

# Self-Archiving of Research Output



David Nutter

Department of Computing and Informatics

# What is Self Archiving

Self archiving is the practice of authors submitting their scholarly output to an online archive which makes the items available for free download.

- Part of wider Open Access movement
- Examples include the Physics preprints service <http://www.arxiv.org> and the Southampton Eprints System <http://eprints.soton.ac.uk>.
- “Output” includes papers, posters/presentations, journal articles, book chapters etc etc.
- This differs from traditional publication models.
  - Publishers are no longer gatekeepers
  - Knowledge is available to all, not just the well-heeled.
  - Access to ongoing research

# arXiv.org

## The Computing Research Repository (CoRR)

Welcome to the Computing Research Repository (CoRR), sponsored by [ACM](#), the [arXiv.org e-Print archive](#), NCSTRL (Networked Computer Science Technical Reference Library), and [AAAI](#). From here you can

- [browse](#) the Repository and peruse recent submissions or browse the entire [NCSTRL collection](#)
- [search](#) the Repository or the [NCSTRL collection](#)
- [subscribe](#) to the Repository or the [NCSTRL collection](#), and get regular email notification of new submissions
- [submit](#) a document to the Repository using web upload, if you have already registered.
- [get help](#) on many topics, including
  - [submission](#), including registration and using email or ftp submissions (which do not require registration). (Click on [web upload help](#) in the submission help file to get more help on submitting by web upload and author registration.)
  - [searching](#)
  - [submitting Latex source](#)
  - [downloading papers](#)
  - [other frequently-asked questions](#)
- [CoRR Advisory Board](#)
- go to [mirror sites](#), for potentially faster access to CoRR if you are not in North America.

Papers in CoRR are classified in two ways: by subject area from a list of subject listed below and by using the [1998 ACM Computing Classification System](#). The ACM classification scheme provides us with a relatively stable scheme that covers all of computer science. The subject areas are not mutually exclusive, nor do they (yet) provide complete coverage of the field. On the other hand, we hope that they better reflect the active areas of research in CS. We expect to add more subject areas and subdivide current subject areas according to demand. Authors who cannot find an appropriate subject area should use subject area Other. We welcome constructive comments and suggestions; send them to [www-admin@arXiv.org](mailto:www-admin@arXiv.org).

Here is the current list of subject areas and their moderators:

- **AR - Architecture** - William Waite
  - Covers systems organization and architecture. Roughly includes material in ACM Subject Classes C.0, C.1, and C.5.
- **AI - Artificial Intelligence** - Erik Sandewall
  - Covers all areas of AI except Vision, Robotics, Machine Learning, Multiagent Systems, and Computation and Language (Natural Language Processing), which have separate subject areas. In particular, includes Expert Systems, Theorem Proving (although this may overlap with Logic in Computer Science), Knowledge Representation, Planning, and Uncertainty in AI. Roughly includes material in ACM Subject Classes I.2.0, I.2.1, I.2.3, I.2.4, I.2.8, and I.2.11.
- **CC - Computational Complexity** - Lane Hemaspaandra

# Southampton CogPrints

Cogprints

[Home](#) [About](#) [by Year](#) [by Subject](#) [Search](#) [Register](#) [User Area](#) [Help](#)

Welcome to Cogprints



Welcome to **CogPrints**, an electronic archive for [self-archive](#) papers in any area of [Psychology](#), [neuroscience](#), and [Linguistics](#), and many areas of [Computer Science](#) (e.g. [artificial intelligence](#), [robotics](#), [vision](#), [learning](#), [speech](#), [neural networks](#)), [Philosophy](#) (e.g., [mind](#), [language](#), [knowledge](#), [science](#), [logic](#)), [Biology](#) (e.g., [ethology](#), [behavioral ecology](#), [sociobiology](#), [behaviour genetics](#), [evolutionary theory](#)), [Medicine](#) (e.g., [Psychiatry](#), [Neurology](#), [human genetics](#), [imaging](#)), [Anthropology](#) (e.g., [primatology](#), [cognitive ethnology](#), [archeology](#), [paleontology](#)), as well as any other portions of the [physical](#), [social](#) and [mathematical sciences](#) that are pertinent to the study of cognition.

## [Browse](#)

Browse the archive by [Subject](#) or [Year](#).

## [Latest Additions](#)

View items added to the archive in the past week.

## [Simple Search](#)

Search the archive using the most common fields.

## [Advanced Search](#)

Search the archive using a more complex range of fields.

## [Registered Users Area](#)

Registered users may submit items to the archive.

## [Register](#)

Register for an account (or reset a forgotten password).

## [Lookup by ID](#)

Find a record by it's unique cogprints ID number.

[More information](#) is available about the **CogPrints** archive and open archiving in general.

Cogprints is running on *EPrints2* archive-creating software, which generates eprints archives that are compliant with the [Open Archives Protocol for Metadata Harvesting OAI 2.0](#).



# Google Scholar (aggregation service)

The image shows a screenshot of the Google Scholar search results page for the query 'Cornelia Boldyreff'. The page features the Google Scholar logo, a search bar with the text 'Cornelia Boldyreff', and a 'Search' button. To the right of the search bar are links for 'Advanced Scholar Search', 'Scholar Preferences', and 'Scholar Help'. Below the search bar, a green header bar displays 'Scholar' on the left and 'Results 1 - 10 of about 157 for Cornelia Boldyreff. (0.05 seconds)' on the right. The main content area lists ten search results, each with a title, author information, citation count, and a brief abstract. The results are: 1. 'The Evolution of Websites' by P. Warren, C. Boldyreff, and M. Munro, cited by 38. 2. 'Distributed group working in software engineering education' by S. Lees, M. Gumbley, C. Boldyreff, and S. Drummond, cited by 20. 3. 'Reverse Engineering to Achieve Maintainable WWW Sites' by C. Boldyreff and R. Kewish, cited by 17. 4. 'Using application understanding to support impact analysis' by MJ Fyson and C. Boldyreff, cited by 15. 5. 'The AMES approach to application understanding: a case study' by C. Boldyreff, E. Burd, RM Halber, and RE Mortimer, cited by 14. 6. 'Active Artefact Management for Distributed Software Engineerin' by C. Boldyreff, D. Nutt, and S. Rank, cited by 14. Each result includes a link to the full text or a PDF, and a link to 'all 5 versions' or 'all 7 versions'.

**Google Scholar**   [Advanced Scholar Search](#)  
[Scholar Preferences](#)  
[Scholar Help](#)

**Scholar** Results 1 - 10 of about 157 for Cornelia Boldyreff. (0.05 seconds)

[The Evolution of Websites](#)  
P. Warren, C. **Boldyreff**, M. Munro - Cited by 38 - Web Search  
Abstract One major attraction of the World-Wide Web is the ease with which a relatively untrained user can create or modify a Web hypertext document. ...  
IWPC, 1999 - doi.ieeecomputersociety.org - doi.ieee.org - ieexplore.ieee.org - portal.acm.org - all 5 versions »

[Distributed group working in software engineering education](#)  
S. Lees, M. Gumbley, C. **Boldyreff**, S. Drummond, P. ... - View as HTML - Cited by 20 - Web Search  
P. Brereton, S. Lees and M. Gumbley Department of Computer Science Keele University Keele Staffordshire ST5 5BG, UK Tel: +44 1782 583079 fax: +44 ...  
Information and Software Technology, 1998 - co.umist.ac.uk - csa.com

[Reverse Engineering to Achieve Maintainable WWW Sites](#)  
C. **Boldyreff**, R. Kewish - Cited by 17 - Web Search  
ABSTRACT The growth of the World Wide Web and the accelerated development of web sites and associated web technologies has resulted in a variety of ...  
WCRE, 2001 - doi.ieeecomputersociety.org - doi.ieee.org - ieexplore.ieee.org - portal.acm.org - all 5 versions »

[Using application understanding to support impact analysis](#)  
MJ Fyson, C. **Boldyreff** - Cited by 15 - Web Search  
With the ever-growing cost of development of software systems it becomes increasingly necessary to extend the life of existing systems, adding new ...  
Journal of Software Maintenance Research and Practice, 1998 - doi.wiley.com - portal.acm.org - portal.acm.org - csa.com

[The AMES approach to application understanding: a case study](#)  
C. **Boldyreff**, E. Burd, RM Halber, RE Mortimer, M. ... - Cited by 14 - Web Search  
Page 1. The AMES Approach to Application Understanding: a case study C. **Boldyreff**, EL Burd, RM Halber, RE Mortimer, M. Munro, EJ Younger ...  
ICSM, 1995 - doi.ieee.org - doi.ieeecomputersociety.org - ieexplore.ieee.org - portal.acm.org

[Active Artefact Management for Distributed Software Engineerin](#)  
C. **Boldyreff**, D. Nutt, S. Rank - Cited by 14 - Web Search  
We describe a software artefact repository that provides its contents with some awareness of their own creation. "Active" artefacts are ...  
COMPSAC, 2002 - ieexplore.ieee.org - doi.ieee.org - doi.ieeecomputersociety.org - eprints.lincoln.ac.uk - all 7 versions »

# Why should we be involved?

Why should we be involved in self archiving?

- At an individual level?
- At an institutional level?

Several reasons. Firstly, As individuals:

- Wider availability of one's research increases its impact.
- Useful for the RAE and similar bean-counting exercises
- Convenience:
  - Nobody need ask you for papers ever again!
  - It is much easier to find papers online than those stored in a library.

# Why should we be involved?

Secondly, as institutions:

- In Europe and now in the UK, increase in signatures to the Berlin Declaration etc.
- Institutional archives are the most common manifestation of a commitment to Open Access
- The Science and Technology Select Committee encouraged Open Access policies in their tenth report
- Increase profile of institution's research
  - Particularly useful for theses, which may not otherwise achieve wide dissemination.

But...

# Example: an archived thesis



## EPrints

- [Home](#)
- [About](#)
- [Browse](#)
- [Search](#)
- [User Area](#)
- [Help](#)

## A Reflective Architecture to Support Dynamic Software Evolution

Rank, Stephen (2002) *A Reflective Architecture to Support Dynamic Software Evolution*. PhD thesis, University of Durham.

Full text available as:



[PDF](#) - Requires [Adobe Acrobat Reader](#) or other PDF viewer.

### Abstract

This thesis presents work which is concerned with the run-time evolution of component-based software systems. In particular, the main result of the research presented here is a framework which is used to model and control the architecture of a software system. This framework allows the run-time manipulation of the components which make up a software system. The framework makes the architecture of software systems visible, and allows interaction with it, using a reflective meta-object protocol. The motivating objectives of this work are providing a framework to support architectural flexibility, higher-level intervention, safe changes, and architectural visibility in software systems. The framework's behaviour and structure was motivated by a set of case-studies which have been used to guide its development and enhancement. The framework was developed iteratively, using each case-study in turn to evaluate its capabilities and to prompt the direction of development. A detailed set of evaluation criteria are developed, and the framework is evaluated with respect to these. The framework was found to meet each of the four objectives fully, with the exception of the aim to allow only safe changes which is only partly satisfied. Ways in which the framework can be improved in order to more fully satisfy its objectives are suggested, as are other extensions to its behaviour.

EPrint Type: Thesis (PhD)

Subjects: Software > SOFTWARE ENGINEERING (K.6.3) > Management (K.6.3, K.6.4) > Software Architectures (NEW) > Patterns (e.g., client/server, pipeline, blackboard) (NEW)  
Software > SOFTWARE ENGINEERING (K.6.3) > Design Tools and Techniques (REVISED) > Object-oriented design methods (NEW)  
Software > SOFTWARE ENGINEERING (K.6.3) > Management (K.6.3, K.6.4) > Software Architectures (NEW)

ID Code: 30

Deposited By: [Dr Stephen Rank](#)

Deposited On: 10 November 2004

Archive Staff Only: [edit this record](#)



# ...what about the publishers?

One might expect publishers to be unhappy about Open Access as:

- Publishers follow a “reader pays” model for access.
- Open Access undermines this.

However, the reality is somewhat different:

- Many publishers permit self-archiving of pre and post prints
- They recognise that academics like Open Access
  - Lawsuits etc are counterproductive: do not bite the hand that feeds!
- Publishers are evaluating changes to their business models to permit Open Access.

# ...what about my institution

Obviously Open Access is easier with institutional support:

- Usual method is to establish a *self-archiving policy* for reasons of:
  - Resourcing
  - “Encourages” individual academics to submit
  - Lobbying of publishers etc.
- There are no problems unless a commitment to open access breaks contract (unlikely).
- Most UK prestige universities now have institutional repositories and archiving policies.
  - Do we want to be left behind?

# ...won't this break copyright?

The simple answer: not if the publishers permit it.

- Many do (see SHERPA/ROMEEO)
- A policy of “don't ask, don't tell” is successfully used by the <http://www.arxiv.org> site.
  - A (comparatively) venerable self-archiving system
  - Has *never* been sued.
- Even if a publisher's policy prevents submission of a full print, even bibliographic data is useful
- The moral argument: consider who pays for the research.
  - If publically funded, arguably the public should have unfettered access to the research.
- However: *always check the legal position first.*

# DCI ePrints

Faculty of Technology ePrints Service



## EPrints

- [Home](#)
- [About](#)
- [Browse](#)
- [Search](#)
- [User Area](#)
- [Help](#)

## Latest Additions to Faculty of Technology ePrints Service

### Today

Nutter, David and Boldyreff, Cornelia  
[Evaluation of an awareness distribution mechanism: A simulation approach](#)  
13 June 2005, Conference or Workshop Item. [Deposited 10 May 2005]



### Thursday

Adams, Paul  
[Towards Supporting Agile Practice Within The Libre Software Paradigm](#)  
11 July 2005, Conference or Workshop Item. [Deposited 05 May 2005]



Adams, Paul and Boldyreff, Cornelia and Nutter, David and Rank, Stephen  
[Adaptive Reuse of Libre Software Systems for Supporting On-Line Collaboration](#)  
17 May 2005, Conference or Workshop Item. [Deposited 05 May 2005]



Adams, Paul and Boldyreff, Cornelia and Nutter, David and Rank, Stephen  
[Using Plone To Support Collaborative Research](#)  
11 July 2005, Conference or Workshop Item. [Deposited 05 May 2005]



Adams, Paul and Nutter, David and Rank, Stephen and Boldyreff, Cornelia



# Existing system: Overview

At Lincoln, the Department of Computing and Informatics already has a repository

<http://eprints.lincoln.ac.uk>

- Small (only 46 publications). However, this is more than the Open University!
- Built using the Open Source ePrints software.
- Registered as an active ePrints archive and indexed by Google Scholar, CiteSeer etc.
- However, not all academics in the department make use of the service:
  - Most submissions from the DSE group.
  - Tedious submission process.

# Existing system: Submissions

The submissions process attempts to capture bibliographic information alongside facsimiles of papers, presentations etc.

1. Select a type of ePrint (e.g. journal article, conference paper etc)
2. Fill in basic bibliographic information (authors, title etc)
3. Publication dates and other bibliographic information
4. Details of publication venue (conference, periodical etc)
5. Item status (published/unpublished)
6. Unformatted abstract, references and keywords to feed the text search system
7. Categorisation according to a subset of the 1998 ACM categories
8. Any additional comments
9. Upload document(s) associated with the print
10. Review and finalise submission

# Customising the repository(1)

Customisation of the repository falls into three categories:

**Styling** ● Controlled by a number of XML files

- These contain XHTML and a custom markup language.

**Authentication** By default ePrints uses a MySQL database and Apache to authenticated users. This is replacable with e.g Active Directory+NTLM:

- Replace handler in config file with custom Apache auth module (hard!)
- Create an `LDAP::Auth` object to replace `Apache::AuthDBI`
- Plenty of help is available

# Customising the repository(2)

**Repository Functionality** This includes modifications to the database schema, categorisation of papers and generation of reports

- These are explained on subsequent slides

A brief specification list for a repository machine:

- A PIII with 1GB of fast RAM and >120Gb of disk will be ample
- Apache and `mod_perl` are required
  - Linux is probably the path of least resistance: however may run on Windows
- 1 virtual host per archive



# Categories

Each print is assigned to a number of categories:

- By default, the Library of Congress classification is used
- DCI ePrints uses a cut-down version of the ACM classification
- Custom categories may be loaded via an XML file using `import_subjects`
- Most institutional repositories use a subject tree that reflects the organisational structure
- This is generally easier for the users to understand than a complex bibliographic classification
- It can be used to generate useful reports too!

# Schema extensions

It is possible to extend the database schema with further information. This is useful for reports and richer citations

- The DCI repository uses some extra metadata
  - Project ID, Group ID etc
- For the library, such information might include
  - Executive summary
  - Shelfmark information
  - Supervisor (for theses etc)
- New ePrint types may be created (e.g. for course materials etc)

# Reports

Lists of prints matching certain criteria may be generated.

- This is used to provide per-project/per-person etc publications lists
- Very useful for inclusion in other sites!
- Citation style may be customized.

# Access Control

ePrints has a simple workflow for submitted documents:

- By default, all submitted items are placed into a submission buffer, until they are approved by an *editor* or *administrator*.
- This is turned off for DCI ePrints (simplicity)
- Submitted items may be superseded by later versions, keeping the original record intact (traceability)
- The workflow is somewhat extensible
- Dissemination may also be controlled
  - Items may be made available to all, or just to registered repository users.

# Other issues

There has been much outside interest in the DCI ePrints repository

- It is part of a feasibility study being conducted by Loughborough
- At OSS2005 last week, I was questioned for over an hour on ePrints and our integration of it with an online work environment
- Plus various e-mails from people interested in the contents

Implementation issues:

- Many academics have a large backlog of papers/materials. Data entry support is useful

# For non-DCI researchers

Obviously, the DCI archive is tailored to the needs of computer scientists. Researchers in other disciplines have several options:

- See if your collaborators have a repository that you can use
- Examine the online archiving provided by journals.
  - Is it suitable?
  - Will the lack of centralisation be a problem?
- Otherwise, lobby for an institutional archive!