

EVENTOS

III WORKSHOP DE PROYECTOS DIGITALES

2003



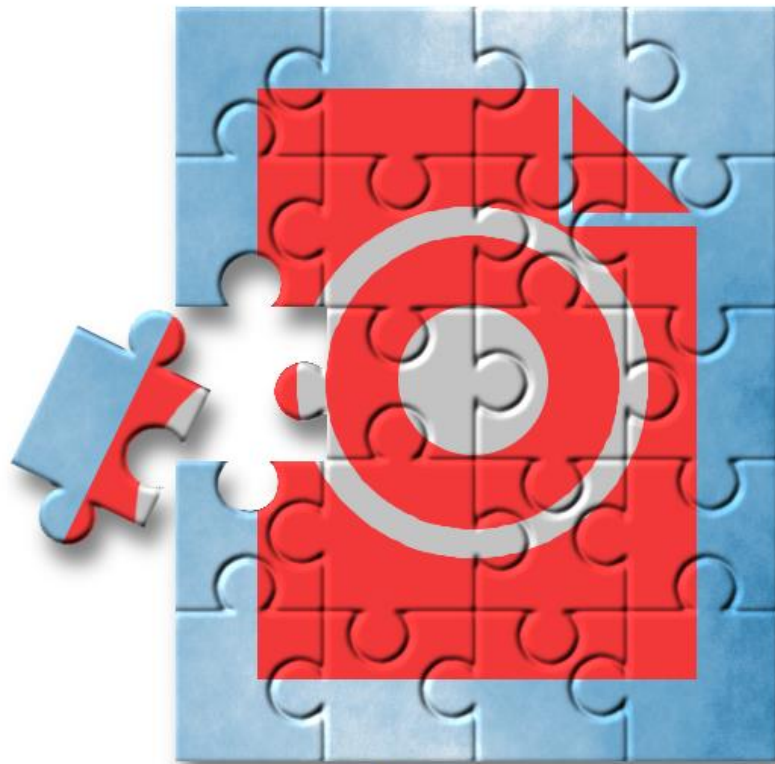
SELF ARCHIVING, OAI AND OPEN ARCHIVES,
INSTITUTIONAL ARCHIVES & AN OVERVIEW OF GNU EPRINTS2



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN



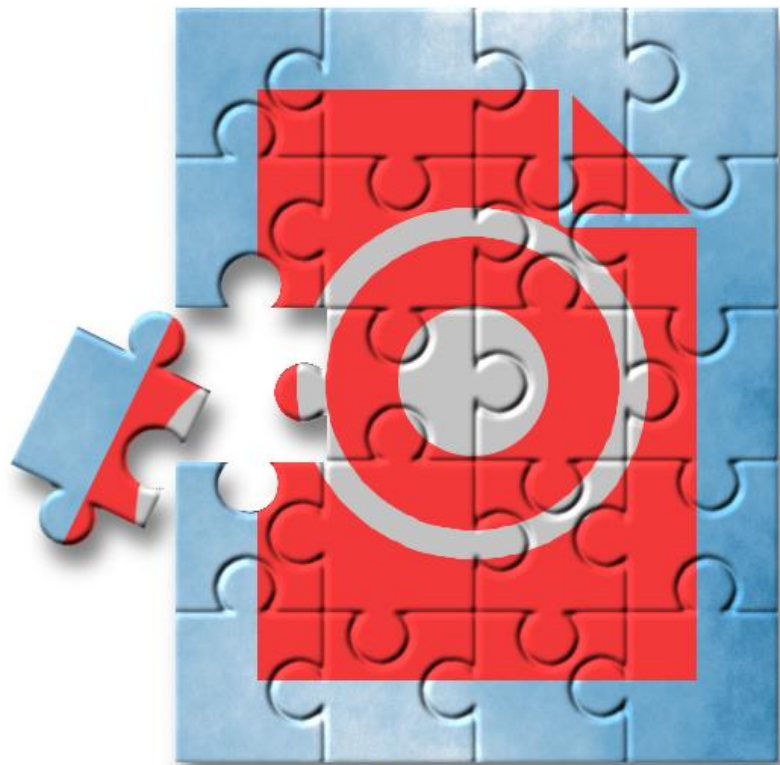
eprints.org

Self Archiving, OAI & Open Archives, Institutional Archives & An Overview of GNU EPrints 2

Christopher Gutteridge
University of Southampton

Christopher Gutteridge

- Systems, Support and Research
- Department of Electronics and Computer Science, University of Southampton
- “Inherited” EPrints Version One
- support@eprints.org



Self-Archiving

eprints.org

What is Self Archiving?

- An author is responsible for submitting their work into the archive
- Sometimes upload is done by student or secretary but it is still the authors *responsibility*

Advantages to the Archive

- It is in the authors interest to have their work in the archive, more likely to actually get done
- Distributes load over many staff
- The author is best person to enter correct information

Advantages to the Author

■ Dissemination

- Increased visibility (Google, OAI...)
- More visibility leads to more citations

■ Preservation

- Able to easily find old papers

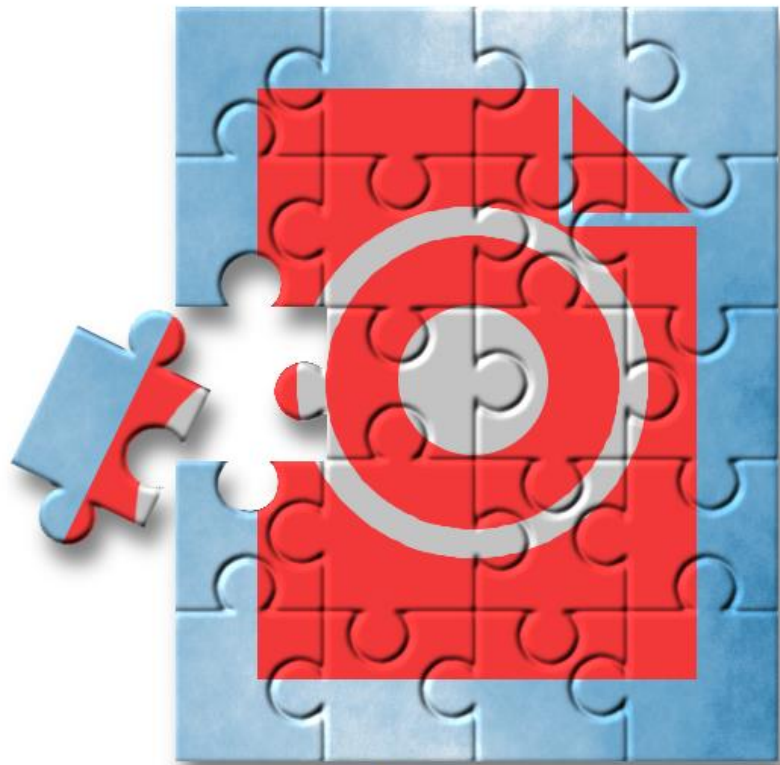
Items placed in arXiv.org take months to get cited, rather than years.



Self-Archiving FAQ

by Professor Stevan Harnad

- <http://www.eprints.org/self-faq/>



eprints.org

Open Archives and OAI-PMH

Open Archives?

- An open archive is an archive which exports its metadata in a useful way (OAI-PMH)
- *Most* also make some or all content available for free

What is OAI-PMH

- The **O**pen **A**rchives **I**nitiative **P**rotocol for **M**etadata **H**arvesting
- A way of asking an archive about the stuff it's got in it.
- This allows services to provide searches and other functionality across the *metadata* from many archives.
- XML over HTTP

What Questions can you ask via OAI-PMH?

- Identify
- GetRecord
- ListIdentifiers
- ListMetadataFormats
- ListRecords
- ListSets

Identify

- Who are you?
- What kind of stuff do you contain?
- What is the copyright of your data and your metadata?

“A collection-level description”

GetRecord

- Give me the metadata of a single record!

ListRecords

- Give me the metadata of all your records!
- May be limited by the date a record was last modified
- May be limited to a subset of the archive (e.g. only physics related records, but only if supported by archive)

ListIdentifiers

- Give me a list of all your records!
- May be limited by date record was last modified
- May be limited to a subset of the archive (e.g. only physics related records, but only if supported by archive)

ListMetadataFormats

- What metadata formats can you supply?

All archives *must* supply Dublin Core but may supply other metadata formats too.

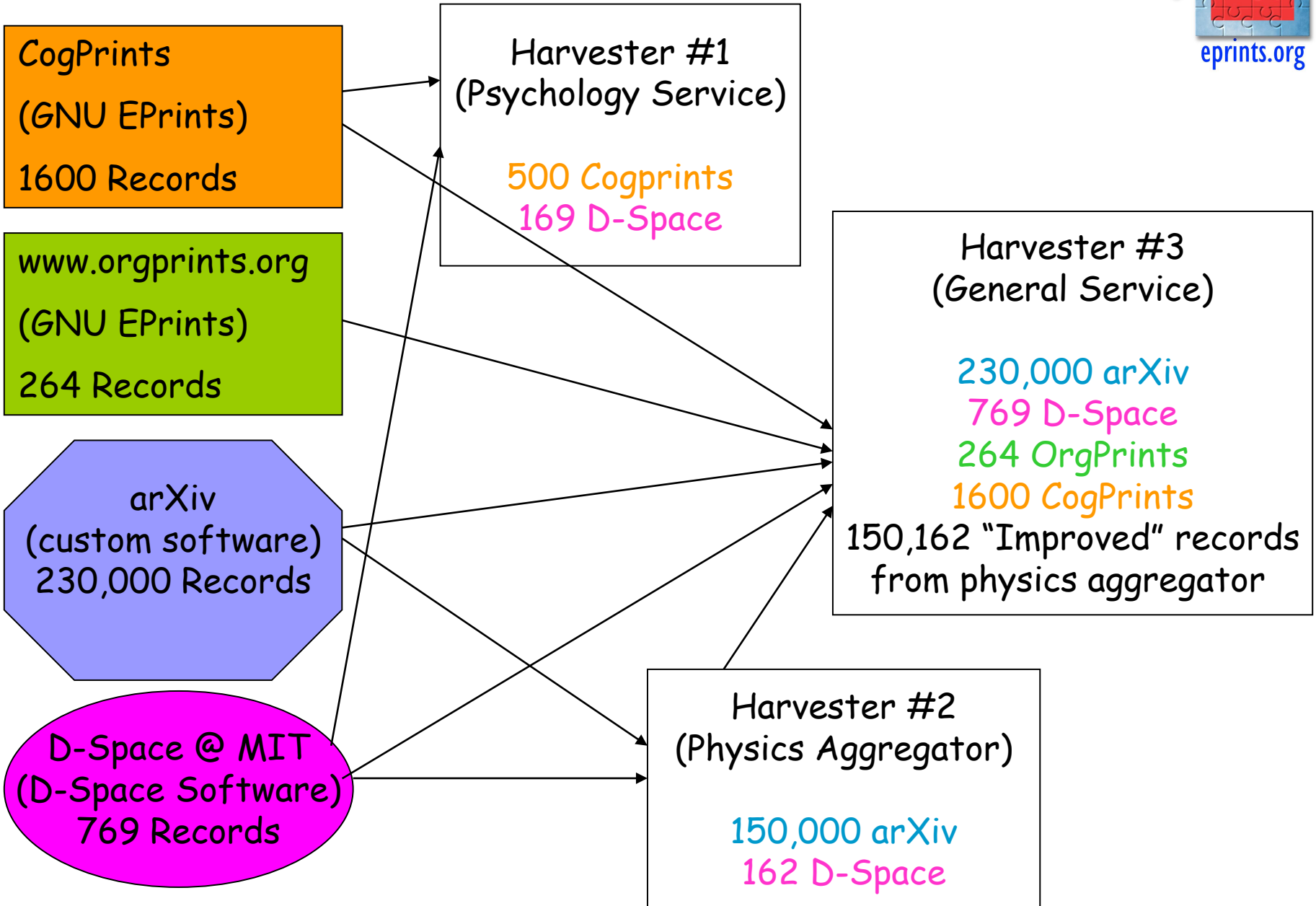
ListSets

- What subsets of your records may I ask for?

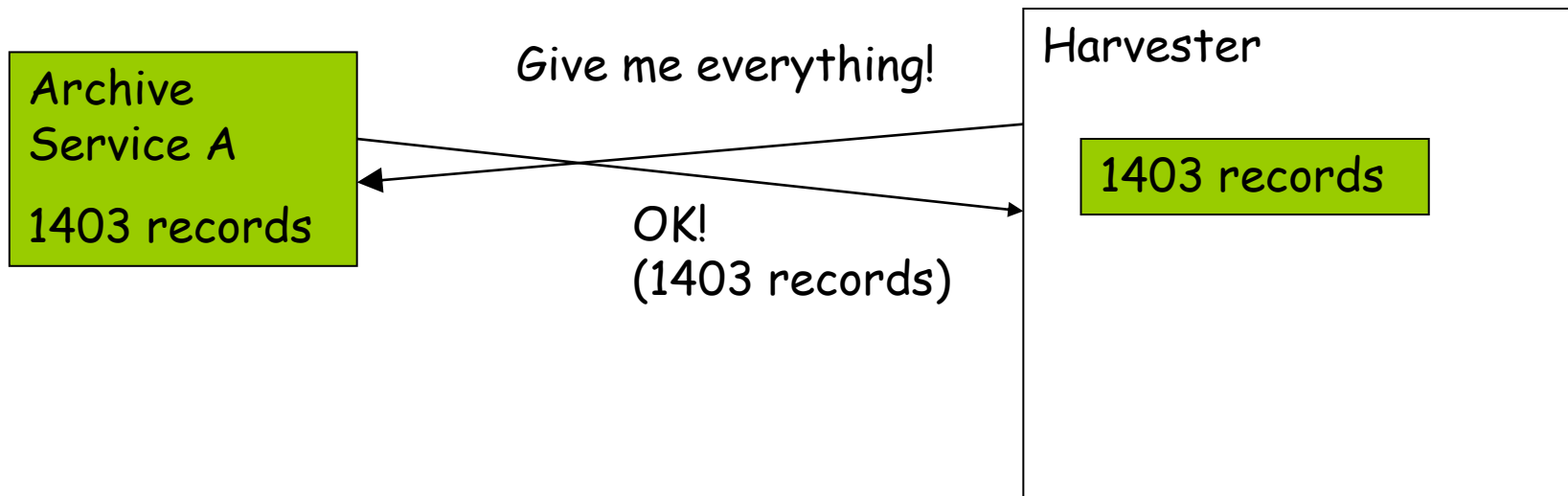
Some archives define subsets, by subject, by rights etc. e.g. Physics related records, or public domain items or peer-reviewed items.

So how does a service query an archive?

- The first time it asks for ALL records.
- Then, every so often (day, week...) it asks for everything that's changed since it last asked.



Day 1

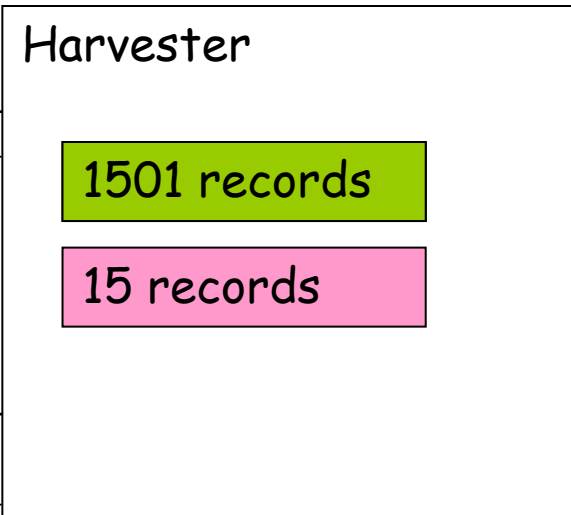


Day 2

Give me all records which were added or changed since yesterday

Archive Service A
1501 records

Archive Service B
123 records



OK!
(102 new records,
4 deleted records,
in 23 "physics" records)

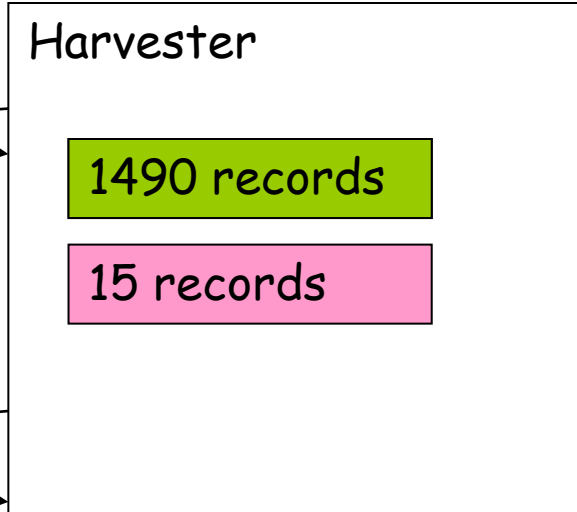
OK!
(15 records)

Day 3

Give me all records which were added or changed since yesterday

Archive Service A
1490 records

Archive Service B
123 records



OK!
Give me everything in set "physics" which were added or changed since yesterday
(25 new records, 36 deleted records, 3 changed records)

OK!
(0 new records, 1 record changed)

What are these records?

- Dublin Core

- Title
- Creator
- Date
- Description
- Identifier (URL)
- ...

- Very simple, but more useful than plain text.

Dublin Core in OAI

- Do I have to use Dublin Core? It's not very good, is it?
 - You *must* provide Dublin Core data via OAI, so that all harvesters can use your data.
 - You may also provide any other metadata formats you want to (MARC, AMF, one you-made-up etc.)

What a user sees...

- Our example user is searching for an item about badgers that they heard about that was written by someone called “Monkhouse”.
- They tried google but only found articles which mentioned the item they were looking for or fan sites about Bob Monkhouse the TV presenter.



C:\Documents and Settings\cjb\My Documents\footle.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help



Address C:\Documents and Settings\cjb\My Documents\footle.html

Welcome to OAI Service Thingy

Your search for **title or description** contains "badger" AND **creator** contains "Monkhouse" return 234 results...

Match	Creator	Title	Source		
#1	Monkhouse, Norman and Smith, Jim	Badgers are my friend	CogPrints	more details	goto item
#2	Monkhouse, Norman	Underground Animals	CogPrints	more details	goto item
#3	N Monkhouse and J Smith	Badgers are my friend	arXiv	more details	goto item
#4	J.K.Monkhouse	The Magic Badgers	FootlePrints	more details	goto item

etc.

Archives Supporting OAI

- **arXiv.org Eprint Archive** (246080 records)
<http://arXiv.org/>
- **The Australian National University (ANU) Eprints** (1585 records)
<http://eprints.anu.edu.au/>
- **BioMed Central (BMC)** (9222 records)
<http://www.biomedcentral.com/>
- **CERN Document Server** (17269 records)
<http://cdsweb.cern.ch/>
- **CogPrints Archives** (1719 records)
<http://cogprints.ecs.soton.ac.uk/>
- **Digital Library of MIT Theses** (7960 records)
<http://theses.mit.edu:80/>
- **Ethnologue: Languages of the World** (7148 records)
<http://www.ethnologue.com/web.asp>
- **Hong Kong University Theses Online** (8747 records)
<http://sunzi.lib.hku.hk/hkuto/>
- **Indiana University Digital Library Program** (2726 records)
<http://dlib.indiana.edu/>

Examples taken from oaister.org statistics.

They list over 200 archives.

OAI Harvesters

- Large and General
- Smaller and Quality Controlling
- Specialized Harvesters
- Value-Adding

General Harvesters

- Harvest all available sources
- Large quantity of data to search
- Unlikely to be able to ensure high quality of data
- <http://www.oaister.org/>
 - **1,723,003 records**
 - from **203 institutions**

Smaller Harvesters

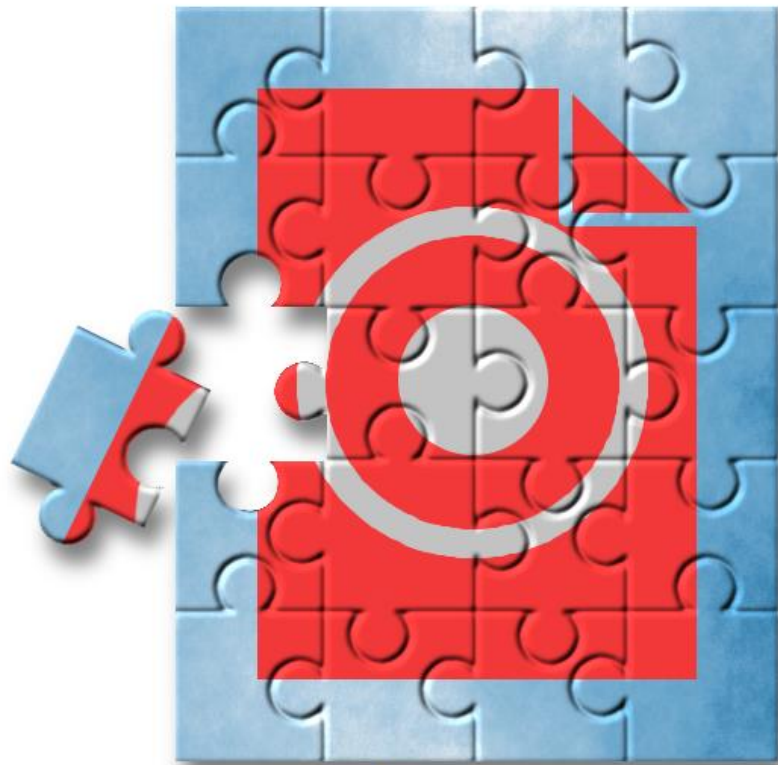
- A harvester collecting data from a smaller number of archives
- Less data
- Easier to maintain a high quality (reject archives with poor data)
- May require certain policies of source archives (e.g. Peer-review only)
- <http://www.myoai.com>
 - Harvests from **17** archives

Specialized Harvesters

- Similar to “Smaller Harvesters”
- Only harvest records related to a single theme
- Based in a single “community” so easier to arrange compatible metadata
- <http://digital.library.ucla.edu/sheetmusic/>
 - Only harvests from archives containing sheet music.

Value-Adding Harvesters

- Provide more than simple search facility
- Can also provide OAI export of “improved” or collated metadata
- <http://citebase.eprints.org/>
 - Uses database to find cited papers
 - Builds statistics and tools using citation information
 - Requires either citation data or full text access

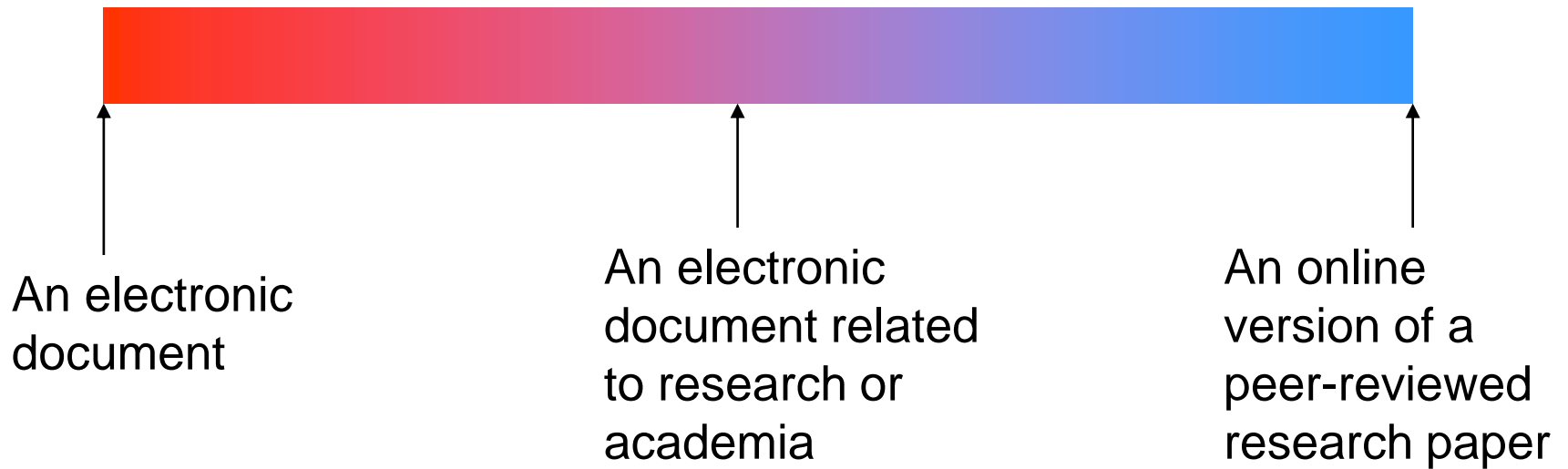


eprints.org

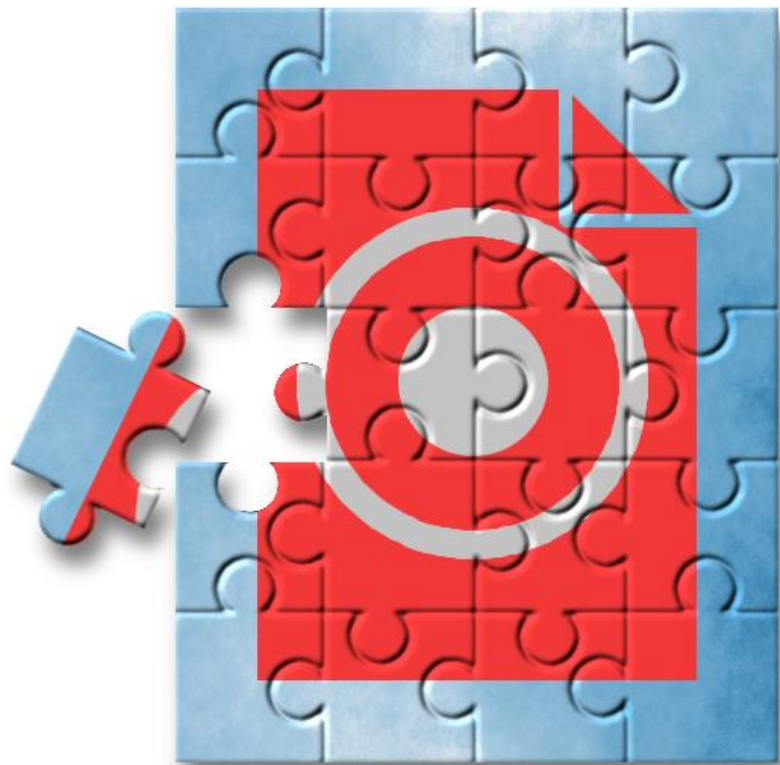
What exactly is
an E-Print,
anyway?

What is an E-Print?

Definitions vary...



Also... GNU EPrints is a piece of software developed at Southampton University, UK, to assist creating archives of E-Prints.



eprints.org

Subject Archives and Institutional Archives

Subject Archives

- Discipline Based
- Accept items relevant to the subject of the archive
- Impossible to be complete
- No way to force authors to submit their work

Example Subject Archives

- Cogprints: Cognitive Sciences EPrint Archive
<http://cogprints.ecs.soton.ac.uk/>
(Using GNU EPrints software)
- arXiv: A large archive of mostly physics and mathematics papers.
<http://www.arXiv.org/>
(Running their own software)

Institutional Archives

- Based around an institution
 - University (or department of University)
 - Other research organization
- Contains only content from members
- Possible to make policy that members must add
- Benefits clearer to administrators!



Example Institutional Archives

- Australian National University
<http://eprints.anu.edu.au/>
(Using GNU EPrints software)
- DSpace at MIT
<https://dspace.mit.edu/>
(Using MIT's DSpace software)

Goals of Institutional Archives

- Dissemination of research
 - Preservation of research
 - Lists and Websites
 - Collecting statistics
-
- Most people can only see one or two of these goals, and get frustrated when other people do not see their goals for the archive.

To achieve the goals...

- Dissemination of research
 - OAI Interface, Freely available documents
- Preservation of research
 - Quality metadata, good choice of document formats
- Lists and Websites
 - Complete metadata
- Collecting statistics
 - Complete and accurate metadata

Too many Goals?

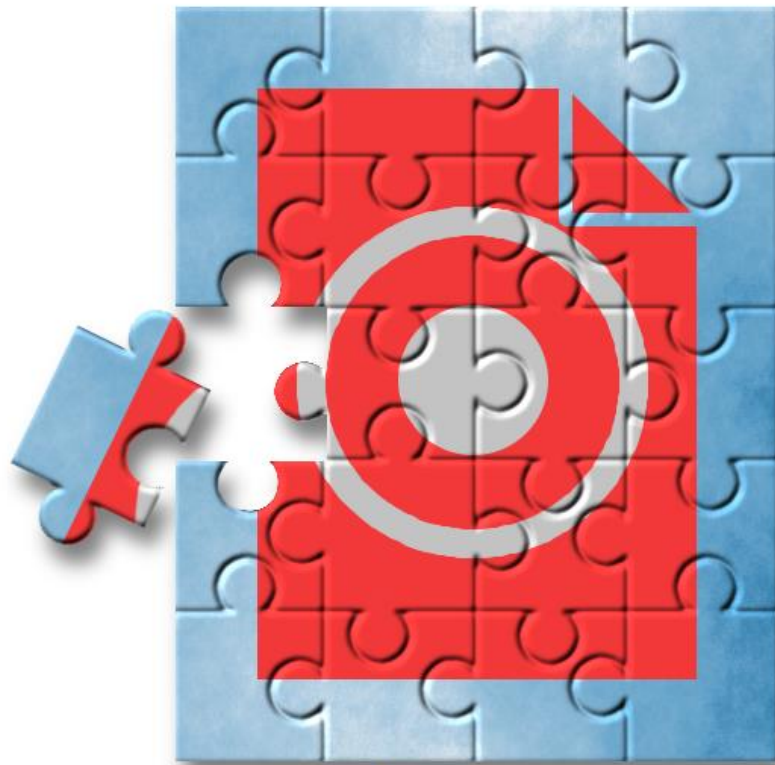
- Goals are not mutually-exclusive
- Each goal increases cost and complexity of archive
- Decide clear policy of goals right at the start!
- Quality of metadata depends on training and time taken to enter and check it – potentially expensive!

Archive Requirements Checklist

- What metadata you want to collect
- How you will ensure quality
- What should not go in your archive (do you allow unpublished work, software manuals, mp3's of speeches?)
- What formats you will allow/require. PDF, MS Word, Powerpoint, ASCII etc.
- Will you allow records without documents attached?
- If your goals include preservation; you need to have a long-term strategy of funding and support.
- Any other special features which you need (for example importing bibtex data, automatic conversion of formats etc)

Software for Institutional Archives

- **GNU EPrints** - University of Southampton
(my own!) <http://software.eprints.org/>
- **DSpace** - from M.I.T. / H.P.
<http://dspace.org/>
- Other tools are available for things such as thesis archives, library collections etc.



eprints.org

GNU EPrints 2 Overview

Christopher Gutteridge

cjg@ecs.soton.ac.uk

University of Southampton



What is GNU EPrints 2?

- Creates Online Archive
- Free Software
- OAI Compliant
- Targeted at Scholarly Material
- Adaptable
- Extendable

What is an EPrint?

(in the GNU EPrints System)

- System Metadata
 - Eprint ID Number
 - Deposit date
- Archive Metadata
 - Title
 - Year
 - Authors
- Zero or more Documents

What is a Document?

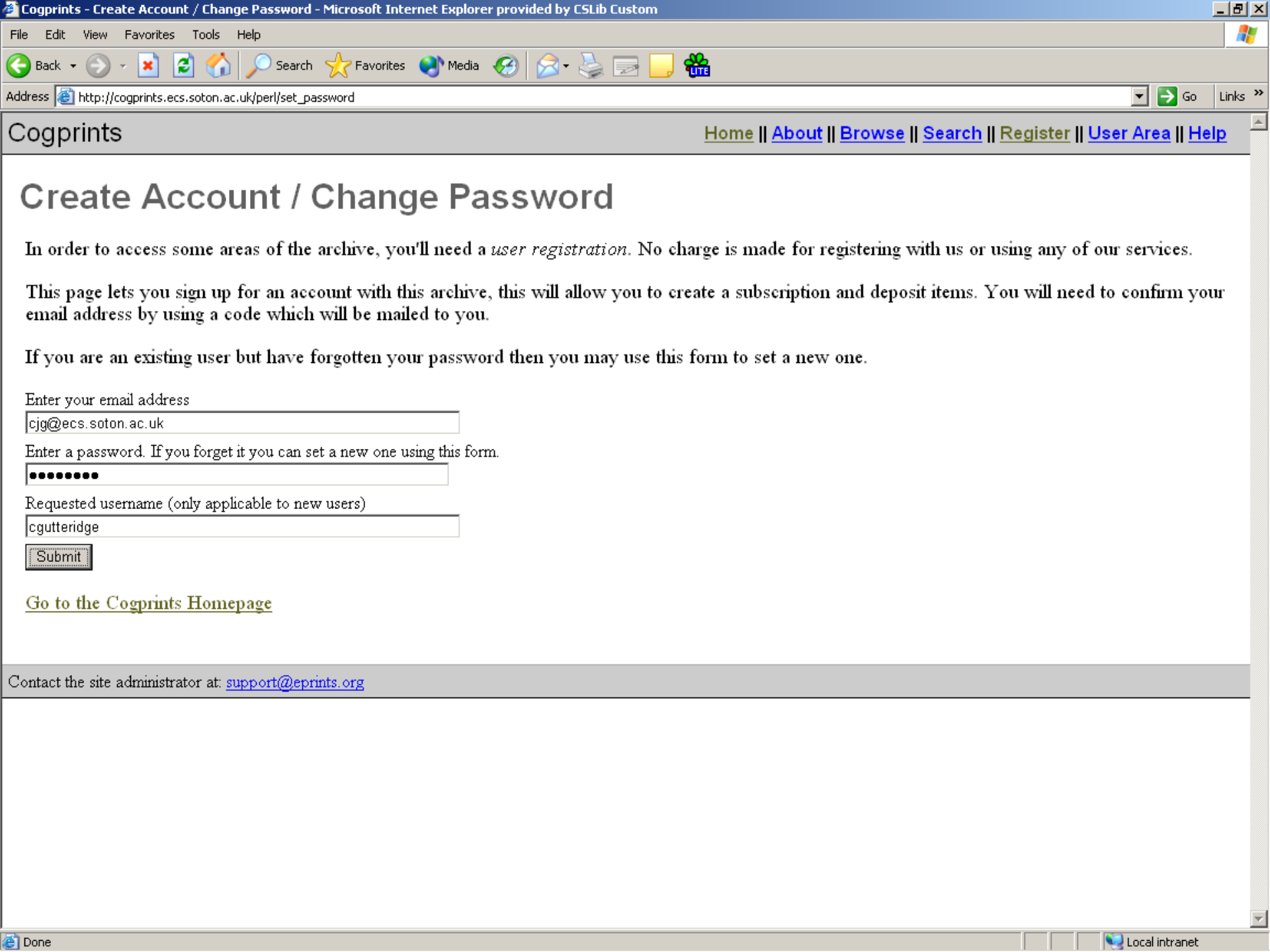
- System Metadata
 - Document ID
 - Format

- One or more files (in the filesystem)

Workflow 1

- New user registers via web
- User is emailed a confirmation URL
- User visits URL to activate account

An alternative is to auto-create user accounts from your own institution database.



Create Account / Change Password

In order to access some areas of the archive, you'll need a *user registration*. No charge is made for registering with us or using any of our services.

This page lets you sign up for an account with this archive, this will allow you to create a subscription and deposit items. You will need to confirm your email address by using a code which will be mailed to you.

If you are an existing user but have forgotten your password then you may use this form to set a new one.

Enter your email address

Enter a password. If you forget it you can set a new one using this form.

Requested username (only applicable to new users)

[Go to the Cogprints Homepage](#)

Workflow 2

- User configures personal metadata
 - Name
 - Address
 - Research interests

The Administrator may add remove and modify user-fields.

Record for Gutteridge, Christopher

Please enter correct information about yourself for our records. This information will be useful to us and readers of your items. You don't have to supply all this information if you don't want to; you need only fill out those fields marked with a * to start using the archive.

For instructions on how to change your e-mail address, [click here](#).

Password

Leaving this field blank will not affect the password.

Hide Email

Select this item if you do not wish your email to be displayed on public pages. The EPrints administrators will still be able to see it.

Name *

Title	Given Name(s)/Initials	Family Name(s)	Lineage
<input type="text"/>	<input type="text" value="Christopher"/>	<input type="text" value="Gutteridge"/>	<input type="text"/>

Department

Organisation

Address

Country

Operating System

Workflow 3

- User creates a new EPrint
- Enters metadata
- Uploads documents
- Views preview
- User “deposits” EPrint

EPrint is now in “submission buffer”

Deposit Type

EPrint Type *

Please select the most appropriate type for your deposit.

- Book Chapter
- Conference Paper
- Conference Poster
- Departmental Technical Report
- Journal (On-line/Unpaginated)
- Journal (Paginated)
- Newspaper/Magazine Article
- Other
- Preprint
- Thesis

Cancel Next >

Contact the site administrator at: support@eprints.org

Bibliographic Information

Please enter the bibliographic data about your deposit. Fields marked with a * are fields that must be filled out before your deposit will be accepted.

Status *

Please state here whether your deposit has been published, is currently in the process of being published (in press), or has not been previously published.

Unpublished
In Press
Published

Refereed

Please indicate whether this version of the work been refereed below.

- Yes, this version has been refereed.
- No, this version has not been refereed.

Public Domain

If the document you are deposit is not your own but rather an old document that is now in the public domain, then please tick the following box. This will prevent your own name and address appearing with the document as the address for correspondence.

- Yes, it is public domain.
- No.

Authors *

Please enter the authors below. If there are more authors than available spaces, click on the 'More Spaces' button. To remove an author, just remove their surname from the surname box.

1.

Title	Given Name(s)/Initials	Family Name(s)	Lineage
	Christopher	Gutteridge	

Person ID:

cjg@ecs.soton.ac.uk

2.

Title	Given Name(s)/Initials	Family Name(s)	Lineage

Person ID:

Workflow 4

- Editor views deposited EPrints

- Editor may:
 - Approve EPrint into main archive
 - Edit EPrint
 - Return EPrint to user with email explaining problem
 - Delete EPrint

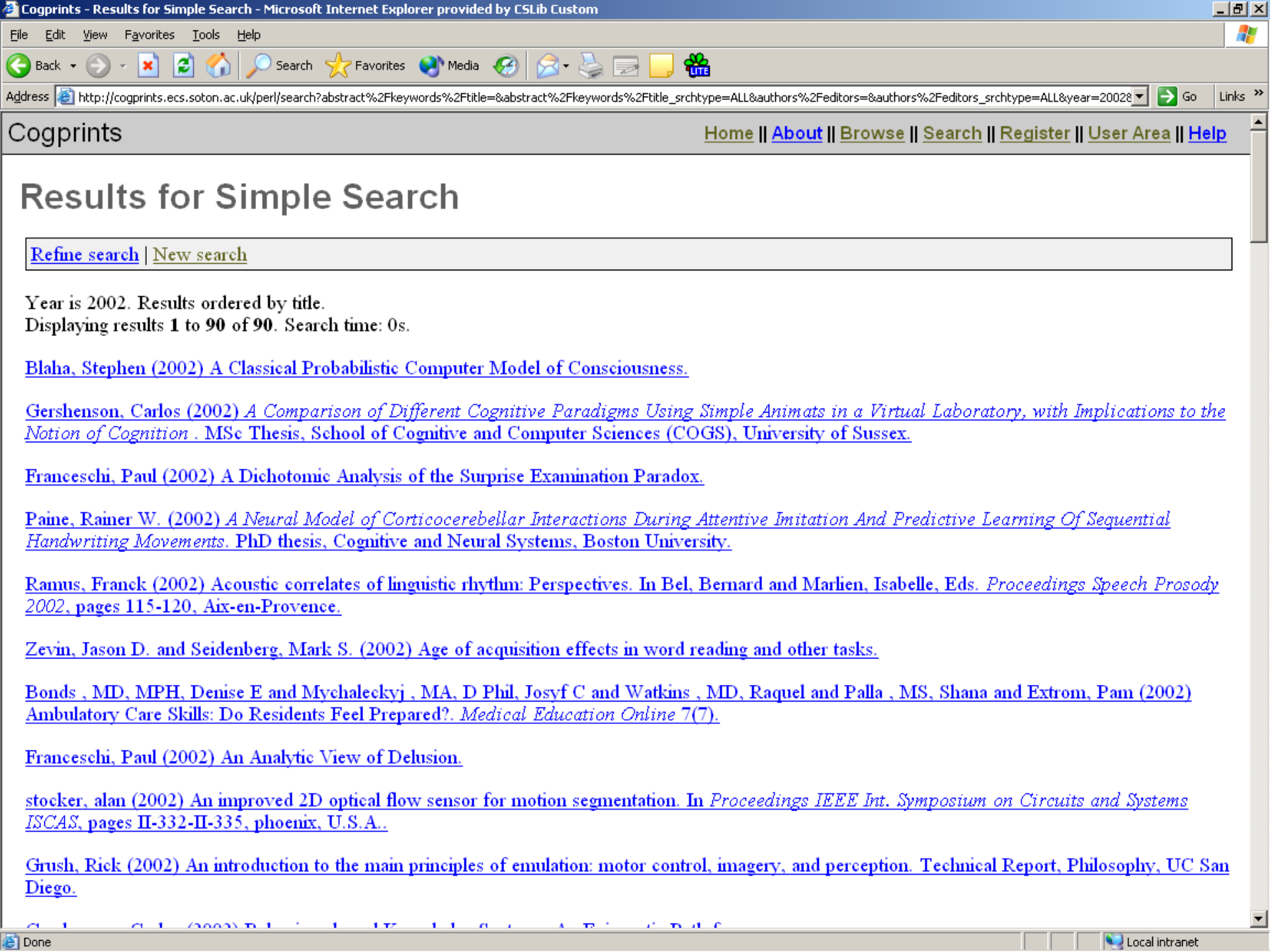
Workflow 5

- EPrint is now “Live”
- EPrint appears in searches
- EPrint appears in “views”
- EPrint has “abstract” page with metadata summary & links to documents
- Abstract page has static URL
- EPrint appears in OAI

Browse by Subject

Please select a value to browse from the list below.

- [Subject Areas \(1538\)](#)
 - [Biology \(275\)](#)
 - [Animal Behavior \(73\)](#)
 - [Animal Cognition \(58\)](#)
 - [Behavioral Biology \(57\)](#)
 - [Ecology \(5\)](#)
 - [Ethology \(58\)](#)
 - [Evolution \(77\)](#)
 - [Population Biology \(8\)](#)
 - [Primateology \(13\)](#)
 - [Sociobiology \(33\)](#)
 - [Theoretical Biology \(73\)](#)
 - [Computer Science \(485\)](#)
 - [Artificial Intelligence \(283\)](#)
 - [Complexity Theory \(37\)](#)
 - [Dynamical Systems \(64\)](#)
 - [Language \(67\)](#)
 - [Machine Learning \(84\)](#)
 - [Machine Vision \(32\)](#)
 - [Neural Nets \(150\)](#)
 - [Robotics \(62\)](#)
 - [Speech \(13\)](#)
 - [Statistical Models \(50\)](#)
 - [Electronic Publishing \(37\)](#)
 - [Archives \(26\)](#)
 - [Copyright \(11\)](#)
 - [Economics \(20\)](#)
 - [Peer Review \(15\)](#)
 - [JOURNALS \(32\)](#)
 - [Behavioral & Brain Sciences \(8\)](#)
 - [Medical Education Online \(21\)](#)



Results for Simple Search

[Refine search](#) | [New search](#)

Year is 2002. Results ordered by title.
Displaying results **1** to **90** of **90**. Search time: 0s.

[Blaha, Stephen \(2002\) A Classical Probabilistic Computer Model of Consciousness.](#)

[Gershenson, Carlos \(2002\) A Comparison of Different Cognitive Paradigms Using Simple Animats in a Virtual Laboratory, with Implications to the Notion of Cognition . MSc Thesis, School of Cognitive and Computer Sciences \(COGS\), University of Sussex.](#)

[Franceschi, Paul \(2002\) A Dichotomic Analysis of the Surprise Examination Paradox.](#)

[Paine, Rainer W. \(2002\) A Neural Model of Corticocerebellar Interactions During Attentive Imitation And Predictive Learning Of Sequential Handwriting Movements. PhD thesis, Cognitive and Neural Systems, Boston University.](#)

[Ramus, Franck \(2002\) Acoustic correlates of linguistic rhythm: Perspectives. In Bel, Bernard and Marlien, Isabelle, Eds. Proceedings Speech Prosody 2002, pages 115-120, Aix-en-Provence.](#)

[Zevin, Jason D. and Seidenberg, Mark S. \(2002\) Age of acquisition effects in word reading and other tasks.](#)

[Bonds , MD, MPH, Denise E and Mychaleckyj , MA, D Phil, Josyf C and Watkins , MD, Raquel and Palla , MS, Shana and Extrom, Pam \(2002\) Ambulatory Care Skills: Do Residents Feel Prepared?. Medical Education Online 7\(7\).](#)

[Franceschi, Paul \(2002\) An Analytic View of Delusion.](#)

[stocker, alan \(2002\) An improved 2D optical flow sensor for motion segmentation. In Proceedings IEEE Int. Symposium on Circuits and Systems ISCAS, pages II-332-II-335, phoenix, U.S.A..](#)

[Grush, Rick \(2002\) An introduction to the main principles of emulation: motor control, imagery, and perception. Technical Report, Philosophy, UC San Diego.](#)

[G...](#)

A Classical Probabilistic Computer Model of Consciousness

Blaha, Stephen (2002) A Classical Probabilistic Computer Model of Consciousness.

Full text available as:

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

Abstract

We show that human consciousness can be modeled as a classical (not quantum) probabilistic computer. A quantum computer representation does not appear to be indicated because no known feature of consciousness depends on Planck's constant h , the telltale sign of quantum phenomena. It is argued that the facets of consciousness are describable by an object-oriented design with dynamically defined classes and objects. A comparison to economic theory is also made. We argue consciousness may also have redundant, protective mechanisms.

Keywords: Consciousness, Statistical Models, brain

[Computer Science: Statistical Models](#)

Subjects: [Philosophy: Philosophy of Mind](#)

[Psychology: Psychophysics](#)

ID Code: 2077

Deposited By: [Blaha, Stephen](#)

Deposited On: 09 February 2002

Contact the site administrator at: support@eprints.org

How much will it cost?

- Free?
- Still need hardware & staff
- Not just \$0.00
- Not free as in “lunch”, but free as in “Willy”

Hardware

- Machine running UNIX (eg. Linux)
- 512M+ RAM
- 20Gb+ of disk space
- Remember to budget for backups!

A demo version could run on a much smaller machine than this.

Staff Costs

- 2 or 3 days of UNIX admin?
 - A few hours of UNIX admin each configuration change
 - Months to be ready, need the admin on-and-off
- Data Entry can be expensive and error-prone
- Low long-term costs are much lower
 - Support to teach staff to enter their own records
 - 2 – 3 minutes to approve/reject each record

Design Philosophy

- It's never going to be perfect!
- Default is a good starting point
- Will need some changes
- EPrints is configurable rather than “perfect” out of the box
- Possible to build new tools & views

Technology

- Any UNIX-like OS
- Developed under RedHat's GNU/Linux distro.
- Written in PERL
 - Not very efficient in terms of RAM
 - Allows rapid development and modification
- XML for import, export and parts of the configuration
- Apache and mod_perl
- MySQL database

Internationalisation

- Core Design Issue
- XML Phrase File for easy translation
- Multilanguage support
- Uses UTF-8 encoding throughout (Unicode)
- Option of Multilingual Metadata



Universitätsbibliothek München

Digitale Publikationen der Ludwig-Maximilians-Universität München

[Home](#) | [Blättern](#) | [Suchen](#) | [Login](#) | [Registrieren](#) | [Hilfe](#)

TITELSUCHE

VOLLTEXTSUCHE

⇒ [Hier](#) können Sie die Dokumente im Volltext durchsuchen.

SUCHE IN ALLEN FELDERN

⇒ [Hier](#) können Sie in allen Feldern eines Dokumentes suchen.

NEWS

⇒ [Hier](#) finden Sie die in den letzten 7 Tagen veröffentlichten Dokumente.

⇒ Die Auflösung der URNs (Persistent Identifier) wird erst funktionieren, wenn alle Dokumente, für die Die Deutsche Bibliothek bereits eine URL archiviert hat, gesammelt bei DDB geändert wurden. Wir bitten Sie dafür um Verständnis.

HOCHSCHULSCHRIFTEN DER LMU

- ⇒ [Medizinische Fakultät](#)
- ⇒ [Tierärztliche Fakultät](#)
- ⇒ [Fakultät für Mathematik, Informatik und Statistik](#)
- ⇒ [Fakultät für Physik](#)
- ⇒ [Fakultät für Chemie und Pharmazie](#)
- ⇒ [Fakultät für Biologie](#)
- ⇒ [Fakultät für Geowissenschaften](#)

HOCHSCHULSCHRIFTEN ANDERER UNIVERSITÄTEN

- ⇒ [TheO: Theses Online](#)
- ⇒ [Karlsruher Virtueller Volltextkatalog](#)
- ⇒ [NDLTD: Networked Digital Library of Theses and Dissertations](#)
- ⇒ [ETD OAI Union Catalog](#)

[Impressum](#)

Der Dokumentenserver der LMU verwendet die Software [EPrints 2.1.1](#)



Blättern nach Fakultäten

Klicken Sie auf einen Eintrag, um sich die Dissertationen anzeigen zu lassen.

- [Hochschulschriften](#)
 - [Evangelisch-Theologische Fakultät](#)
 - [Fakultät für Betriebswirtschaftslehre](#)
 - [Fakultät für Biologie \(82\)](#)
 - [Fakultät für Chemie und Pharmazie \(110\)](#)
 - [Fakultät für Geowissenschaften \(8\)](#)
 - [Fakultät für Geschichts- und Kunstwissenschaften](#)
 - [Fakultät für Kulturwissenschaften](#)
 - [Fakultät für Mathematik, Informatik und Statistik \(6\)](#)
 - [Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft](#)
 - [Fakultät für Physik \(150\)](#)
 - [Fakultät für Psychologie und Pädagogik](#)
 - [Fakultät für Sprach- und Literaturwissenschaften](#)
 - [Juristische Fakultät](#)
 - [Katholisch-Theologische Fakultät](#)
 - [Medizinische Fakultät \(59\)](#)
 - [Sozialwissenschaftliche Fakultät](#)
 - [Tierärztliche Fakultät \(1\)](#)
 - [Volkswirtschaftliche Fakultät](#)



Derived Secondary Classes for Flags of Foliations

**Hoster, Matthias (2001): *Derived Secondary Classes for Flags of Foliations*.
Dissertation, LMU München**

Bitte klicken Sie auf PDF, um sich die Dissertation im Volltext anzeigen zu lassen.

[PDF](#)

Sie benötigen den [Adobe Acrobat Reader](#) oder einen anderen PDF-Viewer, um das PDF-Dokument ansehen zu können.
Beziehen Sie sich bitte beim Zitieren auf die URN oder alternativ auf die URL der Dissertation.

Sachgruppe der DNB: [Mathematics](#)

Sprache der Hochschulschrift: Englisch

Fakultäten: [Fakultät für Mathematik, Informatik und Statistik](#)

Datum der mündlichen Prüfung: 05 Oktober 2001

URL dieser Metadatenseite: <http://edoc.ub.uni-muenchen.de/archive/00000004/>

URL des Dokumentes: http://edoc.ub.uni-muenchen.de/archive/00000004/01/Hoster_Matthias.pdf

URN des Dokumentes: <urn:nbn:de:bvb:19-49>

MD5 Prüfsumme des Dokumentes: 8a399ced7829eee52e4d24a933c4efa2

Signatur der gedruckten Ausgabe: 0001/UMC 11834

ID Code: 4

Other Features

- OAI 1.1 and 2.0 support
- Metadata search
- Subscriptions
- Export metadata as XML

Making it available Online

- Activate the OAI interface
- Easy to view format:
 - Plain Text
 - HTML
 - PostScript
 - PDF
- Visible to Search Engines



More Information

- GNU EPrints Website

- <http://software.eprints.org/>

- EPrints Project Website

- <http://eprints.org/>

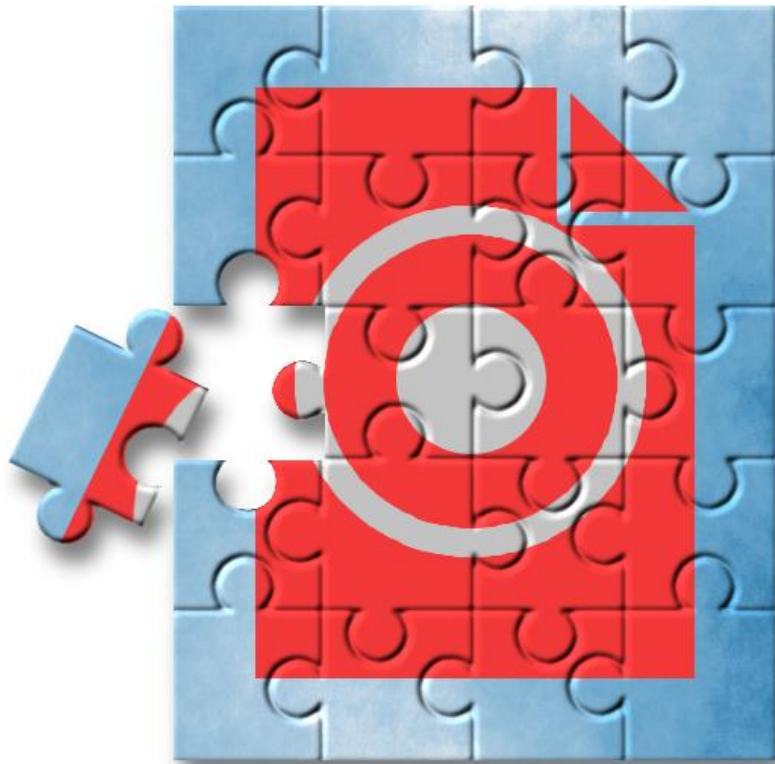
Further Reading

- GNU EPrints 2 Overview

- <http://eprints.ecs.soton.ac.uk/archive/00006840/>

- Applications, Potential Problems and a Suggested Policy for Institutional E-Print Archives

- <http://eprints.ecs.soton.ac.uk/archive/00006768/>



eprints.org

Thank-you.

Christopher Gutteridge
University of Southampton