Open Access by Numbers

RICHARD POYNDER

Few can now doubt that open access (OA) to scholarly research is set to become an important feature of the scholarly communication landscape. What is less certain is how much of the world's research literature is currently available on an OA basis, how fast OA is growing, and what percentage of the world's academic and scientific literature will be OA in the long-term.

Trying to crunch the numbers is complicated by the fact that research papers can be made OA in two ways: Researchers can continue to publish in subscription journals and then make them freely available by self-archiving them in an institutional repository (Green OA), or they can pay to publish their work in an OA journal (either a pure Gold journal or a Hybrid OA journal) so that the publisher will make it freely available for them.

OA enthusiasts like librarian Heather Morrison — who publishes a series called "Dramatic Growth of Open Access" — tend to estimate OA occurrence and growth primarily by the simple counting of things. In March, for instance, Morrison reported that there are now over 6,000 OA journals listed in the directory of open access journals (DOAJ), and implied that the number of OA articles is now growing more quickly than the number of papers being published in subscription journals. As she put it: "Data is presented that strongly suggests that the success rate for open access journals is already higher than that of subscription journals."

In the same post, Morrison argued that by counting the number of papers flagged as OA on the Mendeley research sharing service we could conclude that self-archiving had grown by 171% in the first quarter of 2011.

Counting in this way presents an upbeat picture, suggests that the world is in the process of being flooded with OA, and that universal OA is just around the corner.

Refining the counts

Critics, however, point out that simple counting is too crude when trying to measure OA. Counting Gold OA journals, for instance, is not helpful since many of them publish just a handful of papers a year, if that. Likewise, counting items that have been self-archived can be deceptive: Many records in institutional repositories will consist of metadata alone, or non-target items like presentations and other non-reviewed material.

Certainly publishers describe the incidence and growth of OA in a less upbeat manner. When I spoke to Springer's Derk Haank at the end of last year, for instance, he estimated that only around 2% to 2.5% of the world's papers are being published in Gold or Hybrid journals today. And since the total number of research papers is growing at around 6% to 7% a year, he said, OA remains "just a drop in the ocean".

In fact, predicted Haank, OA publishing will never be more than a niche activity. "I expect it to remain between 5% and 10% at a maximum," he said.

Haank did not provide an estimate of Green OA, but implied that it was relatively low. Pointing out that he would be anxious if it did become commonplace he added, "But we are such a long way from that situation today that we are very easy going about author archiving."

A few researchers, meanwhile, have been busy trying to arrive at more precise figures. When I last wrote on this topic in 2010 I spoke to a number of researchers, including Bo-Christer Björk.

Based at the Hanken School of Economics in Helsinki, Björk has undertaken several studies aimed at sizing the growth of OA, primarily Gold OA. For a variety of reasons, Björk explained, this is not an easy thing to do. Nevertheless, when I spoke to him in January 2010 Björk estimated that Gold OA was probably increasing its share of the market by 0.5% per annum.

He added, however: "I have no evidence to show any acceleration in growth. On the contrary it seems that growth has been relatively stable, after a short expansive period when BioMed Central and PLoS were founded".

"Tremendous growth of Gold OA"

Since then, Björk has taken a closer look at the many new OA journals that have been launched from 1993 - 2009, as well as the many subscription journals that have been converted into Gold journals. There has also been the rise of "mega journals" like *PLoS ONE*, now the largest peer-reviewed journal in the world, and which expects to publish 12,000 papers in 2011 alone. In the wake of PLoS ONE's success a number of *PLoS ONE* clones have recently been launched.

On June 13th 2011 Björk and colleagues published a new paper reporting an average annual growth rate since 2000 of 18% for the number of OA journals and 30% for the number of articles. This, the paper suggests, "can be contrasted to the reported 3.5% yearly volume increase in journal publishing in general. In 2009 the share of articles in OA journals, of all peer reviewed journal articles, reached 7.7%. Overall, the results document a rapid growth in OA journal publishing over the last fifteen years."

And in a note he posted on the American Scientist Open Access Forum (AmSci) Björk said that the results, "show the tremendous growth of gold OA over the past decade".

As we said, Björk's primary focus is on Gold OA. What about Green OA? This is an area that Yassine Gargouri, a postdoctoral researcher who works with OA advocate Stevan Harnad at the Université du Québec à Montréal (UQAM), has been working on for the past four years.

Gargouri's numbers suggest that between 2005 and 2010 the percentage of Green OA rose from about 15% per year to about 21%, which amounts to an increase of about 1% per year. His numbers also suggest that introducing a Green mandate (requiring all an institution's researchers to self-archive their papers) triples the yearly percentage of OA papers from the mandating institution.

Taken together with Björk's work, this would seem to suggest that around 30% of the academic and scientific literature published in 2011 worldwide may now be freely available on the Web, two thirds of it as Green OA and one third of it as Gold OA.

Can Gold alone buy OA?

Nevertheless, it remains difficult to be precise about OA numbers, and especially difficult to make accurate predictions about future growth. Like all attempts to understand and predict the world by means of numbers and statistics, much depends on how one derives them in the first place, how one crunches them, and how one subsequently interprets the results. In the case of OA, a key question

that emerges is whether Gold OA is able on its own to accelerate the growth of OA to the degree that the OA movement would wish.

Why is it necessary to fret over such things? It is necessary for a number of reasons, but above all because if OA advocates knew exactly what was happening, and why, they would be able to put their main effort into those activities most likely to achieve their goal. Vitally, they would be better able to answer a question that has plagued the movement for many years: Should the priority be given to Green or to Gold OA?

Below I publish a Q&A interview with Gargouri. With a PhD in cognitive informatics, Gargouri has also participated in projects dealing with knowledge management, semantic web applications and ontologies. He has also taught in the computer science department at UQAM.

The interview includes contributions from Harnad — a leading OA advocate and self-styled archivangelist.

Yassine Gargouri

RP: Yassine, one of the things you have been working on is trying to size OA. Based on that work what percentage of scholarly papers do you estimate are freely available on the Web today as a result of self-archiving (Green OA)?

YG: Our more recent estimates of Green OA self-archiving levels are averaging at about 20-22% of publications in a given year, somewhat higher than the longstanding 15% level where they had been hovering for years.

RP: There has been a recent uptick then?

YG: It seems to have been rising, beginning somewhere around 2006 to 2009 (see Figure 3, but note that much of this self-archiving was probably done retrospectively, not within the year of publication).

RP: To what should we attribute this uptick?

YG: There are at least four possibilities (not mutually exclusive), but we don't have the data to determine which (if any) is the cause:

(1) Maybe the percent of Green OA has not increased, but the search engines are detecting it better.

(2) Maybe the news of OA and its benefits is inspiring authors to self-archive more, for current as well as older articles.

(3) Maybe the effect of the small but growing number of funder and institutional self-archiving mandates is beginning to make itself felt (Figure 7).

(4) Maybe the effect of librarians encouraging and helping authors self-archive is making itself felt.

RP: How are your figures calculated, and what degree of accuracy can we assume them to have?

YG: We had a software robot search the web automatically for matches with the bibliographic metadata of articles published in the journals indexed in the Thomson-Reuters-ISI database.

RP: Clearly when counting records in institutional repositories you need to be careful to distinguish between full-text files and bibliographic data right? The figures you cited relate to full-text documents only right?

YG: Yes. Whether an item retrieved by the robot was an OA version of the full text was computed by an algorithm. Based on manual evaluation of a random sample of 100 articles, our robot's accuracy was 98%. (But the robot retrieved any matching article that was accessible on the web – not just those in institutional repositories.)

RP: I assume that self-archiving rates vary between disciplines. Can you say which disciplines are more inclined to self-archive (and by how much), and which ones are less inclined, and can you speculate on the reasons for these differences?

YG: Based on our Thomson-Reuters ISI sample of the bibliographic data for 110,212 articles published between 1998 and 2006, covering 11 fields, Figure 1 shows that the self-archiving rate is low in clinical medicine, chemistry and bio-medical research (3%-11%), relatively higher in engineering, health research, and biology (17%-21%) and still higher in psychology, physics, earth science, social science and mathematics (25%-33%). (The subject classifications are also derived from ISI.)

Although the relation may not be causal, we note that the percentages seem to be inversely related to the yearly number of articles published in the field: more articles, lower percentage self-archived. For example, clinical medicine, the most prolific field in terms of articles per year, has about 3% self-archiving. In psychology, earth science and social science, where articles are fewer per year, the respective percentages are around 25%, 27% and 28%.

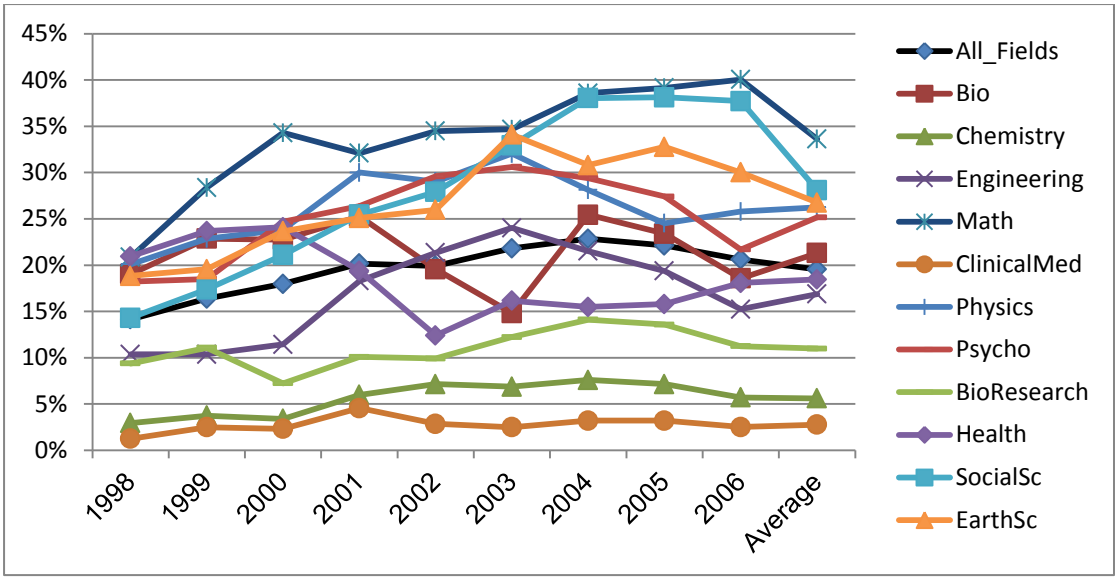

Figure 1: *Percent Green OA self-archiving by field and publication year (1998-2006), measured in 2009. Fields, bibliographic data and totals are derived from* Thomson-Reuters ISI.

RP: What is the current rate of growth of self-archiving?

YG: We have no data on self-archiving in the same year that an article is published. We have only retrospective estimates, made a few years after the publication date. The annual percentage for unmandated self-archiving hovered around 15 % in the first half of the decade and began to rise in the second half, to about 20%-22% since 2010. Overall, it looks as if the yearly self-archiving percentage was increasing from 2005 to 2010 by about 1% per year. The only systematic exceptions are institutions where self-archiving is mandatory (see ROARMAP, and the growth charts in ROAR): Mandates triple the percent self-archived (Figure 2).

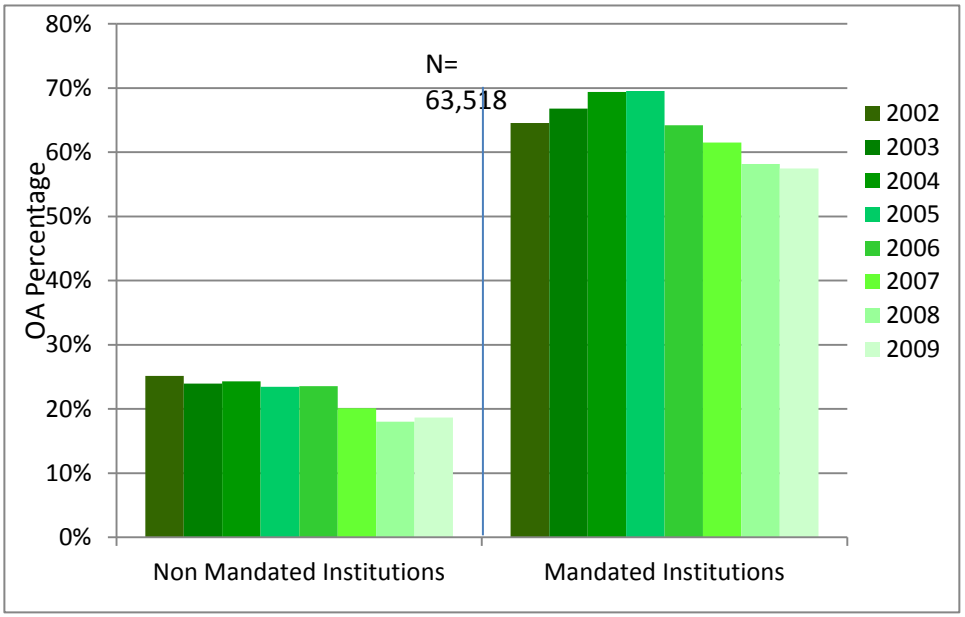

Figure 2: *Percent green OA self-archiving averaged for the four institutions with the oldest self-archiving mandates, compared to the percentage for control articles from other institutions published in the same journals (for years 2002-2009, measured in 2011). Mandates triple the percent Green OA. Respective totals are derived from Thompson-Reuters-ISI index.*

RP: Can you express current OA levels in terms of article numbers? I. e. the total number of papers currently available on an OA basis compared to the total population of papers produced?

YG: Only very approximately. These are all estimates based on samples, using bibliographic data from the ISI-indexed subset (about 10,000) of all journals. But if we ignore discipline differences and assume that the ISI-based counts also generalise to Ulrich's estimate of a total of about 25,000 refereed journals in all, and we guesstimate that ISI + non-ISI journals publish a total of 2.5 million articles per year, then 20% of that would be about 500,000 articles per year. 500,000 Green OA articles per year may sound like a lot; but this also means that 2,000,000 articles (80%) per year remain non-OA.

RP: Has the archiving rate fallen in any discipline during the period you examined?

YG: Our published study was a retrospective sample of articles published between 1998 and 2006, collected and analysed in 2008 (since extended to include articles published between 2007 and 2009, collected and analysed in 2010). In order to find out whether the archiving rate has risen or fallen in any discipline, we would have needed a sample, for each individual publication year, collected and analysed prospectively, in a series of successive years. Such a prospective study has not yet been conducted for Green OA, to our knowledge. (It is much easier to do for Gold OA than for Green OA, because the Gold OA publishers' websites provide the data immediately: no robots or successive yearly searches needed. See Figures 4-6.)

Mandating Green

RP: You indicated that self-archiving rates increase if researchers are mandated to deposit. Do you have any figures to demonstrate the extent to which a self-archiving mandate is likely to increase the percentage of papers made OA? In other words, is it possible to quantify the booster effect provided by a mandate?

YG: Definitely. Figure 2 showed that when measured in 2011, the overall yearly average percent OA across 2002-2009 was about 22% whereas for our four mandated institutions across the same period it was about 64%. So, for this sample of four institutions, their self-archiving mandates approximately tripled their percent OA.

We also found that the percent Green OA self-archiving was increasing somewhat from year to year (Figure 3), probably because of retrospective self-archiving. For publication years 2002-2006, when measured in 2009 and measured again in 2010 the percentage Green OA had increased (though, as I said before, this might reflect a combination of factors, including better detection of OA content by search engines, an overall increase in unmandated self-archiving, or an increase in the number of mandates at other institutions).

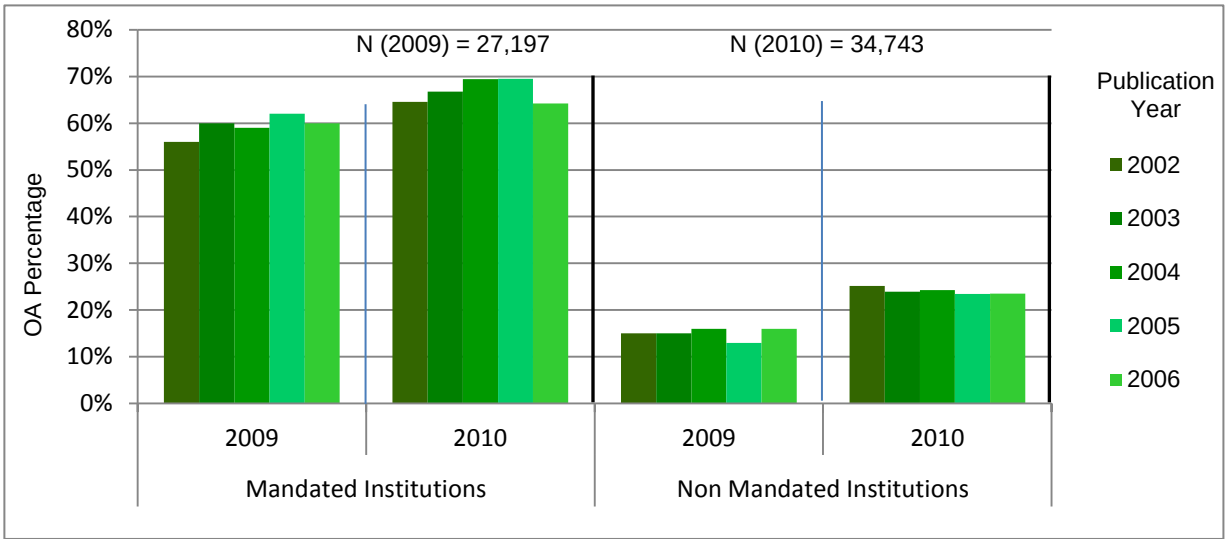

Figure 3: *Percentage Green OA self-archiving for articles published in years 2002-2006, as measured in 2009 and again in 2010. Percentage OA is about three times as high for institutions with self-archiving mandates (left) as for non-mandated control articles (right). The percentage for both increases from the 2009 measurement to the 2010 measurement. For unmandated articles the increase is about 9%.*

RP: Stevan, not all mandates are created equal, not least in the extent to which they oblige researchers to comply. Is it possible to correlate the degree of compulsion a mandate requires with the percentage of an institution's research made OA?

SH: It *is* possible, and we are now beginning to do it.

Stevan Harnad

RP: Can you say how?

SH: ROAR indexes institutional repositories by total records and deposit rates and ROARMAP indexes institutional mandates.

(1) As a first approximation, we will test the correlation of each institution's ROAR total record count (as well as its average daily deposit rate) with whether or not it has a mandate (ROARMAP).

(2) The next approximation will be to test the same correlations, but with the mandates classified by their effectiveness and their age (University of Liege's ORBi, with an ID/OA mandate, has the most effective mandate, with the top deposit rate out of 1,414 institutional repositories in the 10-100 deposits per day range). The most effective mandates require immediate unconditional deposit; the weaker mandates allow delays and opt-outs. A mandate's percentage success increases with age. (Our data, above, come from the four oldest mandates.)

(3) The next approximation will be to refine and weight both the total yearly record count and the yearly deposit rate with an estimate of (1) institution size (2) institution research output, (3) proportion of deposits that are full-text and (4) proportion of deposits that are OA's specific target (refereed research).

This will give a much richer and more nuanced picture of the relation between records/rate and mandate (and mandate-type).

(By the way, ROAR and ROARMAP are open access databases. Analyses like the ones I've sketched above can be done much more promptly, thoroughly and extensively if they are done in parallel by multiple research groups. All interested parties — including sceptics! — are invited to data-mine these two rich, growing and increasingly important data gold-mines along the lines I've just described — as well as to design and conduct any further measurements that they might think useful and informative.)

RP: I'm conscious that in addition to adopting a mandate, a number of institutions — including the University of Liège and Edinburgh Napier University — have introduced a policy in which performance evaluation for the purpose of promotion and tenure is limited to works on deposit in the institutional repository. In other words, any papers a researcher does not deposit in his IR will simply not be considered during performance evaluation. You say the University of Liege has the most effective mandate and that this has seen it achieve the top deposit rate of its type. But I think you are saying that you are not currently able to quantify the effectiveness of that type of mandate?

SH: Not yet, but we're working on it. This is really a question about (2) and (3) above: such analyses can and will be done now. (ORBi's top rank suggests that making deposit the mechanism for submitting papers for performance review might be an important feature: we will have to try to tease apart institutions with ID/OA mandates that are and are not linked to performance review, but the sample is still small and young.)

Waiting for Gold

RP: Yassine, your figures relate to self-archived papers alone. What estimates of Gold OA are available to help us arrive at a total figure for OA?

YG: We are not the ones to whom this question needs to be addressed. Our studies systematically exclude articles published in Gold OA journals because our primary objective was not to quantify percent OA but to test the OA impact advantage by comparing the citation counts of self-archived (Green OA) and non-self-archived articles within the same (*non-OA*) journals. The percent Gold OA in our own samples is hence low (about 2% of our ISI sample).

Having said that, Bo-Christer Björk and colleagues have done the most comprehensive estimate of the respective percentages of Green OA, Gold OA and non-OA by field (see Figure 4, derived from

Björk et al's Table 3). For publication year 2008 (measured in 2009), Björk et al found about 20% OA overall, one third of it Gold and two-thirds of it Green in ISI-indexed journals (hence, the "core" journals) and the reverse (two-thirds Gold, one-third Green) in non-ISI journals. They also have break-downs by discipline, with biomedicine heavier on Gold than Green and the reverse in all other fields. This too may be related to the discipline differences we found for the percentage of Green OA in our sample (Figure 1).

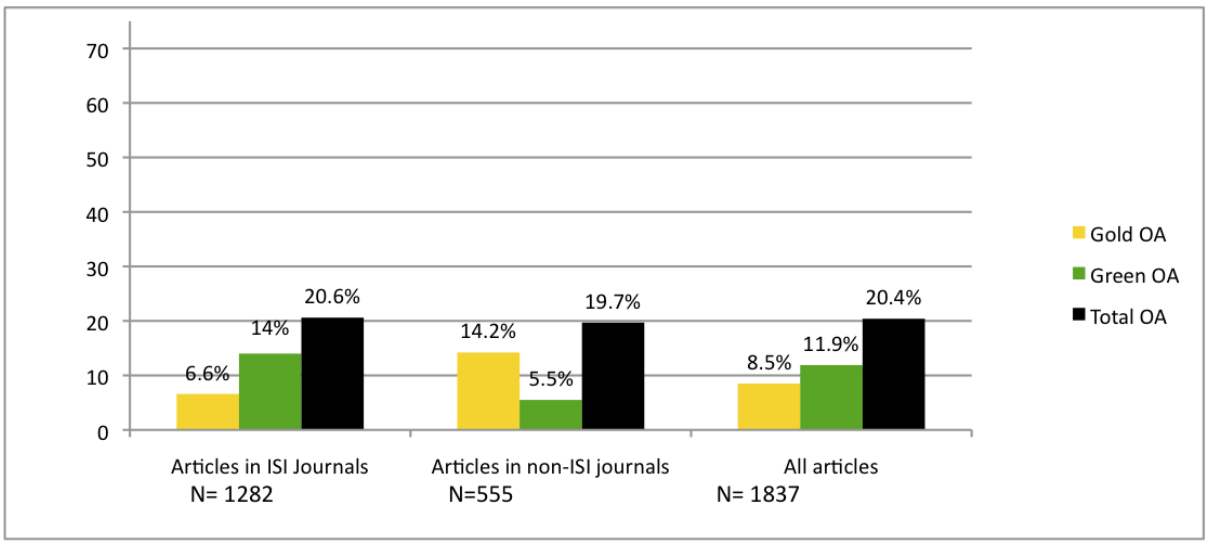

Figure 4. *Percentage Green and Gold OA for 2008 articles, measured in 2009, for ISI and non-ISI journals. (Figure derived from Table 3, Björk et al 2010.)*

RP: Putting your findings alongside Björk's do you think it would be fair to say that around 30% of the world's academic and scientific literature is freely available on the Web today?

YG: Yes.

RP: Is it possible to say anything about the rate of growth of Gold OA, and thus of the overall OA growth rate?

YG: We have no data on Gold OA growth, but I think you were recently given an estimate by Springer?

RP: Correct. When I spoke to Springer CEO Derk Haank at the beginning of the year he said, "Currently the industry publishes about 2% of all articles under the OA model, although uptake is increasing rapidly so it might be 2.5% in 2010." He added, "I expect it to remain between 5% and 10% at a maximum." He was talking about Gold rather than Green OA, but does the data currently available support his claim that Gold OA is likely to remain at between 5% and 10%?

SH: We have not tested this. Springer has since updated its estimates for ISI-indexed journals (Figure 5). According to their data, the Gold OA growth is compounding at 20% per year for ISI journals. This means that even in 10 years from now, in 2020, percent Gold OA will only be about 27% for ISI journals. We have already waited 10 years for OA: The prospect of having only 27% Gold OA (for ISI journals) in 10 years from now is nothing to trumpet triumphantly about.

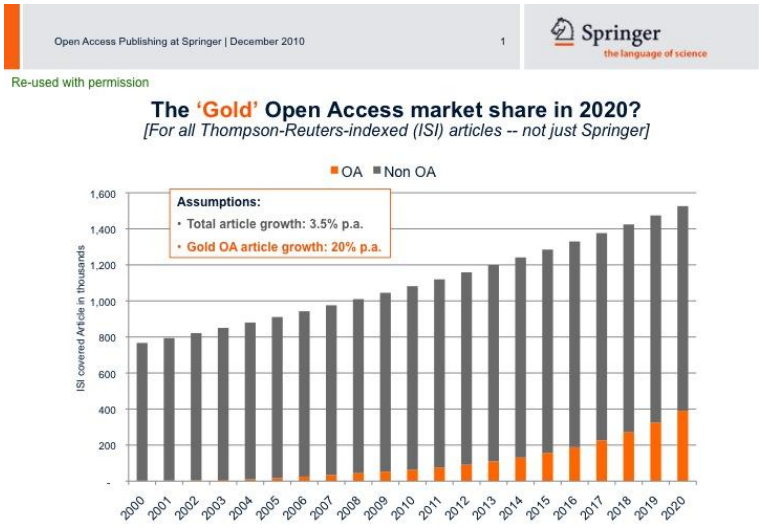

Figure 5. *Projections of Gold OA Growth for ISI-indexed journal articles (data from Springer publishers). Growth will reach 27% of all journal articles by 2020.*

But if Green OA mandates by institutions and funders grow, that could be a game-changer (Figure 2), first ushering in universal Green OA and then hastening the transition to Gold OA (Figure 7).

(The Springer estimates are only for ISI-indexed journals. Those are the top 10,000 journals. But according to Ulrich's there are about another 15,000 journals that are not indexed by ISI. Laakso & Björk's latest data suggest that the Gold OA growth rate amongst those non-ISI journals might be faster, yielding an overall Gold OA growth rate of 30% across all journals, ISI + non-ISI, from 2000-2005). But there's an increasingly blurry boundary between lower-standard peer-reviewed journals and non-peer- reviewed journals...)

Counting on Gold

RP: Do you think there is a tendency for some OA advocates to exaggerate progress (making repeated claims, for instance, that OA is undergoing dramatic growth)? If some claims are indeed exaggerated is there a danger that they could prove counterproductive?

SH: It is definitely counter-productive to exaggerate progress -- especially OA progress, which has been so slow in coming: It just amounts to crying wolf, and the effect is that when there is a genuine sign of progress, people ignore it, as just another false alarm. Whereas if the empty and superficial figures — mostly gold OA journal counts — were not being regularly trumpeted triumphantly, then news of the occasional genuine milestone (the UK select committee recommendation, the NIH, RCUK and EC mandates, the Harvard and MIT mandates) might have a chance of inspiring further acceleration.

RP: What's wrong with counting gold journals, which as you say is where most of the claims about dramatic growth come from?

SH: There is nothing at all wrong with counting gold journals. But once you reckon it as the percentage of all journals published each year (as both Björk and Springer quite properly did) rather than just as absolute numbers, neither the percentage nor its growth rate turns out to be anything to crow about.

The truth, after all, is that although OA itself (whether Gold or Green) is optimal and inevitable, its progress has been appallingly slow, even though it has been fully within reach for well over a decade now.

OA is coming, but it is coming far, far too late to warrant any triumphalism. We should rather hang our heads in shame for having been so sluggish in grasping what has been so long within our reach, despite its enormous benefits.

OA Advantage

RP: Yassine, I believe you have also done some research on the so-called OA Advantage. This is a controversial area, and some studies appear to suggest that there is no such thing as an OA advantage, but how would you characterise your findings on the OA Advantage?

YG: According to our own recent study – as well as the vast majority of all studies to date – OA articles are both downloaded and cited more than non-OA articles. This is called the OA impact advantage.

We have shown that for citations the advantage is statistically significant and independent of other factors that increase citations (such as article age, journal impact factor, number of authors, number of pages, number of references cited, Review article, Science, USA author).

All these other variables are correlated with increased citation counts, so the fact that OA continues to correlate significantly with an independent positive increase in citation counts, even when the contributions of all these other correlates are calculated independently, suggests that the OA Advantage is not just a bias arising from either a random or a systematic imbalance in the other enhancers of citations.

RP: Other researchers, including Phil Davis, have argued that the phenomenon is simply a product of "author bias", or "self-selection bias". This argument, I think, says that authors choose to make their best papers OA — papers, that is, that would attract higher citations anyway.

YG: That's exactly the hypothesis we tested, and showed to be wrong, in a much bigger, longer and broader sample than Davis's: He found a download advantage but no OA advantage when the OA was assigned randomly instead of by author self-selection. We found that the OA citation advantage is just as great when the OA is mandatory as when it is self-selective (non-mandated).

This makes it highly unlikely that the OA advantage is either entirely or mostly the result of an author bias toward selectively self-archiving higher quality – hence higher citeability – articles (though self-selection is probably one of the lesser contributing factors too).

RP: Is there any other data available that could be said to support those who argue that OA (Green or Gold) is desirable — because, for instance, it has a positive effect on research progress and/or on researchers themselves?

SH: First, increased citations themselves have a positive effect on researchers' careers, performance evaluations and funding. There are also data on the OA download advantage. And Alma Swan has a large file of testimonials about other OA advantages to research and researchers. We will also soon be analysing data on industrial uptake of OA (vs. non-OA) research.

Eventually there will also need to be studies on educational uptake of OA content. And there are the important studies of John Houghton, on the substantial economic benefits of OA – and especially Green OA.

RP: There are also benefits to the public I assume?

SH: Well, the special case of health-related research is obvious: The general public as well as health care practitioners benefit from OA there. Publishers implicitly acknowledge this, e.g., Springer's recent OA release of articles relevant to the Fukushima disaster[1] and lately also about the European E. Coli problem.

But the real point is that the benefits of having open access to research findings are there, potentially, whether or not there is an acute crisis, or even when it is not a question of public health and safety. Research progress itself, and researchers' productivity, benefit from researchers having maximal access to one another's findings. That is what maximises the return on tax-payers' investment in research of all kinds.

Priorities

RP: Would you say that the figures Yassine cites about the current percentages of Green and Gold OA, and current growth rates, offer any insight into where OA advocates should be putting their main efforts today?

SH: They do indeed (although it was already evident without the data): The current yearly percent Green OA is higher than Gold OA but both are growing far too slowly. The growth rate of Gold OA is in the hands of the publishing community, but the growth rate of Green OA is entirely in the hands of the research community. Research institutions and funders merely need to mandate Green OA. That's where OA advocates need to focus their efforts.

RP: There are those who argue that it is important to support Gold and Green equally. You I think believe that this could slow progress. Why?

SH: My own longstanding (and so far largely ignored!) hypothesis has been that as long as most journals are subscription journals — meaning that institutions' serials budgets are committed to paying their yearly subscriptions, because there is no other way the institutions' users can get access to the articles in those journals — the growth of Gold OA publishing will be slow (as Springer's chart indicates): there will be no appreciable acceleration in Gold OA growth.

The Gold OA acceleration — a very big and rapid acceleration — can and will come only after (and because) the authors' versions of all or most of those subscription journal articles have been made accessible via Green OA, thereby allowing institutions to cancel their subscriptions, which in turn releases their subscription money to be used instead to pay publishers the (much-reduced) Gold OA fees for the service of refereeing their outgoing articles — instead of paying for access to the incoming refereed articles from other institutions:

It is only then, once we have got past the sticking point, that most or all publishers — under the joint influence of the stick of cancellation pressure and the carrot of the windfall funds suddenly available to pay Gold OA fees —will rapidly convert to Gold OA.

[1] The Fukushima initiative was coordinated through the NIH's Emergence Access Initiative (EIA) and included other publishers.

Everybody who is today obsessed with Gold OA has been systematically overlooking this crucial sticking point, which is that *the money to pay for Gold OA is not available till we have 100% or near-100% Green OA to release it.* Without that, all we have is the current slow growth of Gold OA (dependent on the availability of spare, scarce extra funds from other sources, such as research itself) that would take till 2029 to approach 100% Gold OA (according to Springer's ISI estimate – 2022 according to the *Laakso et al 2011* estimate; see Figure 6).

And meanwhile research usage, impact and progress continue to be lost, day after day, year after year, as they have been for the past decade and a half — completely needlessly, because mandating Green OA is fully within the global research community's immediate reach, and has been ever since the onset of the online era. I've been urging that we grasp what is already within immediate reach in order to stanch the needless daily research access/usage/impact loss. But if the foolish, foolish research community instead needs to see it all in terms of a transition to Gold OA (rather than just a transition to OA), then, fine, so be it: let them see that *it is mandating Green OA that will induce (and fund!) the transition to Gold OA*.

So far, the research community has displayed an embarrassing credulousness when confronted with the following kind of doomsday FUD from the publishing community:

> "If you mandate Green OA, you will destroy publishing! Settle instead for the status quo — or buy into one of our Gold OA offers — and wait patiently for an 'orderly' transition when the 'market' decides it. Or get your research funders to come up with the extra money to pay for Gold OA now, and we'll offer it to them now! Meanwhile, make do with your current on-going access/usage/impact losses — which Davis, in the sole scientifically rigorous study, has shown to be non-existent or negligible in any case — consoled by the knowledge that this is the only way to preserve the refereed publishing industry that has served you so well for centuries. The alternative is chaos and anarchy and the destruction of all you hold dear."

2022 or 2029?

RP: Björk has just published a new paper on the growth of OA. I wonder if perhaps his findings challenge your argument. The abstract of his paper says, "Since the year 2000, the average annual growth rate [of OA] has been 18% for the number of journals and 30% for the number of articles. This can be contrasted to the reported 3.5% yearly volume increase in journal publishing in general. In 2009 the share of articles in OA journals, of all peer reviewed journal articles, reached 7.7%."

This appears to contrast with what Björk told me in January 2010, when he estimated Gold OA to be increasing its overall share by around 0.5% a year. He also said, "I have no evidence to show any acceleration in growth. On the contrary it seems that growth has been relatively stable, after a short expansive period when BioMed Central and PLoS were founded".

It would seem that things may have changed. What in your view does Björk's new paper add to our understanding of the growth of OA, what new data does it provide, and how do those data compare with your data and with Springer's data?

SH: Björk's new data (*Laakso et al 2011*) are very timely and interesting. Springer's data are based exclusively on the ISI-indexed subset of all journals, which is arguably the top 10,000 of the planet's total of about 25,000 peer reviewed journals, and includes the "core" journals that most subscribing institutions and their users most need and want. Björk's new data are based on sampling all journals,

but much more of the non-ISI journals (519, “small to medium-sized journals”) than the ISI journals (44, “large journals”).

It is not yet clear what the difference between Björk’s new figures — 18% growth rate for Gold OA journals and the 30% growth rate for Gold OA articles — means. Gold OA journals had formerly published fewer than the average number of articles per year. It has to be analysed whether their 30% article growth is partly catch-up with the average of non-OA journals (1993-2005), or whether submissions to OA journals are really growing faster than to non-OA journals. One contributing factor may be *PLOS ONE*, which is now a single Gold OA mega-journal, publishing more articles than any other journal.

But if we were to take Björk’s estimate of 30% annual Gold OA article growth for all journals since 2000 on a par with Springer’s estimate of 20% annual Gold OA article growth for ISI journals, then there would be a bit of a contradiction, because Springer’s 20% estimate would reach only about 38% gold OA among the ISI journals in the year 2022, whereas Bjork’s 30% estimate would lead to 100% OA among *all* journals (ISI + non-ISI) in 2022 (see Figure 6, drafted by Yassine to illustrate both the Springer ISI growth curve and the *Laakso et al 2011* ISI + non-ISI total growth curve)!

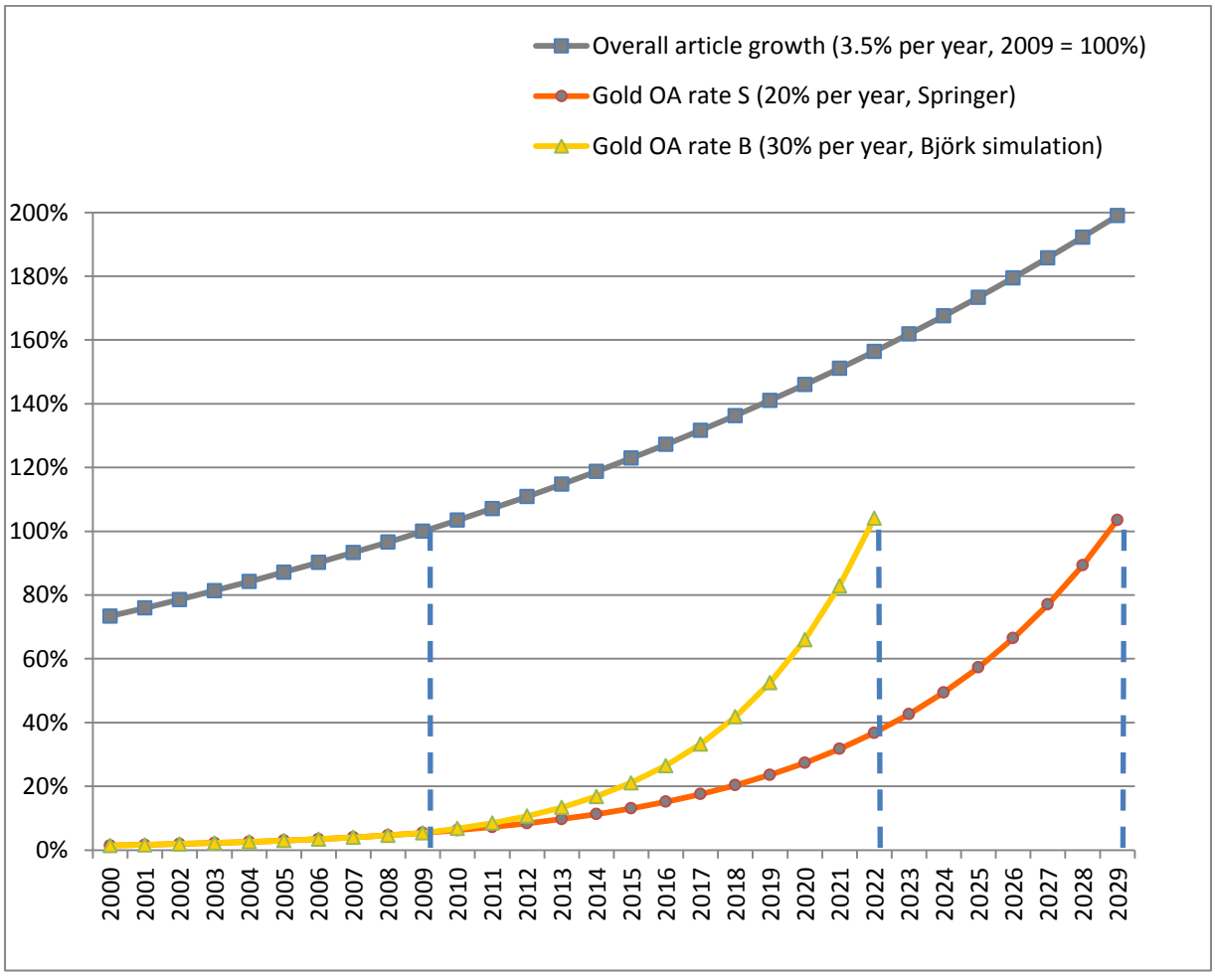

Figure 6: *Springer Gold OA growth curve S (20% per year) and simulated Björk growth curve B (30% per year) (Laakso et al 2011) equated for year 2009. Note that the Björk curve would reach 100% Gold OA for all journals (ISI + non-ISI) in 2022, at a time when the Springer curve would not yet have reached 40% for ISI journals. Laakso et al's estimate of 30% Gold OA growth and Springer's estimate of 20% Gold OA growth can be reconciled if we note that the 30% rate was as of 2000, and has slowed to 20% as of 2005. More important, either way, the Björk curve would not reach 60% till 2019, and the Springer curve would not reach 60% till*

2025, whereas the four mandated repositories had already reached 60% in 2004-2006, within two years of having adopted their mandates (Figure 2).

One possible explanation for a larger OA growth rate among the non-ISI journals may be that most Gold OA today is not fee-based: It comes from either subscription-based or subsidised journals simply making their online version free for all.

This form of Gold OA probably predominates among the 15,000 non-ISI journals, but among the 10,000 ISI journals, where there are already much fewer Gold OA journals, publication fees are much more likely, especially for the "core" journals that institutions most want and need.

So it could be that Bjork's 30% estimate for article growth since 2000 comes predominantly from the journals that are providing Gold OA without charging a publication fee. Hence Björk's overall 30% growth rate was bound to decline as it came closer to the sticking point of the price barrier for the ISI and core journals (as it seems to have done as of 2005).

Sticking point

RP: Björk argues that there are no guarantees that the more or less linear growth rate of Gold OA that we have seen until now will continue. He says: "Most innovations, if successful, follow an S-curve of adoption, and one scenario is that Gold OA is in the beginning of that curve". He adds that several major publishers are currently starting OA journals and so growth might now accelerate. Could he be right?

SH: First, a fixed percentage compounding annually is already accelerating growth. Linear growth would be a fixed-volume increase annually. But the point is that the Gold OA growth rate is far too slow, if we are serious about wanting and needing OA: The Björk curve would not even have us at 50% till 2019, and the Springer curve would not have us there till 2024! Björk is certainly right that spontaneous acceleration in the future is one possible scenario — but there is not yet any sign of that in the actual data — nor in any of the other OA growth data (Green OA, Green OA Mandates).

The analogy with S-shaped innovation curves is probably misleading, because that is predicated on the possibility of unimpeded growth. But (fee-based) Gold OA has a sticking point: *The money to pay for it is currently stuck in subscriptions.* Scarce research money is certainly not available to pay extra for publishing (especially when institutional subscriptions are already paying for it). That is the price barrier.

Institutional journal subscriptions cannot be cancelled while users have no other means of accessing those articles. Nor will it work to re-baptise the existing annual institutional subscriptions as "memberships" in exchange for making the journals Gold OA, because once a journal is Gold OA, there's no need for an institution to renew its membership. Cancelling an institution's annual subscription to a subscription-based journal loses all institutional users' access to that journal's articles, in a non-OA world. But cancelling an institution's "membership" in a Gold OA journal loses the institution nothing: it just saves their cash.

This is because Gold OA is not a service to the user institution, paid for per incoming journal; it is a service to the author institution, paid for per individual outgoing article. Institutions don't have the money to pay for Gold OA while they are paying for subscriptions, and once they need no longer subscribe, they only need to pay for their own outgoing articles, by the article. An institution cannot be an annual "member" of all the journals (up to 25K in all) that its authors might one day submit one article to!

So if anything could transform the Gold OA growth curve into an S-shaped one, it's mandating Green OA, which provides access to subscribers and nonsubscribers alike. If and when Green OA is universally mandated, it can free institutions to cancel their annual subscriptions to incoming journals and use their annual windfall subscription savings instead to pay the Gold OA publishing fees for their outgoing articles – by the individual article.

But whether universally mandated Green OA will eventually make subscriptions unsustainable, generating a global transition to 100% Gold OA, is a speculative matter, whereas what is virtually certain is that it will generate 100% OA — even if subscription publication remains sustainable.

RP: You Stevan are a long-standing and passionate advocate for Green OA. I assume there is always a danger when one has a strong emotional commitment to a certain belief that one might organise a research project, and then interpret the resulting data, in such a way as to confirm your beliefs. How does one avoid that danger?

SH: You're quite right. But sometimes data speak for themselves, with no need of interpretation; and effects should always be re-tested and replicated, repeatedly, to make sure they are reliable. That's the best way to keep oneself honest. Besides, I am not impressed at all with the percentages or growth rates of either Gold OA (which most of the drum-beating tends to be about) or Green OA (which is the one that I favour, and that I argue needs to be given priority).

I think the growth rates of both are pathetic, especially in light of where they will still leave us in 10 years — and the fact that we could already have had 100% (Green) OA a decade ago.

But as long as we're talking about growth curves for Green OA and Gold OA, let me mention one last growth curve: the ROARMAP curve for (Green OA) mandate adoption growth – not in order to crow about any "dramatic" growth rate of adoption itself, which is so far nothing to write home about, but to point out the tremendous potential power behind mandating Green OA.

For whereas the research community cannot accelerate the Gold OA growth rate – that depends on the publishing community and on finding the funds to pay for Gold OA — it can definitely accelerate the Green OA growth rate, in fact, push it toward 100% within a few years: research institutions and funders merely need to mandate Green OA self-archiving.

As Yassine mentioned, our studies have shown that mandates triple self-archiving soon after adoption – especially if self-archiving refereed papers in the institutional repository is made the official mechanism for submitting them for institutional performance review, as the Liège ID/OA mandate has done (thereby raising Liège's daily self-archiving rate to #1 among the 1447 institutional repositories indexed by ROAR in the range of 10-100 deposits daily).

So it is not so much the growth rate of Green OA mandate adoptions (which is still unstable and unimpressive: accelerating 2008-2010, linear 2010-2011) that is noteworthy (Figure 7) but their potential to generate over 60% OA almost immediately (Figure 2) and 100% not long thereafter.

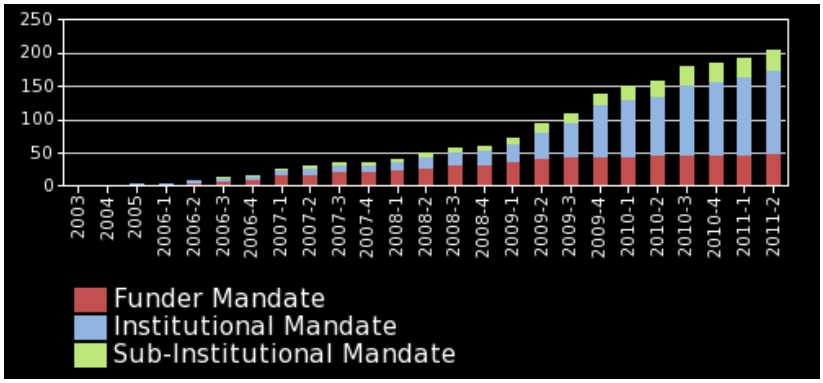

Figure 7. *Yearly increase in the number of Green OA self-archiving mandates adopted by institutions and funders (data from ROARMAP).*

RP: Ok. Let's leave it there for now and see what happens next. Thank you both very much for your time.

Richard Poynder