Underreporting research relevant to local needs in the global south

Database biases in the representation of knowledge on rice

Ismael Rafols, Tommaso Ciarli, and Diego Chavarro

Ingenio (CSIC-UPV), Universitat Politècnica de València

SPRU (Science Policy Research Unit), University of Sussex, Brighton, UK
Observatoire des Sciences et des Téchniques (OST-HCERES), Paris
Misalignment between ST&I outputs and societal “needs”

Science, Technology and innovation “supply” poorly aligned with societal needs (Sarewitz and Pielke (2007))

- Well established in health research (Evans, Shim, Ioannidis (2014))
- Widely discussed in agriculture (Contested agronomy, Sumberg, 2012)

How are STI priorities set? Knowledge politics

Political economy of Science, Technology and Innovation

- Financial and market drivers
- Scientific / Technological opportunities
- Institutional drivers
  - (...)
  - STI Indicators, one of the factors in incentive structure
Priority Setting Type I: Between different “problems” *(Not addressed here)*

Mis-alignment between health priorities and research efforts
Evans, Shim, Ioannidis (2014)

Figure 1. 2004 global disability-adjusted life years (DALYs) and 2005 research articles categorized by 19 broad WHO disease and disability categories. This correspondence suggests the loose relationship between burden of disease and health knowledge (see Figure S1 in File S1 for the distribution of different types of articles by disease).
Priority Setting Type II:
Between different approaches or solutions to “same” problem or challenge. (Addressed HERE)

Influenza A research landscape

Is this aligned with societal needs?

Wallace and Rafols (under review)
Visualising “research portfolios” with overlays of organisations on landscapes

GlaxoSmithKline (Influenza A, 2010-12)

Welcome Trust (Influenza A, 2010-12)

U.S. Centers for Disease Control (Influenza A, 2010-12)
Pressing demands of research management and evaluation

Can S&T indicators help?

Yes, indicators can help make decisions…

- Reduce time and costs
- Increase transparency and sense of objectivity
- Reduce complexity, accessible to managers

The dream of rationality, “the science of science policy”
(De Solla Price, Garfield, 1960s….Marburguer, Julia Lane, 2000s)

**but do they lead to the “right” decisions?**

Evaluation gap (Wouters):
“discrepancy between evaluation criteria and the social and economic functions of science”
Problems, research, indicators and peripheries

- Space of problems
- Space of research
- Space of STI indicators
Problems, research, indicators and peripheries

Space of problems

Space of research

Research well illuminated by indicators
Problems, research, indicators and peripheries

**Space of problems**

**STI Peripheries:**
- Research spaces not well captured by indicators
- Research well illuminated by indicators
Streetlight effect in indicators: mistaking light with “problems”
Bias in bibliometric databases

- Web of Science is biased (English, biomed over health, bench over field…)
- Scopus has a broader coverage, but similar ranks
- In international benchmarking, major int’l organisations continue to use the main databases WoS (e.g UNESCO, 2010) and Scopus (e.g Royal Society, 2011).

What is the extent of bias?

Country and Topic bias
Databases compared

Publications on rice were downloaded from:

- **WoS** (including SCI-Expanded, SSCI, A&HCI, CPCI-S and CPCI-SSH) searching “rice” or “oryza” in the field “topic”.
  
  » **99,500 records**

- **Scopus** searching in title, abstract or keywords, i.e. TIT-ABS-KEY ("rice" OR "oryza").
  
  » **95,701 records**.

- Database **CAB Abstracts**, documents with “rice” or “oryza” were searched in title and abstract.
  
  » **227,873 records**!

Coverage bias against developing countries (rice pubs)

Ciarli and Rafols (2014)
Coverage comparison in other countries

2000-2009

- CABI
- Scopus
- WoS

Publications

Brazil, Pakistan, Iran, Nigeria, Egypt, Sri Lanka, Cuba, South Korea, Thailand, Philippines, Taiwan, Indonesia, UK, Germany, Australia, France, Italy
Publications all databases (2002-2012)
Coverage by database and research topics (2002-2012)
Findings regarding topic coverage

• Significant differences of coverage between research topics by database
  ▪ Conventional databases (WoS and Scopus) have a larger coverage of molecular biology, traditional genetics and consumption
  ▪ CABI has much better coverage about productivity, plant nutrition, plant characteristics and plant protection.

• High coverage appears to be related to
  ▪ Research interests of actors in developed countries such as seed companies, food & industry

• Lower coverage appears to be related to
  ▪ Potential interests of small farmers, local contexts.

• Need to contrast results with stakeholders (inaccurate methods)
Knowledge politics in publication databases

- Lack of coverage makes some research “Invisible”

- Uneven coverage makes production (citation?) indicators biased against certain types of knowledge

- Under-represented research is less valued in research assessment

- Research evaluation is part of the political economy, database coverage is one of the loci where there is “shaping”

- **Broadening out** data and **opening up** methods as a means to a more pluralistic research (two processes) (Leach et al. 2010)
Findings regarding country coverage

• Assumption on the stability of indicators of scientific production are incorrect (Archambault et al., 2009).

• Number of publications is very dependent on the database when one analyses low and middle income countries.

• Important result for international organisations such as FAO, IFRI or UNESCO (UNESCO, 2010) that aim to work on human development.

• Proliferation in the last two decades of journal indexing systems at the regional level, such as Scielo or Redalyc that aim to provide visibility to local journals, often in languages other than English (Chavarro, 2013)
Under-representation of locally relevant knowledge

• Local knowledge important for:
  ▪ Supporting local communities in specific contexts
    – Agriculture, health
  ▪ Global challenges need local knowledge
    – Climate change, pandemics…understanding local conditions is crucial to explaining global effects and trends.

• Mapping research landscape of a topic (science supply)
  ▪ We need a representation of the knowledge on research topics relevant for a problem.

• Conventional databases (WoS, Scopus) only have limited local literature coverage.
  ▪ How can this effect the representation of the knowledge landscape?
Uneven coverage of databases

“When comparing databases one easily forgets that each database has a different purpose. Thus, most of the subject specific databases (including CAB) aim for data completeness, whereas others like Web of Science, following Garfield’s original idea, consider only the “core” journals, which are responsible for 80% of the citations in each discipline.

Thus, it is obvious that the coverage is biased in favour of journals published in industrialised countries, because these normally have a higher impact. (…)

Considering the conclusions, it is alarming to see how often scientometric analyses are performed without even the correct choice of adequate data sources for the required purpose.”

Reviewer of an earlier version of this paper
WoS publications (2002-2012)
## Comparison of document types

<table>
<thead>
<tr>
<th>Language</th>
<th>CABI # docs</th>
<th>%</th>
<th>WoS # docs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>148577</td>
<td>71.84%</td>
<td>92554</td>
<td>94.93%</td>
</tr>
<tr>
<td>Chinese</td>
<td>20544</td>
<td>9.93%</td>
<td>490</td>
<td>0.50%</td>
</tr>
<tr>
<td>Japanese</td>
<td>13844</td>
<td>6.69%</td>
<td>2032</td>
<td>2.08%</td>
</tr>
<tr>
<td>Portuguese</td>
<td>5356</td>
<td>2.59%</td>
<td>1015</td>
<td>1.04%</td>
</tr>
<tr>
<td>French</td>
<td>3942</td>
<td>1.91%</td>
<td>560</td>
<td>0.57%</td>
</tr>
<tr>
<td>Spanish</td>
<td>3320</td>
<td>1.61%</td>
<td>307</td>
<td>0.31%</td>
</tr>
<tr>
<td>Korean</td>
<td>3018</td>
<td>1.46%</td>
<td>31</td>
<td>0.03%</td>
</tr>
<tr>
<td>Russian</td>
<td>2396</td>
<td>1.16%</td>
<td>162</td>
<td>0.17%</td>
</tr>
<tr>
<td>Italian</td>
<td>1546</td>
<td>0.75%</td>
<td>22</td>
<td>0.02%</td>
</tr>
<tr>
<td>German</td>
<td>1462</td>
<td>0.71%</td>
<td>214</td>
<td>0.22%</td>
</tr>
</tbody>
</table>