A new grip on publication data as essential instrument for the transformation

13th Berlin Open Access Conference, 21–22 March 2017
Dirk Pieper, Bielefeld University
Agenda

1. **Introduction: “MPDL Whitepaper” and “Pay It Forward”**
2. Data basis and methodology of the “new grip”
3. Test calculations
4. Discussion
The global perspective of the MPDL Whitepaper

Worldwide Publishing Market

Market today subscriptions: € 7.6 bn
Current worldwide spending on subscriptions # 2m

Market transformed open access: € 4.0 bn
Estimated worldwide spending on open access publications after transition € 2,000 x 2m

Number of scholarly articles possible within the current financial system € 3,800

Current price per article publication £ 2,000

Estimated realistic price per article publication

(Source: MPDL)
From macro-level to micro-level

– The **Whitepaper** has shown, that there is enough budget in the market to transform a majority of scholarly journals from subscription to OA publishing

– Transfering the approach to country and institutional levels is one of the major tasks of „National Open Access Contact Points“
Some major outcomes of “Pay It Forward”

- Less research-intensive US institutions could fund APC costs from their library journal budgets.
- Most research-intensive US research could fund APC costs with additional use of grant funds.
- The “break-even” APC for research institutions has a wide range (depending on individual subscription expenditures, various institutional journal article output, and calculated average APC).
Agenda

1. Initial position: “MPDL Whitepaper” and “Pay It Forward”
2. Data basis and methodology of the “new grip”
3. Test calculations
4. Discussion
Basic quantitative data components

- Publication output (articles, conference proceedings, letters)
- Actual share of gold OA
- Library journal subscription expenditures
- APC data
Comprehensive approach or looking at data subsets?

- Publication output of an institution can not be determined completely
- DOAJ is incomplete, ISSN usage in DOAJ is incoherent
- Data quality of library subscription expenditures has to be improved (e.g. normalized publisher mapping)
- Open APC covers mostly institutional monitored APC expenditures, not those by individual researchers
Looking at data subsets!

- Most publication output numbers depend on WoS-/Scopus data only
- Open APC has enough evidence
- OA2020 aims at transforming the majority of scholarly journals
- Concentration on those publishers, who obtain the biggest part of subscription expenditures
“If gold OA is to take place in the next few years it can only come about via the major publishers massively converting their portfolios of established journals, not via authors choosing outlets among newly started OA journals.”

The “new grip” on publication and cost data

- WoS-/Scopus raw data for calculation of publication output
- Improving precision, integrity and normalization of affiliation names by using institution codings of the “Institute for Interdisciplinary Studies of Science” (I²SoS, Bielefeld U)
- “INTACT gold OA journal title list” (compiled of DOAJ, Open APC, PMC, ROAD)
The “new grip” on publication and cost data

- Institutional journal title lists and library expenditures for journals indexed in WoS/Scopus
- Enriched Open APC data (DOI, UT, PMC-ID, ISSN-L, ...)
- Publisher names normalization
Agenda

1. Introduction: “MPDL Whitepaper” and “Pay It Forward”
2. Data basis and methodology of the “new grip”
3. **Test calculations**
4. Discussion
OA-Analytics (WoS, INTAC project, I²SoS)
Calculating OA transformation for Germany

- Publication output 2015 (WoS): 121,000 articles, 17,000 already gold OA
- Journal expenditures 2015 (German library statistics): DBS 123 (print journals) = 47,881,000 € and DBS 134 (electronic journals) = 94,730,000
- Average APC costs 2015 (Open APC): for Germany = 1,425 € (mainly pure OA), all countries = 2,153 € (mainly hybrid)
Test calculation OA transformation for Germany (2015)

- **52,000** articles to be financed (assumption: 50% german corresponding/ reprint authors from 104,000 articles)

  - 142,611,000 € / 1,425 € = **100,078**
  - 142,611,000 € / 2,153 € = **66,238**

- APC break even for 52,000 articles: **2,742 €**

- (APC break even for 104,000 articles: 1,371 €)
Conclusions from a country-level perspective

- In 2015 the library journal expenditures in Germany seemed high enough for a cost saving transformation
- Methods problems:
  - Total library journal expenditures are related to WoS
  - Estimated share of corresponding/reprint authors correct?
  - Published articles do not equally distribute over institutions and publishers
Testing the „new grip“ at institutional level: example Bielefeld U

- **Method (I²SoS):**
  - Determination of Bielefeld U publication output using WoS raw data and normalization of institutional affiliations including every publication type and articles, proceedings, letters only
  - Complete payment by Bielefeld U author, Complete payment by Bielefeld U reprint author
  - Calculation of the library expenditures for WoS journals
Testing the „new grip“ at institutional level: example Bielefeld U

- Out of 2,724 subscription titles 1,336 titles are indexed in WoS (49.05%)
- Expenditures 2015 for WoS titles: 1,282,659 €

<table>
<thead>
<tr>
<th>Year</th>
<th>Number all pub types in WoS</th>
<th>Number articles, proceedings, letters in WoS</th>
<th>Number w. UNIBI repr. auth</th>
<th>Already gold OA</th>
<th>Ø APC per Article UNIBI</th>
<th>Ø APC all countries incl hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>1,112</td>
<td>989</td>
<td>550</td>
<td>103</td>
<td>1,315 €</td>
<td>2,153 €</td>
</tr>
</tbody>
</table>
Test calculation at institutional level for WoS articles, proceedings, letters: example Bielefeld U 2015

- **886** publications to be financed
- 1,282,659 € / 1,315 € = **975**
- 1,282,659 € / 2,153 € = **596**
- APC break even for 886 articles: **1,448 €**
Test calculation at institutional level for Bielefeld U corresponding/reprint authors 2015

- **447** articles to be financed
- $1,282,659 \text{ €} / 1,315 \text{ €} = \textbf{975}$
- $1,282,659 \text{ €} / 2,153 \text{ €} = \textbf{596}$
- APC break even for 447 articles: $\textbf{2,869} \text{ €}$
Top 10 publishers for Bielefeld U authors 2014-2016 (WoS, normalized publisher names by I²SoS)
Distribution of Bielefeld U journal articles over publishers 2014-16

- Elsevier, Springer Nature and Wiley represent about 56% of the top ten publishers for Bielefeld U authors, but collect more than 80% of the subscription expenditures.
- Already 3 out of top 10 publishers are pure OA publishers (share about 14%).
- If we treat SCOAP3 as a publisher, the percentage is even higher.
- Long tail of small publishers.
Agenda

1. Introduction: “MPDL Whitepaper” and “Pay It Forward”
2. Data basis and methodology of the “new grip”
3. Test calculations
4. Discussion
Discussion

- Using WoS raw data, institution codings of I²SoS, the INTACT gold OA journal title list and evidence based APC cost data provide a sufficient data basis for calculating and monitoring OA transformation

- Challenge for libraries is to provide valid cost data for their subscriptions of WoS indexed journal titles
Discussion

- 100% data perfection cannot be achieved, but looking at data subsets is the pragmatic approach for OA transformation.
- A major task of National Open Access Contact Points is to provide data, the institutions themselves cannot collect.
- Need for standardized data submissions, automatic enrichment and normalization workflows for subscription and cost data similar to Open APC.
Discussion

- APC costs are not sacrosanct! APC prices can decrease, because the big publishers have decreasing margin costs per article
- If traditional publishers don’t use the chance of OA2020, funders and libraries could switch their institutional support to pure OA to foster OA transformation
Thanks to the INTACT team (Bielefeld UL, l²SoS and MPDL) and you for your attention!

dirk.pieper@uni-bielefeld

base-search.net
intact-project.org
oa2020-de.org (coming soon)