UK Institutional Repository Activity

A longitudinal study

Tim Brody

2/11/2008

A longitudinal study of UK Institutional Repository activity based on data from the Registry of Open Access Repositories (ROAR).

Revision History

Revision	Notes
1.0	Initial version
1.1	Readability revisions, added caveat about full-text vs. metadata-only to conclusions
1.2	Removed erroneous Geometric Mean graph/section

Definitions, Acronyms and Abbreviations

Term	Definition
UK	Refers to any Institutional Repository that caters to UK authors – covering the regions of England, Scotland, Wales and Northern Island.
Institutional Repository	Loosely defined in ROAR as being a repository that restricts depositions to a UK HE institution or research department. Contains some form of primary research output (journal or conference articles, mongraphs, etc.).
IR	Institutional Repository
ROAR	Registry of Open Access Repositories (http://roar.eprints.org/)
OAI-PMH	Open Access Initiative's Protocol for Metadata Harvesting (http://www.openarchives.org/)
OpenDOAR	Directory of Open Access Repositories (http://www.opendoar.org/)

Contents

Brief

Perform a longitudinal comparison of UK institutional repository activity using our 'activity' metric.

Design

The Activity Metric

In Carr 2007 we describe a method of evaluating the degree of *activity* an IR has based on the frequency of record deposits. The goal of measuring activity is to determine the degree to which faculty are using their IR, which is critical to the long-term success of IRs and to getting Open Access to UK research output. We argue that activity is a better measure of IR success than the total number of records, because often the total can be inflated by the bulk import of data sets or existing bibliographic data.

Instead of measuring the total number of records in an IR, activity counts the number of days in which deposits have been made in. This is further broken down into three broad categories of 'low', 'medium' and 'high' daily activity which correspond respectively to 1-9 deposits, 10-99 deposits or 100 deposits or more. Broadly speaking an IR will want to have consistent 'medium' activity, which reflects ongoing, substantial use of the IR by a significant number of faculty. Of course these boundaries are artificial and, ideally, would be adjusted for the amount of research activity in each institution.

Longitudinal Study

The goal of this study is to develop a mechanism for assessing the success (or not) of the investment in UK IR capability. That investment has two outputs: the setting up of IRs and the deposit of research outputs in those IRs. Identifying the number of IR installations is relatively easy as both of the main IR registries (OpenDOAR and ROAR) can provide this data. Figure 1 shows the number of UK IRs registered in OpenDOAR and Figure 2 for ROAR, which have 84 and 65 registered IRs respectively[1]. In the RAE 2001 there were submissions from 173 HE institutions, which suggests there remain a number of research-active HE institutions that have not yet implemented an IR.

Applying the activity metric at the level of the UK has required aggregating the activity of all UK IRs. This presents some problems, for instance if one normalises the total activity by dividing by the number of active repositories, then the rapid increase in the number of repositories (but lack of corresponding **use** of those repositories) would appear to show a *decrease* in activity. This is obviously not the case – that there is increasing IR capacity doesn't infer that activity is dropping.

In order to evaluate activity across the UK we have provided a series of metrics. The last of these is to simply count the total number of repository-active days (i.e. across all active repositories, how many days have been 'active' in the given year). This is the current metric we would like to use, as it - with the low/medium/high breakdown - gives a clear view of total IR activity across the UK. As UK IR activity approaches saturation we expect to see close to 220 (available work days) times the number of HE institutions' (about 173 based on RAE 2001) repository-activity days i.e. 38,060 repository-active days per annum.

---

[1] The different number of entries in OpenDOAR and ROAR are due to different editorial policies (ROAR separates out demonstrations/prototypes and funding-body IRs).
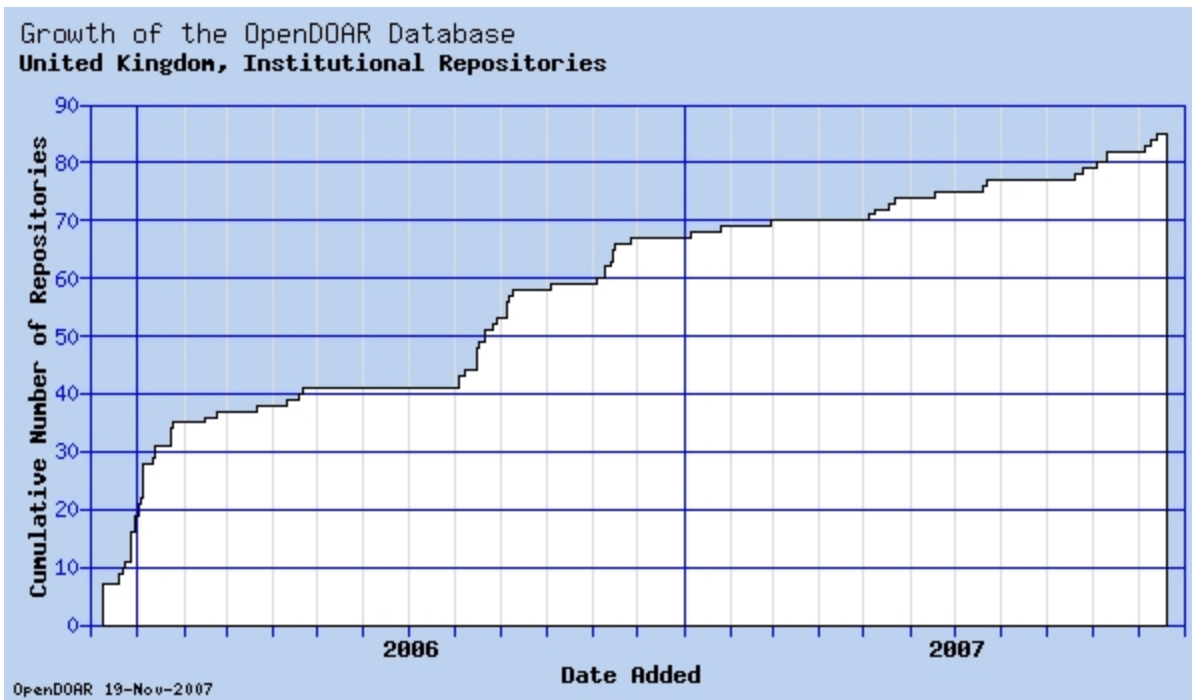
Figure 1 Number of UK Institutional Repositories registered in OpenDOAR over time (84 as-of 19th November 2007).
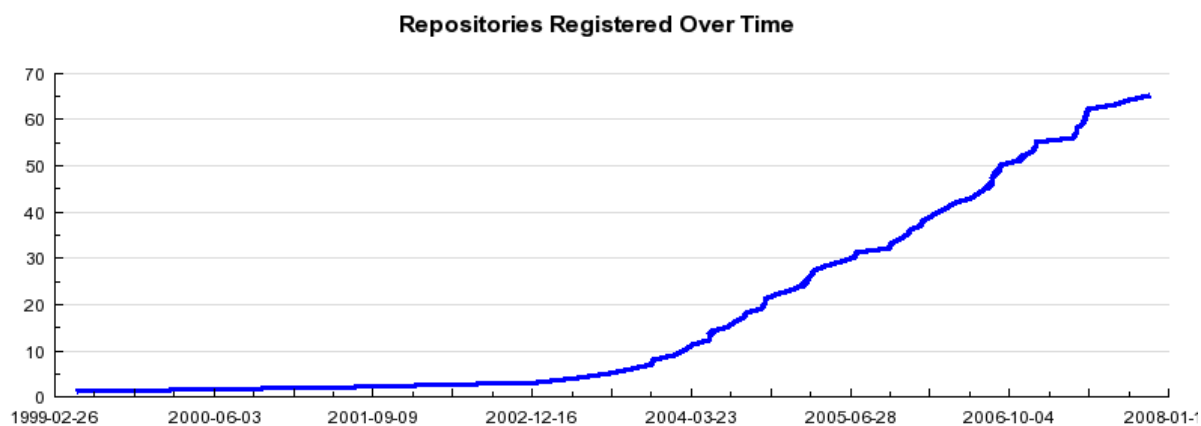

Figure 2 Number of UK Institutional Repositories in ROAR over time (65 as-of 19th November 2007).

Results

All the data used in this study comes from ROAR. A total of 65 UK IRs are registered, of which 4 have no data available due to their OAI-PMH interface being unavailable or unknown or just having no data for the period looked at. Figure 3 shows the number of active repositories over time and Figure 4 the total number of OAI-PMH records harvested from them.

The total UK repository-active days were calculated for each year there has been an active repository (starting 1999). Two averages were also calculated: the average activity per-repository normalised by the number of days available (220) and the geometric-mean. These are described in the following sections.
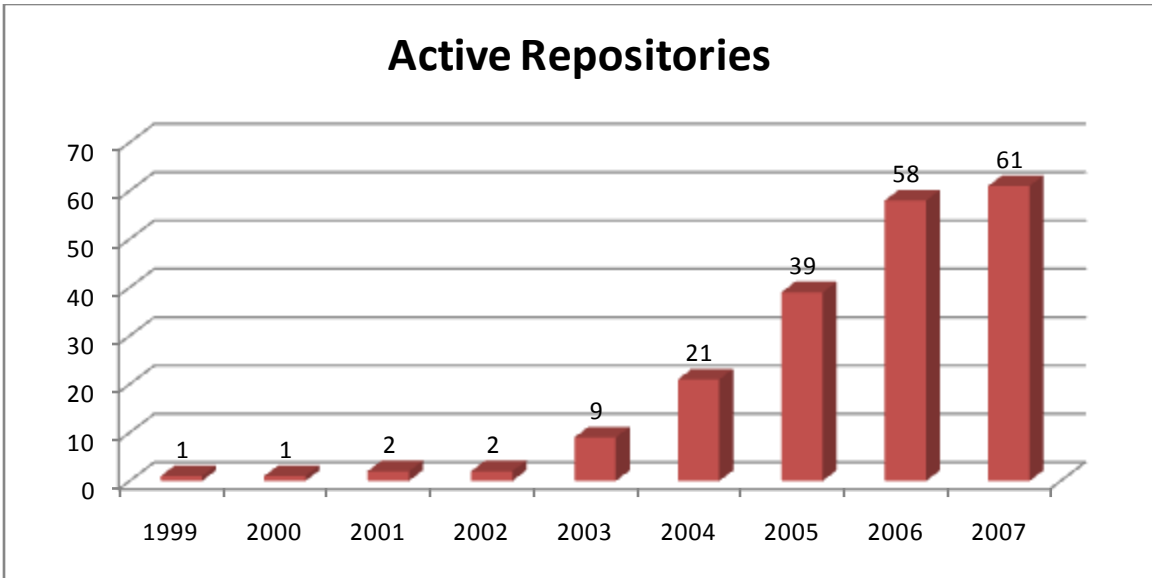
Active Repositories

Figure 3 Active UK IRs registered in ROAR. Each repository must have had at least one record deposited in the year to be classed as 'active'. 2007 is upto November 19th.
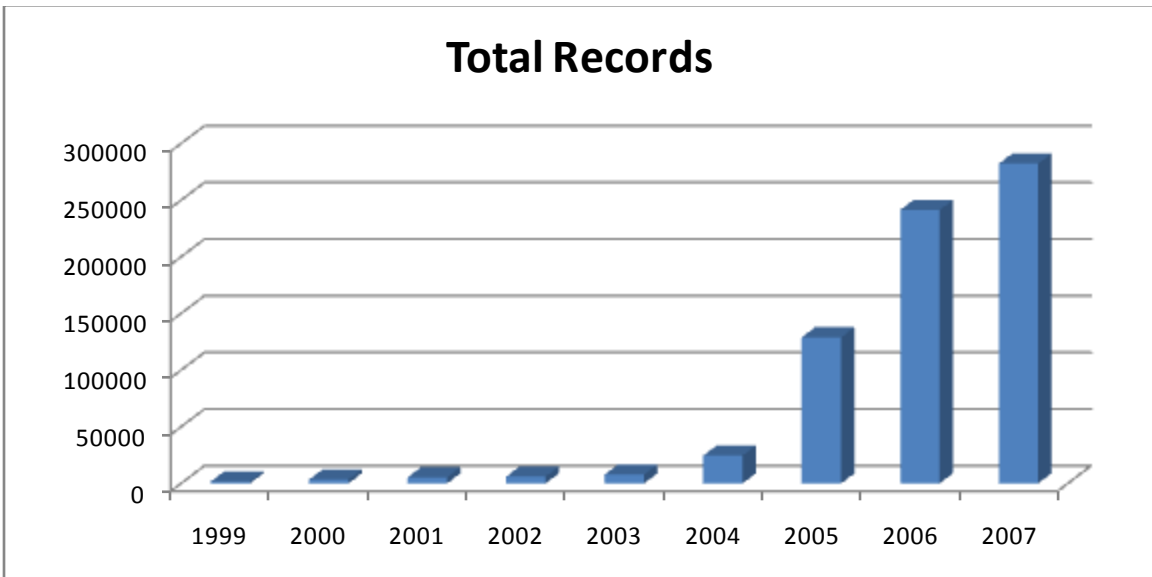
Total Records

Figure 4 Total number of records deposited in UK IRs registered in ROAR. A *record* is an OAI-PMH metadata record, hence may only contain metadata and no digital object(s). 2007 is upto November 19th.

Raw Data Table

Year	Repositories	Low	Low GeoMean	Low Mean	Med	Med GeoMean	Med Mean	High	High GeoMean	High Mean
1999	1	53	53	0.24	32	32	0.15	3	3	0.01
2000	1	119	119	0.54	57	57	0.26	1	1	0.00
2001	2	136	41	0.31	22	5	0.05	1	1	0.00
2002	2	107	23	0.24	19	19	0.04	0	0	0.00
2003	9	190	8	0.10	39	3	0.02	2	1	0.00
2004	21	607	13	0.13	175	5	0.04	6	2	0.00
2005	39	1251	22	0.15	372	4	0.04	29	2	0.00
2006	58	1923	15	0.15	712	8	0.06	81	3	0.01
2007	61	2543	27	0.19	925	10	0.07	18	2	0.00

Figure 5 ROAR does not provide aggregated activity data – activity metrics are only available per-repository and only for the last year. Therefore additional data was generated for this study based on the OAI-PMH records.

Average Activity per Repository

This is an attempt to normalise for the increase in the number of IRs over time. It is intended to show how high the activity is on a per-repository basis i.e. while it may be relatively easy to install and set up IR software, how far have repository managers succeeded at getting their faculty to actually *use* the IR?

The total repository-active days for the year were divided by the number of active IRs multiplied by the number of working days available (220). In 1999/2001 there was only one active IR (*University of Southampton: Department of Electronics and Computer Science*). The *decrease* in activity post-2000 is due to new IRs coming online, but without the level of activity that the existing repository had. The level of activity has steadily increased from 2003 onwards, suggesting that as IRs mature they attract more deposits. The majority of activity is still constrained to 'low', suggesting there is a lot of scope to increase the *number* of deposits as well as the consistency.
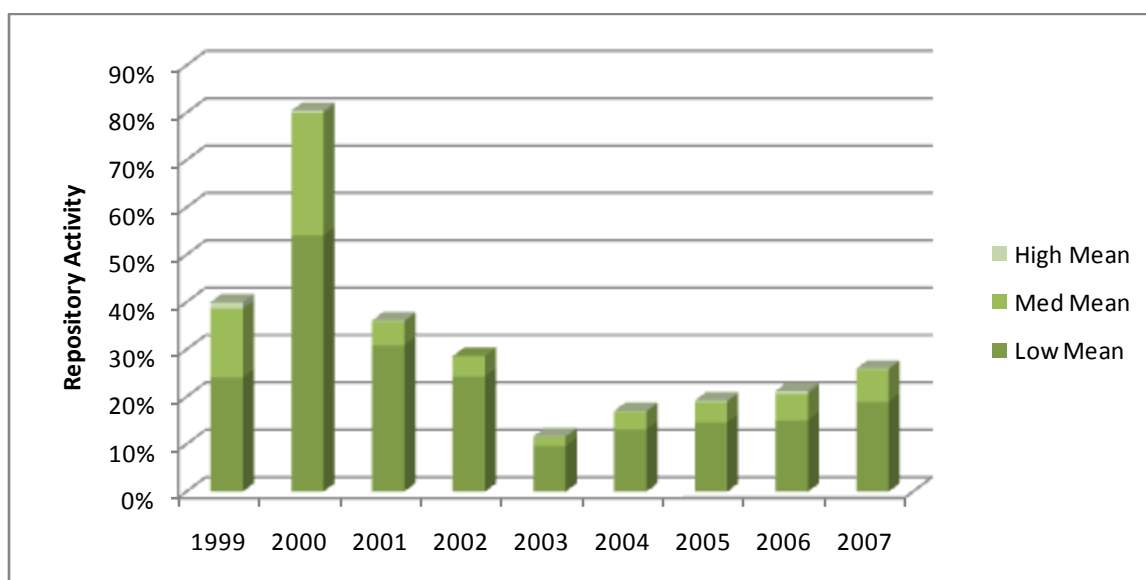

Figure 6 The average active days per IR, as a percentage of total work days available (220).

Total UK Repository Active Days

Each IR studied has low, medium and high days of activity. Totalling these gives us an approximation for the amount of use of those repositories by depositing users. Figure 7 shows the total repository-active days across all UK IRs. The number of IRs has only significantly increased from 2003 onwards. As the number of IRs has steadily increased so has the total amount of activity. Most of this activity is less than 10 deposits per day, however.
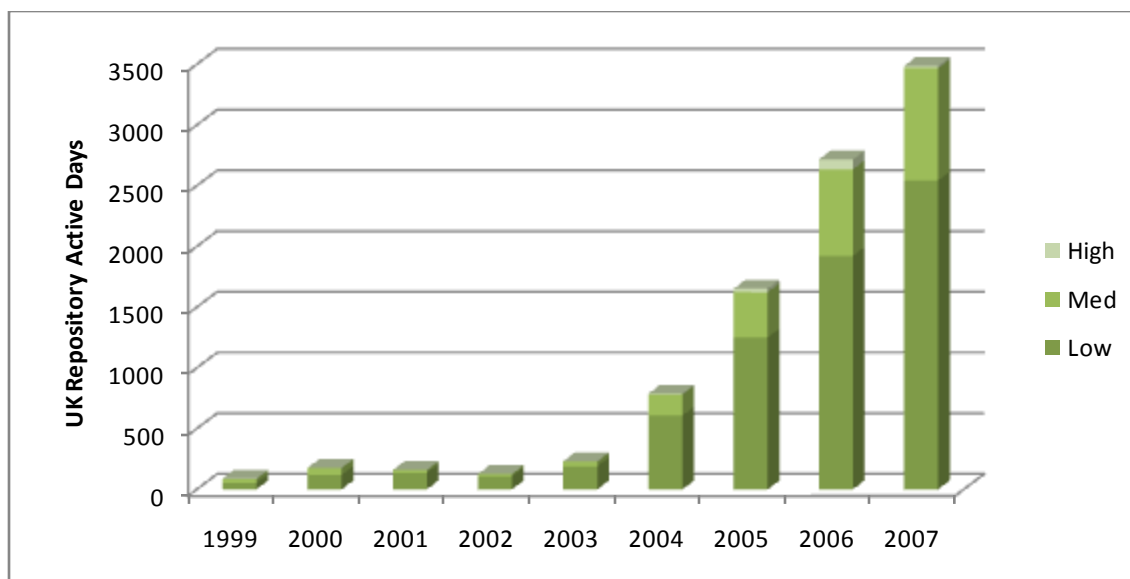
Figure 7 The total number of repository-active days per annum in the UK.

Conclusions

As more IRs are brought online so the *capacity* of the IR network increases. The maximum activity of an IR has been defined as 220 (working days). Each of those working days can have any number of deposits, but we have assumed somewhere between 10 and 100 is 'normal' (less than that is too quiet, more than that is likely to be bulk imports from other sources).

The total potential UK capacity is therefore 220 times the number of institutions for whom an IR is appropriate. Based on the RAE 2001[2] we estimate there are 173 potential (research-based) IR locations. That means the total UK capacity is close to 38,060 repository-active days – somewhere to aim at.

In 2007 there have been approximately 3,500 repository-active days – representing less than a tenth of our estimate of the potential UK IR activity.

It should also be kept in mind that these figures are only data for OAI-PMH records – most of which do not have an associated 'full-text' (or other digital object). To achieve Open Access users must also deposit a freely accessible version of their research, but we don't currently have the capability to distinguish Open Access records from metadata-only records. It is assumed, however, that as IRs become a natural part of academics' work-flow so it will become natural for them to deposit the full-text as well as entering the bibliographic data.

References

Carr, L. and Brody, T. (2007) Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. DLib Magazine, 13 (7/8). ISSN 1082-9873

---

[2] Institutions that submitted to the 2001 RAE:
http://www.hero.ac.uk/rae/rae_dynamic.cfm?myURL=http://195.194.167.103/submissions/inst.asp