Open Access & research metrics

Establishing reliable baselines for science policy

Thed van Leeuwen, Rodrigo Costas, Alfredo Yegros-Yegros, and Ingeborg Meijer
CWTS, Leiden University

STI Conference 2017, Paris, France,
6-8th September, 2017
Outline of the presentation

• Study 1: 2014 Dutch baseline analysis
• Study 2: 2015-2017 OpenAIRE Impact analysis
• A current national discussion on OA scores
• Study 3: Finding new solutions
• Results of the new approach
• Conclusions and discussion
Pre-ambule
Study 1:
Baseline OA measurement of Dutch science

- The debate on *Openness* and *Open Access* started some years ago in the Netherlands. In the light of this debate, our Ministry of Science, Education & Culture wanted to know about the then current situation on *Open Access* publishing by Dutch academics (“nul-meting”). *

* The paper related to this research is accepted for publication by JASIST
Defining Open Access in Web of Science

Method I: By selecting OA output from the WoS desktop interface:

• Select and download OA output from WoS on internet;
• Link that to the CWTS in-house version of WoS
  – *Older papers that are backwardly labeled as OA!*

Method II: By linking the DOAJ list with the CWTS WoS database:

• Select the papers from journals on the DOAJ list;
• Use the doi’s to link both DOAJ and WoS to each other
  – *Many older papers in WoS do not carry DOI’s yet!*

Method III: By linking the DOAJ list via ISSN with the CWTS WoS database

• Use the ISSNs on both ends (DOAJ) and WoS
  – *Journals that contain only a few OA format papers are considered completely OA, and*
  – *... this applies in a backward fashion, thereby re-labeling former non OA journals to now OA format journals*
Study 2: Impact measurement of European science

• OpenAIRE is a European project consortium focusing on Openness in science, in the current phase of the study we were asked to analyze openness with bibliometric means, in other words, to measure the impact of Open Access.

• In the study we focus on FP-7 projects as visible through the OpenAIRE database. *

* The paper related to this research is in preparation
Issues with linking pre-defined OA labels from OpenAIRE to Web of Science

Data arrive pre-formatted

- No control over data structure (errors, duplicates, lacking basic bibliographic info)

OA is pre-defined

- No control over OA label definitions (Green, based on OpenAIRE preferences)
Conclusion based upon studies

- Currently OA is not well defined within the databases used for bibliometric studies (Study 1)
- A pre-defined OA labeling requires insight in the methods of constructing that label’s definition (Study 2)
- Comparing such studies is also complicated
  - National vs selected output
  - Defined in various ways
- Hampering definitions and lack of control over the raw data get bibliometricians nervous!
A recent policy debate in the Netherlands
A recent debate on OA in the Netherlands

• Only Gold OA is policy relevant, only 12% of the Dutch output is published in OA (Rathenau Institute)
  – Stated in the recent AWTI report on the state of affairs of Dutch science

• Dutch OA output amounts up to 30-40% (UU Library)
  – Reaction on that report, based upon Canadian research in support of the EUOA policy

• Actually, these two statements are both right and wrong!
The OA situation in the Netherlands

• **Claim #1** (Rathenau Institute)
  - **Right**: Gold OA is about 12-14% in 2014/2016
  - **Wrong**: ignores the initiatives for Green OA, as labeling Green OA as ‘non policy relevant’

• **Claim #2** (UU Library)
  - **Right**: the total Dutch OA output is about 43% in 2016
  - **Wrong**: as it is based upon Canadian research, which is random-sampling/harvester based
Indicators used in the study
Indicators selected for the analysis of OA in WoS

**Number of publications: P**
- The number of normal articles, reviews, and letters as processed for journals covered in the Web of Science database

**Field normalized citation impact: MNCS**
- The comparison of the real impact of a set of publications with expected citation scores, based upon output similarity in the exact same fields, years, and documents.

**Scientific cooperation is defined by mutually exclusive classes**
- $SI = \text{Single Institute, only one address mentioned}$
- $IC = \text{International Cooperation, two country names are mentioned}$
- $NC = \text{all remaining publications}$
Defining OA in Web of Science via several sources
Define sources for creation of OA labels

Data sources should comply with two criteria:

- Sources have to be sustainable
  - Data are in the public domain, without direct risk of disappearing behind a pay-wall
- Sources need to be legal
  - Inclusion in the data source should not be based on ‘illegal acts’ by individual researchers

Data sources that did not comply with the second requirement are:

- ResearchGate
- SciHub
Sources that comply with both criteria

The DOAJ list ➔ Gold OA
The ROAD list ➔ Gold OA
CrossRef ➔ Green OA
PubMedCentral ➔ Green OA
OpenAIRE ➔ Green OA

Currently, we are working on oaDOI for inclusion in the process

A next challenge is to define Hybrid OA in the database:
• Sherpa-Romeo as a source
• Identify from the database, based on bibliographic information
Updating the database and challenges ...

The DOAJ list ➔ Decided to change the contents of the list

OpenAIRE ➔ Has changed the document types re-labeling articles, to pre-prints

In updating the database we were challenged to consider changes in the sources underlying the database:

• *Remain faithful to the principles (Legal & Sustainable),*

• ... *make our outcomes reproducible, ...*

• ... *and as such follow the developments of the selected sources!*
Results of defining OA in Web of Science
Results of the OA labeling analysis

<table>
<thead>
<tr>
<th>Country</th>
<th>%OA in 2015</th>
<th>%OA in 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATVIA</td>
<td>20%</td>
<td>21%</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>BULGARIA</td>
<td>23%</td>
<td>25%</td>
</tr>
<tr>
<td>GREECE</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>LITHUANIA</td>
<td>23%</td>
<td>22%</td>
</tr>
<tr>
<td>MALTA</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>CYPRUS</td>
<td>27%</td>
<td>30%</td>
</tr>
<tr>
<td>ITALY</td>
<td>27%</td>
<td>31%</td>
</tr>
<tr>
<td>POLAND</td>
<td>27%</td>
<td>33%</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>28%</td>
<td>30%</td>
</tr>
<tr>
<td>SLOVAKIA</td>
<td>28%</td>
<td>26%</td>
</tr>
<tr>
<td>FINLAND</td>
<td>30%</td>
<td>33%</td>
</tr>
<tr>
<td>GERMANY</td>
<td>30%</td>
<td>33%</td>
</tr>
<tr>
<td>ESTONIA</td>
<td>31%</td>
<td>33%</td>
</tr>
<tr>
<td>HUNGARY</td>
<td>31%</td>
<td>35%</td>
</tr>
<tr>
<td>AUSTRIA</td>
<td>32%</td>
<td>36%</td>
</tr>
<tr>
<td>DENMARK</td>
<td>32%</td>
<td>38%</td>
</tr>
<tr>
<td>LUXEMBOURG</td>
<td>32%</td>
<td>40%</td>
</tr>
<tr>
<td>PORTUGAL</td>
<td>32%</td>
<td>36%</td>
</tr>
<tr>
<td>SLOVENIA</td>
<td>32%</td>
<td>37%</td>
</tr>
<tr>
<td>SPAIN</td>
<td>32%</td>
<td>37%</td>
</tr>
<tr>
<td>FRANCE</td>
<td>33%</td>
<td>36%</td>
</tr>
<tr>
<td>BELGIUM</td>
<td>34%</td>
<td>41%</td>
</tr>
<tr>
<td>GREAT BRITAIN</td>
<td>34%</td>
<td>44%</td>
</tr>
<tr>
<td>IRELAND</td>
<td>34%</td>
<td>39%</td>
</tr>
<tr>
<td>SWEDEN</td>
<td>34%</td>
<td>38%</td>
</tr>
<tr>
<td>NETHERLANDS</td>
<td>37%</td>
<td>43%</td>
</tr>
</tbody>
</table>

Output for EU countries:

- Cover the period 2009-2016
- WoS articles, reviews, letters
- Rather arbitrary threshold of 25%
- Color indicates penetration of OA
  - Blue, OA < than 25%
  - Red, OA >= than 25%
- Europe is becoming more OA focused!
Distinguish between Gold and Green OA

Output for EU countries:

- Green Focus is in North-Western Europe
- Gold OA focus is in many Eastern European countries
- Reasons for this:
  1) development of infrastructure in North West Europe
  2) Stronger grip of the publishing industry on Eastern European countries
Smaller EU countries:

- First time ‘proof’ of effect of OA, on the national scale, with a full set of WoS papers
- Green OA is a game changer when included in the analyses (contrary to Study I)
- International cooperation OA output reaches even higher levels
Eastern European situation

- The overall impact scores still lag behind the north west of Europe
- However, OA published output has a strong effect, as it is on worldwide impact level or above ...
- As we have seen, this is mostly based upon Gold OA output
- International cooperation OA output reaches even higher levels
Conclusions and discussion
Conclusions based upon new approach

• *The use of the WoS provides a definite universe of publications that correspond to the mainstream international journals, while at the same time allows for advanced citation analyses*

• *Our operationalization of ‘Open Access evidence’ is flexible and exportable* (to other datasets, but also to fields and countries)

• *The methodology will be based on open, transparent and accessible tools, which makes it cost-efficient, replicable and exportable to any set of publications*
Discussion

• Urgent need for reliable analyses of OA
  – Based upon trustworthy data sources:
    • Legal
    • sustainable
  – Based upon clear definitions on OA and its various appearances
  – To support science policy and research management in making decisions on OA publishing
  – On various levels within the science system:
    • National
    • Institutional
    • disciplinary

• What else do we need for expanding our OA labeling?
Thank you for your attention!
Any questions?
Ask me now, or mail me
Leeuwen@cwts.nl

Inspiration for the studies came from
Clifford Tatum, Paul Wouters,
Wouter Gerritsma, Ron Dekker
OA, international cooperation, & research impact