

The Dramatic Growth of Open Access: Rationale & Method

By: Heather Morrison, MLIS
 Doctoral Candidate, Simon Fraser University School of Communication
<http://pages.cmns.sfu.ca/heather-morrison/>

The *Dramatic Growth of Open Access* (DGOA)
<http://poeticeconomics.blogspot.com/2006/08/dramatic-growth-of-open-access-series.html> is a quarterly series designed to capture at a macro level the best indications of growth of open access scholarly literature and related metrics (such as open access mandate policies). DGOA is available in open data editions (see the DGOA dataverse <http://dvn.iq.harvard.edu/dvn/dv/dgoa>).

About the researcher

My approach to research holds that it is important for the reader to be aware of the perspective of the researcher, and so I disclose at the outset that I am an open access advocate, and that this informal research project, in which I aim for the greatest accuracy possible, forms a part of my OA advocacy. I am a PhD candidate at the SFU School of Communication, where my research will focus on scholarly communication and open access, including Dramatic Growth. My work on the *Dramatic Growth of Open Access* began long before I thought of applying for the PhD program.

Rationale: seeing the forest and the opportunities

The need for a macro level approach to assessing the growth of open access is apparent from my perspective, for three major reasons. First, the strong growth of open access is important to understand because of its implications for the work of scholars and those who work with scholars, including librarians and publishers. The strong growth of DOAJ came as a pleasant shock to me about 2004, when I compared the number of titles in the DOAJ with the number of DOAJ titles in our local journal knowledgebase, CUFTS, and found that we were behind by several hundred titles in a short period of time. Even though I am a *very* optimistic open access advocate (from my perspective, OA is not only necessary, it is inevitable), I am constantly amazed at the breadth and depth of the OA movement and the growth in open access materials; this is why the series is named “Dramatic Growth”. It is about this time that I wrote the first iteration of this series, a peer-reviewed article for the Journal of Interlibrary Loan, Document Delivery & Electronic Reserve, designed to alert my librarian colleagues to the extent of the material already available, and the implications and opportunities this presented for our work (Morrison, 2006).

Second, the macro view of constant growth is important to counter misperceptions about open access growth. For the institutional repository manager with an IR that is indeed growing slowly, or not at all, it is very easy to miss the big picture, which now includes a global IR movement that is growing by the millions of items every quarter. When we focus on the details, it is easy to see the occasional OA setback (as when one title reverts from open access to toll access) as a sign of failure, when the macro level view at that very moment is that the world likely saw a net increase of open access journals, even on the very day that the setback occurred; the Directory of Open Access Journals (DOAJ) is adding an average of 4 titles per day; this is a net gain, after the DOAJ weeds all of the one-off setbacks for that year.



This work is licensed under the Creative Commons

Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA

Finally, I *am* and always have been an optimistic open access advocate. Even when I first began this series in 2005, I foresaw that it would yield beautiful charts to illustrate the growth of OA, and indeed for several years now the data has been useful to create such charts to illustrate the dramatic growth of open access. Impressive though growth to date is, it is still the case that most of the world's scholarly literature remains behind toll barriers. From my perspective, what this means is that there will be a need for this series for years to come. Others may see this not as dramatic growth, but rather as painfully slow growth of open access. Fair enough; the facts are as they are (elusive to capture though they may be), but when it comes to perspectives, there is no one viewpoint that is correct, and our world is richer when there are more ways of seeing things, rather than less.

Data collection

Most data is captured from the website of the initiative in question on the date of the issue of the DGOA in question. Data has also been provided by Peter Suber and Tim Gray (of Homerton College Library). PubMedCentral article data is captured using the search which can be found on the PMC free tab of the DGOA spreadsheet. Older data has been gathered from the Internet Archive's *The Wayback Machine*.

Open access journals

The *Directory of Open Access Journals* (DOAJ) <http://www.doaj.org> has been selected as the best available surrogate for the total number of fully open access, peer-reviewed scholarly journals in the world. The DOAJ is a vetted list of fully open access, active, peer-reviewed journals. The DOAJ list is an imperfect measure of OA journals, as the vetting process tends to result in delay of inclusion of new titles, and the comprehensiveness of discovery of new titles is unknown. For example, it is not clear to me whether all Chinese open access journals (particularly those in Chinese languages) have been reported to DOAJ. The DOAJ title count does not include hybrid journals (although DOAJ *search* services do include hybrids), with some articles open access and others toll access, or journals that provide free back access. In summary, the DOAJ title count is used as a surrogate for the total number of open access journals, although DOAJ understates this number to an unknown extent.

The *Highwire Free* <http://highwire.stanford.edu/lists/freeart.dtl> collection and the *Electronic Journals Library* <http://rzblx1.uni-regensburg.de/ezeit/index.phtml?bibid=AAAAA&colors=7&lang=en> are more inclusive lists, that do include journals with free back issues as well as fully open access journals. The *Electronic Journals Library* includes journals of interest academically that are not necessarily peer reviewed. Important as peer review is in an academic context, there have always been non-peer-reviewed sources included in academic libraries for good reason (consider magazines such as *Wired* or *Adbusters*, for example), and it is important to be aware that while peer review and open access are most compatible, the universe of quality free material is much larger than the OA peer review literature reflected in DOAJ.

Open J-Gate <http://www.openj-gate.com/Search/QuickSearch.aspx> is a search service for English-language open access journals, including both peer-reviewed and non-peer-reviewed journals, with separate title counts made available for peer-reviewed titles and total titles.



This work is licensed under the Creative Commons

Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA

The PubMedCentral <http://www.ncbi.nlm.nih.gov/pmc/> title list count is included as an interesting (to me) case study. While the NIH *Public Access Policy* <http://publicaccess.nih.gov/> applies only to authors of NIH-funded authors, not journals at all, the number of journals voluntarily participating in PMC continues to grow.

Open Journal Systems <http://pkp.sfu.ca/?q=ojs> is included because this free, open source software is in use by more than 9,000 journals around the world, about half of which are fully open access. Including OJS in this series is a way of recognizing that this software has been instrumental in the dramatic growth of open access. Data is gathered by hand by Public Knowledge Project staff and/or research associates, and may be understated; as OJS is open source software, those using OJS have no obligation to report.

Open access archives (repositories) and articles

OpenDOAR <http://www.opendoar.org/>, as a vetted list of open access archives (repositories), is a standard for the number of archives (repositories). The Registry of Open Access Repositories (ROAR) <http://roar.eprints.org/> is included since I began tracking this earlier, and it remains useful for comparison purposes. Both services provide access to a great deal of growth data and charts for repositories.

As a surrogate for the number of open access articles, I currently use and the Bielefeld Academic Search Engine (BASE) <http://base.ub.uni-bielefeld.de/en/index.php>, and I have used Scientific Commons <http://en.scientificcommons.org/> in the past, and may do so again in the future (Scientific Commons' numbers are not being updated at the moment). Limitations of these services for this purpose are that the archives (repositories) searched include items with metadata only, lacking fulltext; these archives contain a variety of materials, ranging from scholarly articles and theses to data to material that is less scholarly in nature; and the extent of overlap (duplication, for example if multiple authors each submit to their local repository) is unknown. In spite of these limitations, the sheer numbers, both in size (well over 28 million items) and growth (millions per quarter), are a strong indication that collectively these repositories are full of *stuff*, even if it isn't absolutely clear what that *stuff* is.

Elusive as the total number of open access or freely available articles is, there are a number of indicators of strong growth in open access articles. As of December 2010, I began tracking Mendeley, as this is a popular service that appears to be growing quickly. Mendeley is an interesting service to study; while institutional repository managers gripe about the difficulty of recruiting content, researchers are flocking to services like Mendeley and voluntarily uploaded articles by the droves.

Several archives are tracked separately, illustrating the growth in open access articles available through OA archives. The total number for PubMedCentral (close to 2 million as of December 2010) is taken from ROAR, even though this number is an underestimate, for comparison purposes. PMC data is tracked in depth (see the third tab on the spreadsheet), in total and by NIH funding (for external and internal researchers and in total), and by time. This provides a very rough estimate of the success of the NIH Public Access policy, from a public view-point. In December 2010, I began to add some preliminary data for CIHR and Wellcome Trust, as small indicators of the value to all of expanding PMC internationally, as well as total OA articles in PMC. arXiv <http://arxiv.org/> and RePEc <http://www.repec.org/> both represent relatively



This work is licensed under the Creative Commons

Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA

well-established, mature archives; E-LIS <http://eprints.rclis.org/> is included as the major archive for LIS.

Open access policy

Data are taken from the *Registry of Open Access Material Archiving Policies* (ROARMAP) <http://www.eprints.org/openaccess/policysignup/>. ROARMAP relies on self-reporting, and so policy numbers may be understated; this is probably the case with theses open access mandate policies.

Open data

Open data (to scholarly research data, government data) is closely related to the open access movement, and also appears to be growing rapidly. The number of journal open data policies in the Open Access Directory is included as of December 2010. Other macro level metrics for open data is one area for possible future exploration.

Baseline growth of scholarly journals and articles

As reported by Ware (2006, p. 3), citing Mabe (2001), there are about 23,000 scholarly journals in the world, collectively publishing 1.4 million articles a year. The number of articles published each year and the number of journals have both grown steadily for over two centuries, by about 3% and 3.5% per year respectively. The reason is the equally persistent growth in the number of researchers, which has also grown at about 3% per year and now stands at around 5.5 million.

Data, commentary and review

Data is provided for downloading from the *Dramatic Growth of Open Access Dataverse* <http://dvn.iq.harvard.edu/dvn/dv/dgoa> (courtesy of Harvard), and posted to Google docs for easy viewing. Each issue includes a full data edition and a show growth edition (generally illustrating growth over the previous quarter and year).

The spreadsheet for the full data edition includes 8 sub-sheets (tabs):

- DGOA (Dramatic Growth of Open Access)
- DGOA Notes (detailed notes to accompany DGOA)
- PMC Free (PubMedCentral free)
- Wayback Machine Data
- PMC Individual Journals (occasional analysis of articles in selected biomedical journals)
- PMC Individual Notes (accompanies PMC Individual Journals; also instructions for subscription-based publishers wishing to evaluate their authors' compliance with NIH policy)
- Journal Collection Comparison: occasional comparison of large journal collections (e.g. DOAJ, Highwire Free, Science Direct)
- Article metasearch comparison (occasional comparison of large article search tools, e.g. Scientific Commons, Science Direct)

Commentary is posted to my blog, *The Imaginary Journal of Poetic Economics*; links to all issues of the series, and occasional between-series notes, is available here: <http://poeticeconomics.blogspot.com/2006/08/dramatic-growth-of-open-access-series.html>

The Dramatic Growth of Open Access series is not peer-reviewed in a traditional sense, except for the original article for Journal of Interlibrary Loan, Document Delivery



This work is licensed under the Creative Commons

Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA

& Electronic Reserve. However, this series is well read by many, including experts in the area of open access, who occasionally provide comments, corrections, and suggestions. Not peer review, but perhaps review by peers? This is of interest to me, as someone who sees scholarly communication as in a time of transformation. While I see traditional peer review and other forms of academic quality control as vital and not to be dismissed until better alternatives are found, I would suggest that the informal research project that is *The Dramatic Growth of Open Access* is more valuable as an ongoing quarterly series than it would be if I had stopped with the one peer reviewed article. Peer review is indeed necessary and desirable, but if we relied solely on peer review, would we be basing our knowledge on data that is largely out of date in this rapidly changing area?

Final note

This June 30, issue is a third version of the rationale and method for The Dramatic Growth of Open Access (the first issue was Dec. 31, 2010 and the second March 31, 2011).

References & Bibliography

- Björk, B., Roos, A., & Lauri, M. (2008). Global annual volume of scholarly peer reviewed journal articles and the share available via different open access options. Paper presented at the *ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing Held in Toronto, Canada 25-27 June 2008*. Edited by: Leslie Chan and Susanna Mornati. Retrieved from http://elpub.scix.net/cgi-bin/works/Show?178_elpub2008
- Björk, B., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & et al. (2010). Open access to the scientific journal literature: Situation 2009. *PLoS ONE*, 5(6)
- Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, et al. 2011 The development of open access journal publishing from 1993 to 2009. *PLoS ONE* 6(6): e20961. doi:10.1371/journal.pone.002096. Retrieved June 2011 from <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0020961>.
- Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147-162.
- Morrison, H. (2006). The dramatic growth of open access: implications and opportunities for resource sharing. *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 16:3 <http://eprints.rclis.org/handle/10760/6680>
- Ware, M. (2006). *Scientific publishing in transition: An overview of current developments*. Bristol, UK: Association of Learned and Professional Society Publishers; STM. Retrieved March 27, 2011 from <http://pdf.aandamar.com/pdf/scientific-publishing-in-transition-an-overview-of-current.html>



This work is licensed under the Creative Commons

Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA