PEER-REVIEWED JOURNAL ON THE INTERNET

The acquisition of open access research articles
by Arthur Sale

Abstract

The behavior of researchers when self–archiving in an institutional repository has not been previously analyzed. This paper uses available information for three repositories analyzing when researchers (as authors) deposit their research articles. The three repositories have variants of a mandatory deposit policy.

It is shown that it takes several years for a mandatory policy to be institutionalized and routinized, but that once it has been the deposit of articles takes place in a remarkably short time after publication, or in some cases even before. Authors overwhelmingly deposit well before six months after publication date. The OA mantra of 'deposit now, set open access when feasible' is shown to be not only reasonable, but fitting what researchers actually do.

Contents

Introduction

Motivation
This paper was written to understand researcher behavior in depositing research articles in open access institutional repositories.

Acquisition policies
Two types of policies are prevalent in open access research repositories:

- *Voluntary deposit*, where the decision to deposit a research article is made voluntarily by the author/researcher, and
- *Mandatory deposit*, where the deposit of research articles is required by the employing institution.

In the future, there may be examples of mixed policies, where some authors are under no obligation to deposit, but others are required to do so by their research funder. However, these are not yet widespread.

Three universities with mandatory policies were approached; all agreed to participate. The criteria also required that the research repository and its policy have been operational for several years, which limited the field very markedly. Universities with mandatory deposit policies have *all* researchers in the university as depositors, and the results should therefore apply to most universities with similar policies.

■ ────────────────────────────

Acquisition over time

Queensland University of Technology, Australia
The Queensland University of Technology (QUT, 2006a) is a medium to large university situated in the heart of Queensland's capital city, Brisbane. It is notable, so far, as the only university in Australia that has adopted a mandatory deposit policy for all members of its staff (QUT, 2006b). This far–sighted policy commenced effect on 1 January 2004, with the repository starting at the same time. The software used is EPrints (http://www.eprints.org/).

The University provides an invaluable testbed for analysing the effects of the introduction of a 'mandatory' deposit policy, since both early–phase and late–phase deposit rates can be observed.

The first study of QUT focused on the acquistion rate of documents with a selected publication year. With the assistance of the repository manager, data were extracted from the repository over its lifetime, and segregated by the stated publication year. The deposit date was then used to show how articles were deposited for each publication year.

Before looking at the results, consider a thought–experiment as to what might be expected. Suppose that under a mandatory policy *all* published research articles are deposited in the repository. Suppose further that this occurs *exactly* on the date of publication. And suppose as a third assumption that the publication dates are *uniformly distributed* over the calendar year. Then the repository document count will rise from zero on 1 January, approximately linearly, to the total publication count (journal articles and conference papers) at 31 December, which for QUT was 1,013 in 2004. Now with that thought in mind, look at what actually happened at QUT for 2004–2006 (Figure 1).

Figure 1 - QUT deposits over time
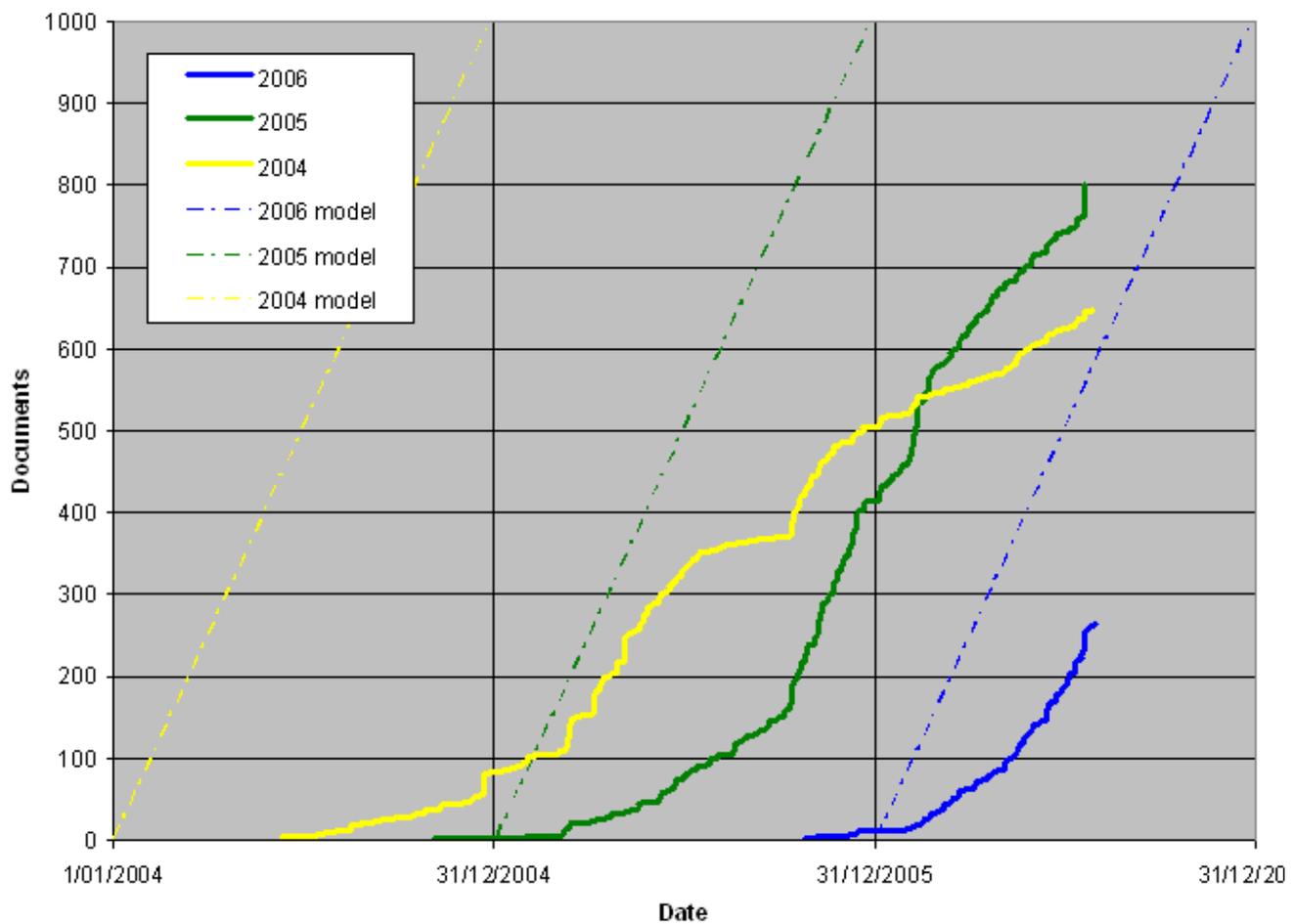
Clearly these graphs do not fit the model. However, they are difficult to compare, so it was decided to bring all years back to a common origin so that the differences between years could be more easily seen, as in Figure 2. This convention will be used throughout this section of the paper.

Figure 2 - QUT deposit rates
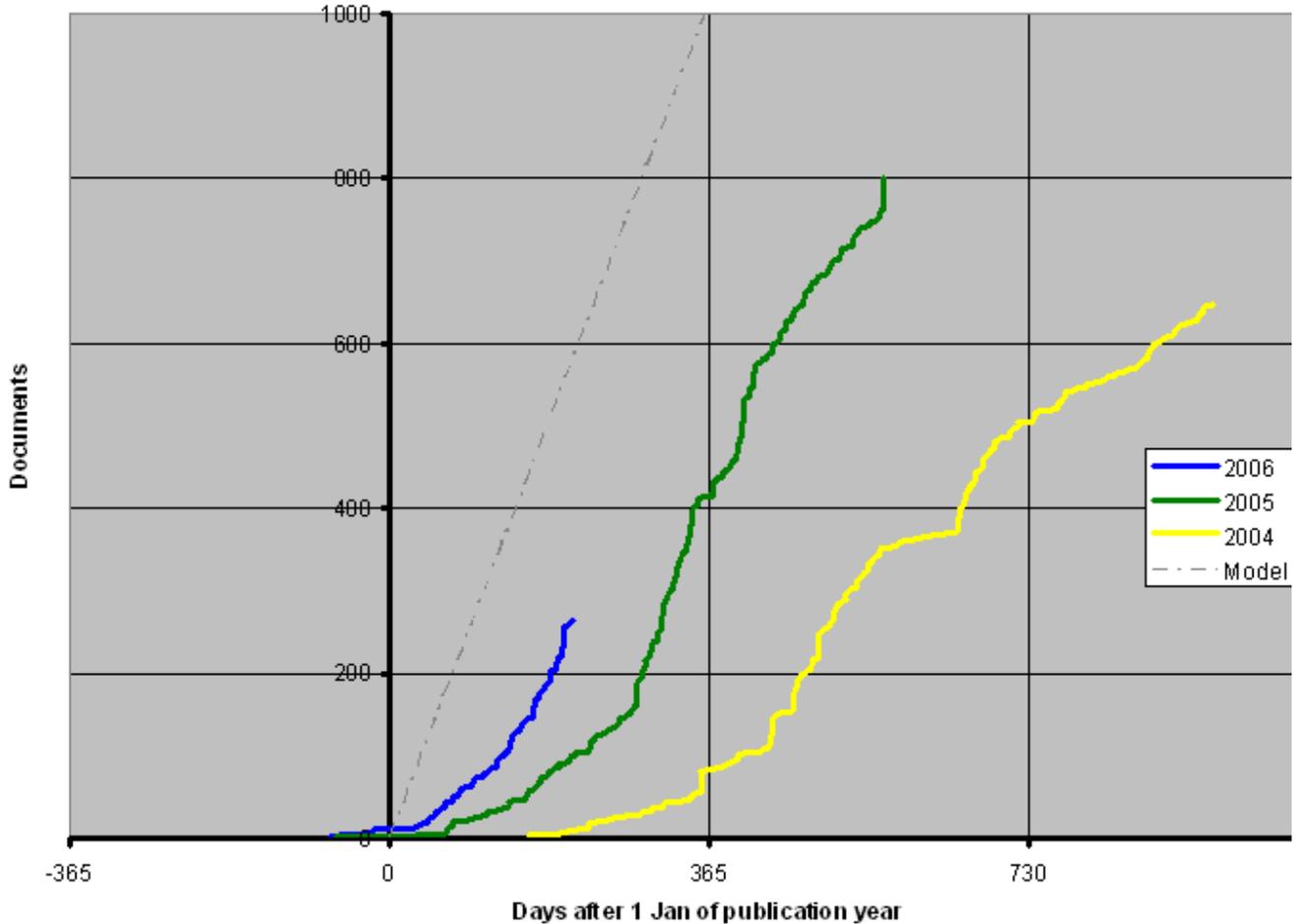
It can be seen that during 2004 (0–365 days of the yellow line) the mandatory policy did not bite in any real sense. Maybe 10 percent of the documents published in that year were collected by year's end. The librarian responsible for the repository stated that the low acquisition rate during 2004 acted as a wake–up call, and midway through 2005 the QUT Library commenced a campaign of publicity, and gentle follow–up with chairs of departments. In Australia, each university must report to the federal government around March of every year on its refereed publications in the preceding year, so what should be in the repository is known for the preceding year. However, no penalties were ever implied for non–compliance.

University of Southampton, United Kingdom

The University of Southampton (Soton, 2006) is a medium–large university situated in the City of Southampton, Hampshire, U.K. It has very recently adopted a university–wide mandate. However, since 2002 the Department of Electronics & Computer Science (ECS) has operated a repository and had a departmental deposit mandate. Looking at the same type of data, the acquisition of research articles is shown in Figure 3. The software is again EPrints.

Figure 3 - Soton deposit rates
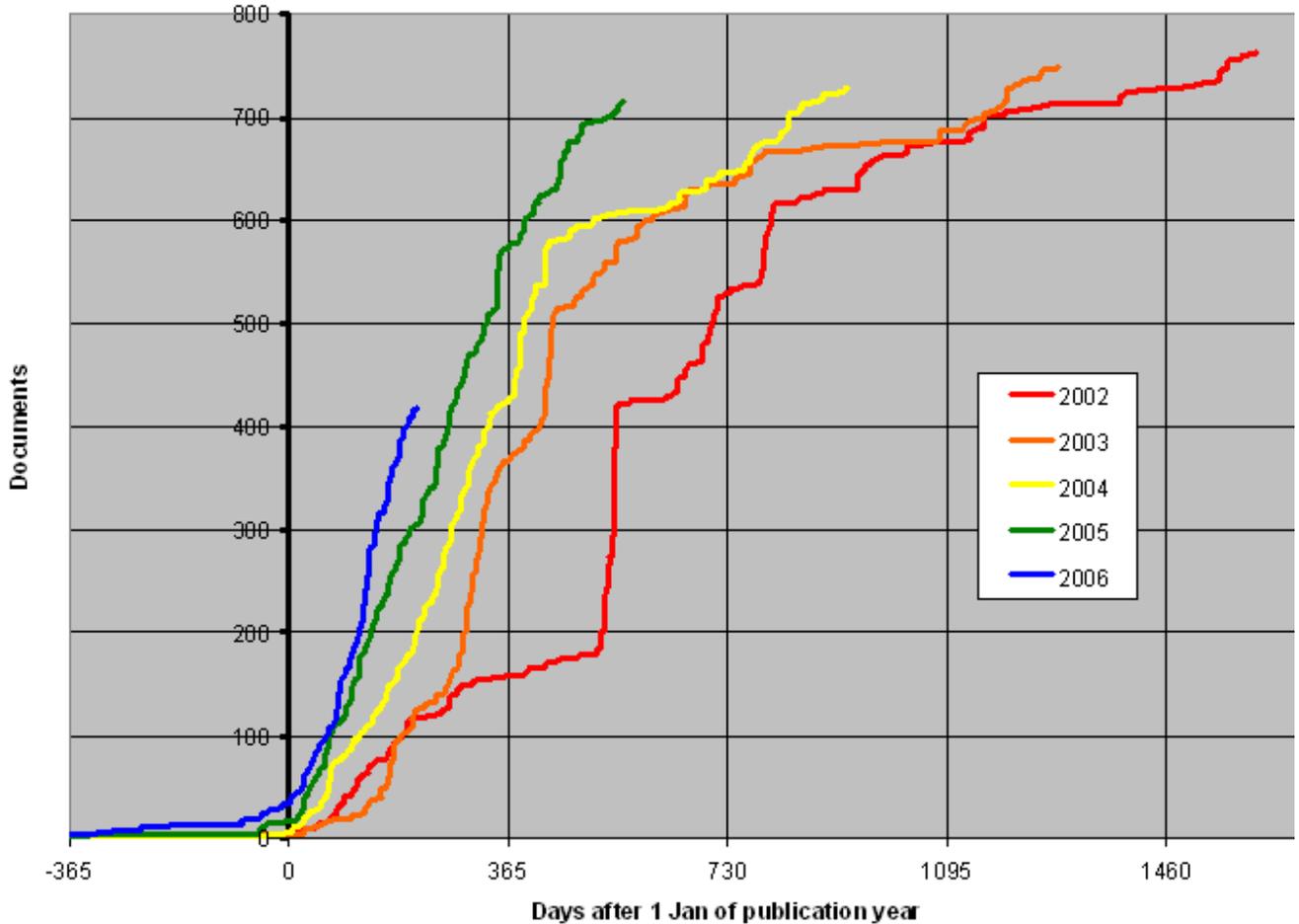
The same issues and the same trend are evident. In the first full year available (2002), acquisitions were slow, but continue to be received over four–five years. Moving ahead to 2005 (the most recent full year available), the same level was achieved within six months after the close of the publication year. Intervening years show a clear progression to this result. The data for 2006 is not yet final, but shows continuance of this trend though the improvement in acquisition rate is slowing.

Why choose six months after the end of the publication year as a significant date? This allows for delays in deposit, especially for those publications that occur in the closing months of the year such as November or December. This issue is taken up in the next section of the paper.

University of Tasmania, Australia
The University of Tasmania (UTas, 2006) is a small–medium university situated on three campuses in Australia's island state, and is generally regarded as being in the top ten Australian universities in research performance relative to its size. The School of Computing at the University of Tasmania is in a position similar to that of ECS at Southampton. A mandate exists at the school (departmental) level, but does not extend to the whole university. The pattern shown in Figure 4 is similar to the two previous cases, differing only in scale and implementation (which was almost immediate in 2004). EPrints software is used.

Figure 4 - UTas deposit rate

■ —————————————————————————
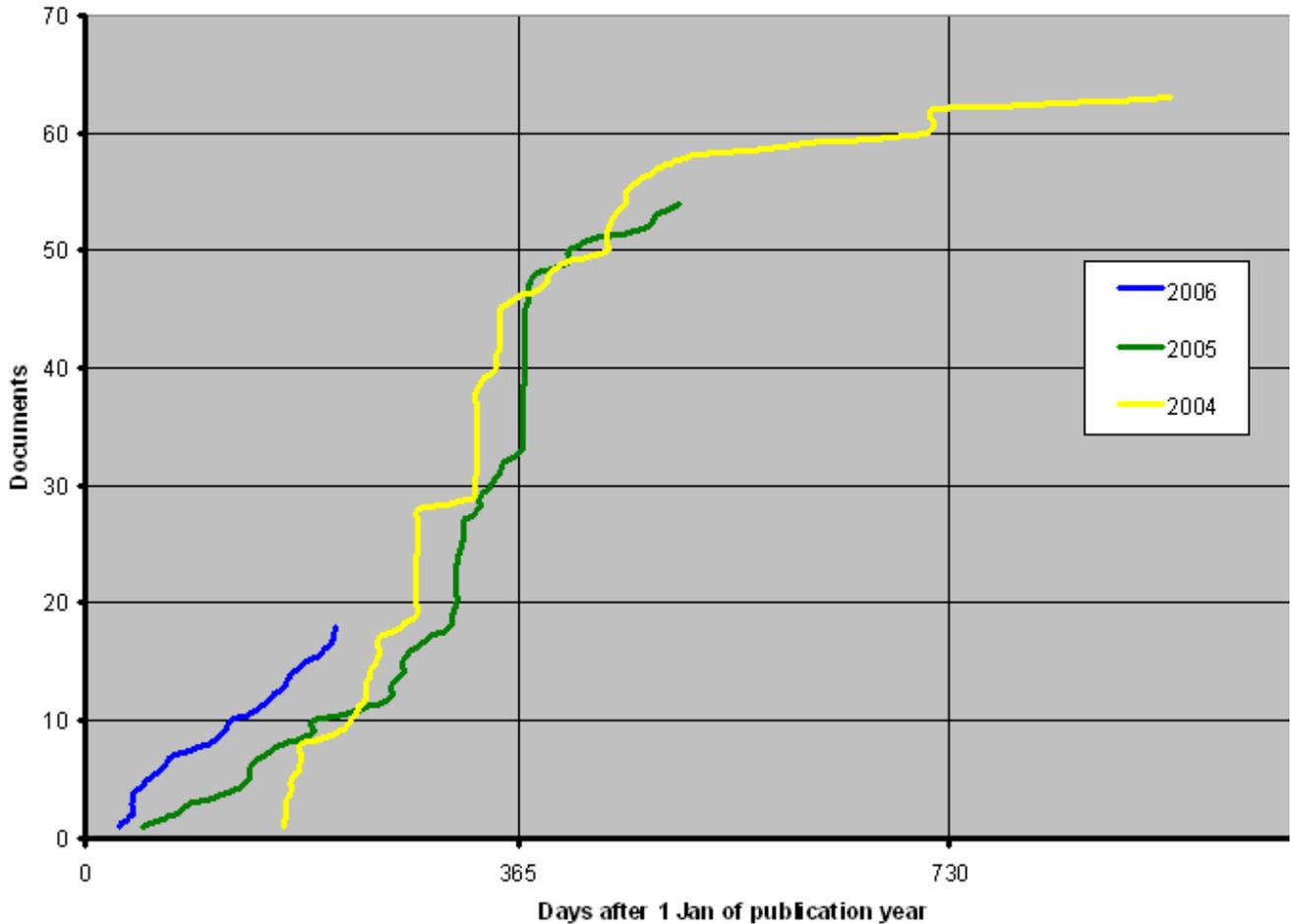
Deposit delays

Queensland University of Technology, Australia
Having analysed what happens over a window of a year, this immediately raises questions about one of the assumptions: the delay between the publication date and the deposit date. Do authors delay depositing even if required to deposit? By how much? When is it reasonable to expect all of a year's publications to have been archived?

While the deposit date is always available to the precision of a day, the publication date is not always available. The year of publication is required metadata, but the month is optional. Consequently only a fraction of deposits can be used to analyse the delays.

With this caveat, the same data could be easily analysed for delay information. To indicate how deposit behavior changed with time, Figure 5 shows the delay distribution for QUT, again presented by publication year. The granularity chosen is one month, since publication dates are not specified to greater accuracy, and smaller granularity has little meaning anyway.

Figure 5 - QUT deposit delays
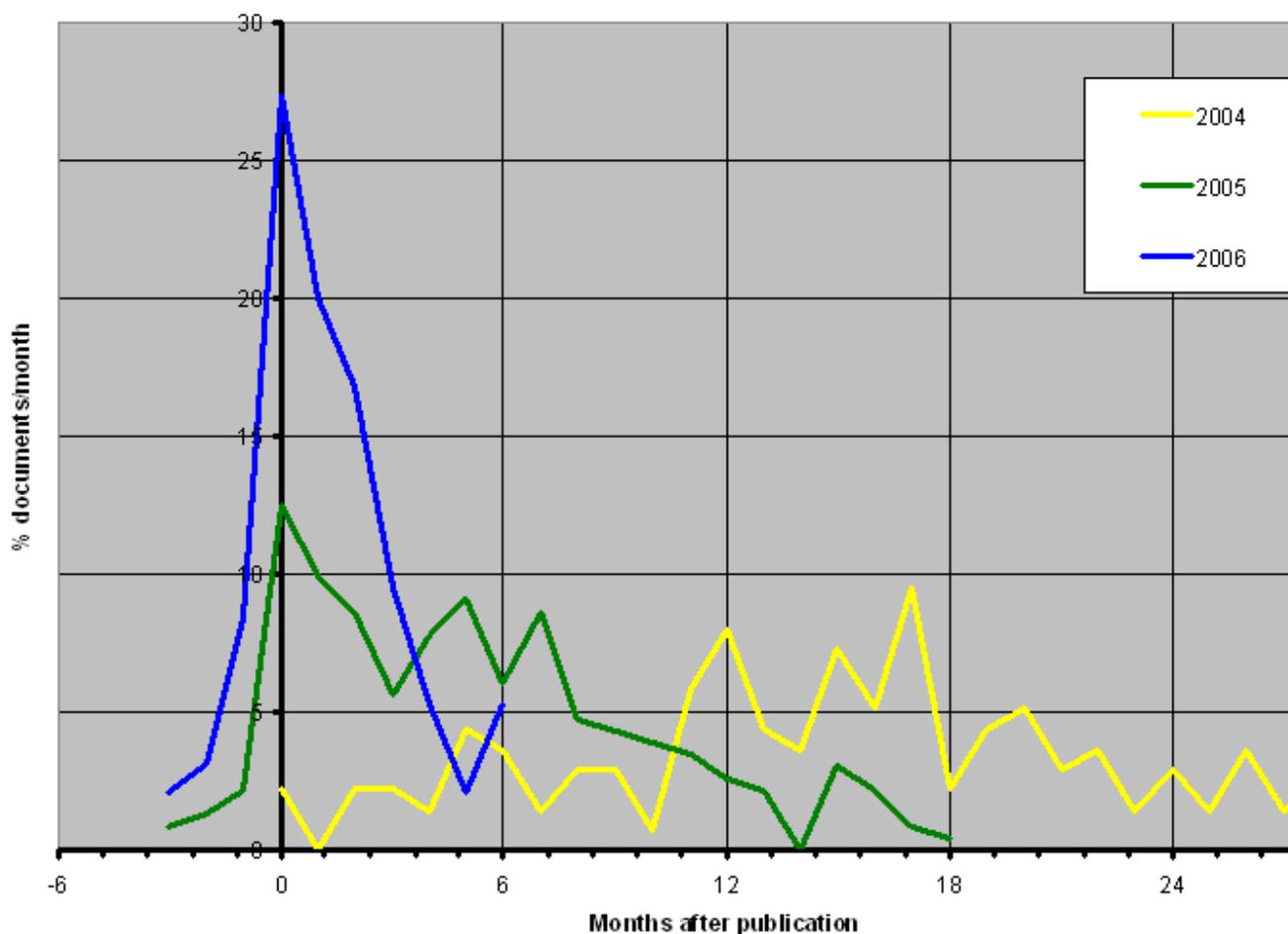
In 2004, articles dribbled in at a more or less steady and low rate, around three percent per month. This picked up a little in the first half of 2005 (for 2004 articles), but declined thereafter. The brief upturn is attributed to Library initatives to publicize the mandate.

The data for 2005 are strikingly different. Articles were deposited more frequently around the publication date and by six months 64 percent had been deposited. Many articles are deposited before publication, some up to three months before (presumably around acceptance date, or from preprint–familiar disciplines). The data for 2006 shows this even more strikingly. The change from 2004 behavior is attributable to the mandatory policy gaining acceptance and beginning to be effective in 2005, and routinized in 2006.

University of Southampton, United Kingdom
As before, the data from the departmental mandate at ECS at Southampton University confirm the foregoing analysis (Figure 6).

Figure 6 - Southampton deposit delays
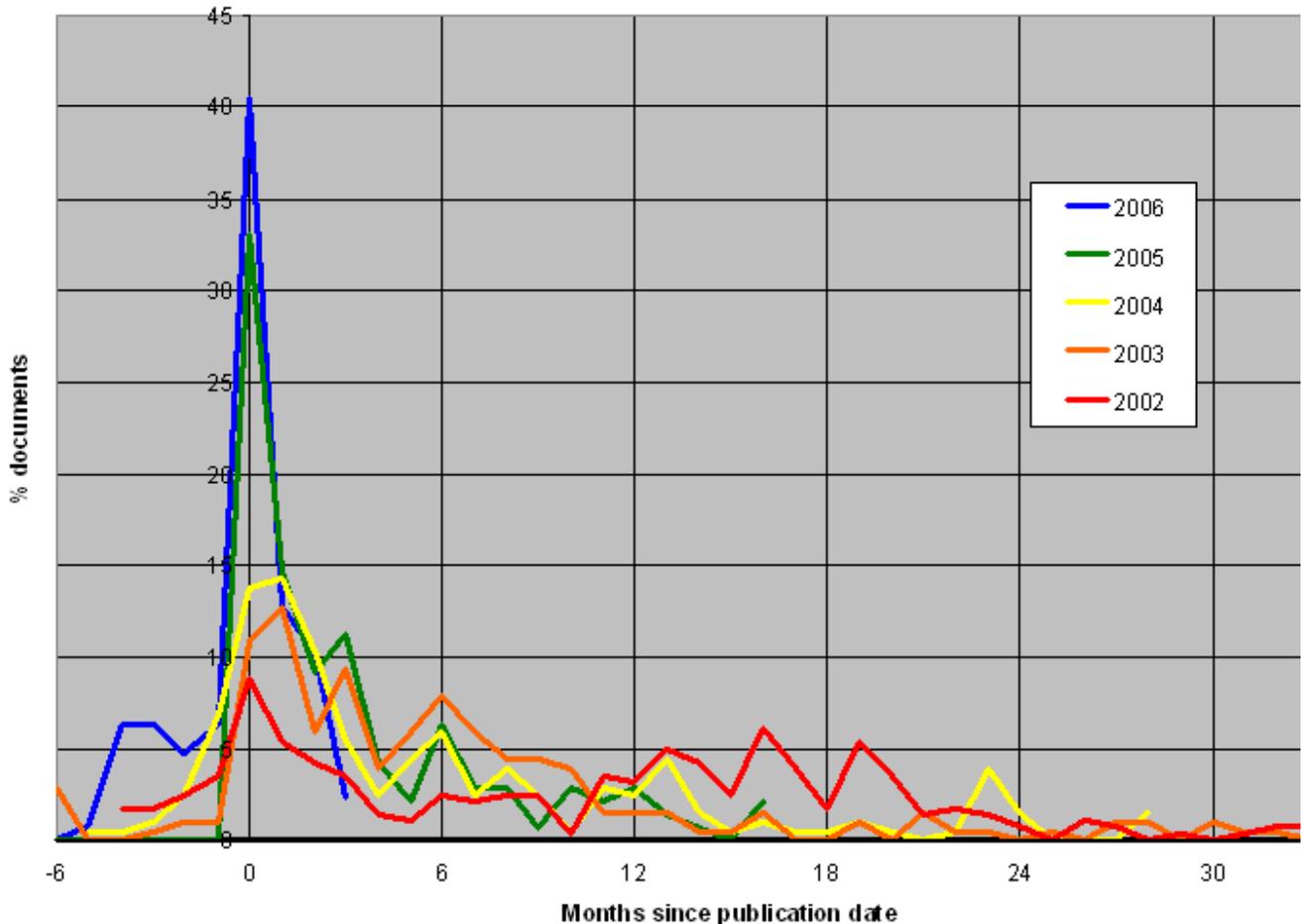
These data cover a longer time span. The transition to an effective mandatory policy was probably complete at the end of 2002. However, continuing evolution in author behavior is still evident. Focussing on the publication date, with every new year the deposit distribution:

- becomes more peaked around the publication date, and
- pre–publication deposits beome more established.

Deposit rates (in percentage) over time are:

| 2002	2003	2004	2005	2006
9%	11%	14%	33%	40%

By 2005 (the last full year on record), 82 percent of articles were deposited by six months after the publication month.

University of Tasmania, Australia (UTas)
The University of Tasmania again shows a similar pattern, though with a smaller sample the

distribution is more noisy (Figure 7).

Figure 7 - UTas deposit delays
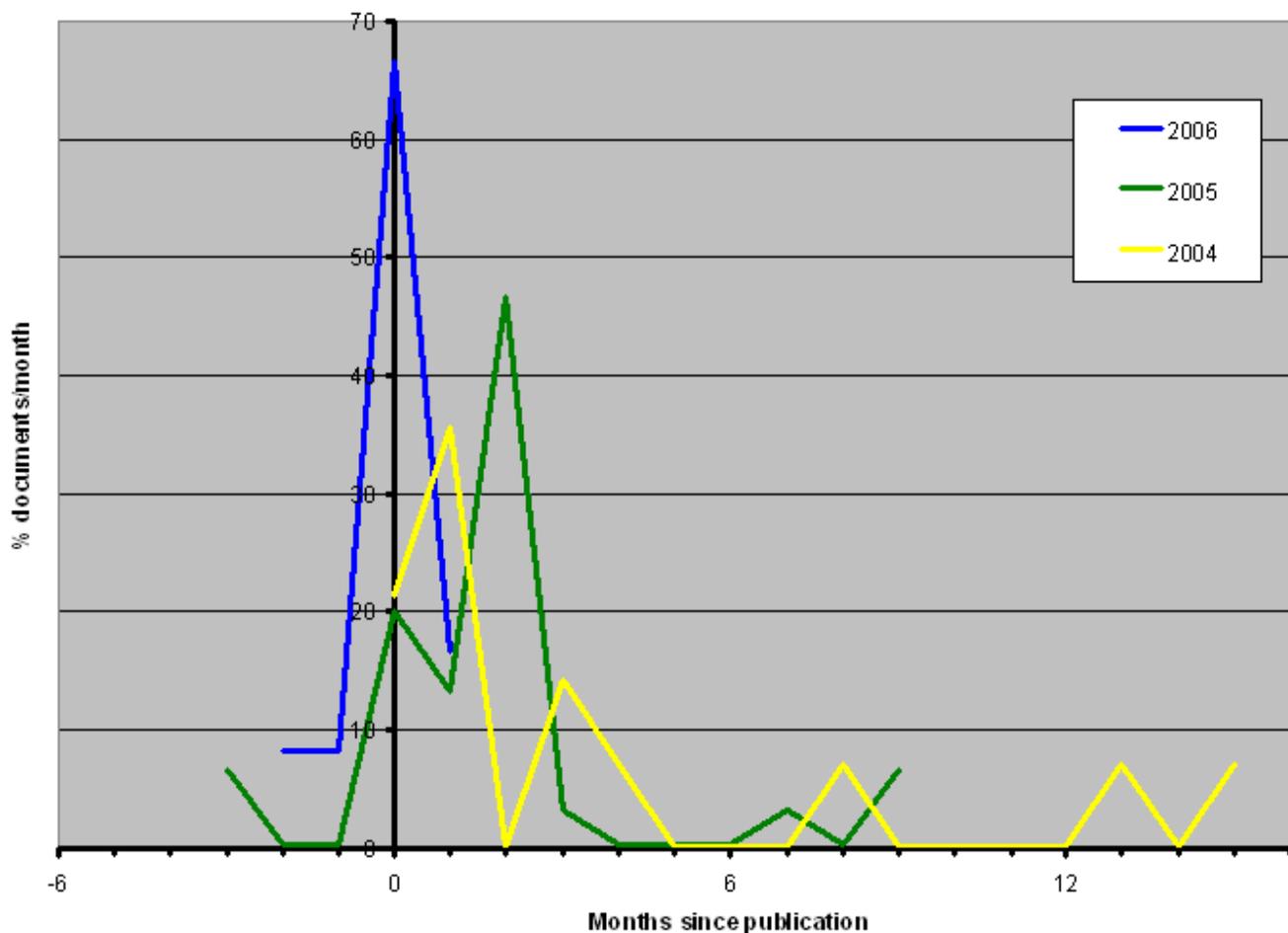
While the sample is small, 90 percent of all documents were deposited in 2005 in three months or less after publication. It appears that 2006 will repeat or better this performance.

■ ————————————————————————

Content

Mandatory policies are now widely recognized as the only way to achieve close to 100 percent of content in institutional repositories. How do these three universities shape up?

To show this information, the publication count was requested from the repository managers for all relevant years. In the case of QUT, officially government–reported data were also available for 2004 (Australian Vice–Chancellors Committee, 2006), and this was used to cross–check accuracy. Where known, the count was of refereed journal articles and refereed conference papers. Whole books and book chapters were not counted as they are subject to publisher agreements. Publication counts for Tasmania are derived from the official departmental returns. Publication counts for Southampton are estimated by the repository manager at 740/year.

The previous analysis has suggested that deposit is essentially complete by six months after the publication date, and therefore by six months after the calendar year almost everything that is likely to be deposited has been deposited. Table 1 shows the content percentage of each of the three repositories for the years on record.

Table 1 – Content percentages				
	Year			
	2002	**2003**	**2004**	**2005**
QUT			32%	73%
Southampton	57%	91%	83%	95%
Tasmania			105%	80%

The data are self–explanatory and consistent with other studies (Sale, 2006a and 2006b). Content greater than 100 percent, such as for Tasmania in 2004, reflects deposit differences not complying with the model regarding multiple authorship. Again, it is reinforced that mandatory policies result in high content rates (70–90 percent), compared to voluntary deposit policies which tend to capture only 10–20 percent of the available research output.

■ ————————————————————

Methodology issues

This study examines repositories which are still developing. Some identified methodological issues are listed below.

- Estimates of the total annual publication output (refereed journal articles and research papers published per year) are subject to some interpretational variation among those supplying the data. However, the data are believed to be accurate within ±5 percent.
- In the Australian total publication output, multi–author papers are apportioned proportionally to institutional affiliation. Thus a three–author paper with two authors from University A and one from University B would be credited as 2/3 of a publication to A and 1/3 to B. On the assumption that the author who does the archiving is randomly self–selected, and the paper is archived only once, this count is taken as the expected count of papers.
- For Southampton, the repository entry is sometimes just a metadata stub and the full–text has not been not uploaded to the repository. However, the full–text would have had to be available to the researcher when entering the metadata, and the deposit information is therefore regarded as equally significant, differing from a full–text submission by only a few clicks. In the case of the Australian repositories, metadata–only entries are held in separate reporting databases (*WARP* in Tasmania, *Research Master* in QUT) and the open access repositories hold 100 percent full–text items.
- While mostly holding School of Computing items, the Tasmanian repository also contains items contributed by a few researchers in other schools who discovered the repository and asked to be included. These fall outside the departmental mandate.

■ ————————————————————

Conclusions

Policies
A separate and more complex study is being undertaken of universities with voluntary deposit policies. As contributors to such repositories are self–selected, their characteristics may differ from those of the group of repositories studied here, as local factors may play a larger part. A university with a mandatory deposit policy for its repository includes all researchers in its ambit, and the behavior is expected to be generalizable to most universities.

Time to be effective
The time required for a mandatory deposit policy to become effective varies with the scale of the enterprise, as would surprise no one in management. In departmental mandates, such as Tasmania and Southampton, the uptake appears to be swifter than in university–wide mandates such as QUT. At the departmental level a few years – or even one – suffices to reach close to 100 percent capture, though Southampton keeps showing improvement in the *rapidity* of acquisition over at least five years.

At a university level, however, there is as yet insufficient data. What can be estimated is that a university–wide mandatory deposit policy takes at least three years to be (say) 80 percent effective, if it is the authors themselves who provide their documents. If the repository managers adopt a proactive policy of actively uploading missing documents on behalf of the authors, as at CERN (http://public.web.cern.ch/), then the *apparent* transition will be faster, but the rise of self–archiving might be slowed due to lack of direct author incentive and involvement. Repository managerial promotion and assistance, such as that undertaken by the Library in QUT, matters very significantly under a mandatory policy, although under voluntary policies it seems to be largely a waste of money (Sale, 2006b).

Conclusion 1
1. Repository managers should invest in promotion and follow–up for two–three years after a mandatory policy is promulgated, after which the behavior becomes routinized.

Deposit behavior
Before a mandatory policy is established, documents dribble in to the repository even many years after the date of publication. Once a mandatory policy is established, the pattern changes dramatically, and deposit occurs around the date of publication. The publication month is the peak month for deposits, and the size of this peak grows and phase–advances with time. Even the data for Southampton do not yet show clear evidence of this peak or phase–shift stabilizing.

In this regime, a fraction of deposits occur even before the publication date. These are either early adopters or persons used to a paper preprint culture who mount their papers on submission to a journal or conference, and subsequently insert the publication date and page numbers; or researchers who deposit at or around acceptance of the paper for publication. This fraction is estimated at 15–25 percent.

... researchers are not favorably inclined towards the six–month embargos adopted by some publishers ...

Substantial numbers of papers are deposited in the months following publication, and by six months, over 80 percent of all documents that will be acquired have been deposited. This clearly indicates that researchers are not favorably inclined towards the six–month embargos adopted by some publishers, if their normal behavior is to deposit so closely after publication date. Of course, this is entirely natural behavior: the longer a researcher waits to deposit a document, the higher the probability that it will be lost or mislaid, or that the researcher will forget to deposit it at all. Indeed the most natural time to deposit a research article is at the time the final manuscript is delivered to the publisher – at that time the electronic copy is at hand, and has not yet been filed away.

Further conclusions
2. No especial activities need to be undertaken to convince researchers to deposit research articles soon after publication – this seems to happen naturally under mandatory policies.
3. Six–month embargos by publishers are likely to be unpopular with researchers, since in the absence of constraints they deposit earlier than this.
4. The recommendation widely adopted by the open access movement and summarized as *'deposit immediately, and make open access as soon as legally possible'* is shown to be excellent advice for any university or funding agency considering adopting a mandatory policy. ▥

About the author

Arthur Sale is currently Professor of Computing Research at the University of Tasmania, and Research Coordinator of its School of Computing. From 1993–99 he was a member of the University's Senior Executive as Pro Vice–Chancellor, and from 1974–93 Chair of the Department of Computer Science. Arthur Sale has published extensively in the ICT literature, and is internationally known for his work on programming languages and computer architecture. His current research interests extend to bioinformatics, mobile computing, and Internet technologies. He has been described as Australia's archivangelist for open access.

Acknowledgements

References

Australian Vice–Chancellors Committee (AVCC), 2006. "Higher Education Research Data Collection Time Series 1992–2004," at
http://www.avcc.edu.au/documents/publications/stats/HERDC%20TimeSeries%20Data%201992-2004.xls, accessed 5 August 2006.

Stevan Harnad, 2006. "Generic Rationale for University Open Access Self–Archiving Policy," *Open Access Archivangelism* (13 March), at
http://openaccess.eprints.org/index.php?/archives/71-guid.html, accessed 5 August 2006.

Queensland University of Technology, 2006a. QUT Queensland University of Technology Web site, at http://www.qut.edu.au/, accessed 5 August 2006.

Queensland University of Technology 2006b. "Policy F/1.3 E–print repository for research output at QUT," at http://www.mopp.qut.edu.au/F/F_01_03.html, accessed 5 August 2006.

Arthur Sale, 2006a. "The impact of mandatory policies on ETD acquisition," *D–Lib Magazine*, volume 12, number 4, at http://www.dlib.org/dlib/april06/sale/04sale.html, accessed 5 August 2006

Arthur Sale, 2006b. "Comparison of IR content policies in Australia," *First Monday*, volume 11, number 4 (April), at http://www.firstmonday.org/issues/issue11_4/sale/, accessed 5 August 2006.

University of Southampton, 2006. University of Southampton Web site, at http://www.soton.ac.uk/, accessed 5 August 2006.

University of Tasmania, 2006. University of Tasmania Web site, at http://www.utas.edu.au/, accessed 5 August 2006.

Editorial history

Paper received 22 August 2006; accepted 20 September 2006.

Contents Index