The role of higher education institutions in scholarly publishing and communication

Stuart M. Shieber
Welch Professor of Computer Science
Paulson School of Engineering and Applied Sciences
Faculty Director
Office for Scholarly Communication
Harvard University
Presented at the 10th Annual Munin Conference on Scholarly Publishing, University of Tromsø – The Arctic University of Norway, Tromsø, Norway, 1 December 2015.

Author contact:
Stuart M. Shieber
Maxwell-Dworkin Laboratory — 245 Harvard University
33 Oxford Street
Cambridge, MA 02138
shieber@seas.harvard.edu

This work is licensed under a Creative Commons Attribution 4.0 International License.
Stipulations

Toll-access (subscription) journal system is dysfunctional
Open-access (article-processing fee) journal system is preferable
The two roles of higher ed institutions

Provision of results to local researchers

Communication of results to the public
The two roles of higher ed institutions

**Provision of results to local researchers**

“Implement a Harvard Library collection and content development strategic plan in support of University-wide research, teaching and learning for today and the future.”

**Communication of results to the public**

“Enable effective access to the world of knowledge and data through intuitive discovery, networks of expertise and global collaborations.”
The two roles of higher ed institutions

Provision of results to local researchers

journal subscriptions and collecting

Communication of results to the public

scholarly publishing and dissemination
The two roles of higher ed institutions

Provision of results to local researchers

Communication of results to the public
Communication of results to the public

**Green open access:**
- Rights retention through open-access policies
- Repository distribution

**Gold open access:**
- Journals provide immediate online access
- Funding via author-side article-processing charges
The Harvard rights-retention open-access policy

**Default rights retention**

- “Each Faculty member grants to the university permission to make available his or her scholarly articles and to exercise the copyright in those articles.”

- “The Dean or the Dean's designate will waive application of the policy for a particular article upon express direction by a Faculty member.”

**Commitment to deposit:**

- “Each Faculty member will provide an electronic copy of the author’s final version of each article no later than the date of its publication.”
Articles posted in DASH repository per year
Good Practices for University Open-Access Policies

Edited by Stuart Shieber and Peter Suber

Produced by the Harvard Open Access Project and the Berkman Center for Internet & Society at Harvard University.

The wiki for this document, which should be considered the most up-to-date and authoritative version, is available at http://cyberlaw.harvard.edu/wiki/good_practices_for_university_open_access_policies or bit.ly/goodocs
Good Practices for University Open-Access Policies

Edited by Stuart Shieber and Peter Suber

Produced by the Harvard Open Access Project and the Berkman Center for Internet & Society at Harvard University.

The wiki for this document, which should be considered the most up-to-date and authoritative version, is available at http://cyber.law.harvard.edu/hoap/good_practices_for_university_open-access_policies or bit.ly/goodoa

Association of Research Libraries (ARL)
Australasian Open Access Support Group (AOASG)
Coalition of Open Access Policy Institutions (COAPI)
Confederation of Open Access Repositories (COAR)
Electronic Information for Libraries (EIFL)
Enabling Open Scholarship (EOS)
Harvard Open Access Project (HOAP)
Library and Information Association of South Africa (LIASA)
Mediterranean Open Access Network (MedOANet)
Oberlin Group
Open Access Directory (OAD)
Open Access Policy Alignment Strategies for European Union Research (PASTEUR4OA)
Open Access Scholarly Information Sourcebook (OASIS)
Right to Research Coalition (R2RC)
Scholarly Publishing and Academic Resources Coalition (SPARC)
SPARC Europe
UK Open Access Implementation Group (OAIG)

Communication of results to the public

**Green open access:**
- Rights retention through open-access policies
- Repository distribution

**Gold open access:**
- Journals provide immediate online access
- Funding via author-side article-processing charges
Sample university

*Publishing*: 8,000 articles per year
Sample university

Publishing: 8,000 articles per year
Open-access APC: $3,000 per article

“Subsequent reports also suggest that the costs for open access journals average between £1.5k and £2k.”

Sample university

Publishing: 8,000 articles per year
Open-access APC: $3,000 per article
Sample university

*Publishing*: 8,000 articles per year

*Open-access APC*: $3,000 per article

<table>
<thead>
<tr>
<th>Journal</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLoS Biology</td>
<td>$2,950</td>
</tr>
<tr>
<td>Science Advances</td>
<td>$3,000 ($4,000)</td>
</tr>
<tr>
<td>typical hybrid fee</td>
<td>$3,000</td>
</tr>
</tbody>
</table>
Sample university

*Publishing:* 8,000 articles per year

*Open-access APC:* $3,000 per article

*Total outlay:* $24,000,000
Sample university

*Publishing:* 8,000 articles per year

*Open-access APC:* $3,000 per article

*Total outlay:* $24,000,000

*Subscription outlay:* $5,000,000
Sample university

Publishing: 8,000 articles per year
Open-access APC: $3,000 per article
Total outlay: $24,000,000
Subscription outlay: $5,000,000

Cost increase: ~5 times
“Our Task Force has concluded that what appears at present to be the most viable route for sustaining Open Access to peer-reviewed scholarship – a model in which institutions pay for their faculty to publish in refereed OA journals – would not bring about cost savings for Cornell. In fact, taking into account the number of articles published by Cornell researchers each year and the average cost to publish a single refereed article, CUL would likely see its serial expenditures rise significantly if the library used its current subscription funds to pay for author fees instead – even in scenario in which the majority of publishers switch overnight to a producer-pays OA business model.”

The macro view

Industry revenue: $9.4 billion

Industry articles published: 1.8 million

Average revenue per article: $5,222

(Data from Outsell for 2011, via Nature News)
The macro view

*Industry revenue*: $9.4 billion

*Industry articles published*: 1.8 million

*Average revenue per article*: $5,222

(Data from Outsell for 2011, via Nature News)
1. Readers vs. writers

Serials expenditure ($) per published article
2. Who should fund?
2. Who should fund?

A Principle:
Dissemination is an intrinsic part of the research process.

Those that fund the research should be responsible for funding its dissemination.
2. Who should fund?

A Principle:
Dissemination is an intrinsic part of the research process.

Those that fund the research should be responsible for funding its dissemination.

- Granting agencies and foundations
- Other universities (on behalf of coauthors)
3. High-end is above average
4. Ceteris non paribus
4. Ceteris non paribus
Review

Publishing: 8,000 articles per year
Open-access APC: $1,000 per article
Review

*Publishing*: 8,000 articles per year

*Open-access APC*: $1,000 per article

*Not externally funded*: 25%
Review

*Publishing*: 8,000 articles per year
*Open-access APC*: $1,000 per article
*Not externally funded*: 25%
*Internal authors*: 60%
Review

*Publishing:* 8,000 articles per year

*Open-access APC:* $1,000 per article

*Not externally funded:* 25%

*Internal authors:* 60%

*Total outlay:* \(8,000 \times $1,000 \times .25 \times .6 = $1,200,000\)
Review

*Publishing:* 8,000 articles per year
*Open-access APC:* $1,000 per article
*Not externally funded:* 25%
*Internal authors:* 60%

*Total outlay:* $7,200,000

*Total outlay:* $8,000 × $1,000 × .25 × .6 = $1,200,000

*Subscription outlay:* $7,200,000
Review

*Publishing:* 8,000 articles per year

*Open-access APC:* $1,000 per article

*Not externally funded:* 25%

*Internal authors:* 60%

*Total outlay:* \(8,000 \times $1,000 \times .25 \times .6 = $1,200,000\)

*Subscription outlay:* $7,200,000

*Cost decrease:* 6 times
After MPDL analysis of spring 2015, corrected as per Ralf Schimmer, pc.
Flag icons CC by-nc-nd 3.0 courtesy of Alpak http://alpak.deviantart.com/art/European-Flags-91396569
Promoting the long-term transition

1. Our goal should be to establish an environment in which publishers are enabled to change their business model from the unsustainable toll-access model based on reader-side fees to a sustainable open-access model based on author-side fees.

2. Dissemination is an intrinsic part of the research process. Those that fund the research should be responsible for funding its dissemination.
The two roles of higher ed institutions

Provision of results to local researchers

Communication of results to the public
Allocating resources optimally

Subscriptions purchase download access

Goal:
  • Maximize downloads covered under a fixed budget
14 @ €2

ranking by downloads
ranking by downloads

14 @ €2

12 @ €6
<table>
<thead>
<tr>
<th>Rank</th>
<th>Price</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>€2</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>€6</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>€3</td>
<td>9</td>
</tr>
</tbody>
</table>

Ranking by downloads
ranking by downloads
ranking by downloads

14 @ €2

12 @ €6

9 @ €3

8 @ €2

5 @ €1

4 @ €2
ranking by downloads

14 @ €2

12 @ €6

9 @ €3

8 @ €2

5 @ €1

4 @ €2

26 @ €8
ranking by cost / download

14 @ €2
5 @ €1
8 @ €2
9 @ €3
4 @ €2
12 @ €6

36 @ €8
ranking by downloads

14 @ €2

12 @ €6

9 @ €3

8 @ €2

5 @ €1

4 @ €2

36 – 24 @ €8
Allocating resources optimally

Subscriptions purchase download access

Goal:

• Maximize downloads covered under a fixed budget = minimize regret

Method:

• Rank journals by cost per download (cpd)
• Select journals with cpd under a threshold determined by the budget
How to make subscription judgements?

- journal cost
- cost per article
- cost per citation
- cost per download
- past usage
- Impact Factor™
- where faculty publish
- faculty input
- librarian expertise
- pertinence to current research
- predicted future research needs
- prestige of requesting faculty
- frequency of champion’s librarian visits
How to make subscription judgements?

- journal cost
- cost per article
- cost per citation
- cost per download
- past usage
- Impact Factor™
- where faculty publish
- faculty input
- librarian expertise
- pertinence to current research
- predicted future research needs
- prestige of requesting faculty
- frequency of champion’s librarian visits

predicted future
cost per
download
Setting the cpd threshold

![Graph showing savings and download reductions vs. threshold (in dollars).]
Can we predict future downloads better than past history?

Mean “regret”
Can we predict future downloads better than past history?

Mean “regret”

Baseline: 47%
Past year: 2.5%
Clairvoyant: 0.99%
Can we predict future downloads better than past history?

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean &quot;regret&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>47%</td>
</tr>
<tr>
<td>Past year</td>
<td>2.5%</td>
</tr>
<tr>
<td>AR(2)</td>
<td>1.4%</td>
</tr>
<tr>
<td>Random forest</td>
<td>.99%</td>
</tr>
<tr>
<td>Clairvoyant</td>
<td></td>
</tr>
</tbody>
</table>

Mean "regret"
Can we predict future downloads without past history?

Mean “regret”
Can we predict future downloads without past history?

Mean "regret"
The two roles of higher ed institutions

Provision of results to local researchers

- optimize subscriptions by predicting future cost per download
- so as to allow canceling journals

Communication of results to the public

- open-access policies and repositories
- underwrite fair share of open-access APCs
- don't support unsustainable models