To be published in *Program August 2006*

Repositories for research: Southampton's evolving role in the knowledge cycle

Pauline Simpson and Jessie Hey

Author 1: Pauline Simpson is Head of Information Services and Project Manager TARDis Project, National Oceanography Centre, University of Southampton, UK. E-mail: ps@noc.soton.ac.uk

Author 2: Jessie Hey is Research Fellow, University of Southampton Libraries and School of Electronic and Computer Sciences, University of Southampton, UK. E-mail: jmnh@ecs.soton.ac.uk

Abstract

Purpose – To provide an overview of how open access repositories have grown to take a premier place in the e-Research knowledge cycle and offer Southampton's route from project to sustainable institutional repository.

Design/methodology/approach – The evolution of institutional repositories and open access is outlined raising questions of multiplicity of repository choice for the researcher. A case study of the University of Southampton Research Repository (e-Prints Soton) route to sustainability is explored with a description of a new project that will contribute to e-Research by linking text and data.

Findings – A model for IR sustainability.

Originality/value – The TARDis Project was one of the first IRs to achieve central university funding in the UK. Combined with increased visibility and citation, the Research Assessment Exercise route has become the 'hook' on which a number of IRs are basing their business models.

Keywords: Institutional repositories; E-print archives; Digital content repositories; Open Access; TARDis; CLADDIER.

Paper type: Case study

Word count: 2,442

1. Introduction

Subject-based repositories of e-prints were pioneered in 1991 by Paul Ginsparg at the Los Alamos National Research Laboratory in New Mexico with a collection of preprints of articles in the subject area of high energy physics. This collection, known as arXiv, is now based at Cornell University (http://arxiv.org) and has grown to include materials in Atmospheric and Oceanic Physics, Mathematics, Computer Science and Quantitative Biology. In 1995, Stevan Harnad (who is a professor in the School of Electronic and Computer Science at the University of Southampton) made his 'subversive proposal' leading to the open access vision for scholarly material – the idea being that, in an ideal world of scholarly communication, all research should be freely available (Harnad,

1995). From this open access vision, repositories have developed from being subject based to include the complementary institutional-based model and their growth has been fuelled by timely project funding from a variety of sources. This welcome increasing number also creates a dilemma of repository choices for the researcher. However, despite the success of arXiv and others like RePEc for the Economics community (http://repec.org), there has been only varying success in other subject communities and some have even been terminated (e.g. Chemistry Preprints (http://www.sciencedirect.com/preprintarchive).

The University of Southampton has been involved with Open Access (OA) since its inception and benefited from project funding to implement its own institutional research repository. The Targeting Academic Research for Deposit and Disclosure (TARDis) project at Southampton was funded as part of the Joint Information System Committee (JISC)'s Focus on Access to Institutional Repositories (FAIR) Programme. TARDis was influenced by internal and external drivers and has now evolved into a university funded service which additionally provides the underpinning management for the Research Assessment Exercise (RAE) for 2008 (http://tardis.eprints.org). The changing paradigms for repositories has given the movement new horizons that include IRs becoming one of the building blocks of e-Research.

2. Open Access and repositories

Open Access literature is digital, online, free of charge and free of most copyright and licensing restrictions. Ideally, full text should be available immediately, rather than through delayed mode. However, although some of the vocabulary has changed, authors still write, review, journal edit and transfer copyright for free whilst we still have the problem of rising journal subscriptions far outstripping the Retail Price Index.

Over the ensuing years, two complementary solutions have evolved :

- Open Access Repositories where articles, conference papers, books, book sections, reports, theses, learning objects and multimedia are deposited in open electronic archives which conform to the Open Archives Initiative (OAI) standards.
- Open Access Journals where publishers do not charge subscriptions or online access fees but instead look to other publishing models, including author pays for publication, to ensure OA.

From 2000 onwards a complementary implementation of Institutional Repositories worldwide began, powered by the information community and fuelled by project funding from bodies such as the Andrew Mellon Foundation, Howard Hughes Institute, and the Open Society Institute and, in the UK, by the JISC FAIR Programme (http://www.jisc.ac.uk/index.cfm?name=programme_fair).

The University of Southampton carried out a survey of IRs in 2002 for the purposes of finding out if subject classification schemes were in general use (http://tardis.eprints.org/discussion/) and at that time there were 112 IRs around the world (of varying 'flavours'). At the time of writing, in March 2006, checking the

Registry of Open Access Repositories (ROAR - previously Institution Archives Registry) there are some 637 IRs, with the USA and UK at the top of the list, but many countries are now catching up (http://archives.eprints.org/). Table I gives details of some of the 40 countries included in ROAR.

Take in Table I

Table I Number of IRs in specific countries as provided on ROAR (March 2006)

US	176
UK	68
Germany	60
Brazil	42
Canada	32
France	29
Australia	25
Sweden	25
Italy	22
Netherlands	18
India	15

Alongside the ROAR is the newly developed Directory of Open Access Repositories – OpenDOAR (http://www.opendoar.org/) which also evidences the increasing number and diversity of repositories: subject, institutional, national, national/subject, international, regional, consortia, funding agency, project, conference, personal, media-centric, publisher and data archives. The dilemma for the researcher depositor is that the above are not mutually exclusive; there is a problem of repository choice.

Mandates from funding agencies, such as the Wellcome Trust (http://www.wellcome.ac.uk/doc_WTX025197.html), and the much awaited final version of the draft Research Councils UK (RCUK) Position Statement (http://www.rcuk.ac.uk/access/statement.pdf) will require grantees to deposit funded research output in a 'designated repository' - but in which one should the researcher chose to deposit the full text? A researcher wants to enter metadata and deposit full text only once but may want the record to be in more than one repository At present it is not possible to target multiple repositories with one exercise and this is achieved by harvesting (but the harvester is not the choice of the depositor) or duplicate keying of metadata into repositories of choice. The question is perhaps whether it matters where a researcher deposits, since search engines like Google, Google Scholar, Yahoo and Scopus will pick it up wherever it is. On mailing lists, such as AMERICAN-SCIENTIST-OPEN-ACCESS-FORUM@LISTSERVER.SIGMAXI.ORG, and the new JISC-REPOSITORIES@JISCMAIL.AC.UK, discussion on deposit in subject or institutional repository has long been debated. I venture to put the case for deposit in Institutional Repositories.

Subject-based or project repositories are often linked to an individual or group which can be transitory. Collections can be at risk and sustainability an issue. Institutions are the logical implementers of repositories because they can take responsibility for centralising a distributed activity, provide the framework, infrastructure and permanence to sustain change. They have an acknowledged responsibility for stewardship, including preservation of their digital assets, and for providing a showcase for the research, teaching and scholarship of the institution.

3. The TARDis project

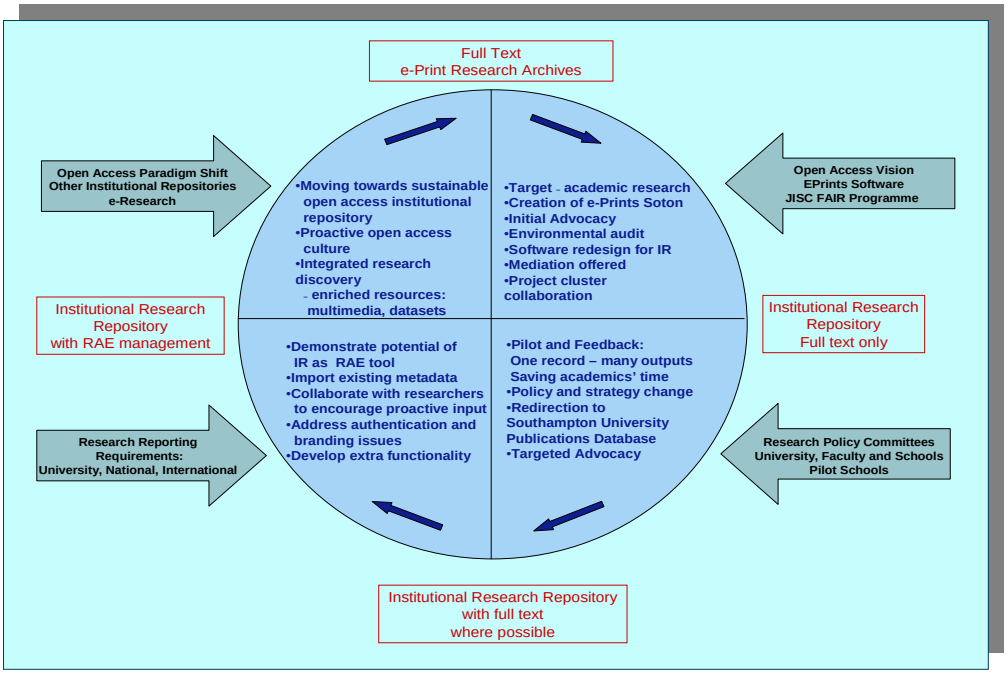
The University of Southampton's TARDis Project (http://tardis.eprints.org) implemented its institutional research repository - e-Prints Soton (http://eprints.soton.ac.uk) with funding from the JISC FAIR Programme, along with other universities such as Glasgow, Nottingham and Cambridge. It was a collaborative project with the University Library, Information Systems Service, the School of Electronics and Computer Science and, of course, the academics. However, the project was based on a long history of Southampton open access support from the following:

- Stevan Harnad
- EPrint software development. The EPrints open source software (http://www.eprints.org/software/) was developed at Southampton and is used by many IRs throughout the world.
- early adoption of an IR by the School of Electronics and Computer Science as well as the National Oceanography Centre.

The TARDis Project targeted academic research for its IR in its first stage, as a manageable goal with key benefits for the institution. The implementation of the Southampton University Research Repository followed a route based on studying current practices and needs and on acting on feedback from both the institution and individual faculty members. The series of steps which were taken to build a framework (or a route map) for a sustainable repository for a large multidisciplinary institution is illustrated in Figure 1.

Take in Figure 1

Figure 1 The TARDis Route Map



The TARDis Route Map was an effective model to showcase the research of all disciplines not just in science, technology and medicine but also in the arts and humanities where the range of publication types is quite different. The Route Map demonstrates the key interactions that have influenced the development and the strategic direction of the Southampton University Research Repository (e-Prints Soton) which we believe will lead to open access to research results, in a sustainable way.

4. The University of Southampton Research Repository

The original intent at Southampton was to provide a full-text publications database in the spirit of open access following the pioneering work already done at Southampton and as envisioned by the FAIR Programme. The pilot was set up and demonstrated and current practices investigated. The University, however, had a tradition of recording publications for research assessment and for promotion of the University. Although there was a need to update the mechanism for obtaining this metadata it was made evident to the project that the University would encourage the 'repository' principle provided the publications recording could be improved and authors would not have to duplicate effort. This led to a distinct change in policy to create a publications database with the capacity to add full text as academics felt comfortable with copyright and became familiar with the deposit process. The next phase involved more targeted advocacy so that the model was developed with the specific needs of the different 'schools' in mind by close dialogue with academics and research managers. The development moved through the first and second quadrants of the circle in Figure 1.

The third phase involves more detailed thinking about research reporting whether at individual or group level or university and national level. The Research Assessment Exercise (RAE) – so core to the UK environment – was likely to be of particular

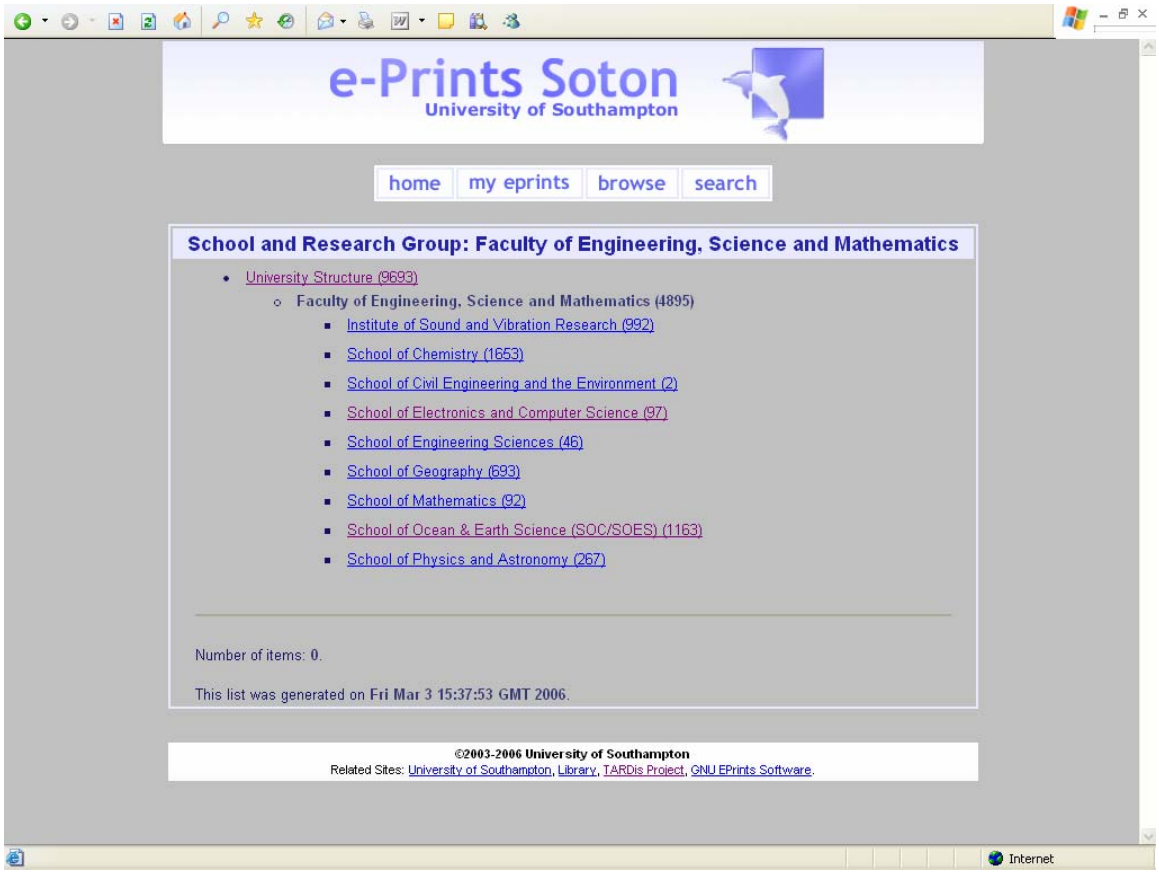
importance. There was a fundamental need to improve a process which in the past had produced warehouses full of papers which first had to be gathered (Day, 2005). The next stage was to demonstrate on a pilot database the input of publications which could be selected or deselected along with measures of esteem such as involvement in conferences. This work on an RAE module has now been made available to all universities through the Institutional Repositories and the Research Environment (IRRA) Project (http://irra.eprints.org). Depositing metadata and preferably full text where possible for the RAE gives a strong incentive to authors and encourages familiarity with the process of deposit. Groups which are doing this then frequently begin to add other material. Along with an ever more positive external environment and other universities creating their own repositories, we have an appropriate climate to build up a practice of open access in a sustainable fashion.

In December 2004, Southampton University issued a press release to announce the decision to provide core funding for its IR. This established it as a central part of its research infrastructure, managed by the University Library.

The Southampton Research Repository now offers enhanced visibility to the research profile of the university, schools and academics, but at the same time the one record provides for the population of individual CVs, Web pages, and school RSS (Really Simple Syndication) feeds. It provides secure storage of the full text of research publications and other outputs, both current and legacy, which can be repurposed for learning and teaching and importantly underpins all the research reporting requirements. Figure 2 shows a screenshot of the departments within the Faculty of Engineering, Science and Mathematics and the numbers of items included within the IR per department, for example 1163 in the School of Ocean and Earth Sciences.

Take in Figure 2

Figure 2 A browse of e-Prints Soton



5. Building on TARDis and e-Research

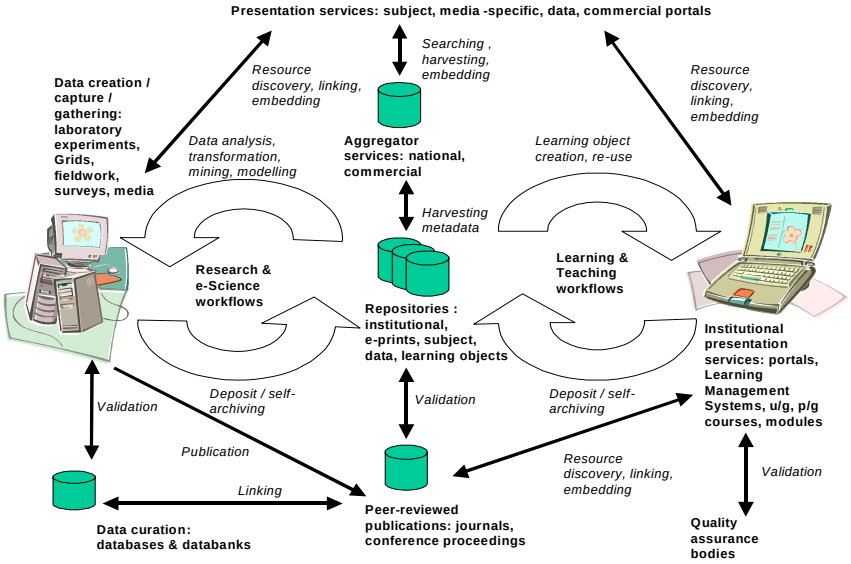
The TARDis Project was successfully completed during 2005 and we are now building on the implementation of the Southampton Research Repository which has provided a platform on which to contribute to further JISC Digital Repositories Projects:

- Preservation Services for EPrints – Preserv (http://preserv.eprints.org/)
- Scoping a Geospatial Repository for Academic Deposit and Extraction – GRADE (http://edina.ac.uk/projects/grade/)
- Citation, Location and Deposition in Discipline and Institutional Repositories- CLADDIER (http://claddier.badc.ac.uk/).

Within information cyberspace, IRs now contribute to the vision of 'joined up research'. or e-Research. In the digital world the availability of original data, together with the ability to track its use in subsequent research work, scholarly publications or learning materials, will have a significant long-term impact on the whole scholarly knowledge cycle as can be seen in Figure 3.

Take in Figure 3.

Figure 3 The Scholarly Knowledge Cycle

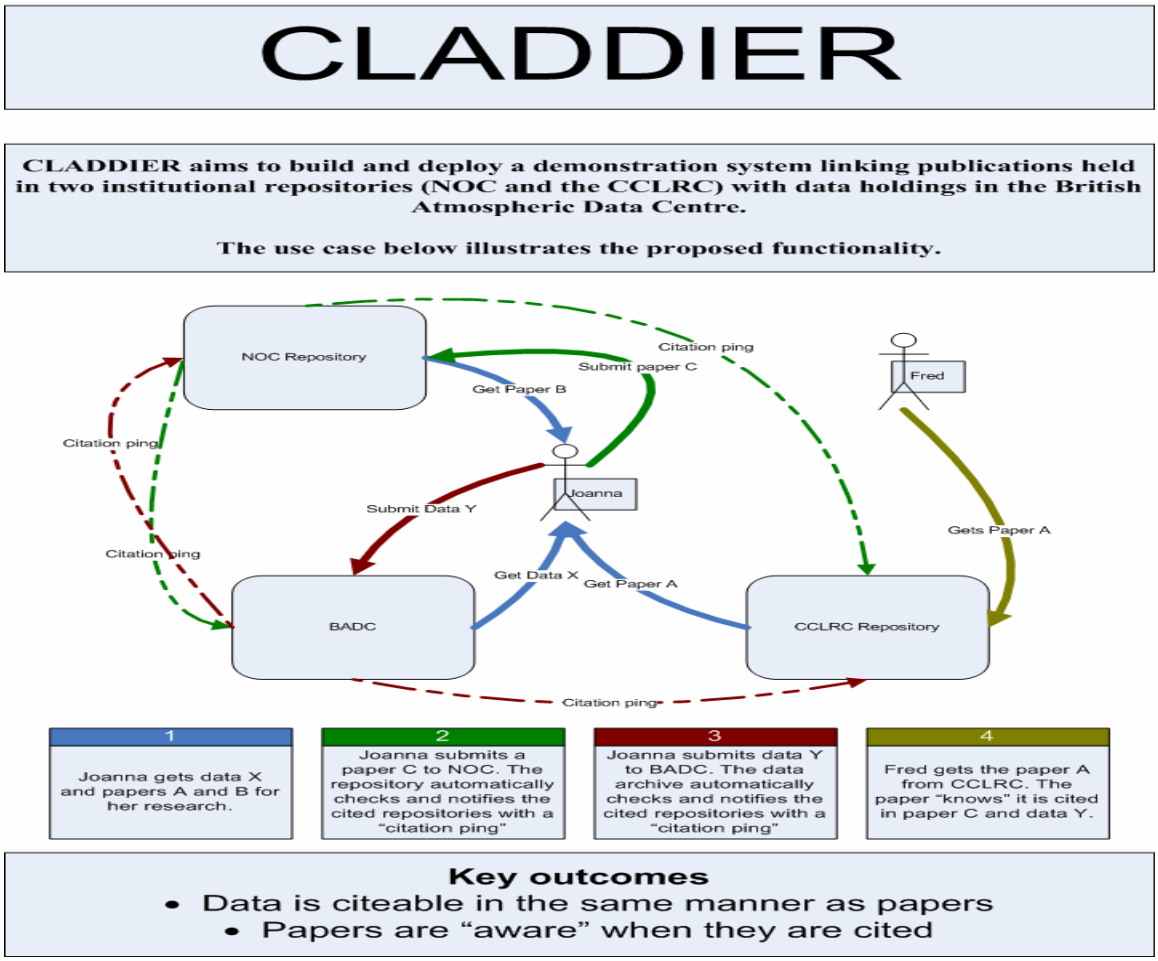


(Reproduced with permission, Lyon 2003.)

CLADDIER links e-Research, encompassing experimentation, analysis, publication, research and learning. The CLADDIER system will be a step on the road to linking text and data in a situation where researchers (which in this project are environmental scientists) will be able to move seamlessly from information discovery (location), through acquisition to deposition of new material with all the digital objects correctly identified and cited as shown schematically in Figure 4.

Take in Figure 4

Figure 4 CLADDIER Schematic



An unexpected discussion area within the CLADDIER Project which could be an important, outcome for IRs, is that as well as harvesting ('pulling' the record from other repositories), it could be possible to 'push' the record. This would enable the researcher to deposit in an IR and choose to upload (push) the metadata to other repositories of choice with the link back to the full text in the IR. There may be redundancy of record but no longer would there be a need for the debate about where researchers should deposit - institutional or subject-based repository.

The TARDis Project completed its transition to invisibility in 2005, but the sequel is -- back to the future!

Editor's Note

This paper is an updated version of a presentation given at the Joint Internet-Based Sources (JIBS) User Group meeting "Are institutional repositories taking over the world?" held at the British Geological Survey, in September 2005. Available at: http://www.jibs.ac.uk/meetings/workshops/repositories/SimpsonJIBS2005.pdf as well as at: http://eprints.soton.ac.uk/19339/.

References (All URLs were checked 25 March 2006)

Day, M. (2004), *Institutional Repositories and Research Assessment. Supporting Study No. 4.* UKOLN, ePrints UK Project, Bath. Available at: http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/rae/rae-study.pdf.

Harnad, S. (1995), "A subversive proposal", in: Okerson, A. and O'Donnell, J. (eds), *Scholarly Journals at the Crossroads*. Association of Research Libraries, Washington DC. Available at: http://www.ecs.soton.ac.uk/~harnad/subvert.html

Lyon, L. (2003), "eBank UK: building the links between research data, scholarly communication and learning", *Ariadne*, No.36. Available at: http://www.ariadne.ac.uk/issue36/lyon/intro.html