Scientometric Indicators in Use: An Overview

Sybille Hinze¹, Wolfgang Glänzel²

¹Institute for Research Information and Quality Assurance (iFQ), Berlin, Germany
²Centre for R&D Monitoring and Dept. MSI, KU Leuven, Belgium

Berlin, September 2013
Structure of the presentation

1. Introduction
2. Types of indicators
3. Bibliometric indicators
   i. Document types
   ii. Counting schemes
   iii. Subject classification
   iv. Citation based indicators
Introduction

- Indicators = Proxies
  - represent a highly complex reality
  - empirically ascertainable variables and factors, that are used to reflect aspects that cannot be directly measured
  - terms are rather vaguely defined – quality, performance, progress, usefulness, importance…

- Accommodate the need for “objective“ data but, also the interest to better understand developmental processes and contexts of science itself
  - Indicators used as analytical tools but, also as information to inform science policy decisions
Introduction

„on the applied side, the demand from science policy for ‘objective’ data and specific manipulations of data (‘science and technology indicators’) is a continuous driving force for the development of quantitative studies of science and technology. (...) on the basic side, science is a complicated system of knowledge production and knowledge exchange, and the use of empirical methods in which sophisticated data-collection and data-handling techniques play a substantial role, is undoubtedly a prerequisite for the advancement of our understanding“

(van Raan 1988, p. 1)
Introduction

Systematic development of *quantitative and evaluative science studies* since the mid-20th century

- Introduction of the term “bibliometrics” by Pritchard, 1969:
  
  “the application of mathematical and statistical methods to books and other media of communication”

- Introduction of the term “scientometrics by Nalimov & Mulchenko, 1969:
  
  “the application of those quantitative methods which are dealing with the analysis of science viewed as an information process”
Introduction

Scientometrics / Bibliometrics

- depicts essential aspects of scientific activities by *quantitative* and *statistical* methods, and its output proved to be a valuable supplement to qualitative methods such as peer reviews

- has developed tools to quantify that part of research output, which is documented in the framework of scholarly communication
**Types of indicators**

**Input**
- Human resources
- Financial resources
- Infrastructure (equipment, laboratory space etc.)
- Third party funding
- Other PhD, Habilitation, Speeches, Board Memberships, Stipends etc.

**Output**
- Prizes
- Bibliometric indicators
- Patent indicators
- Performance
- Structure
- Activity
- Reception
- Cognitive
- Collaboration
- Authorship
- Citation

**Efficiency**
- Various Input / Output relations
- Inter and intra institutional comparisons

Based upon Hornbostel 1999, p. 59
Types of indicators

- Other
- Management
- Admin
  - Acquisition of third party funding
  - Supervising students and PhD candidates
- Teaching
  - Peer Review (papers, proposals, eval)
- Research

Böhmer et al. 2011, p. 129
Types of indicators

S&T resource indicators

R&D personnel
Financial resources
Infrastructures / Labs etc.

Based upon Schmoch et al. 1996, p.118
Types of indicators

Makro
- global developments
- national R&D systems
- policies
- cross-sectional fields
- research and grant programs
- academic fields
- universities, research institutes, funding agencies
- university institutes/departments
- target/status groups
- research groups
- individuals

Meso

Mikro
Bibliometric indicators

- Productivity / Activity: publication output
- Collaboration: co-authorship
- Reception / Impact: citation rates
- Cognitive structures: co-occurrences of words / classifications / citations

or the combination thereof.

Most indicators are derived from simple counts of items extracted from various bibliographies and databases. Advanced measures are “network indicators” derived from the analysis of co-authorship-, citation-, co-word- etc. networks.
Bibliometric indicators – Document types

Data sources: usually papers published in periodicals and serials.

Only conveyors of original scientific information are included. These are considered as citable items.

Citable items include research articles, short communications and notes, letters, reviews, and proceedings papers.

Book reviews, editorials, corrections/errata, meeting abstracts, and reprints are not considered original research output.
### Bibliometric indicators

#### Disciplinary coverage in the ISI citation indices

<table>
<thead>
<tr>
<th>EXCELLENT (&gt; 80%)</th>
<th>Good (60-80%)</th>
<th>Good (40-60%)</th>
<th>MODERATE (&lt;40 %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular biology &amp; biochemistry</td>
<td>Applied physics &amp; chemistry</td>
<td>Mathematics</td>
<td>Other social sciences</td>
</tr>
<tr>
<td>Biological sciences primarily related to humans</td>
<td>Biological sciences primarily related to animals and plants</td>
<td>Economics</td>
<td>Humanities and arts</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Psychology and psychiatry</td>
<td>Engineering</td>
<td></td>
</tr>
<tr>
<td>Clinical medicine</td>
<td>Geosciences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics &amp; astronomy</td>
<td>Other social sciences primarily related to medicine and health</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bibliometric indicators

Bibliometric indicators – Document types

Data sources: usually papers published in periodicals and serials. Only conveyors of original scientific information are included. These are considered as citable items:

- research articles, short communications and notes, letters, reviews, and proceedings papers.

Book reviews, editorials, corrections/errata, meeting abstracts and reprints are not considered original research output.
### Bibliometric indicators

National publication counts and percentages by document type – Web of Science (2007)

<table>
<thead>
<tr>
<th>Country</th>
<th>All papers</th>
<th>Article</th>
<th>Letter</th>
<th>Review</th>
<th>Book review</th>
<th>Meeting abstract</th>
<th>Editorial abstract</th>
<th>Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>392,488</td>
<td>66.5%</td>
<td>2.4%</td>
<td>5.0%</td>
<td>0.6%</td>
<td>5.4%</td>
<td>19.4%</td>
<td>0.7%</td>
</tr>
<tr>
<td>UK</td>
<td>104,561</td>
<td>65.9%</td>
<td>4.6%</td>
<td>5.9%</td>
<td>1.2%</td>
<td>5.5%</td>
<td>16.2%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Germany</td>
<td>95,892</td>
<td>72.3%</td>
<td>1.7%</td>
<td>4.6%</td>
<td>0.1%</td>
<td>3.2%</td>
<td>17.4%</td>
<td>0.6%</td>
</tr>
<tr>
<td>China PR</td>
<td>95,231</td>
<td>92.0%</td>
<td>0.6%</td>
<td>1.5%</td>
<td>0.0%</td>
<td>0.7%</td>
<td>4.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Japan</td>
<td>89,575</td>
<td>78.8%</td>
<td>1.5%</td>
<td>2.6%</td>
<td>0.0%</td>
<td>1.2%</td>
<td>15.4%</td>
<td>0.4%</td>
</tr>
<tr>
<td>France</td>
<td>63,656</td>
<td>77.6%</td>
<td>2.0%</td>
<td>4.4%</td>
<td>0.1%</td>
<td>2.8%</td>
<td>12.5%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Canada</td>
<td>57,500</td>
<td>71.7%</td>
<td>2.1%</td>
<td>4.9%</td>
<td>0.5%</td>
<td>3.9%</td>
<td>16.2%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Italy</td>
<td>55,223</td>
<td>72.7%</td>
<td>3.5%</td>
<td>4.6%</td>
<td>0.1%</td>
<td>2.6%</td>
<td>16.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Spain</td>
<td>41,274</td>
<td>75.9%</td>
<td>3.2%</td>
<td>4.0%</td>
<td>0.1%</td>
<td>2.5%</td>
<td>13.8%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Australia</td>
<td>35,327</td>
<td>72.4%</td>
<td>3.3%</td>
<td>5.9%</td>
<td>0.6%</td>
<td>4.0%</td>
<td>13.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td>India</td>
<td>32,842</td>
<td>86.4%</td>
<td>3.5%</td>
<td>2.6%</td>
<td>0.0%</td>
<td>2.1%</td>
<td>4.7%</td>
<td>0.6%</td>
</tr>
<tr>
<td>World total</td>
<td>129,9678</td>
<td>68.6%</td>
<td>2.8%</td>
<td>3.7%</td>
<td>0.5%</td>
<td>4.8%</td>
<td>16.6%</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

Source: Zhang et al. 2011
Counting schemes

= method according to which publications are to be assigned to contributing units

- The fractional counting scheme:
  if \( n \) units (authors, institutions, countries, etc.) have contributed to the paper in question, each contributing unit takes the value \( 1/n \) for this paper (partially additive)

- The first address count:
  a paper is assigned to one unit only, on the basis of the first address in the address list of a paper (additive)

- The full or integer counting scheme:
  assigns a co-publication fully to each contributing unit (non-additive)
Counting schemes

Nowadays only *fractional and full counting* are used while the first-address count is obsolete; in the past it was used due to the coverage of bibliographic databases, which as a rule recorded only one address.

Databases providing incomplete author address information are not appropriate for bibliometric analyses.
Counting schemes - Example

SCI CDE with Abstracts (Jan 93 - Jul 93) (D4.0)

Authors: Prassides-K Kroto-HW Taylor-R Walton-DRM David-WIF Tomkinson-J Haddon-RC Rosseinsky-MJ Murphy-DW
Title: Fullerenes and Fullerides in the Solid-State - Neutron-Scattering Studies
Address: UNIV-SUSSEX, SCH CHEM & MOLEC SCI, BRIGHTON BN1-9QJ, E-SUSSEX, ENGLAND
RUTHERFORD-APPLETON-LAB, DIDCOT OX11-0QX, OXON, ENGLAND
AT&T-BELL-LABS, MURRAY-HILL, NJ07974, USA

Source: Glänzel, Bibliometrics as a Research Field, 2003
Example: Counting options by level of aggregation

- 9 co-authors, 3 institutions, 2 different countries

<table>
<thead>
<tr>
<th></th>
<th>Full Count</th>
<th>Fractional Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prassides K</td>
<td>1</td>
<td>0,111</td>
</tr>
<tr>
<td>Kroto HW</td>
<td>1</td>
<td>0,111</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Murphy DW</td>
<td>1</td>
<td>0,111</td>
</tr>
<tr>
<td>UNIV SUSSEX</td>
<td>1</td>
<td>0,333</td>
</tr>
<tr>
<td>RUTHERFORD APPLETON LAB</td>
<td>1</td>
<td>0,333</td>
</tr>
<tr>
<td>AT&amp;T BELL LAB</td>
<td>1</td>
<td>0,333</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1</td>
<td>0,5</td>
</tr>
<tr>
<td>USA</td>
<td>1</td>
<td>0,5</td>
</tr>
</tbody>
</table>

- Apply fractional counting only within the same level of aggregation, otherwise inconsistencies will arise e.g. national versus supra-national level.

Source: Glänzel 2011
Example: Counting options by level of aggregation

Example: national versus supra-national level analysis

- 3 addresses, 3 countries: DE / FR / USA
  - National level = 1/3 for each country
  - Supra-national level (EU versus USA)
    - 2/3 : 1/3 (based on sum of individual country contributions)
    - 1/2 : 1/2 (delimitation of the regions)

Source: Glänzel 2011
# Counting schemes

## Shares of publications by country and counting method

<table>
<thead>
<tr>
<th>Country</th>
<th>Full count</th>
<th>Fractional count</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>28.0 %</td>
<td>23.9 %</td>
</tr>
<tr>
<td>Germany</td>
<td>7.3 %</td>
<td>5.4 %</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>7.8 %</td>
<td>4.9 %</td>
</tr>
<tr>
<td>France</td>
<td>5.2 %</td>
<td>3.9 %</td>
</tr>
<tr>
<td>China</td>
<td>11.0 %</td>
<td>9.9 %</td>
</tr>
<tr>
<td>Other</td>
<td>36.1 %</td>
<td>27.3 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>126.3 %</strong></td>
<td><strong>100.0 %</strong></td>
</tr>
</tbody>
</table>

Data source: Web of Science, calculations by Fraunhofer ISI 2011
Subject classification

- Disciplines / fields analysis based on classification schemes
  - Specialized databases often offer hierarchical subject classification at the document level e.g. Medline (MeSH – Medical Subject Headings), Mathematical Reviews (MSC - Mathematics Classification System)
  - Multidisciplinary databases often assign whole journals to their subject classification scheme

Due to multiple assignments of papers / journals to classification categories indicators are not additive over subject categories, sub-fields, fields etc.
Bibliometric indicators

Their use of bibliometric data to inform science policy

- National and international reporting
- Research Evaluation and Monitoring
- Identifying research priorities
Chapter 3. The scientific and technological outputs of R&D activities and their high-tech outcomes

3.1 Has the EU increased its efficiency in producing scientific publications since 2000?

3.2 Has the EU's inventiveness, as measured by patent applications, improved since 2000?

3.3 Has the EU moved towards a more knowledge-intensive economy since 2000?
National and international reporting

Figure Q-13
S&E journal articles produced by selected regions/countries: 1988-2008

Thousands:

EU
United States
Rest of world
Asia-8
Japan
China

Figure Q-14
Field shares of research articles for selected countries/economies: 2007

Source: NSF Science and Engineering Indicators 2010
### Productivity and developmental dynamics

**Publication shares of Countries and Regions 2000 to 2009 (Index 2000 = 100)**

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>98</td>
<td>96</td>
<td>94</td>
<td>92</td>
<td>89</td>
<td>87</td>
</tr>
<tr>
<td>JP</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>94</td>
<td>90</td>
<td>86</td>
<td>81</td>
<td>76</td>
<td>73</td>
</tr>
<tr>
<td>DE</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>95</td>
<td>93</td>
<td>91</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>GB</td>
<td>100</td>
<td>97</td>
<td>9</td>
<td>93</td>
<td>91</td>
<td>90</td>
<td>89</td>
<td>85</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>96</td>
<td>93</td>
<td>92</td>
<td>89</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>99</td>
<td>102</td>
<td>101</td>
<td>104</td>
<td>103</td>
<td>102</td>
<td>105</td>
</tr>
<tr>
<td>CA</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>103</td>
<td>104</td>
<td>108</td>
<td>109</td>
<td>108</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>SE</td>
<td>100</td>
<td>103</td>
<td>102</td>
<td>98</td>
<td>97</td>
<td>96</td>
<td>64</td>
<td>92</td>
<td>88</td>
<td>89</td>
</tr>
<tr>
<td>IT</td>
<td>100</td>
<td>104</td>
<td>105</td>
<td>109</td>
<td>110</td>
<td>109</td>
<td>112</td>
<td>111</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>100</td>
<td>100</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>106</td>
<td>105</td>
<td>104</td>
<td>104</td>
<td>107</td>
</tr>
<tr>
<td>FI</td>
<td>100</td>
<td>104</td>
<td>101</td>
<td>100</td>
<td>99</td>
<td>95</td>
<td>97</td>
<td>94</td>
<td>92</td>
<td>91</td>
</tr>
<tr>
<td>KR</td>
<td>100</td>
<td>117</td>
<td>126</td>
<td>142</td>
<td>159</td>
<td>165</td>
<td>170</td>
<td>167</td>
<td>182</td>
<td>194</td>
</tr>
<tr>
<td>CN</td>
<td>100</td>
<td>118</td>
<td>129</td>
<td>150</td>
<td>176</td>
<td>203</td>
<td>233</td>
<td>247</td>
<td>264</td>
<td>296</td>
</tr>
<tr>
<td>EU15</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>96</td>
<td>95</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>EU27</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>World</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Productivity and developmental dynamics

**Sharpe Ratio** – Adjusted Growth Rate

- growth indicator borrowed from financial economics/stock market analysis

\[
BW = \left( W_F - W_G \right) / S_{WF}
\]

- \( W_F \) = Growth of a field
- \( W_G \) = Growth of all fields
- \( S_{WF} \) = Standard deviation of growth of a field

Based on Fischer 2001, p. 271
Specialization / Profiles

Activity Index - AI

$$AI = \left( \frac{P_{ij}}{\sum_i P_{ij}} \right) / \left( \frac{\sum_j P_{ij}}{\sum_{ij} P_{ij}} \right)$$

Revealed Literature Advantage - RLA

$$RLA = 100 \tanh \ln \left( \frac{P_{ij}}{\sum_i P_{ij}} \right) / \left( \frac{\sum_j P_{ij}}{\sum_{ij} P_{ij}} \right)$$

- $P_{ij}$ = No. of Publications of a country $i$ in field $j$
- $\sum_i P_{ij}$ = No. of Publications of all countries in field $j$
- $\sum_j P_{ij}$ = No. of all Publications of country $i$
- $\sum_{ij} P_{ij}$ = No. of all Publications of all countries
**Specialization / Profiles**

**Revealed Literature Advantage - RLA**

Publications in WoS differentiated by Science Fields, 2009

![Graph showing RLA for various science fields](image)

Collaboration

- **Co-authorships used as proxi**
  - Partial indicator due to the fact that neither all collaboration lead to a common publication nor all co-authored papers are based on collaboration (see Katz & Martin 1997)
  - But, the higher the level of collaboration the better the approximation by „Co-publication“ (see Glänzel & Schubert 2004)
Collaboration

- Aggregation levels used in co-authorship analysis
  - Individual authors
  - Institutions - domestic
  - International collaboration – institutions and countries
  - Collaboration between sectors
Collaboration

Share of worldwide S&E articles coauthored domestically and internationally

Source: NSF Science and Engineering Indicators 2008
Collaboration

Source: NSF Science and Engineering Indicators 2010
Collaboration

Research collaboration in the field future internet – 2008-2010

Citation based indicators

- Citation used as proxy to reflect reception or impact

Basically two notions of citations have become prevalent in bibliometrics,
  — the information science related and
  — the sociological approach.

According to the first notion, citation is “one important form of use of scientific information within the framework of documented science communication”. (Glänzel & Schöpflin, *Information Processing & Management*, 1999)


Holmes & Oppenheim found that citations are not primarily a measure of quality, though they significantly correlate with other quality measures. (Holmes & Oppenheim, *Information Research*, 2001)
Citation based indicators

- 15 reasons to cite other’s work
  1. Paying homage to pioneers
  2. Giving credit for related work (homage to peer)
  3. Identifying methodology, equipment, etc.
  4. Providing background reading
  5. Correcting one’s own work
  6. Correcting the work of others
  7. Criticising previous work
  8. Substantiating claims
  9. Alerting to forthcoming work
  10. Providing leads to poorly disseminated, poorly indexed, or uncited work
  11. Authenticating data and classes of facts – physical constants, etc.
  12. Identifying original publications in which an idea or concept was discussed
  13. Identifying original publications or other work describing an eponymic concept or term
  14. Disclaiming work or ideas of others (negative claim)
  15. Disputing priority claims of others (negative homage)

Source: Garfield, Current Contents, 1970
Citation based indicators

A bibliometricians’ view

“if a paper receives 5 or 10 citations a year throughout several years after its publication, it is very likely that its content will become integrated into the body of knowledge of the respective subject field; if, on the other hand, no reference is made at all to the paper during 5 to 10 years after publication, it is likely that the results involved do not contribute essentially to the contemporary scientific paradigm system of the subject field in question.”

Source: Braun et al., Scientometric Indicators, 1985
Figure O-18

Figure O-21
Share of region’s/country’s papers among world’s most cited S&E articles: 2007
Citation based indicators

- Self citations
  - Author self-citation
    Inevitable part of scholarly communication
  - Journal self-citation
    Large share of journal self-citations hints to the fact that a journal might be highly specialized while a low share of self citations is, for example, characteristic for review journals (Schubert & Braun, 1993)

Increasing extent of journal self-citations has been reported in the context of possible manipulation of the Journal Impact Factor (Smith, BMJ, 1997; Weingart, Scientometrics, 2005)
Citation based indicators

- Factors influencing citation impact
  - Subject area or discipline
  - Age of the paper
  - "Social status" of the paper (through co-author(s) and the journal)
  - Document type
  - Observation period
Citation based indicators

- Citation window

  Rule of thumb: the larger the citation window the more reliable the results.

  But: Science policy is interested in timely analysis and reporting

  As a compromise usually citation windows between three and five years are applied.

  (see also Moed, Scientometrics 1996; Glänzel Scientometrics, 1997; van Raan, JASIST, 2006)
Standard citation indicators

The following notations are used:

- $c_i$: number of citations to paper $I_i$
- $n$: number of publications
- $x_i$: impact of journal $J_i$, where the paper is published
- $F_i$: impact of the subject $F_i$ the paper belongs to

see also Braun et al., Scientometric Indicators, 1985; Braun & Glänzel, Scientometrics, 1990; Moed et al, Scientometrics, 1995
Standard citation indicators

- Observed citations
  - Total citations (within a defined citation window)
  - Share of uncited papers
- Mean Observed Citation Rate (MOCR)

\[
MOCR = \frac{\sum_{i=1}^{n} c_i}{n}
\]
Standard citation indicators

— Expected citation rates
  — Mean Expected Citation Rate (MECR)

\[
\text{MECR} = \frac{\sum_{i=1}^{n} x_i}{n}
\]

— Field Expected Citation Rate (FECR)

\[
\text{FECR} = \frac{\sum_{i=1}^{n} f_i}{n}
\]

The ratio of these two indicators MECR / FECR reflects whether a unit publishes in higher / lower impact journals than it would be expected taken the field the unit is active in.
Relative citation indicators

— Normalized citation rate

NMCR = MOCR / FECR

Indicates whether a paper is cited above / below average compared to the field it is assigned to.

— Relative citation rate

RCR = MOCR / MECR

Indicates whether a paper is cited above / below average compared to the journal it appeared in.
References

Thank you for your attention!