

Dave Puplett

The Economists Online subject repository: using institutional repositories as the foundation for international Open Access growth

**Article (Accepted version)
(Refereed)**

Original citation:

Puplett, Dave (2010) The Economists Online subject repository: using institutional repositories as the foundation for international Open Access growth. [New review of academic librarianship](#), 16 (Supp1). pp. 65-76. ISSN 1361-4533

DOI: [10.1080/13614533.2010.509490](https://doi.org/10.1080/13614533.2010.509490)

© 2010 [Taylor & Francis](#)

This version available at: <http://eprints.lse.ac.uk/29768/>

Available in LSE Research Online: October 2010

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 accepted version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this version and the published version may remain. You are advised to consult the publisher's version if you wish to cite from it.

The Economists Online Subject Repository – Using institutional repositories as the foundation for international open access growth

Dave Puplett

Information Services, Library, London School of Economics, London, England.

d.puplett@lse.ac.uk

Abstract – A new subject repository, Economists Online (EO), has recently been launched. The pioneering model upon which it is built, aggregating the subject specific content of a consortium of participating institutions and their repositories, is examined in this article. An overview of existing subject repositories is given, along with an analysis of the scholarly communications landscape in economics and how the new EO subject repository fits into this environment. This paper makes a case for collaboration between institutional repositories as a way of increasing Open Access (OA) access to research.

Keywords: subject repositories; open access; economics, institutional repositories

Introduction

Successful examples of subject repositories are rare, and limited to a few scientific disciplines. Given the thousands of institutional repositories (IRs) that now exist around the world this looks iniquitous. However, while it is widely understood who owns IRs (mostly universities and librarians), where their content comes from (academics at any given university) and who set them up (librarians, often funded by national initiatives), there are no such uniform answers if you ask the same questions of the few subject repositories (SRs) that have been established.

Subject repositories

Definitions of ‘subject repository’ are as rare as successful subject repositories themselves. The definition given in the report of JISC’s *Subject and Institutional Repositories Interactions Study* is useful:

‘Subject repository: This is a collection of research outputs with a common link to a particular subject discipline. Subject repositories are likely to cover one broad-based discipline, with contributors from many different institutions supported by a variety of

funders; the repositories themselves are likely to be funded from one or more sources within the subject community.

Deposit of content is voluntary. These repositories are usually concerned with dissemination; for example the emergence of the arXiv repository replacing the practice of circulating paper preprints in the particle physics community.’ (Jones et al, 5)

This definition does not cover bibliographies that only collect references to research outputs and not full text or content. However, because most repositories contain both full-text content and citations, citation-only bibliographies and repositories have much in common and will be examined here as appropriate. It is also not uncommon to see references to a Central Repository (CR). This could include subject based repositories, but also include funder repositories or collections based around another notion or purpose.

ArXiv, a pre-print repository for Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics looms large over the landscape of existing subject based repositories, because of its penetration in its field, its size and its longevity. It grew organically and quickly, and is widely credited with being one of the major factors behind the burgeoning Open Access movement. ArXiv has become an essential place for authors to present their work in the discipline of high energy physics, and Gentil-Beccot et al describe the benefits of those working in high-energy physics in participating in Open Access:

‘There is an immense advantage for individual authors, and for the discipline as a whole, in free and immediate circulation of ideas, resulting in a faster scientific discourse.’ (10)

Other notable subject repositories are BioMed Central, the Open Access publisher of biomedical journals, Cogprints, a cognitive sciences archive, and E-LIS for library and information science. Beyond a few other examples, there are very few subject repositories catering for academic disciplines. Economics has several examples of subject level collections that pre-date EO. Three key examples are

Research Papers in Economics (RePEc), Econstor, and the Social Science Research Network (SSRN).

RePEc describes itself as a subject bibliography, as it does not hold any content, just references. It is very successful, and has achieved a critical mass of references that have made it a key resource for economists. It is populated with references directly from authors or research groups, and holds over 315,000 working papers and links to 490,000 journals articles (based on co-operation with publishers). It is maintained with volunteers, and has no central 'home'.

EconStor contains all the content of the German National Library of Economics (ZBW) catalogue, and also working papers, discussion papers and conference proceedings from economic research institutions in Germany as well as some from the United States. It acts in a way a national subject repository for Germany,

SSRN is a scholarly repository for a variety of social science disciplines, ranging from Accounting to Economics, to Law and Political Science, etc. The service allows academic papers to be uploaded directly by authors and become available worldwide for free downloading. At the same time, SSRN includes articles and working papers from major publishers, which are available only by licence or by paying a fee. In addition to these services, SSRN has a function for ranking of institutions, authors and papers, which is available to registered users.

SSRN finances itself through fees from institutions that outsource the distribution of their research papers to SSRN, subscription fees for SSRN's subject matter abstracting journals, fees for professional and job announcements, conference fees for SSRN's Conference Management System and fees shared with publishers who distribute their papers through SSRN on a pay per download basis (Jensen).

While this cost structure allows SSRN to invest substantially in its infrastructure, the fees limit the access to SSRN services for less well-funded institutions, and thereby augments the existing dissemination gap between European and American institutions. Finally, there is a clear American bias in relation to its content and so the sharing of both European and non-English material in its data collection is limited.

On this evidence, subject repositories that contain significant amounts of full text articles are rare. Economics as a discipline has some laudable examples of citation databases and extensive bibliographies, such as RePEc, but full text articles are hard to find without subscriptions to toll access journals or services such as SSRN.

The subject repositories that do exist do not have a close relationship with institutional repositories. Content may well be duplicated, and there are no relationships where content from a SR will automatically appear in an institutional repository.

Subject repositories offer much to the respective researcher in a field. They offer a window to research as it happens; the inclusion of pre-prints, before their formal publication makes them an indispensable to the researcher looking for the latest research trends. Also, their independence from publishing houses is also a potential benefit for those wishing to see change in the mechanisms of scholarly communication. However, they are also usually patchy in coverage and depth of content, and are often reliant a subject community to sustain them – often on a volunteer basis. The particular needs of different subjects primarily shape the composition of these existing subject repositories. Appreciation of the process of research dissemination within a discipline is essential to understanding where any existing subject collections exist, and how viable a repository based on the Economists Online model would be.

Scholarly communication in economics

The economics publishing market exhibits a number of unique peculiarities. First, the economics research market is dominated by academic institutions in the United States. The top-producing departments and individuals are mainly located in the United States. Despite the considerable amounts of research output produced by European research institutions, their output is underrepresented in top academic journals. While scholars in Europe contributed approximately one quarter of the world's academic publications in the field, only one sixth of the world's citations refer to European publications (Drèze and Estevan 273). This under-representation, however, does not appear to depend primarily on the quality of the actual publications, as the top 10 European economics departments rank high in global comparison, despite considerable disagreement about ranking methods and the relative weighting of ranking criteria (see Kalaitzidakis et al, Coupé and Neary et al). This suggests that there are weaknesses in dissemination practices, access and visibility. Findings from previous studies commissioned by the European Economics Association (EEA) have revealed that there is a European market for subject-oriented information services in terms of both potential end users and under-exploited sources of academic content (Combes and Linnemer).

Second, the landscape of publication types is particular to economics. In contrast to other disciplines, books are reputed not to carry much weight in the discipline. Peer-reviewed journal articles, which are predominantly final, archival publications, therefore are the most visible and well known output of researchers (Lubrano et al 2). A relatively small number of prestigious journals with extremely low acceptance rates dominate the field. Most journal content is available electronically, and a variety of citation indices track the prestige and impact of journal

outlets and articles (Dawson and Rascoff 3). Economists experience extremely long time lags (on average two to four years, and sometimes even more) from completing a research paper to its publication. Given the relatively short 'shelf life' of economics information, journals therefore represent the body of knowledge several years after, rather than at the moment of publication. This has led to a proliferation of sharing work in progress, predominantly in the form of working and conference papers. For scholars, these papers present an important way of claiming a stake in research topics and receiving feedback before publication. These publications are scattered across different online resources, for example international databases, departmental websites and institutional repositories. While there have been some experiments with varying forms and degrees of open access journal publication, progress in this area is still limited (Harley et al 343).

Third, datasets have become of growing importance in economics research communication. Economists derive or collect data predominantly from public sources, and in fewer cases generate the data themselves. While data-sharing is still relatively limited (McCullough et al 2), some journals, such as the *Canadian Journal of Economics*, and funding bodies now mandate the release of associated datasets. Apart from data, material from some subfields of economics appear in an increasing number of popular press books, newspaper editorials or economics blogs; however, these are traditionally considered outside the realm of standard scholarship. Fourth, concerns about the high costs of commercial publications impede access to final publications for some segments of the research community, especially scholars situated in less well resourced institutions.

Economics therefore has a fertile environment for change and development in the future of scholarly communication. The existing infrastructure features some

important collections and organisations, but prior to the launch of Economists Online, the potential for open access sharing of full text articles and data has not been exploited.

Economists Online – a new subject repository

The Economists Online service, launched in January 2010, allows access to citations and a large number of full text copies of economics research carried out in some of the world's leading economics research institutions. In a pioneering development, Economists Online also contains datasets linked to the publications that reference them. EO was developed by the Network of European Economists Online (NEEO) project and is now supported by the Nereus consortium. Nereus members include libraries from around the world within academic institutions carrying out significant economics research.

EO uses a model that has previously been untried, aggregating subject specific content from IRs into one subject collection. EO achieves this by harvesting metadata and links to economics material held in the institutional repositories of the academic institutions involved in the EO service, making them cross-searchable and accessible from a central portal. This architecture exploits OAI-PMH harvesting technology, and is able to draw content from a range of different repository systems. A strength of this model is that it is easily scaleable – new IRs could be added to be harvested by EO with relative ease. Institutions who join the Nereus consortium can have their content added quickly to EO.

There are currently around 24 contributing member institutions, who provide access to a variety of material types:

- **Journal articles**, primarily the accepted versions (i.e. final author versions after peer-review) or publisher's final portable document formats (PDF) - where publishers allow;

- **Working papers**, preferably from institutional working paper series;
- **Conference proceedings**;
- **Books and book chapters**, book references and the books themselves digitally available, or parts thereof where publishers allow and where viable;
- **PhD theses**;
- **Datasets** linked to the publications that reference them.

The intuitional repositories of these institutions variable contain much more content than just economics. However, the EO harvester is able to identify via local institutional metadata only the items that are of importance to economics. By harvesting all of this subject-specific material into one collection, Economists Online provides an interface that is deigned to service the economics research community specifically – something that the broader institutional repositories cannot do as specifically. To meet this opportunity, Economists Online features a number of facilities requested by the research community when the portal was being developed:

- Simple and advanced search options;
- Facets, to refine search results;
- Multilingual searching (Spanish, German, French, English);
- RSS feeds;
- Export search results (e.g. for storage in a reference manager);
- Journal of Economic Literature (JEL) code searching;
- Author publication lists.

The development of EO was funded by the European Commission's eContentplus programme. Library staff in the NEEO project member institutions have led the development of EO, and these library staff consulted their economics faculties regularly for guidance, using both focus groups and surveys, to ensure that authors were happy with how their work is presented in the portal and that the portal would include features that are useful to them as researchers.

Alongside the material harvested from the EO contributing members the portal contains the contents entire RePEc economics database. This is a very significant benefit for economics researchers, who can use the advanced features of the EO

interface to search all of RePEc's content as well and the new material that EO has made open access. Material harvested by EO is also preserved. The Nereus consortium has concluded a landmark agreement with the Koninklijke Bibliotheek (the National Library of the Netherlands) for the preservation of the content available via EO.

This model allows provides a high quality, tailored service for economists in the form of the portal itself, and relies upon the submission of content to be managed at a institutional level. This is a clear departure from previous subject repository collections, and relies upon local mechanisms for collecting information about publications and research.

A new library-led model for subject repositories

Why are librarians, and the institutional repositories that they have built and populated, well placed to found a subject repository upon?

There are several reasons, but most salient is that the library profession has embraced the Open Access movement. Unfettered access to academic research is a goal that university librarians have been working hard to attain. The work invested in developing IRs in universities around the world demonstrates the power of their conviction in OA, and forms a stable foundation to build upon. Charles Bailey Jr, describes libraries as

...no longer simply consumers of scholarly information. A growing number of libraries have become digital publishers, primarily offering free/open access journals and institutional repositories. (21)

The role that IRs play at the institutional level is also changing, and repositories are becoming authoritative sources of information about institution's research outputs. Many UK universities have used information from their IR for their

Research Assessment Exercise (RAE) submission and many other institutions, such as the University of Southampton, use their IR as a publication database (Simpson and Hey 5).

Subject repositories generally rely on submissions by individuals or bodies. The EO subject repository model relies upon content in IRs, something that there is strong evidence will continue to grow and develop. The OpenDOAR website has reported consistent growth in its database of IRs in existence over the last 5-10 years, and as more IRs contain more content, critical mass will develop, making IR's an even stronger source of information about research outputs.

This new model for populating a subject repository means that authors only need to deposit their work in one repository for it to be visible in others, exploiting the interoperability of the OAI-PMH protocol. Studies show that authors are highly unlikely to deposit their work in more than one place:

'evidence reveals that when an article has been presented in one repository, the author(s) will be hesitant to make it repeatedly available in a second repository. (Xia 494)

Another major benefit of a leading role for librarians in this sort of development is their commitment to quality metadata. As popular as the subject repositories and bibliographies like ArXiv and RePEc are, one area that they lack is high quality, accurate and descriptive metadata. The importance of this information, especially in the linked environment of the web, is increasingly important. Librarians recognise this and have built repositories that use rich metadata schemas and the ability to share this rich information. The metadata that EO contains originates from the host IR, and is then enriched by the EO service to include information like JEL codes.

IRs are also strongly focussed on benefits for end users, and not just the authors who contribute to them. In the DRIVER's Guide to European Repositories, it is claimed that the:

“...greatest opportunities for abundant and diverse services are where these relate to user needs (such as) discipline or subject-specific portals” (Weenink et al 44)

Economists Online fundamentally relies upon IRs that provide its foundations, and the EO model is predicated on the success and future of repositories at an institutional level. However, where there is the risk of duplication of effort content between institutional and existing subject repositories (Harnad), SR's using the EO model can only benefit the institutional repositories that they harvest. Institutional repositories hardly need any configuration to be successfully harvested into a SR in this infrastructure, and continue to hold and full text items on their servers, meaning that downloads will still be counted by the individual institutions. Institutions will receive more traffic to the repositories because of the greater exposure their content will receive by being exposed in a subject level repository like EO. Authors will also be able to deposit locally , knowing that their work will then be harvested to appropriate collections like EO. This type of distributed deposit, rather than a centralised model used by other subject repositories, utilises the enormous expertise being developed by the growing institutional repository movement in libraries and institutions around the world.

Conclusions

Is this new model sustainable in the long term? To be truly sustainable, the identity of the subject repository comes from the subject, not from the contributing members. No existing repository can yet claim to be a complete representation of research in a subject (including all types of material, from papers to datasets). A subject repository

is defined by its content and those who supply that content, and while a SR based on the EO model could survive the loss of a few contributing members, critical mass of content is essential to their success. Subject repositories should strive to transcend the individual sources of its material if they are to become relevant at an international, subject level.

That is not to say that the reputation for quality in such a subject repository does not rely upon authoritative sources. Ironically however, subject repositories built on this model could turn out to be less sustainable than individual IRs, because of the lack of a specific body to run them. EO is the first subject repository to be managed by an international consortium of libraries like Nereus, and this model will now be tested in particularly hard economic times.

While the scope for the role of IRs is clearly established, and the case for individual institutions to support them has been proven, it remains to be seen whether subject repositories built on top of this stability will benefit from the same support. Ultimately, engagement with researchers in a given field of research is key to the success of a SR, and EO is the first international SR by expert library staff for the benefit of a particular field.

Economists Online therefore marks an exciting moment in the history of the Open Access movement. Leveraging the potential of repositories set up by libraries from around the world, it brings a unfettered access to a great deal of excellent economics research to a wider audience than ever before. In the fast changing world of scholarly communication, SRs that harvest full text content, datasets and citations from existing IRs can play a key role in the dissemination of research.

Acknowledgements

Thanks to Tobias Thiel, PhD candidate at LSE who contributed material for the overview of the economics scholarly communication environment. Also thanks to Hans Geleijnse, Tilburg University, from whom analysis of the existing economics bibliographies and was taken.

References

- Bailey, C.W.. *Open Access and Libraries*, Collection Management 32.3 (2008). 09 Apr. 2010
- Coupé, T.. *Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990-2000*. Journal of the European Economic Association 1 (6) (2003): 1309-1345.
- Combes, P.P. and Linnemer, L.. *Where are the Economists who Publish? Publication Concentration and Rankings in Europe based on Cumulative Publications*. Journal of the European Economic Association 1(6) (2003): 1250–1308.
- Dawson, M., Rascoff, M. (2006). *Scholarly Communications in the Economics Discipline. A Report Commissioned by JSTOR, Ithaka Strategic Services*.
- Drèze, J.H. and Estevan, F. *Research and Higher Education in Economics: Can we deliver the Lisbon objectives?* Journal of the European Economic Association 5 (2-3) (2007): 271-304.
- Gentil-Beccot, A and Mele, S and Brooks, T. *Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories*. ArXiv. 25 Nov 2009. Web 15 Mar 2010
- Harley, D., Acord, S.K., Earl-Novell, S., Lawrence, S., King, C.J.. *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. UC Berkeley: Center for Studies in Higher Education. (2010)
- Harnad, S. *Central versus institutional self-archiving*. Technical Report , ECS, University of Southampton. (Unpublished) (2006). Web. 20 June 2010
- Jensen, M.C. *SSRN's Objectives and Commitments to Users*. Social Science Research Network. 2007. Web. 15 Mar 2010
- Jones, C and Darby, R and Gilbert, L and Lambert, S. *Report of the Subject and Institutional Repositories Interactions Study*. JISC. Nov 2008. Web. 15 Mar 2010
- Kalaitzidakis, P., Mamuneas, T.P., and Stengos, T.. *Rankings of Academic journals and Institutions in Economics*. Journal of the European Economic Association 1 (6) (2003): 1346-1366.
- Lubrano, M., Bauwens, L., Kirman, A., Protopopescu, C.. *Ranking Economics Departments in Europe: A statistical approach*. Journal of the European Economic Association, 1 (6) (2003): 1367-1401.
- McCullough, B.D., McGeary, Kerry Anne, Harrison, Teresa D. *Do economics journal archives promote replicable research?* Canadian Journal of Economics, 41, 4, (2008): 1406-1420
- Neary, P.J., Mirrlees, J.A., Tirole, J.. *Evaluating Economics Research in Europe: An Introduction*. Journal of the European Economic Association 1(6) (2003):1239–1249.
- Simpson, P and Hey, J *Repositories for research: Southampton's evolving role in the knowledge cycle*.

Program, 40 (3) (2006). 224-231.

Weenink, K. and Waaijers, L. and van Godtsenhoven, K. (editors). *A DRIVER's Guide to European Repositories*, Amsterdam University Press. Web. 15 Mar 2010

Xia, J. *A Comparison of Subject and Institutional Repositories in Self-archiving Practices*, *The Journal of Academic Librarianship*, 34 (6) (2008), 489-495.