



Sally Rumsey

The purpose of institutional repositories in UK
higher education : a repository manager's view

Originally published in International journal of information management,
26 (3). pp. 181-186 © 2006 Elsevier B.V.

You may cite this version as:

Rumsey, Sally (2006). The purpose of institutional repositories in UK
higher education : a repository manager's view [online]. London: LSE
Research Online.

Available at: <http://eprints.lse.ac.uk/archive/00000800>

Available online: June 2006

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this version and the publisher's version remain. You are advised to consult the publisher's version if you wish to cite from it.

The purpose of institutional repositories in UK Higher Education: a repository manager's view

Sally Rumsey, LSE, UK

Open access institutional repositories (IRs) in UK Higher Education (HE) have been the topic of much discussion between the various interested parties. There have been disagreements and arguments as to whether or not they will either solve a multitude of problems or threaten commercial scholarly publishing and learned societies. Topics of debate are commonly the 'serials crisis' and library funding (Suber, 2003), the future of scholarly communication, new publishing models and the roles of learned societies.

HE institutions across the UK and internationally are developing IRs and have been encouraged to do so by the UK Government Science and Technology Committee report of July 2004 (Great Britain, 2004) and the expected mandate from the Research Councils UK (RCUK)¹. In addition, the JISC (Joint Information Systems Committee) is funding development of IRs in the UK (for example under the Digital Repositories Programme²). Open access repositories are nothing new: the oft-quoted physics repository, ArXiv³, has been in operation since around 1994 and the repository of economics working papers, RePEc (Research Papers in Economics)⁴ which grew out of its predecessor, WoPEc, was created around 1996.

Firstly, it is important to clarify the difference between open access journals and open access institutional repositories. David Prosser of SPARC Europe defines open access journals as those which 'do not charge for access to the papers, but make the papers available to all electronically and look to other financial models to cover the costs of peer-review and publishing. They do not invoke copyright or exclusive licenses to restrict access to the papers published within them, rather they encourage the dissemination of research limited only by the reach and extent of the internet' (Prosser, 2004). An IR is an open access (ie freely accessible by anyone with internet access), searchable, digital archive of materials emanating from an institution, usually scholarly but not limited to journal articles, which are usually (but not always) available in their entirety. An IR does not offer peer-review.

This article focuses on IRs dedicated to storing research output: others are designed to include content such as teaching and learning materials or administrative documents. The example used here is the IR at LSE (London School of Economics and Political Science) which is designed to enable the collection and management of research output of the institution and to help academics make the most of their research publications as well ensure their work is preserved for the future. The IR at LSE is called LSE Research Online (LSE RO) and items have been deposited in it since March 2005 (see <http://eprints.lse.ac.uk/>). It is similar to repositories in many other institutions.

Open access publications may be a means of increasing availability of scholarly materials. However, there are other drivers which prompt HE institutions to spend staff time and investment on developing IRs. If the publishing arguments are put to one side, there is a separate and strong case for developing an IR which is justified by a library's functions of collecting, managing and curating materials.

The reasons why an institution may wish to provide an IR are:

¹ See <http://www.rcuk.ac.uk/access/index.asp>

² See http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories

³ See <http://arxiv.org/>

⁴ See <http://www.repec.org/>

- the management of digital objects and preservation of research
- to enable mechanisms for dealing with different publication types, multiple versions and relationships
- branding and increased visibility
- compliance with requirements of external bodies

Management of digital objects and preservation of research output

HE institutions vary in their provision of a publications database to record the publishing success of their researchers. Those which have such a facility may not record all publications and the databases contain metadata-only records. The IR includes the full text of each work. An IR may or may not be intended to replace a publications database.

Repositories do not include the full content of any single journal issue, only those items contributed by the institution's authors. It is interesting to take the case of journals published by LSE. The percentage of articles published by LSE authors in three of the five titles between the years 2002 and 2004 is shown in Table 1. It is only this small proportion of articles that the School wishes to retain for the reasons stated above.

There may be copies of research output retained in various formats, scattered around any number of locations within the institution; on personal computers, departmental websites and so on. Alarming, a copy of all items may not even exist at the institution. An IR is a means of gathering the entire oeuvre of that institution in one place as a record of that institution's research output. Not only is this useful for storage and management purposes, but also for searching. It is easier to search across all items at once, rather than having to select the correct location before searching – even if that location exists. This search facility is also useful to the institutional outreach department which liaises with external bodies, for identifying relevant experts at that institution.

An IR can have a wide scope: not only can it include any 'publications' (records and full text), but it may also be used as a store for items from diverse departments around the institution. For example, LSE is considering including items held by the Archives division (items used for research support and digitised and 'owned' by the institution rather than 'created' by members of the institution). Others are storing institutional administrative or teaching materials. Coupled with this, IRs may offer the ability to include information about an individual's 'esteem,' ie information about research grants awarded, the numbers of PhD students supervised and details of consultancy work. Prestige is another aspect of IRs. Services such as *Economists Online* (see <http://www.nereus4economics.info/economistsonline.html>) aim to include internationally respected experts in their fields under the one brand and in one subject area by harvesting and enhancing metadata from IRs.

Academics are producers of research output; but their role as end users is often overlooked. They may wish to re-use their own publications for a number of purposes. One obvious area is that of teaching. Having their (and their colleague's) work easily accessible in an IR means that their students are able to use it in research-led teaching activities. Other uses may be an academic's support of their doctoral students and the generation of personal CVs and bibliographies.

It would be fair to say that in many HE institutions, there are multiple databases scattered around which work independently and have different functionalities and purpose. These databases have grown piecemeal and institutions are working on methods (if they haven't already) in which information can be shared between integrated systems. True integration results in more efficient work practices and, because information only need be input and stored in one location, more accurate and up to date information. The ultimate aim at LSE is to integrate LSE RO with other school databases such as the human resources, student, publications and research databases. These systems will be centrally linked so that data such as an academic's details (for example,

name, title and department) would be stored in one location, but able to be used by other systems. It is easy to imagine how information about the items stored in the IR may be used when dealing with information, say, about research activity which required details of funding awards and publications.

Traditionally it has been libraries (and archives and museums etc.) which have provided long-term storage and preservation of publications and other research output. Academics are often surprised to learn that their publisher has no commitment to the long-term preservation of their publications. And, although making their publications available via personal and departmental websites is convenient, the authors/creators have usually not contemplated the fact that such sites are not stable in the long-term. Although the digital preservation issue is a difficult one, it is one which librarians and repository managers are at least aware of, and at best, committed to finding a means of preserving items in their IR. This commitment also implies digital curation which is necessary to ensure availability of eligible items in the future and good housekeeping of the IR.

Setting in place curation and preservation mechanisms for items held within an IR means that, providing the metadata are also preserved, the provenance of any item can be established. Related issues are the use of digital signatures and persistent identifiers which repository managers are investigating and implementing in order to support 'long-term accessibility, re-use and authenticity' (DCC, 2005).

Different publication types, multiple versions and relationships

The debate surrounding institutional repositories has generally focused on journal articles and models of paying for scholarly publishing, but IRs include many more item types. Book chapters, working and discussion papers, datasets, questionnaires, doctoral theses, conference and seminar papers, slides from presentations and any other items can be included. Because they provide a means to store, search for and access all types of research output, IRs have an important role to play for creators and users of such materials. Multimedia and teaching materials may be added, and repository managers are ensuring that metadata for all item types is of high quality and to international standards.

One particular attraction for LSE is that of being able to provide a means of discovery and access to book chapters. The social sciences produce many book chapters (sometimes in preference to journal articles) which become hidden within the complete book. Such chapters may have the added advantage of offering a tempting taster for the whole volume.

Traditional publishing has not yet fully grasped the facility that the online environment makes possible, ie that of linking and providing access to and manipulation of related relevant items⁵. Linking should not be confined to within the one work, or even a single publisher: the possibilities are infinite. Links may be to original data, dynamic digital models, primary source documents and relevant commentary as well as other related items. From another perspective, research is underpinned by questionnaires and other work which contribute to the final findings and the final publications may have been preceded by conference presentations. It may be important that the author is able to link an item to other relevant related items. Published articles usually allow a limited word count: authors may significantly have to reduce their arguments in support of their findings in order to meet the publisher's requirements. Being able to link to additional supporting evidence may be crucial for the credibility of the findings.

An interesting aspect is that of linking in **both** directions between say, articles resulting from datasets and the data themselves ie linking to the data which supports the article, but also linking from the data to any articles which have been based on them (see, for example, the StORe project at http://www.jisc.ac.uk/index.cfm?name=project_store).

⁵ An example of a publisher accepting related linked items can be found in the author instructions for the *International journal of information management*. See 'Preparation of supplementary data' at http://www.elsevier.com/wps/find/journaldescription.cws_home/30434/authorinstructions

The commonly accepted term for inter-related and composite items in the digital domain is complex digital objects. It is likely that objects will become more complex over time as authors begin to utilise the digital domain and exploit the many possibilities. Deposit into an IR which has been constructed so that such relationships between objects can be expressed and used will enable rich and imaginative uses of research, with all the different elements available to the end user.

It is common practice in academia that publications progress through a complex series of iterations until the 'final' version(s) is/are created. When considering items such as images, the same image may be stored in different formats (for example, TIFF for preservation purposes and JPEG for access). There may also be instances of small sections or thumbnails derived from the master image. All versions of one item need to be linked somehow so that users can select the appropriate version for their needs and so different versions relate to each other. If the author chooses to retain all versions, this can also provide a record of the development of the topic. There are other instances where expressing relationships is important such as linking the papers from a single conference held at an institution or a series of working papers published by a particular research group.

Branding and increased visibility

An institution's website is probably the single most popular means by which external people find out about it. For many institutions that interest is, to a large part, international, and so this medium becomes of crucial importance. Because the website is a means of attracting new students and staff and is a showcase of institutional activities, a high profile matters to the standing of the institution. Not only can external users find administrative and general information, but the website becomes a means of publicising the quality of the research generated by academics. Funding bodies and sponsors have access and press offices and outreach departments are keen to communicate an institution's strengths to the outside world. An IR is a means of addressing all these issues for the research output.

Visibility and impact of scholarly output is a main concern of most academics: that is the reason for them publishing their work as well as presenting it at conferences and seminars. However, academics themselves play a dominant role in publicising their own work, communicating their research by means of discussion lists, networking with their peers, attendance at conferences and conventions, and delivering conference and seminar papers. An IR will add to that publicity toolkit. Academics are not generally in the habit of seeking monetary reward for the publication of articles.

An IR can aid visibility of its contents because it:

- a) is searchable
- b) is likely to comply with international metadata standards to aid discovery
- c) is probably registered with search mechanisms such as OAIster (see <http://oaister.umdl.umich.edu/o/oaister/>) and Google and
- d) allows harvesting by other information services.

These reasons alone warrant setting up an IR. Repository managers are beginning to record healthy statistics of accesses and downloads to their IRs. Even at LSE, where LSE RO has been running for a mere six months or so, the statistics are encouraging (see Table 2). It is likely that increased visibility results in increased impact and citations, although this premise is the topic of other vigorous debates. Not only does an IR provide increased visibility for the deposited items, but also publicity for the journal in which the item was published, the society with which the item is associated or the book in which a chapter was published.

It is important that the author's (or creator's) affiliation is clear for each item, no matter where it is published or made available, so that the user knows where the item is emanating from and the

institution can lay claim to its output. If, as in many repositories, each item has been clearly 'branded' by the institution, any user accessing that material is aware of its pedigree. For example, LSE RO incorporates a prominent title page for each item stating the item type and status (eg peer-reviewed or draft), that it is work by an academic associated with LSE, and a citation which points the user towards the commercially published version (with no assurance that they will have permission to access it)⁶.

Compliance with requirements of external bodies

The current practice of research assessment in the UK, the RAE (Research Assessment Exercise), places enormous demands on administrative staff within HE. The process of gathering bibliographic information, selecting those records which are to be submitted, sending off the submissions to HEFCE (Higher Education Funding Council for England) and then providing full text is a major headache. Coupled with that, any academics who are reviewers in the process often resort to their own libraries at the last moment to obtain copies of submissions for review. If items were to be deposited as a matter of course in IRs, which were enabled to a) indicate records which were to be submitted (or hold them in a separate collection), b) link to HEFCE software and/or comply with submission technology and c) enable reviewers to obtain full text easily, the process would be streamlined.

At the time of writing, UK academia is awaiting the expected final position statement from RCUK regarding visibility of output resulting from research funded by any of the research councils. It is expected that RCUK will require that a copy of any publication resulting from research funded by one of the research councils is deposited in an institutional repository or similar archive, providing that deposit complies with copyright law. If this is the case, then IRs will become of increased importance for academics.

Conclusions

From the reasons given above for developing an IR, it is obvious that the main drivers are neither the downfall of publishing companies nor the demise of journals. In fact, IRs can act as a pointer to the publisher's version of an item by, for example, clearly citing any commercially published version on the title page of every item. Repository managers encourage compliance with copyright law: IRs contain materials where they are legally permitted and may, as at LSE, give a clear statement of the user's rights and responsibilities in the re-use of those items.

IRs are evolving as managers realise their potential and discover how they fit into their institutional practice. In its purest manifestation, an archive should contain items which are **all** full text and open access. In practice, if records of all research output are included, including all submissions for RAE, then record-only items will have to be accepted, resulting in a 'dark archive' of work not permitted to be freely available. It is for each individual institution to agree its policies defining the purpose of its IR and what will be accepted for deposit. The emergence of IRs will have an effect on scholarly publishing: time will tell what those effects will be. In the meantime, IRs will continue to be set up, based on the needs of scholars and their institutions.

References

DCC (Digital Curation Centre), (2005). DCC Digital Curation Manual Instalments. Persistent Identifiers. [WWW page] URL <http://www.dcc.ac.uk/resource/curation-manual/chapters/persistent-identifiers/>

Great Britain. House of Commons Science and Technology Committee, 2004. Scientific publications: Free for all? 10th report of session 2003 – 04. URL http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories

⁶ See for example http://eprints.lse.ac.uk/archive/00000288/01/Barr_2004a_OXREP040429.pdf

Prosser, D, (2004). The Next Information Revolution - How Open Access Repositories and Journals will Transform Scholarly Communication. Sparc Europe. URL <http://www.sparceurope.org/resources/SPARC%20-%20LIBER.doc>

SUBER, Peter (2003). Removing the Barriers to Research: An Introduction to Open Access for Librarians. [WWW page] URL <http://www.earlham.edu/~peters/writing/acrl.htm>. Originally published as Suber, Peter (2003). Removing the Barriers to Research: An Introduction to Open Access for Librarians. College & Research Libraries News, **64** (2), 92-94. February 2003. URL <http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2003/february1/removingbarriers.htm>

Sally Rumsey
eServices Librarian
Library
London School of Economics and Political Science
10, Portugal Street
London
WC2A 2HD
UK

s.a.rumsey@lse.ac.uk

Journal	% of items by LSE authors	Numbers of items by LSE authors
British Journal of Sociology	3%	1 article out of an average 32 published each year
British Journal of Industrial Relations	7.5%	4.5 articles out of an average 28 published each year
Economica	3%	1 article out of an average 35 published each year

Table 1: Numbers of LSE authors published in LSE journals 2002 - 2004

	June	July	Aug	Sept	Oct	Nov	Dec	TOTAL
Numbers of downloads	926	1293	1321	1409	2001	2035	2382	11,313

Table 2: Statistics for LSE RO for 2005