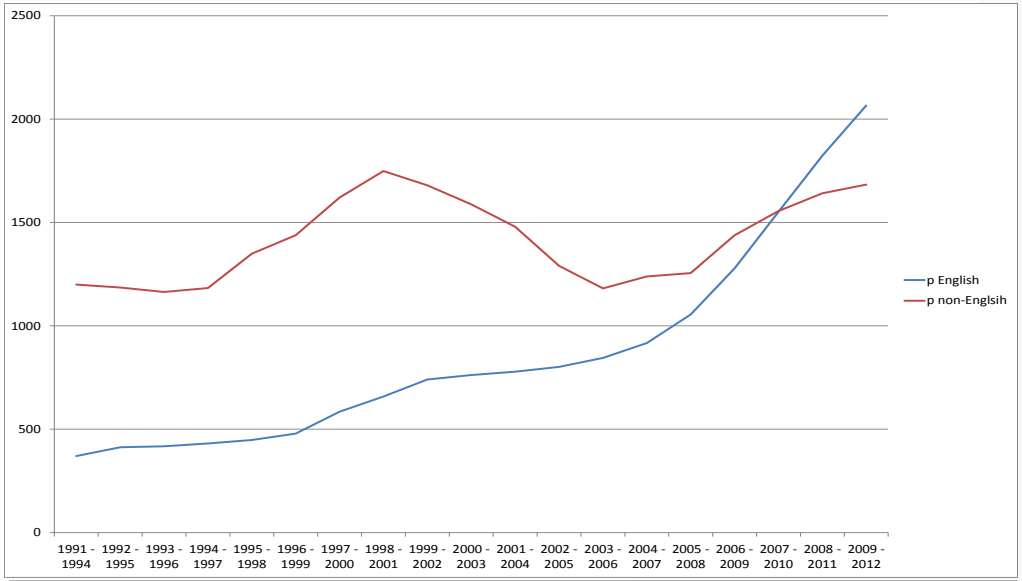
It affects Clinical medicine as well …

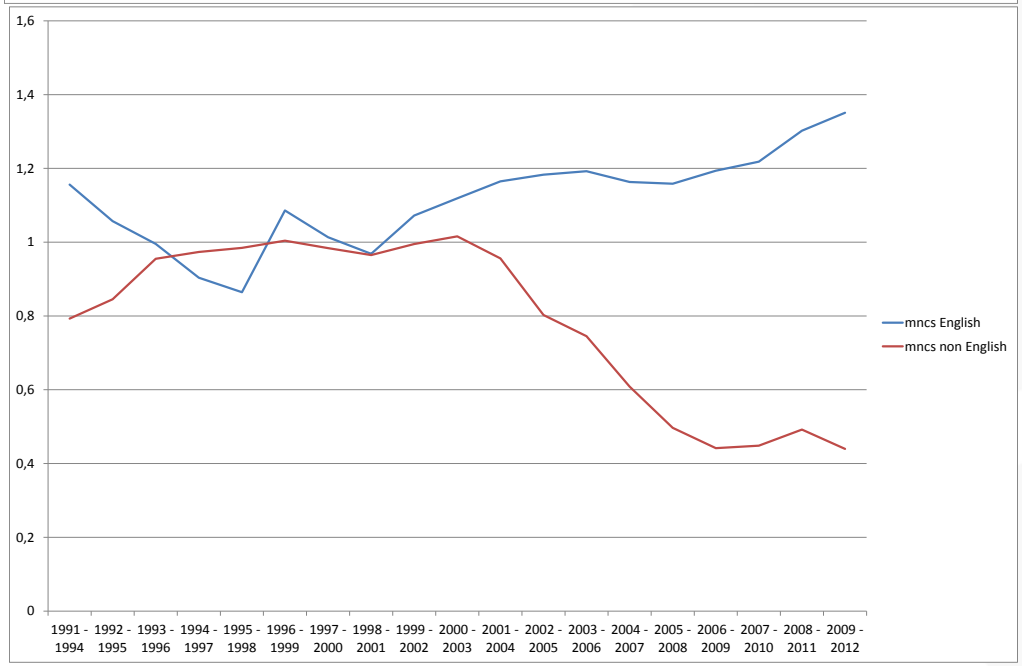




- Output wise, we observe a strong divergence of both types of publications.
- This is less well visible for the impact of both types.

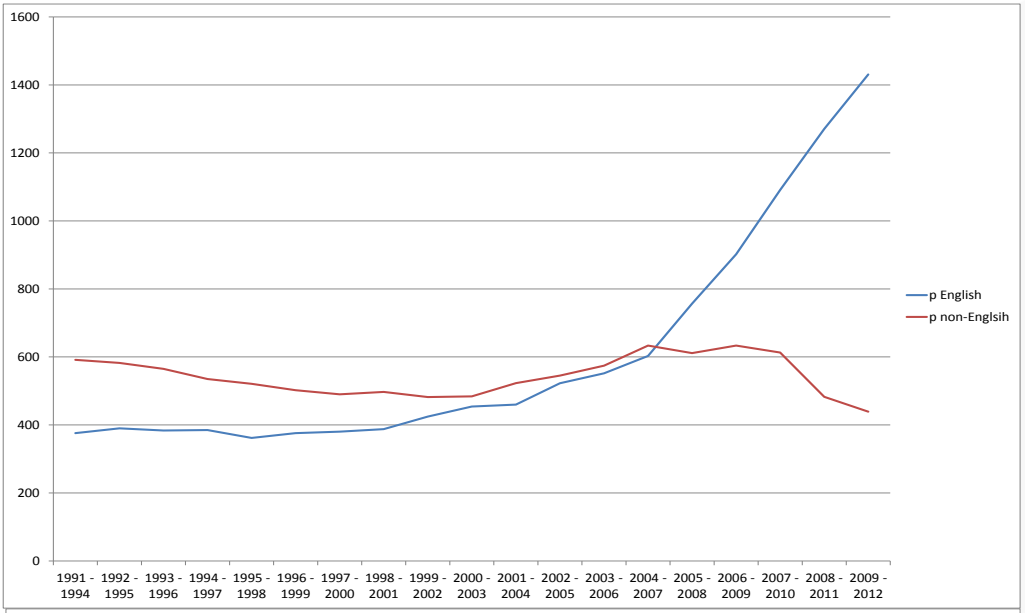
History, Philosophy & Religion

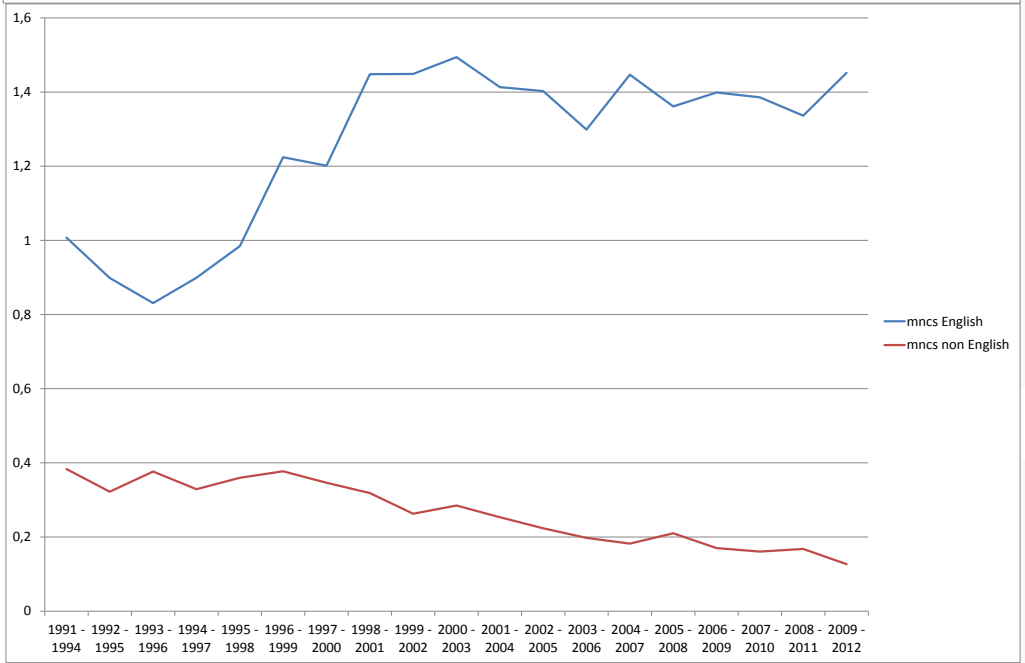




- Output wise, we observe a strong increase of English language output in this discipline, while non English fluctuates.
- Impact wise, we observe a strong divergence from the mid 1990's onwards.

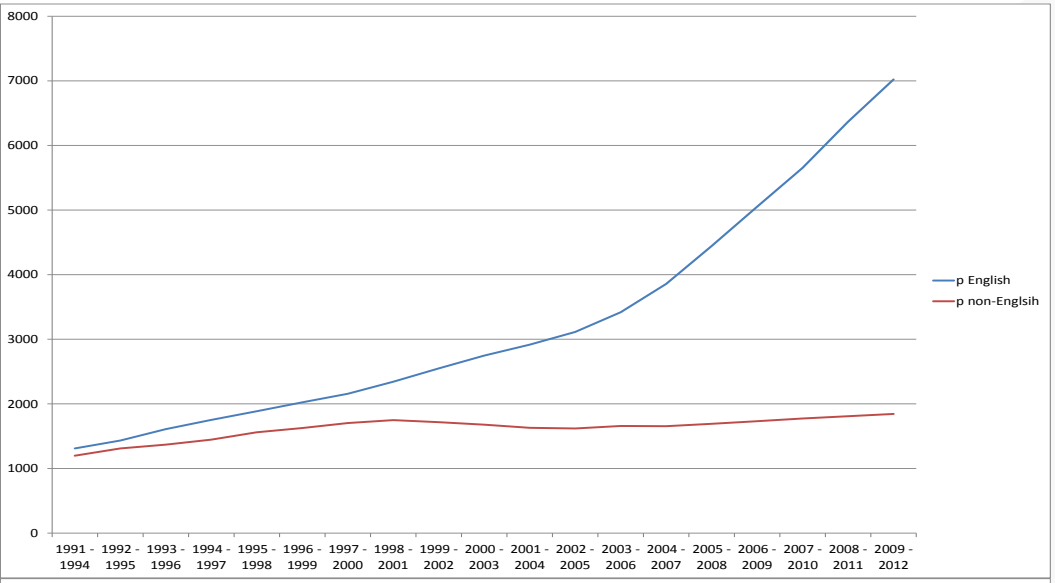
Political Science

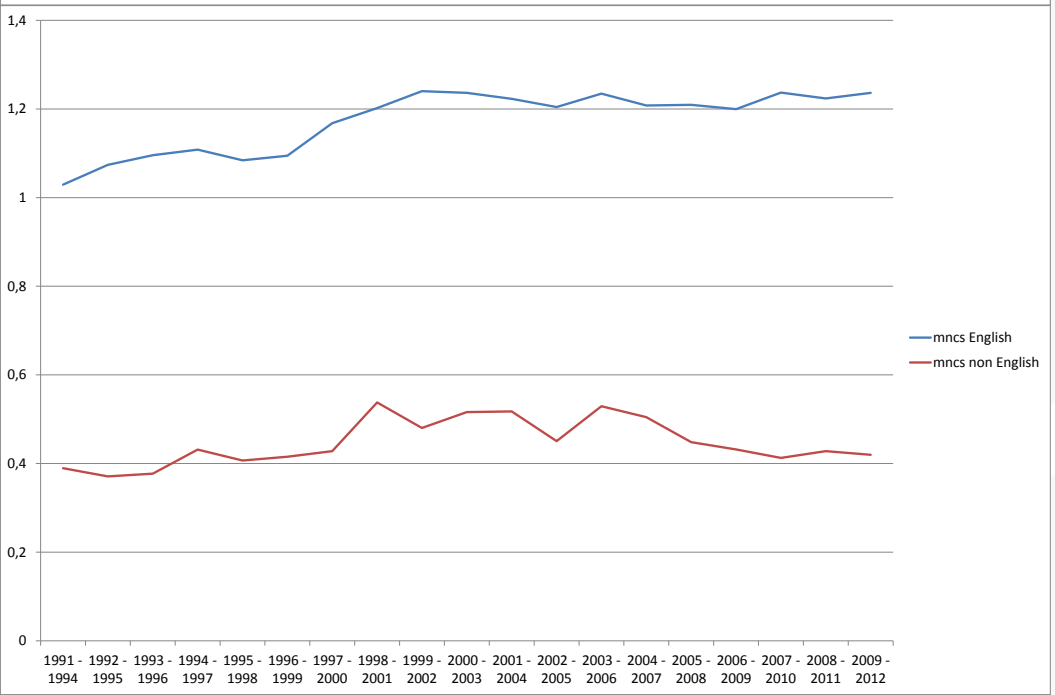




- Output wise, non English language is relatively stable, with an increasing output in English language output.
- Difference in impact scores initially diverges, and stabilizes for English language output, the non English language output decreases.

Psychology





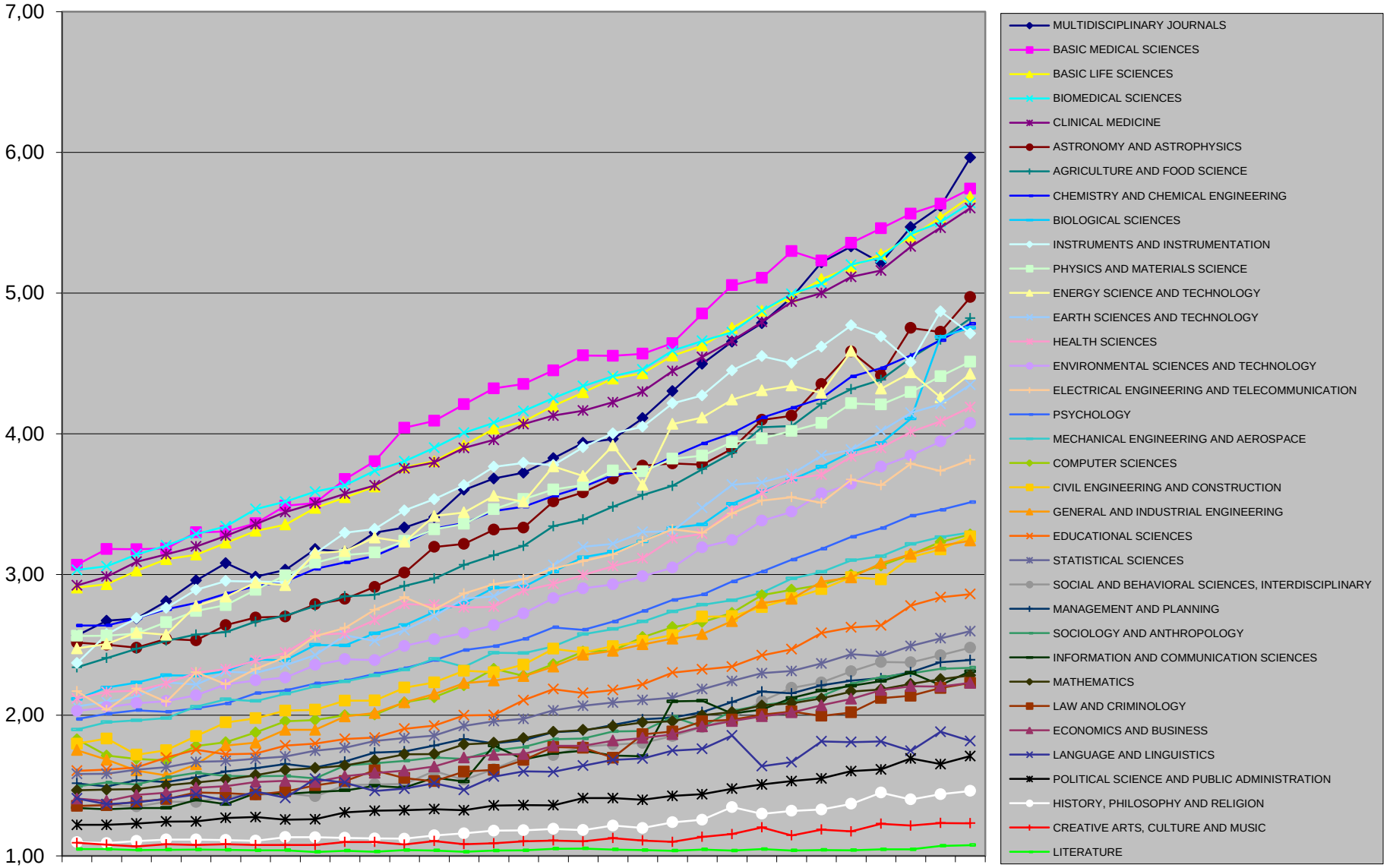
- Output wise, we observe a strong divergence of both types of publications, from the late 1990's onwards.
- Impact scores are stable, and strongly differing.

The authorship issue …

- Authorship is an important element in bibliometric analysis.
- Authorship relates directly to:
  - Scientific collaboration.
  - Mechanisms of credibility.

CWTS

Authorship across disciplines



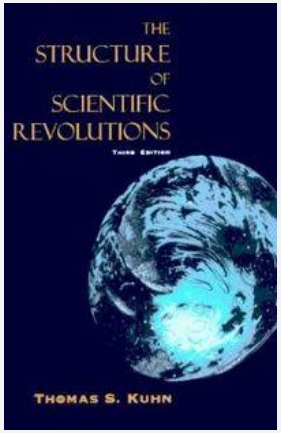
Field-specific referencing practices

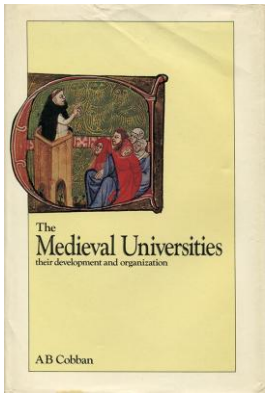
- But how about the usage of other source material through referencing ?
- An interesting starting point could be the Book Citation Index by Thomson Reuters.

- Indications of the usage of primary and secondary material come in various forms:
  - References
  - Endnotes
  - Footnotes
  - Bibliography

CWTS

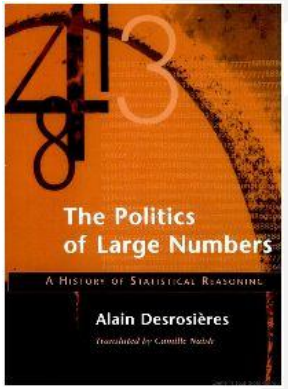
- Thomas Kuhn, "***The structure of scientific revolutions***", 1996
- *Contains footnotes, with extensive references*
- *No bibliography*

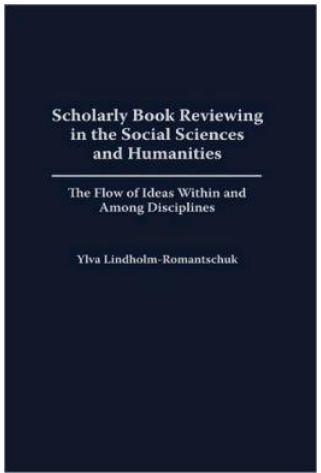




- AB Cobban, "***The medieval universities***", 1975
- *Contains footnotes, function partially as references*
- *Plus a bibliography*

- A. Desrosieres, "***The politics of large numbers***", 1998
- *Contains endnotes, which function partially as reference, partially as explanation/expansion of the main text.*
- *Bibliography labeled as References available.*



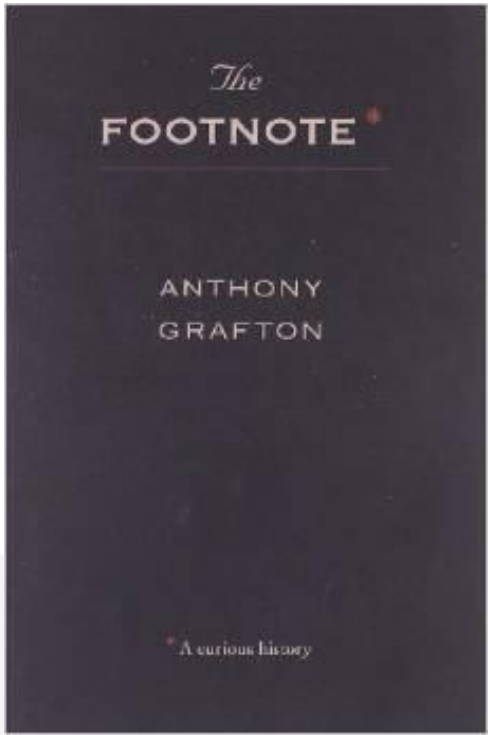


- Y. Lindholm-Romantschuk, "***Scholarly book reviewing in the social sciences and humanities***", 1998
- *Contains endnotes per chapter, references in the text to the bibliography.*
- *Bibliography available.*

Another perspective on referencing practices: *The Footnote. A curious history*

- In history, the references used indicate the novelty of the research indicated, so the number of references used that either completely new, or give a new interpretation on already used archival material is what determines the relevance of the work.

- This book focused on the famous German historian von Ranke, who set new standards in historical research.



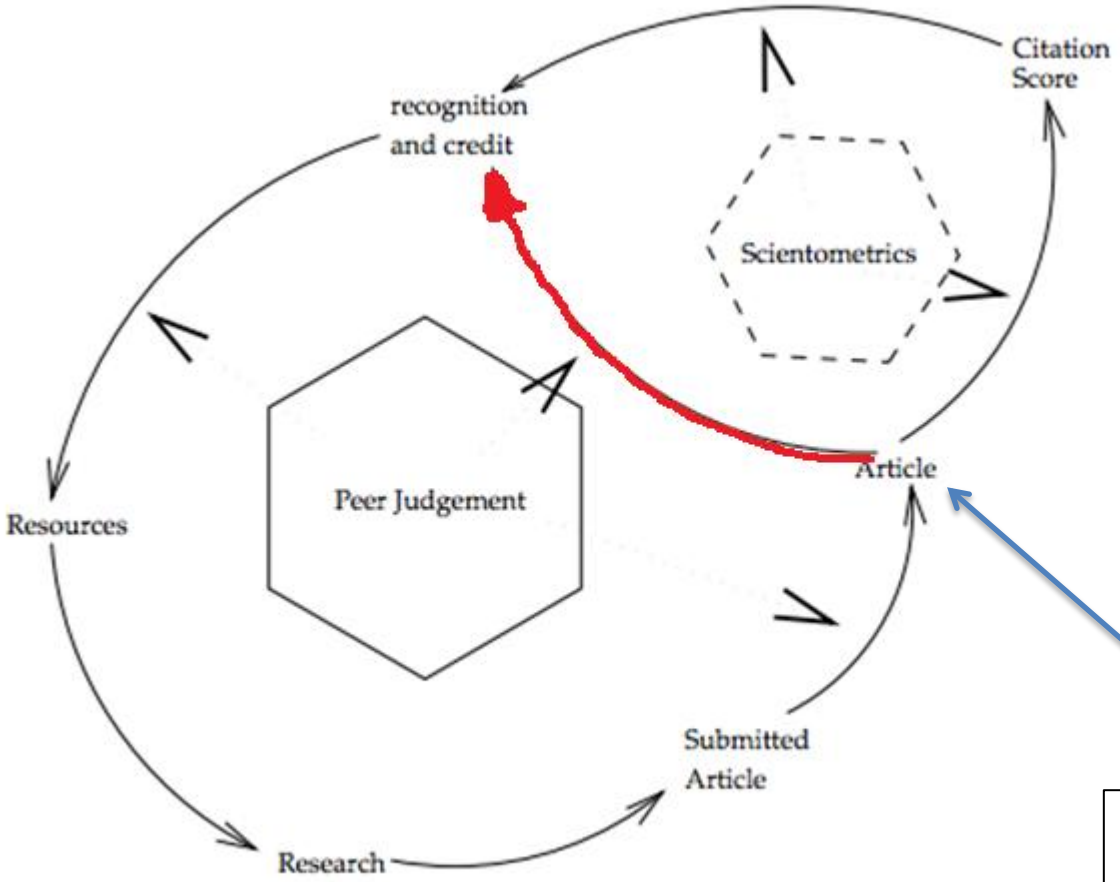
CWTS

Differences between *History* and STM

- A first difference:
  - In history, footnotes serve the purpose to attribute legitimacy to authors, footnotes normally contain new material to show the innovative character of the work presented.
  - In STM disciplines, referencing is based upon known literature, and not so much on new stuff.
- A second difference:
  - Next, the innovative character in history work is made visible through the reference and/or note system, the *value-ing* of that noting system gives credits to the authors ...
  - … while in the STM disciplines, the innovative character becomes clear later on by the references (read: citations) received, being cited determines the innovative character of published work.

CWTS

Credibility cycle in historical research



In History, the inversion of the reference into a citation does not take place in a similar way as compared to the STM domains

CWTS

Some conclusions so far …

- The wider variety of usage of source material urges to re-consider standard bibliometric techniques in SSH&L
- Absence of (received) citations is **not** an indication of absence of influence and /or quality.
- Re-orientate ourselves on the ***usage of source material*** **i**n the SSH&L, and the consequences this might have for research assessment contexts.
- Re-orientate ourselves on the ***meaning of referencing*** in the SSH&L, and the consequences this might have for research assessment contexts.

CWTS

Some consequences …

- If the findings of this exploratory research, and the conclusions from it so far, are correct, then we end up with some serious consequences:

  - *The current use of bibliometric techniques for the SSH&L domains should be considered with even more care.*

  - *There is an urgent need for data sources that cover the communication of the SSH&L domains in a broader sense.*

CWTS

Next steps …

- The findings urge us to conduct further analyses on the usage of primary material and secondary literature by scholars in the SSH&L domains.

- This could be done by:
  - *Analyzing academic products of SSH&L (think about books and chapters, but also see how that works in journal publications);*
  - *Conduct interviews with scholars in the SSH&L domains;*

CWTS

Conclusions, consequences, and next steps

Take-home messages on bibliometrics

- Ask yourself the question "*What do I want to measure ?*"
- And also "*Can that be measured ?*"
- Take care of proper *data collection* procedures.
- Then, always use *actual* and *expected* citation data.
- Apply various *normalization* procedures (field, document, age)
- Always have a *variety of indicators*.
- Always include *various elements* of scholarly activity.

- And perhaps most important, include ***peer review*** in your assessment procedures !!!

CWTS

Thank you for your attention!

Any questions?
Ask me now, or mail us
Leeuwen@cwts.nl

CWTS

Infamous bibliometric indicators

Definitions of Journal Impact Factor & Hirsch Index

- **Definition of JIF:**
  - The mean citation score of a journal, determined by dividing all citations in year T by all citable documents in years T-1 and T-2.

- **Definition of h-index:**
  - The ‘impact’ of a researcher, determined by the number of received citations of an oeuvre, sorted by descending order, where the number of received citations on that single paper equals the rank position.

CWTS

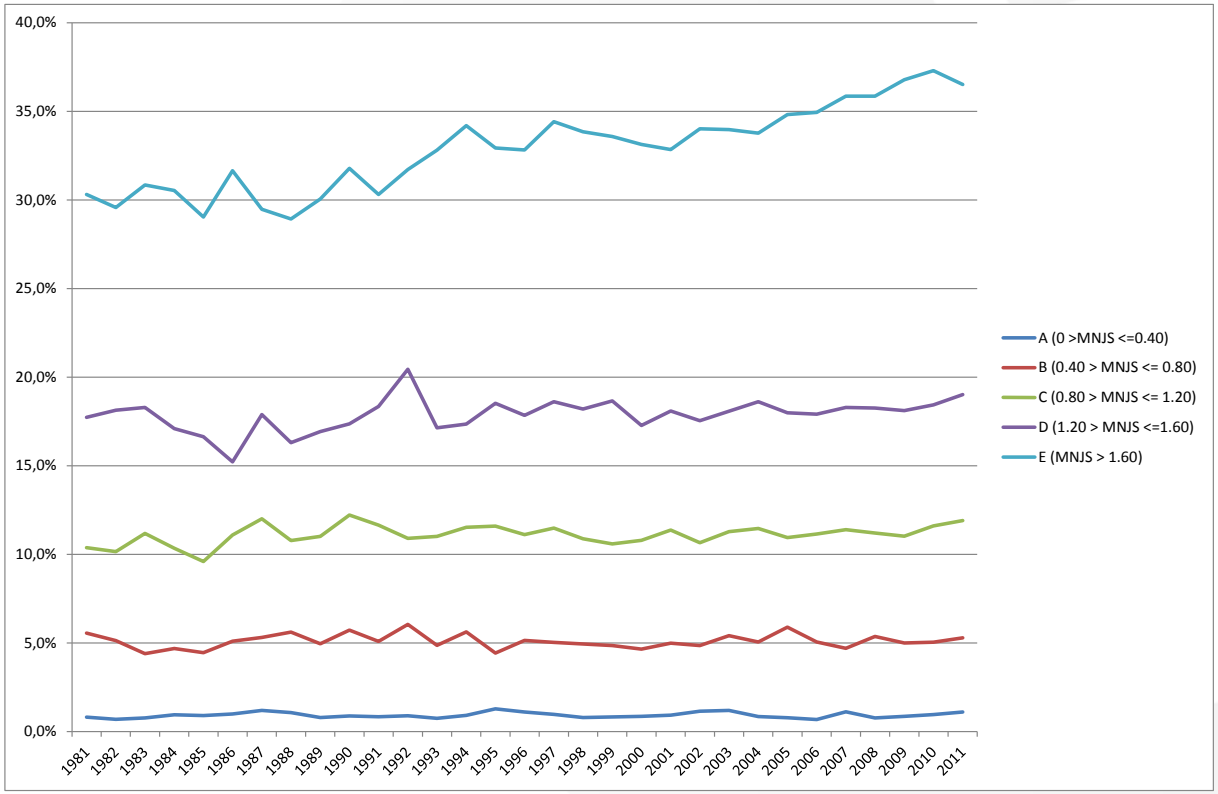
Problems with JIF

- **Methodological issues**
  - Was/is calculated erroneously (Moed & van Leeuwen, *Nature*, 1996)
  - Not field normalized
  - Not document type normalized
  - Underlying citation distributions are highly skewed (Seglen, *JASIS*, 1994)
- **Conceptual/general issues**
  - Inflation (van Leeuwen & Moed, *Scientometrics*, 2002)
  - Availability promotes journal publishing
  - Is based on expected values only
  - Stimulates one-indicator thinking
  - Ignores other scholarly virtues

CWTS

Deconstructing the myth of the JIF…

- Take the Dutch output
- Similar journal impact classes
- Focus on publications that belong to the top 10% of their field



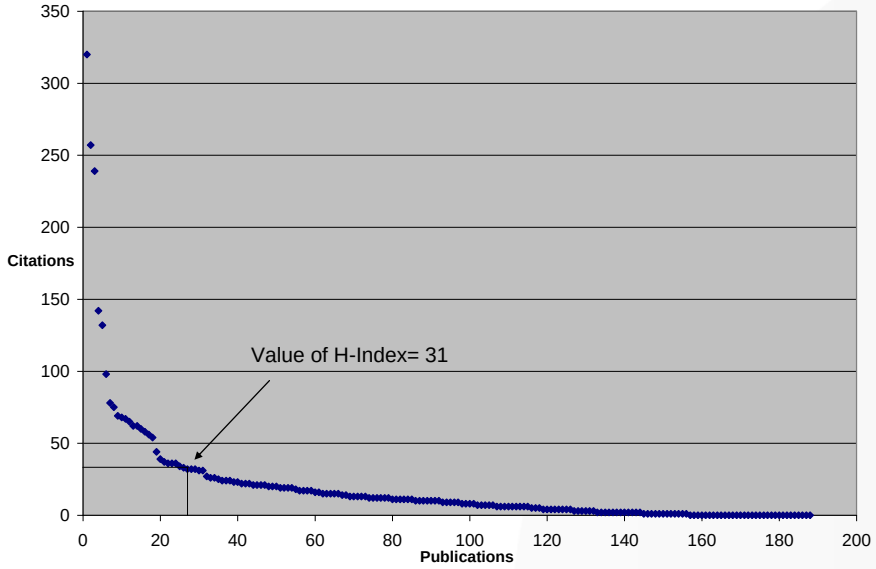
CWTS

Problems with H-index

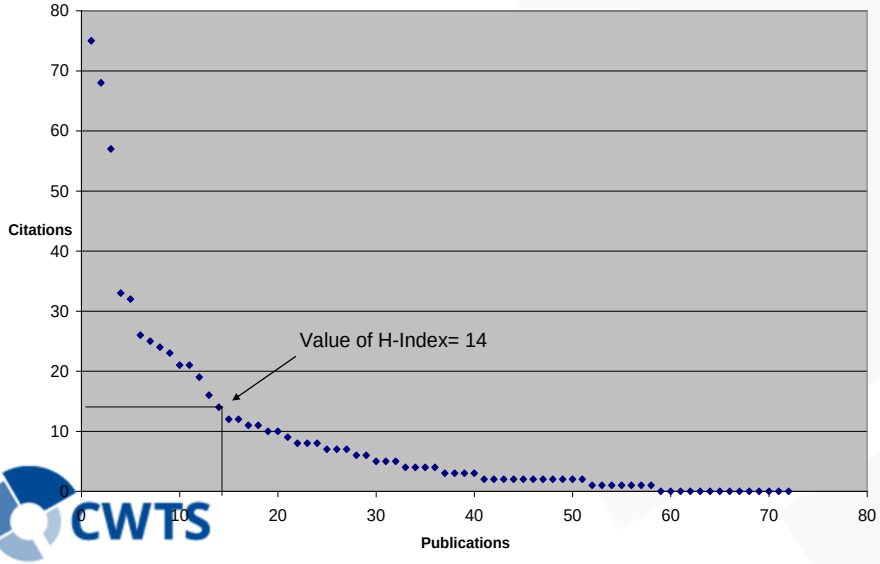
- **Bibliometric-mathematical issues**
  - mathematically inconsistent (Waltman & van Eck, *JASIST*, 2012)
  - conservative
  - Not field normalized (van Leeuwen, *Research Evaluation*, 2008)
- **Bibliometric-methodological issues**
  - How to define an author?
  - In which bibliographic/metric environment?
- **Conceptual/general issues**
  - Favors age, experience, and high productivity (Costas & Bordons, *Journal of Informetrics*, 2006)
  - No relationship with research quality
  - Ignores other elements of scholarly activity
  - Promotes one-indicator thinking

CWTS

Examples of Hirsch-index values



- Environmental biologist, output of 188 papers, cited 4,788 times in the period 80-04.
- Hirsch-index value of 31



- Clinical psychologist, output of 72 papers, cited 760 time sin the period 80-04.
- Hirsch-index value of 14

CWTS

Consistency: Definition

Definition. A scientific performance measure is said to be consistent if and only if for any two actors A and B and for any number $n \geq 0$ the ranking of A and B given by the performance measure does not change when A and B both have a new publication with n citations.

CWTS

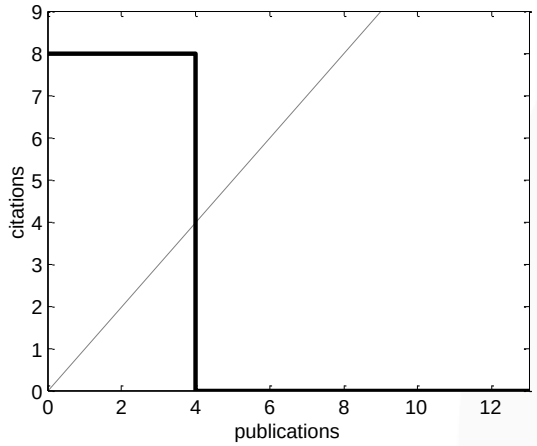
Consistency: Motivation

- Consistency ensures that if the publishing behavior of two actors does not change over time, their ranking relative to each other also does not change
- Consistency ensures that if the individual researchers in one research group **X** outperform the individual researchers in another research group **Y**, the former research group **X** as a whole outperforms the latter research group **Y**.

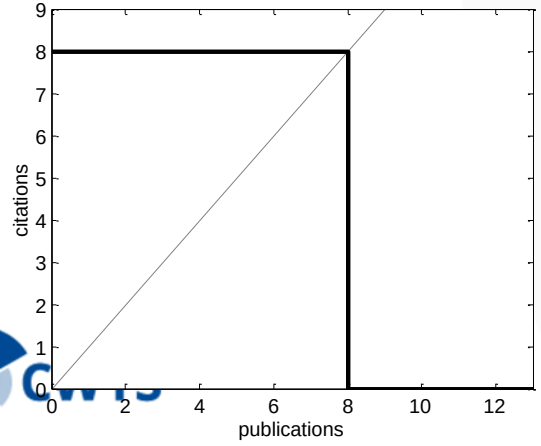
CWTS

Inconsistency of the h-index

Actor *A*

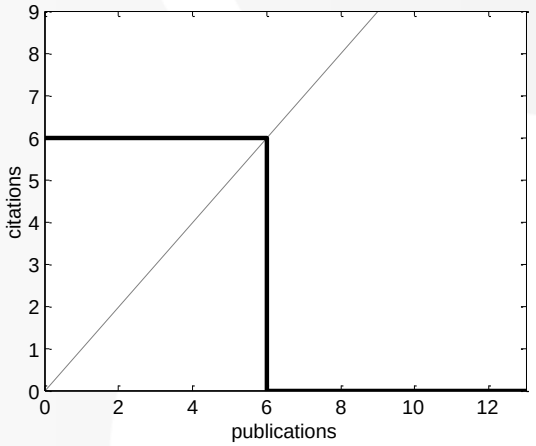


h = 4

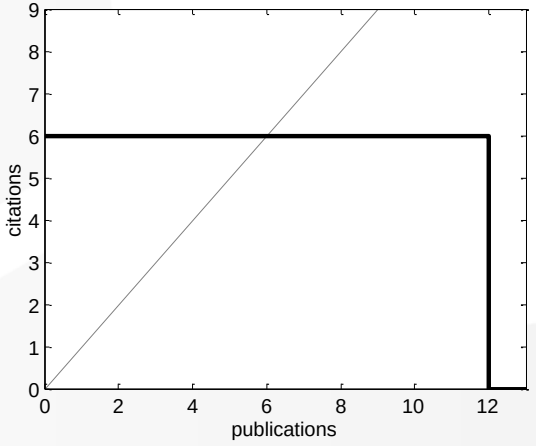


h = 8

Actor *B*



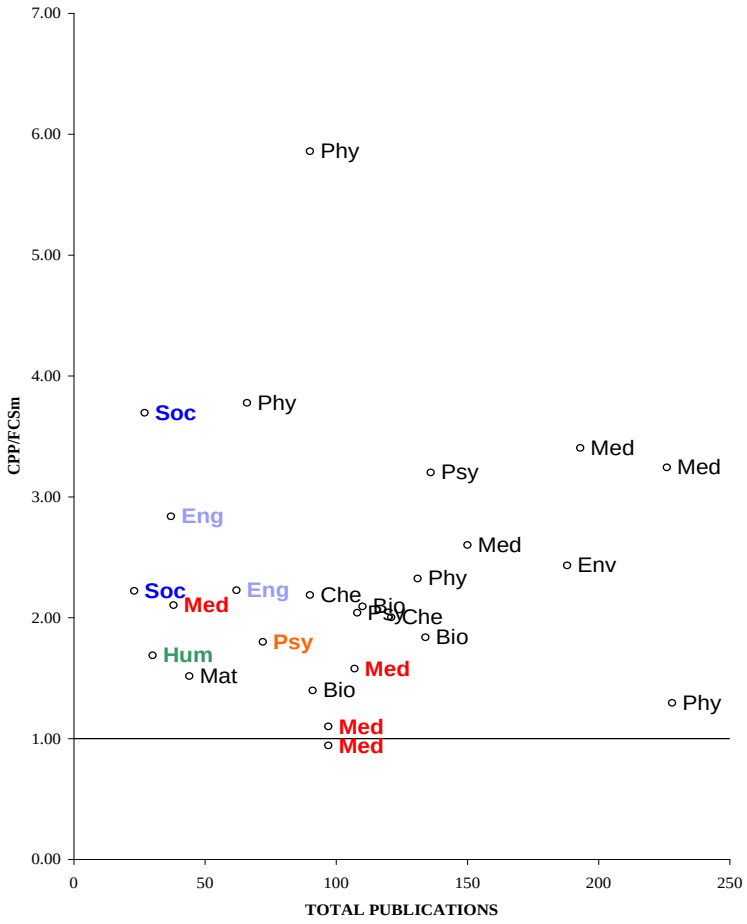
h = 6

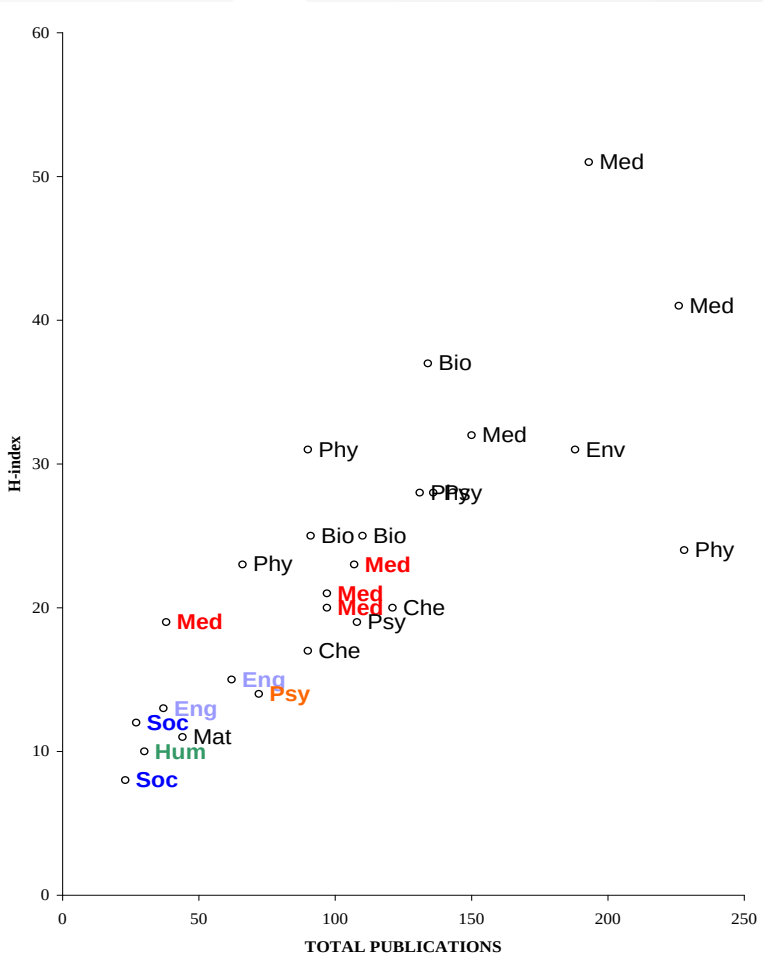


h = 6

The problem of fields and h-index …

- Spinoza candidates, across all domains …
- Use output, normalized impact, and h-index





In what database context …

Selected my own publications in WoS and Scopus, Google Scholar has a pre-set profile.

Database	H-index	Based upon …
Web of Science	14	Article s in journals
Scopus	25	Articles, book (chapters), and conference proceedings papers
Google Scholar	32	All types, incl. Reports

CWTS