The EPrints story: Southampton as the cradle of institutional self-archiving

Die EPrints-Story: Southampton – die Wiege der institutionellen Selbstarchivierung

Abstract

The idea of institutional self-archiving was conceived 15 years ago, has gained momentum and is here to stay. It is inseparably linked to one of the movement’s pioneers, the University of Southampton, UK. This article is about Southampton’s tireless involvement in the self-archiving movement and describes the history of EPrints, the first freely available software designed to set up an institutional repository. EPrints has since evolved into Open Source software for creating most powerful information hubs and remains one of the most used repository platforms worldwide. Institutions with limited knowledge or personnel have the additional choice to use the not-for-profit EPrints Services to get their EPrints repository up and running.

Keywords: EPrints, University of Southampton, School of Electronics and Computer Science, self-archiving, institutional repository, digital repository, EPrints Services

The EPrints story

The grass-roots of institutional self-archiving

Revolutions are inseparably connected with charismatic personalities. This is also true for the Open Access movement in scholarly communication. One of its most famous advocates and promoters is Stevan Harnad [1], Cognitive Scientist at the University of Southampton, UK [2] and UQAM, CA. Being an experienced researcher as well as an editor in traditional scientific publishing he knows both worlds as few others. Apart from his substantial scientific output Harnad is recognized for pioneering new ways of electronic publishing. He created one of the first peer-reviewed scientific electronic journals titled Psycoloquy [3] in 1990 and furthermore introduced the

Christian Gumpenberger

1 EPrints Services, University of Southampton, Southampton, UK

Zusammenfassung


Schlüsselwörter: EPrints, Universität Southampton, School of Electronics and Computer Science, Selbstarchivierung, institutionelles Repositorium, digitales Repositorium, EPrints Services
concept of “open peer review”, whereby publications are directly enhanced with expert comments to stimulate discussion.

In June 1994 – at this time centralized, discipline-based archives like arXiv or CERN preprint server have already been launched – Harnad laid the foundation for what would eventually result in institutional self-archiving. His “Subversive Proposal” [4] was posted to a discussion list devoted to electronic journals. It suggested that scientists should simply self-archive all of their publications solely written for research impact in a local electronic archive. For this type of publications Harnad coined the term of “esoteric knowledge”, meaning the ‘not-for-profit’ academic publications produced by a small group of scientists and restricted to a limited readership. The proposal was of course heavily discussed by proponents and opponents alike, and Harnad himself revisited his ideas a decade later [5], pointing out achieved objectives as well as missed marks. He made it clear that “Self-archiving Proposal” would have been a far more appropriate term, and that the main issue always was (and still is) achieving the maximized research impact by a maximized toll-free online access for all. The term “Open Access” did not exist yet, but the principles did.

From individuality to interoperability

1999 is the next milestone in the history of institutional self-archiving. Paul Ginsparg, Rick Luce and Herbert Van de Sompel founded the Open Archives Initiative (OAI) [6] in order to overcome the lack of interoperability between all the gradually evolving individual self-archiving solutions. The aim was to make the archives visible and searchable collectively rather than individually. In October 1999 the OAI founders held their first successful meeting in Santa Fe [7] for computer scientists, digital librarians and representatives of either existing or emerging e-print archives and of their sponsors. The outcome was the so-called Santa Fe Convention [8], presenting a technical and organizational framework designed to facilitate retrieval of content deposited in individual archives, and enabling a communication mechanism between data providers and service providers. The convention aimed to attract early adoption of already existing archives and also encourage the establishment of new archives supporting the agreed mechanisms. Amongst the meeting participants were influential representatives of existing archives stating their intention to become early adopters. Unsurprisingly Stevan Harnad was one of them and declared to make CogPrints [9] Santa Fe-compliant (the Cognitive Sciences EPrint Archive was modelled after the Los Alamos arXiv, designed by ECS [10] doctoral student Matt Hemus, was launched by Stevan Harnad in 1997 and has been hosted since then by the University of Southampton).

The transformation of CogPrints to EPrints

A welcome side effect of the Santa Fe meeting was the conclusion that the world lacked an easy to implement and free software to create interoperable e-print archives. Consequently Stevan Harnad’s intention was not only to make CogPrints Santa Fe-compliant, but to transform it into a generic software that could be installed by any interested institution for free to create an institutional open archive for all of its disciplines. The transformation was supported by the Joint Information Systems Committee (JISC) [11] in the Open Citation Project [12], a collaboration between the University of Southampton, Cornell University and arXiv. In 2000 it led to the release of the first software called EPrints [13] designed to create an institutional repository (IR), and developed by Robert Tansley (likewise an ECS doctoral student, later recruited by MIT and Hewlett Packard to create DSpace). This was the start of an unprecedented success story, as EPrints would be rapidly adopted by many institutions worldwide. Only in 2001 an OAI-PMH [14] compliant EPrints version was announced immediately after the new protocol’s emergence, and in 2002 an open source EPrints v2.0, now developed by Christopher Gutteridge and affiliated with GNU [15], was released [16].

Southampton University goes Open Access

Throughout the development of the EPrints software JISC has proven to be a strong supporter. EPrints has been deployed in many JISC application projects, such as SHERPA (Securing a Hybrid Environment for Research Preservation and Access) [17] and TARDIs (Targeting Academic Research for Deposit and Disclosure) [18]. The latter project’s goal was to build a sustainable multidisciplinary institutional repository based on EPrints for the University of Southampton [19], which led to the launch of ePrints Soton (the University of Southampton Research Repository) [20] in 2004. At the end of the same year the University of Southampton announced that it would make all of its academic and scientific research output freely available to everybody worldwide [21]. ePrints Soton was regarded as a key tool to promote the institution’s scientific accomplishments and now became a joint service of the university library and the School of Electronics and Computer Science (ECS), the latter having hosted the EPrints development team and having run their own departmental ECS EPrints Repository [22] successfully for a couple of years already. Remarkably ECS was the first academic department worldwide to adopt a self-archiving mandate [23] in January 2003. The press release of December 2004 [21] emphasized once again the University of Southampton’s pioneering status in the self-archiving movement, as it became the very first institution in the UK to make the institutional
repository an integral part of the university’s research infrastructure. In order to attract high submission rates, but also to guarantee compliance with publisher self-archiving policies, Stevan Harnad came up with the ID/OA (immediate deposit, optional access) approach [24], which is embedded in Southampton’s self-archiving policy [25]. Full text and metadata need to be deposited for a complete institutional record keeping and performance evaluation, whereas open access to the full text is optional (originally it was required), but of course highly encouraged.

According to a study most authors are willing to comply with self-archiving mandates [26], and all the institutions following Southampton’s example prove to run far more successful repositories than the ones with only optional policies. As the initiator of the self-archiving mandate the University of Southampton keeps track of worldwide institutional (and also funder) self-archiving policies in its ROARMAP (Registry of Open Access Repository Material Archiving Policies) [27].

From funding to sustainability: a virtuous cycle model

Like every other popular open source software package, EPrints faced the problems of uncertain funding for further development. An increasing number of institutions relying on EPrints were certainly appreciated; however, this also meant a big responsibility to ensure future availability and support. Free software definitely did not produce any income, and obviously JISC could not continue to fund the development indefinitely. Thus at the end of 2004 it was decided to put the EPrints software into the user community. JISC was willing to fund this project called “Community EPrints” through 2005–2007 as an exit strategy [28].

Early in the project the so called EPrints Services [29] were launched, an ECS spin-off organisation offering commercial support. This launch directly affected the anticipated future community development. In spite of the fact the project failed to meet the initial community idea, it nevertheless produced the release of the EPrints v3.0 under the responsibility of lead developer Christopher Gutteridge. This new version was launched in January 2007 (augmented by v3.1 in 2008) and praised as “a significant milestone towards ideal repository software” by Peter Millington and William J. Nixon [30].

EPrints has been financially self-sustaining since the JISC funding ceased two years ago. The revenue generated by providing EPrints Services directly supports the EPrints team on a not-for-profit basis. Institutions or organisations which benefit from the paid-for services guarantee the sustainability of EPrints for the long term.

The EPrints digital repository solution

A broad scope

Almost a decade after its first release in 2000, EPrints has evolved into a piece of software to create powerful information hubs. Initially developed to simply follow the Open Archives Initiative and foster the dissemination of research publications, its scope is now much broader. It can be used for e-science (publications and data sets), resource discovery and delivery, e-learning, research management, research assessment, preservation, conferences and marketing. Therefore a digital EPrints repository appeals to academic, educational and corporate institutions or organisations alike, be it individually or in a consortium.

EPrints is OAI-PMH compliant and comes with a comprehensive set of metadata for a wide range of digital objects including publications, scientific datasets, teaching resources, multimedia items and artistic performances. These metadata are extensible, so repository managers can edit existing or add new metadata fields and define new kinds of material for deposit. Content can easily be moved in and out for reuse, which makes EPrints highly interoperable.

Captivating simplicity and flexibility

A powerful solution does not necessarily need to be complicated. EPrints has been designed with a focus on simplicity for all the involved parties: the developers, the repository managers and the users. Arthur Sale stated: “World’s best practice for an institution commencing an institutional repository is EPrints.” [31]. The software’s core code is written and maintained by a small team of developers in Southampton following a single coding standard. This core takes care of the basic functionality, whereas user interactions with the repository are handled with plugins. Thus the behaviour of the repository can simply be tailored by adding, deleting or substituting alternate plugins. Such plugins can be used for import and export functionalities, interface screens or the creation of input components.

EPrints installation, configuration and management tasks are straight-forward and can be run from a conventional web browser, minimizing the dependence of repository managers on the responsiveness of technical support stuff. Easy to use interfaces allow the change of the templates and schemas that control the appearance and behaviour of the repository at any time.

The EPrints workflow process has been designed as intuitive and user-friendly as possible, in order to encourage further take-up of data creators’ self-deposits. This workflow consists of various default deposit stages (Type → Upload → Details → Subjects → Deposit), but again EPrints is designed to be flexible so repository managers can customise the workflow (add or delete or rename workflow stages) to reflect the individual needs of the users. Depositors definitely appreciate the auto comple-
vation features for vital metadata such as journal title, ISSN and author names. Apart from the user-friendly aspects of auto completion it also helps to ensure the consistency and accuracy of the repository’s metadata. Other popular features are an embargo option for restricted content that cannot immediately be made publicly available (including a “Request a copy”-button initiating e-mail requests to the author of a restricted item), thumbnails of images or PDF front covers on the record display view (enlarged views triggered by mouse-overs), options to turn completed searches into RSS feeds, fast repository population via multiple import formats (e.g. DOI, PubMed ID, XML, etc.) & batch importing, a history tab to track the history of deposits, and a quality assurance tab to highlight potential issues with the eprint.

Guaranteed support and expertise

The open source concept of EPrints allows anyone who wishes to set up an EPrints repository to download the software for free. It works “out of the box”, and people with some knowledge of Perl will find EPrints straightforward to customise. The community-based technical help list [32] as well as the EPrints Wiki [33] is monitored by the EPrints team and advice is freely offered. Despite its simplicity not all institutions or organisations have the necessary resources or expertise to develop their repository in-house. Sometimes it can be more cost-effective and less time-consuming to outsource the processes of building, customising and even hosting a repository. EPrints Services exist exactly to meet the needs of such customers and offers

- Development of new repositories
- Migration of legacy data or whole repositories
- Customising repositories with bespoke functionality
- Branding repositories according to the customer’s corporate identity
- Hosting of repositories, including full security, backup and upgrading services
- Two levels of technical support with specified email and telephone response times
- Advice on advocacy and policy issues as well as on strategies to increase deposit and usage rates of the repository
- Professional Project Management

The EPrints Services team members work alongside the developers of the core software. This advantage provides EPrints Services customers with the reassurance that their repositories are built professionally and efficiently. Full use can be made of the functionality embedded within EPrints, and bespoke plug-ins are written directly by technical EPrints experts to meet particular customer requirements.

EPrints in the Life Sciences

There are currently more than 250 EPrints installations worldwide. A detailed listing can be viewed at the Registry of Open Access Repositories (ROAR) [34], maintained at the University of Southampton, or via the Directory of Open Access Repositories (OpenDOAR) [35], maintained at the University of Nottingham.

To represent EPrints in the Life Sciences two examples are given here: one is an already successfully implemented EPrints solution at an American cancer research centre, and the other one is an exciting international project in the pharmaceutical industry, which is still ongoing.

1.) The Arnold Library at the Fred Hutchinson Cancer Research Center, WA, USA runs a customized version of EPrints to meet the particular needs of their biomedical research community [36]. The reasons for choosing EPrints were manifold, but as a biomedical institution the PubMedID-based import feature has been recognized as extremely helpful. It is efficient and adds quality to the institution’s workflow and the resulting records. Citation data from PubMed can even be batch imported, which is perceived as incredibly convenient for retrospective analysis of faculty publications. EPrints’ short “persistent” URLs are appreciated by faculty using these links in their NIH Biosketches and on several other publication lists. Furthermore the default Library of Congress subject heading schema has been replaced by a subset of NLM’s Medical Subject Headings (MeSH) used by the research centre. Overall the long existence of EPrints, its multitudinous installations worldwide, and the clear support from the University of Southampton gave the Fred Hutchinson Cancer Research Center confidence in the stability and longevity of the repository software solution.

2.) The Novartis Institutes for BioMedical Research are currently developing a highly customised EPrints repository for the pharmaceutical company’s research output in collaboration with EPrints Services [37]. OAK (Open Access to Knowledge @ Novartis) will not only take care of the standard copyright-compliant long term archiving issues. It will also include a sophisticated review and approval workflow for Novartis research publications. OAK will therefore comprise of two sections: a) OAK Internal Archive: containing all publications from “Draft” state to “Published” state (audience restricted to Novartis employees); b) The Novartis Repository: containing all metadata and final drafts of externally published articles cleared for Open Access (to the public). The complexity of running an institutional repository project in a highly competitive corporate environment with simultaneous consideration of information security and protection of the company’s intellectual property make it difficult to predict a final launch date. However, the very fact that a big pharmaceutical company is even engaged in an EPrints project can be judged as a proof of the software’s quality.

Outlook into the future

Institutional Repositories are here to stay and are obviously no longer restricted to academia. They are meant
to be places where institutions can curate their intellectual output. The University of Southampton will continue with its efforts to make EPrints repositories more attractive information hubs to use, share and effectively reuse deposited content. Recent involvement in the CLADDIER project [38] and the EdShare project [39] emphasize the ongoing importance of Southampton’s leading role in the field of institutional self-archiving.

Acknowledgements

The author is grateful to Jessie Hey (University of Southampton) and Sheridan Brown (Key Perspectives Ltd.) for pointing him to valuable information resources to draw from. Many thanks as well to Timothy Miles-Board (EPrints Services) for his critical reading and constructive English language suggestions.

References


12. The Open Citation Project (OpCit) [homepage on the internet]. Southampton: School of Electronics and Computer Science, University of Southampton [updated 2009 March 09; cited 2009 Apr 19]. Available from: http://opcit.eprints.org/


17. Securing a hybrid environment for research preservation and access (SHERPA) [homepage on the internet]. Nottingham: University of Nottingham; 2006 [cited 2009 Apr 19]. Available from: http://www.sherpa.ac.uk/


21. Harnad S. Press Release: University of Southampton to provide open access to academic research online [internet page from mailing list LIBLENCENSE-L (liblencense-l@lists.yale.edu)]. New Haven, CO: Yale University Library [posted 2004 Dec 15; cited 2009 Apr 19]. Available from: http://www.library.yale.edu/~license/ListArchives/0412/msg00097.html


35. The directory of open access repositories (OpenDOAR) [homepage on the internet]. Nottingham University of Nottingham; c2006-8 [updated 2008 Sep 11; cited 2009 Apr 19]. Available from: http://www.opendoar.org/index.html


37. Gumpenberger C. Open access to knowledge @ Novartis: new ways of knowledge sharing in the pharmaceutical industry. In: Third international conference on Open Repositories; 2008 Apr 1-4; Southampton, United Kingdom [cited 2009 Apr]. Available from: http://pubs.or08.ecs.soton.ac.uk/23/2/43.pdf

38. Citation, location, and deposition in discipline & institutional repositories (CLADDIER) [project page on the internet]. British Atmospheric Data Centre (BADC), United Kingdom [cited 2009 Apr]. Available from: http://claddier.badc.ac.uk/trac


Corresponding author:
Christian Gumpenberger
Manhartsbrunnerstr. 40 2111 Seebn, Österreich
christian.gumpenberger@yahoo.com

Please cite as

This article is freely available from

Published: 2009-06-16

Copyright
©2009 Gumpenberger. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en). You are free: to Share – to copy, distribute and transmit the work, provided the original author and source are credited.