BEING A WORLD-CLASS UNIVERSITY

Bibliometric Considerations

Michel ZITT */** Ghislaine FILLIATREAU*

* Observatoire des Sciences et des Techniques (OST)
  93 rue de Vaugirard F-75006 PARIS, tel 33 (0)1 42 22 30 30
** IBIS, INRA-LERE CO, rue de la Géraudière, BP 71627
  F-44316, NANTES Cedex 03, tel 33 (0)2 40 67 51 73/4,

WCU Conference - Shanghai 2005
Context

- France, a medium power in science (5% world pub.), specific balance PROs/universities

- Need for specific S&T indicators, especially for outputs, recognized in the late eighties – creation of OST in 1990, with initial missions aiming at ‘macro’ analysis, national and regional

- Under the umbrella of the Ministry of Education and Research, a program of institutions-level shared indicators is launched in 2000, ‘The Cooperative’ with OST as a technical operator
Dimensions of Players’ activity:
Triple Helices and Compass

Leydesdorff, Etzkowitz
The ‘Research Compass’

Univ/PRO

State

Society

Univ

Industry

NGO

Callon, Laredo
1st Family of Measure: Information Output

1. Output & spectrum
2. Visibility (citations)
3. Web information
2nd Family of Measures: Networks

1. Collaborations
2. Human res. flows, mobility
... Funding

Recurrent funding
Contracts
Multidimensional activity

Research Compass on revealed activity / on missions and strategies: ‘Evaluation mix’

Indicators depending on reference sets

Productivity issues
An exercise of multidimensional evaluation
The ‘Cooperative’

- Objective: indicators developed in interaction with -- and validated by -- specialized services in PROs, based on shared frameworks

- Five packages:
  - financial resources
  - human resources
  - scientific output (publications; collaborations; visibility)
  - ‘valorisation’ (IP forms; spin-offs)
  - support to public policy (expertise)

  ➢ only a partial coverage of the ‘research compass’
1. Dependence on Data Sources

- In most benchmarking studies, ISI databases are accepted for bibliometric counts. They are excellent products in many respects.

- However, this source tells only part of the story (delineation, balance amongst disciplines, ‘biases’ of coverage).

- Corrections of coverage change performance indicators, sometimes in a spectacular way: factor 2 for Russia…

➤ Which map?
2. Dependence on Players’ Identification

- Players’ identification is not trivial, due to the well-known absence of ‘unification’ of names/addresses in ISI databases

- This is particularly true for hybrid forms which are dominant in France: joint laboratories between PROs or between PROs and universities

- In this context, self-identification of players is necessary

➢ Who is the player?
3. Dependence on Players’ Size

Bibliometric benchmarking deals with various kinds of indicators, with respect to their relation with players’ size:

- Some are mechanically dependent on the size of the player (# or world share of publications or citations)

- Some are not size-dependent, at least directly (‘impact’ measures)

- Some explore, with great difficulties, the input-output linkage (‘productivity’ measures; sophisticated techniques, but often poor data)

➢ Is size as such a performance?
4. Dependence on Field delineation / Thematic Breakdown

- Heterogeneity is the rule amongst disciplines, whatever the indicator: production, productivity, visibility through citations

- Necessity of breakdowns
  - Attempt to build references with a sensible degree of homogeneity
  - Access to production spectrum, diversity, differentiation profile of univ.
  - Normalised measures necessary to rebuild global indicators

➢ Is diversity as such a performance?
5. Dependence on Field delineation / Scale issues

- Delineation of fields: universal or tailor-made for each player? Effects on publication shares

- What is the set of reference for citation measures? Global measures and journal-level normalized measures exhibit quite different outcomes

- Impact and especially ‘excellence’ measures are scale-dependent

➢ *Is presence on most visible areas a performance?*
Citation classes - Activity profile

Institut Pasteur
Disc = med res
Citation classes - Activity profile

INRA (Agr Res)
Disc = appl bio
Top 5% impact - Activity index, ranks

Nor_Discipline   Nor_Speciality

IRD  1  1  INRIA
INRA  2  2  CNRS
CNRS  3  3  UNIVERSITE
CEA  4  4  IRD
UNIVERSITE  5  5  CEA
INRIA  6  6  INRA
INSERM  7  7  INSERM

7 signif actors
Citation classes - Activity index

University A
Disc = chemistry
Republic of Science is inegalitarian, ranking is unavoidable, but

Activity and strategies of universities cover many dimensions

‘Ranking bibliometrics’ cannot avoid the issues of data sources, players’ identification, size-dependence, variety and scale of reference sets, and ‘beyond-ranking’ issues of emergence in the scientific system

*Further commitments of OST:* recurrent monitoring indicators; the Observatory of Universities (within the Prime NoE); methodological issues
# Identification of CNRS

## IDENTIFICATION OF CNRS PUBLICATIONS (thousands) - source : UNIPS

<table>
<thead>
<tr>
<th>Publication year</th>
<th>CNRS (a)</th>
<th>CNRS (b)</th>
<th>% missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>15,0</td>
<td>20,6</td>
<td>27,2%</td>
</tr>
<tr>
<td>1997</td>
<td>15,7</td>
<td>21,7</td>
<td>27,7%</td>
</tr>
<tr>
<td>1998</td>
<td>15,8</td>
<td>22,3</td>
<td>28,9%</td>
</tr>
<tr>
<td>1999</td>
<td>16,2</td>
<td>22,7</td>
<td>28,8%</td>
</tr>
<tr>
<td>2000</td>
<td>16,0</td>
<td>23,1</td>
<td>30,7%</td>
</tr>
<tr>
<td>2001</td>
<td>15,6</td>
<td>22,9</td>
<td>31,6%</td>
</tr>
<tr>
<td>2002*</td>
<td>14,2</td>
<td>20,8</td>
<td>31,8%</td>
</tr>
</tbody>
</table>

* incomplete

CNRS (a) = identification based on 'CNRS' in the address field  
CNRS (b) = identification by CNRS indicators unit (UNIPS)
## Institutional overlaps - France

<table>
<thead>
<tr>
<th></th>
<th>CEA</th>
<th>Cemagref</th>
<th>CNRS</th>
<th>Inra</th>
<th>Inserm</th>
<th>IRD-France</th>
<th>Universities</th>
<th>Cirad</th>
<th>Inria</th>
<th>Pasteur</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEA</td>
<td>50</td>
<td>1</td>
<td>2,4</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cemagref</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13,7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRS</td>
<td>5,4</td>
<td>1,1</td>
<td>2,5</td>
<td>0,5</td>
<td>85</td>
<td>0,6</td>
<td>1,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inra</td>
<td>1,4</td>
<td>0,1</td>
<td>15,6</td>
<td>3,5</td>
<td>0,7</td>
<td>38,1</td>
<td>0,8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inserm</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>85,4</td>
<td>2,4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRD-France</td>
<td>42,1</td>
<td>4,6</td>
<td>0,8</td>
<td>33,8</td>
<td>7,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universities</td>
<td>1,6</td>
<td>0</td>
<td>64,4</td>
<td>2,1</td>
<td>16,2</td>
<td>0,3</td>
<td>0</td>
<td>0</td>
<td>0,1</td>
<td>1,1</td>
</tr>
<tr>
<td>Cirad</td>
<td>13,2</td>
<td>19,8</td>
<td>3,3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15,9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pasteur</td>
<td>46,6</td>
<td>18,2</td>
<td>44,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: OST on ISI databases
First row: 50% of CEA publications are from associated labs with CNRS, 1% with Inra, etc.
NB Different from co-authorship
effect of journal set

--- relative ctty indic --- original series SCI/CMCI ---
CTRY = IND

Y = shares - pub, qt

Y = rel impact

ranks of journal impact

thin line: publication share
thick lines: relative impact (dark: actual; grey: expected)
med. lines: citation share (dark: actual; grey: expected)
effect of journal set

--- relative ctry indic --- original series SCI/CMCI ---
CTRY = RUS

thin line: publication share
thick lines: relative impact (dark: actual; grey: expected)
med. lines: citation share (dark: actual; grey: expected)
effect of journal set

--- relative ctry indic --- original series SCI/CMCI --- CTRY= USA

\[ Y = \text{shares - publ, cit} \]

\[ Y = \text{rel impact} \]

ranks of journal impact

thin line: publication share
thick lines: relative impact (dark: actual; grey: expected)
mixed lines: citation share (dark: actual; grey: expected)
effect of journal set

pub. shares and 1st differences (original series)

CTRY=CHN

upper curve: publ.shares
lower curve: 1st differences (enlarged)
effect of journal set

--- pub. shares and 1st differences (original series) ---
CTRY=FRA

upper curve: publ. shares  lower curve: 1st differences (enlarged)
Publication Indicator for India and Brazil

SCI Corrected series

Range R = 768 through 3410
upper curve = India
lower curve = Brazil
Publication Indicator for USA and Switzerland

SCI Corrected series

Range R = 768 through 3410
upper curve = USA
lower curve = Switzerland
Publication Indicator for Germany and France

SCI Corrected series

Range R = 768 through 3410
upper curve = Germany
lower curve = France
Publication Indicator for Spain and Finland

SCI Corrected series

Range R= 766 through 3410
upper curve = Spain
lower curve = Finland