

What is Wrong with Digital Repository Software?

Or why to Archive Now !



Hussein Suleman
hussein@cs.uct.ac.za

University of Cape Town
Department of Computer Science
Advanced Information Management Laboratory
Sivulile

November 2006

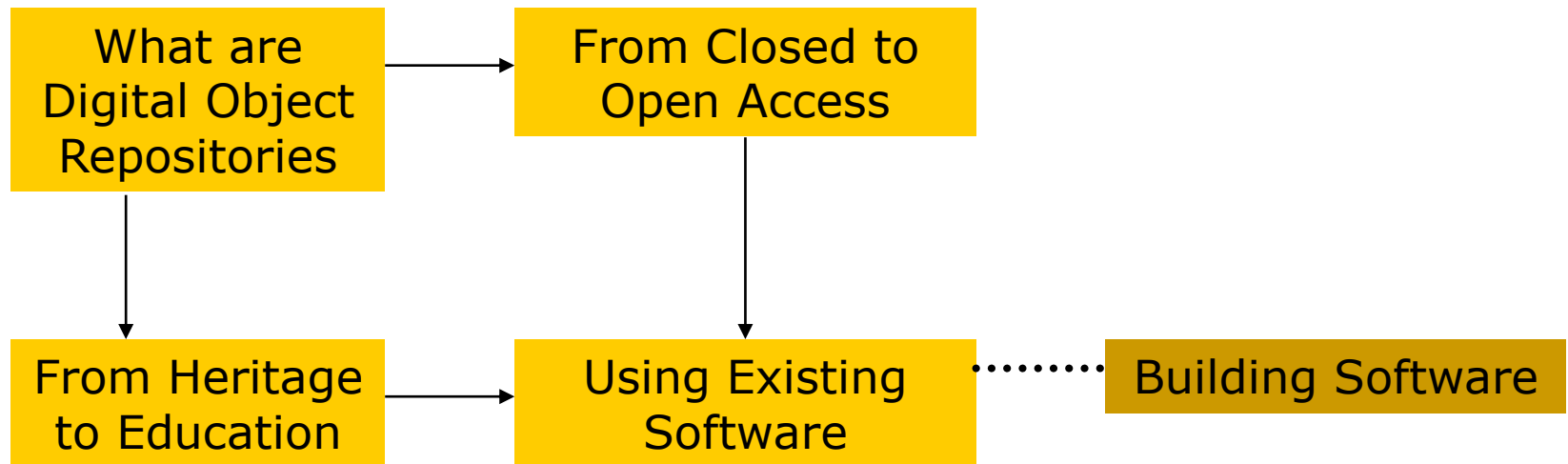


sivulile



open access
south africa

Outline 1



What is a Digital Object Repository?



[Home](#)

[Article title search](#)
[Author search](#)
[Keyword search](#)
[Full text search](#)

[Advanced search](#)

[Acronyms](#)
[Thesaurus](#)

[Need help?](#)

Browse the collection

List of journals:

[Abasebenzi](#)

[Afra Newsletter](#)

[Africanist News and Views](#)

[Amandla-Matla](#)

[Apdusa Views](#)

[Arise! Vukani!](#)

[Black Review](#)

[Clarion Call](#)

[Congress Resister](#)

[Contact](#)

[Cosatu News](#)

[Crisis News](#)

source: DISA, Univ. of KZN
<http://disa.ukzn.ac.za>



Heritage Repository Object

Article : We shall never forget:

We shall never forget

SPECIAL ISSUE

15th November, 1966.

WE SHALL NEVER FORGET.

WE SHALL NEVER FORGET our comrades-in-arms who have paid the supreme price for the lofty cause of freedom for our country.

WE SHALL NEVER FORGET the callous murder of MINI, KHAYINGA, MKABA, BONGCO, NGUDLE, SALOOJEE and many others; their death remains a permanent indictment against the South African Apartheid Rulers.

WE SHALL NEVER FORGET these comrades, who have braved Vorster's torture chambers and faced fascist brutality boldly in the face. Their sacrifices SHALL NOT BE IN VAIN.

*source: Mayibuye, DISA, Univ. of KZN
<http://disa.ukzn.ac.za>*



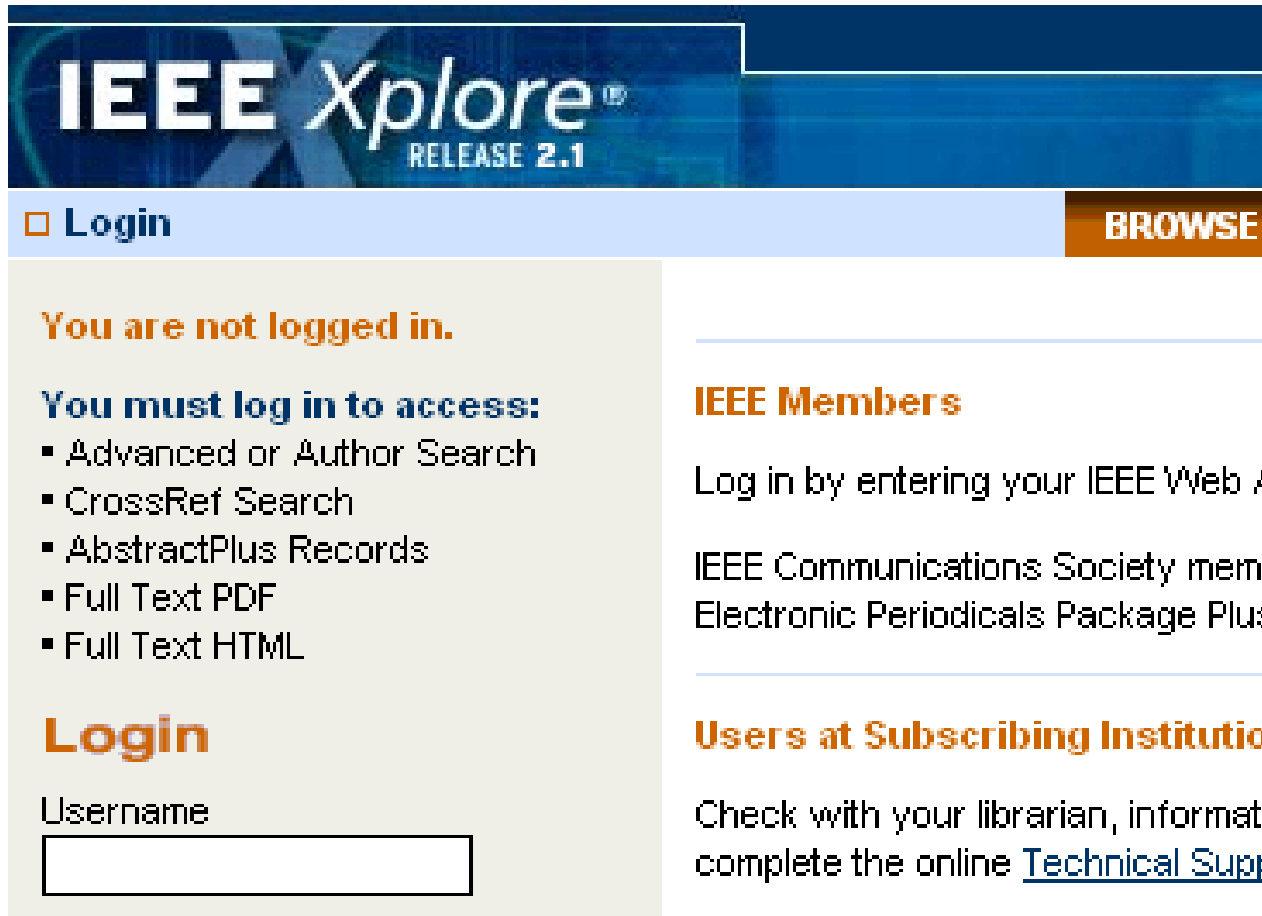
DORs in Education



source: Worldwide Greenhouse Education, Univ. of Vermont
<http://www.uvm.edu/wge/education.htm>



Closed Repositories



IEEE Xplore[®]
RELEASE 2.1

[Login](#) [BROWSE](#)

You are not logged in.

You must log in to access:

- Advanced or Author Search
- CrossRef Search
- AbstractPlus Records
- Full Text PDF
- Full Text HTML

Login

Username

IEEE Members

Log in by entering your IEEE Web A

IEEE Communications Society memk
Electronic Periodicals Package Plus

Users at Subscribing Institution

Check with your librarian, informatik
complete the online [Technical Supp](#)

source: IEEE Explore
<http://ieeexplore.ieee.org/Xplore/home.jsp>



Open Access Repositories - Research



The screenshot shows the top navigation bar of the Cornell University Library website. On the left is the Cornell University Library logo. To its right are two search links: "Search Library" and "Search Cornell". Below this is a red banner with the "arXiv.org" logo on the left. On the right side of the banner is a search form with the text "Search for" followed by an empty input field, a dropdown menu set to "All papers", and a "Go!" button. Above the input field are links for "(Help | Advanced search)".

Open access to 391,238 e-prints in Physics, Mathematics, Computer Science and Quantitative Biology

Subject search and browse:

See cumulative "What's New" pages.

Robots Beware: indiscriminate automated downloads from this site are *not* permitted.

Physics

- [Astrophysics](#) ([astro-ph new](#), [recent](#), [abs](#), [find](#))
- [Condensed Matter](#) ([cond-mat new](#), [recent](#), [abs](#), [find](#))
includes: Disordered Systems and Neural Networks; Materials Science; Mesoscopic Systems and Quantum Hall Effect; Other; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- [General Relativity and Quantum Cosmology](#) ([gr-qc new](#), [recent](#), [abs](#), [find](#))
- [High Energy Physics - Experiment](#) ([hep-ex new](#), [recent](#), [abs](#), [find](#))
- [High Energy Physics - Lattice](#) ([hep-lat new](#), [recent](#), [abs](#), [find](#))
- [High Energy Physics - Phenomenology](#) ([hep-ph new](#), [recent](#), [abs](#), [find](#))
- [High Energy Physics - Theory](#) ([hep-th new](#), [recent](#), [abs](#), [find](#))
- [Mathematical Physics](#) ([math-ph new](#), [recent](#), [abs](#), [find](#))
- [Nuclear Experiment](#) ([nucl-ex new](#), [recent](#), [abs](#), [find](#))
- [Nuclear Theory](#) ([nucl-th new](#), [recent](#), [abs](#), [find](#))
- [Physics](#) ([physics new](#), [recent](#), [abs](#), [find](#))
includes (see [detailed description](#)): Accelerator Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and

*source: arXiv.org, Cornell University
<http://arxiv.org/>*



Open Access Repositories - Teaching

Navigation: [CAUSEweb](#) -> [Resources](#) -> [Statistical Inference and Techniques](#) -> [Significance Testing Principals](#) -> [P-value](#)

CAUSE causeweb.org Search
Advanced Search

Member Login:

Password:

Login

[Create a Log In.](#)
[Forgot password?](#)

[Resources](#)

- Data Collection
- Data Presentation
- Statistical Inference**
- Probability
- Forum
- Professional Development
- Research
- USCOTS
- Workshops
- Webinars
- Calendar
- About CAUSE

 **P-value**

Displaying 1 to 10 of 12 resources:

[Data Matters with Excel](#)
This activity suggests two ways of estimating a population variance of grouped data: the variance within a group and the variance between groups. This activity is performed in Excel and accompanies S...
<http://www.keycollege.com/DM/activities/excel/07.3.html>

[Hypothesis Testing](#)
This PowerPoint lecture presentation explores hypothesis testing. It describes how to use the p-value in making the decision whether or not to reject the null hypothesis as well as considers possible...
http://www.stat.psu.edu/~resources/ClassNotes/ljs_16/index.h...

[Hypothesis Testing for a Single Population Mean](#)
This PowerPoint lecture presentation examines hypothesis testing for a single population mean. It discusses determining the p-value, making the decision, and terminology, illustrated through various ...
http://www.stat.psu.edu/~resources/ClassNotes/ljs_17/index.h...

[Interpreting Statistical p-values](#)
This page discusses the understanding of and interpretation of p-values for those who read articles with statistical information.
<http://www.texasoft.com/pvalue.html>

[P Values and Statistical Significance](#)
This resource defines what a p-value is, why .05 is significant, and when to use it. It also covers related topics such as one-tailed/two-tailed tests and hypothesis testing.
<http://www.sportsci.org/resource/stats/pvalues.html>

[Robustness of t test and ANOVA](#)
This applet lets you explore the effect of violations of the assumptions of normality and homogeneity of variance on the type I error rate and power of t tests (and two-group analysis of variance).
http://www.ruf.rice.edu/~lane/stat_sim/robustness/index.html

Consortium for the Advancement of Undergraduate Statistics Education
<http://www.causeweb.org/resources/>



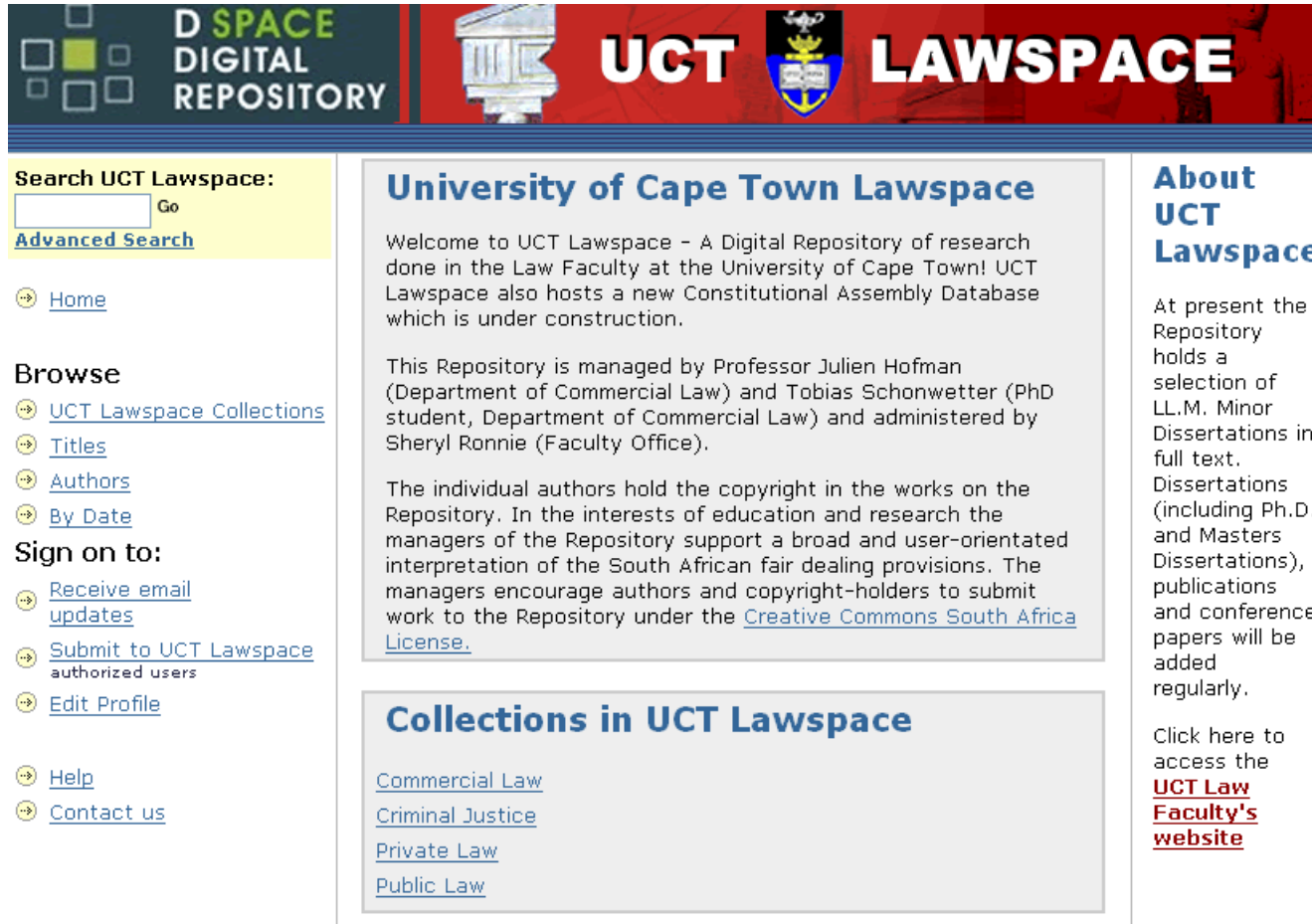
Why Open Access?

- ❑ Lawrence (2001): clear correlation between open access and impact in CS
- ❑ Eysenbach (2006): double citation impact for open access as compared to closed journal

- ❑ and closer to home...
- ❑ Suleman (2006):
 - open access repository part of normal operations of department – formal record of research
 - open access archive used extensively by external parties
 - indexed aggressively by search engines
 - resources accessed almost immediately after deposit and consistently thereafter




OA Repositories in SA



The screenshot shows the UCT Lawspace website. At the top, there is a header with the DSpace Digital Repository logo on the left, the UCT crest in the center, and the text 'UCT LAWSPACE' on the right. Below the header, the page is divided into three main sections. On the left is a navigation sidebar with a search box, 'Advanced Search' link, 'Home' link, 'Browse' section with links to 'UCT Lawspace Collections', 'Titles', 'Authors', and 'By Date', 'Sign on to:' section with links to 'Receive email updates', 'Submit to UCT Lawspace authorized users', and 'Edit Profile', and finally 'Help' and 'Contact us' links. The middle section is titled 'University of Cape Town Lawspace' and contains a welcome message, information about the repository's management by Professor Julien Hofman and Tobias Schonwetter, and a paragraph about copyright and the Creative Commons license. The right section is titled 'About UCT Lawspace' and describes the repository's holdings of LL.M. Minor Dissertations, Ph.D. and Masters Dissertations, and conference papers, along with a link to the UCT Law Faculty's website.

DSpace
DIGITAL
REPOSITORY

UCT  **LAWSPACE**

Search UCT Lawspace:
 Go
[Advanced Search](#)

→ [Home](#)

Browse

→ [UCT Lawspace Collections](#)

→ [Titles](#)

→ [Authors](#)

→ [By Date](#)

Sign on to:

→ [Receive email updates](#)

→ [Submit to UCT Lawspace authorized users](#)

→ [Edit Profile](#)

→ [Help](#)

→ [Contact us](#)

University of Cape Town Lawspace

Welcome to UCT Lawspace - A Digital Repository of research done in the Law Faculty at the University of Cape Town! UCT Lawspace also hosts a new Constitutional Assembly Database which is under construction.

This Repository is managed by Professor Julien Hofman (Department of Commercial Law) and Tobias Schonwetter (PhD student, Department of Commercial Law) and administered by Sheryl Ronnie (Faculty Office).

The individual authors hold the copyright in the works on the Repository. In the interests of education and research the managers of the Repository support a broad and user-orientated interpretation of the South African fair dealing provisions. The managers encourage authors and copyright-holders to submit work to the Repository under the [Creative Commons South Africa License](#).

About UCT Lawspace

At present the Repository holds a selection of LL.M. Minor Dissertations in full text. Dissertations (including Ph.D. and Masters Dissertations), publications and conference papers will be added regularly.

Click here to access the [UCT Law Faculty's website](#)

Collections in UCT Lawspace

[Commercial Law](#)

[Criminal Justice](#)

[Private Law](#)

[Public Law](#)

source: UCT Lawspace, Univ. of Cape Town
<http://lawspace.law.uct.ac.za>



OpenDOAR's view on South Africa

Results **1 - 5** of **5**.

Page: << **Prev**

Electronic Thesis and Dissertation Database

Organisation: [University of Johannesburg \(UJ\)](#), South Africa

UPSpace (Institutional Research Repository - University of Pretoria)

Organisation: [University of Pretoria - Academic Information Service](#), [University of Pretoria \(UP\)](#), South Africa

Rhodes eResearch Repository (ReRR)

Organisation: Rhodes University Library, [Rhodes University](#), South Africa

UCT Computer Science Research Document Archive (UCT CS Archive)

Organisation: Department of Computer Science, [University of Cape Town](#), South Africa

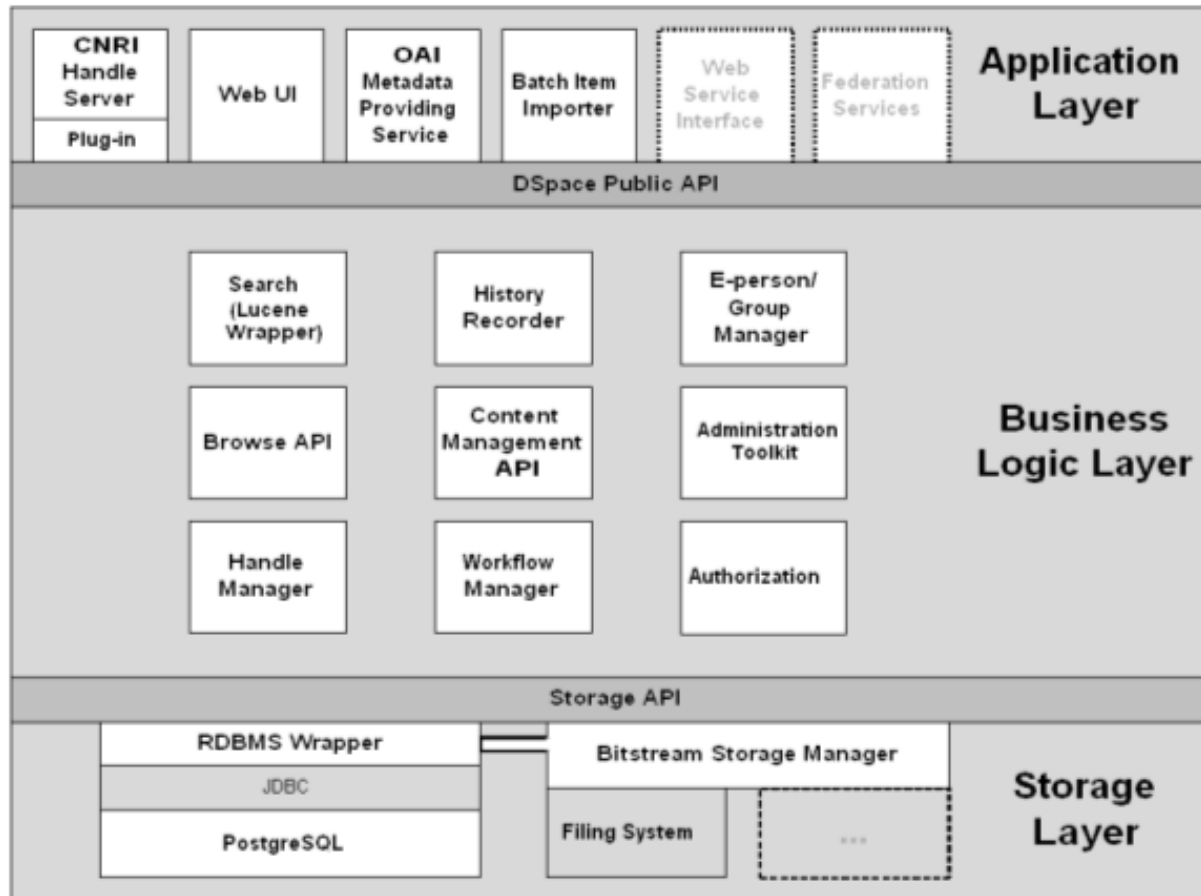
University of Pretoria Electronic Theses and Dissertations (UPeTD)

Organisation: Academic Information Service, [University of Pretoria \(UP\)](#), South Africa

Page: << **Previous** **1** **Next** >>



How to Build a Digital Repository



source: *DSpace: An Open Source Dynamic Digital Repository*, DLib Magazine
<http://www.dlib.org/dlib/january03/smith/01smith.html>



(Open Source) Repository Packages



Welcome to UCT CS Research Document Archive



Welcome to the [UCT Computer Science](#) Research Document Archive, which archives and makes accessible documents that are products and by-products of research in the department.

Search the Titles, Authors, Abstracts and Keywords:

Browse

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

Latest Additions

View items added to the archive in the past week.

Simple Search

Search the archive using the most common fields.



Poverty Alleviation Workshop HOME HELP PREFERENCES about

search titles

Search for documents that contain of the words

About this collection

This collection contains 101 documents. It was last built 2288 days ago.

How to find information in the Poverty Alleviation Workshop collection

There are 2 ways to find information in this collection:

- search for particular words that appear in the text by clicking the Search button
- browse documents by Title by clicking the Titles button

NDLTD DOCUMENT ARCHIVE

Search Documents

[Advanced Search](#)

[Home](#)

Browse

- [Communities & Collections](#)
- [Titles](#)
- [Authors](#)
- [By Date](#)

Sign on to:

- [Receive email updates](#)
- [Submit/Review authorized users](#)
- [Edit Profile](#)
- [Administer Site](#)

[Help](#)

[NDLTD Home](#)

Welcome to the digital repository of documents related to Networked Digital Library of Theses and Dissertations (NDLTD) and Electronic Theses and Dissertations (ETDs) in general.

Search

Enter some text in the box below to search Document Archive.

Communities in NDLTD Document Archive

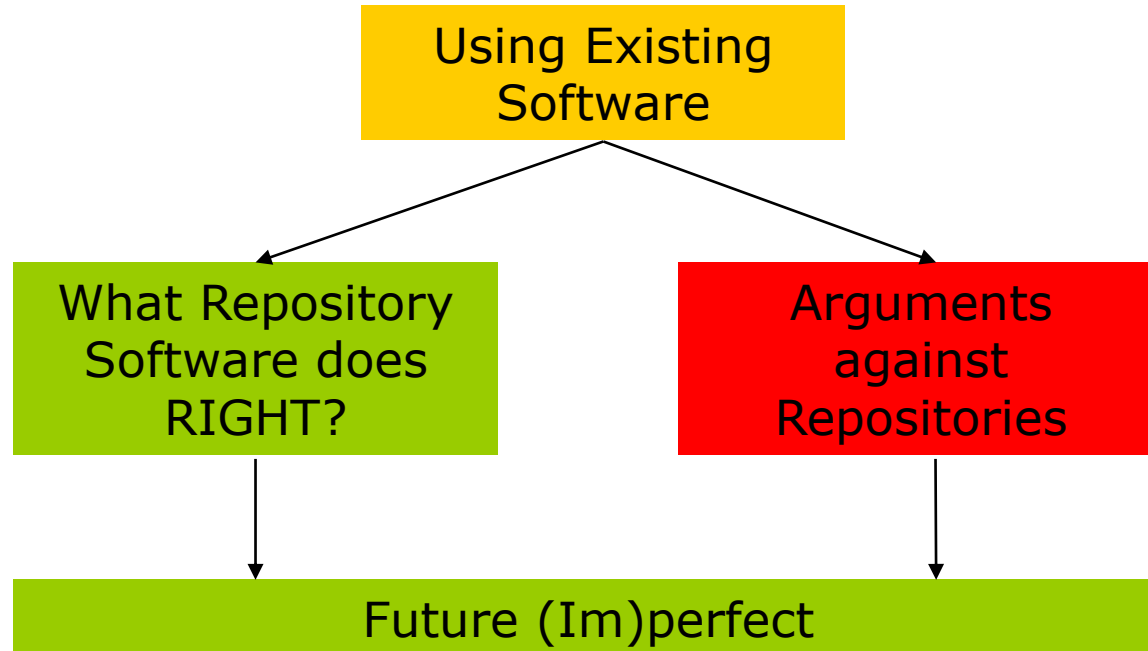
Choose a community to browse its collections.

- [Documents](#) [126]
- [International Conferences, Workshops and Symposia](#) [55]
- [National Conferences, Workshops and Symposia](#) [0]

Networked Digital Library of Theses and Dissertations - Feedback



Outline 2



What does repository software do RIGHT?

□ Infrastructure

- Digital Objects are stored/archived
- Users can access items from the Web

□ Services

- Search full-text and metadata
- Browse by author/title/etc.

□ Interoperability

- OAI-PMH interoperability
- Ability to ingest and export items

□ Security

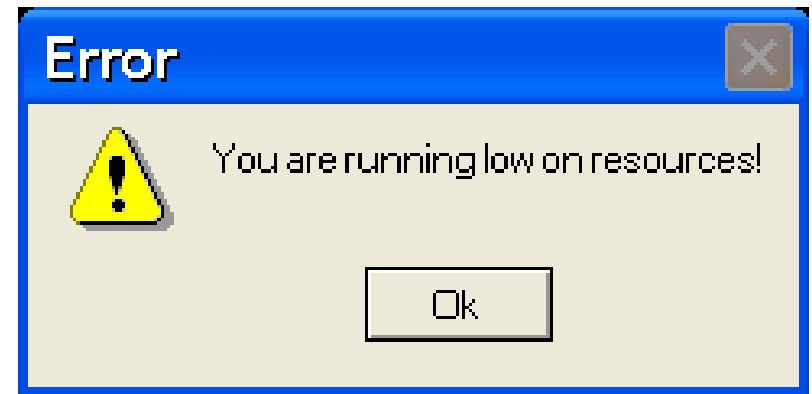
- User roles and authentication



Arguments against OA Repositories

Digital Repositories are a lot of hype – the technology is actually not yet mature enough to be practical or usable!

It is so difficult for us as poor institutions in South Africa, with few staff and inadequate facilities and training.



Training, Policies, Staff, Tools, ...

□ Training

- Have you been to a Sivulile workshop?

□ Policies

- Listen to the other speakers!

□ Staff

- What proportion of your library staff are IT vs. how many customers use IT rather than physical resources?

□ Tools

- Are they any good? or are the tools pure EVIL?

sivulile



open access
south africa



Repository Evils ?

- No integration with Windows/Linux/BSD
 - Modern operating systems have packaging systems – many IR systems are still distributed in “source code”.

Why not?

Applications

- ✂ **Digital Libraries**
 - ✓  **DSpace**
Digital Repository
 - ✓  **Character Map**



Repository Evils ?

□ Low-level Components

- No clean external API to communicate with services within most systems.
- None of DSpace, EPrints, Greenstone ...



Repository Evils ?

□ Customisation requires programmer!

Date: Mon, 23 Oct 2006 20:58:20 +0100
From: Christopher Gutteridge <cjg@ecs.soton.ac.uk>
To: "EPrints.org Technical List" <eprints-tech@ecs.soton.ac.uk>
Subject: Re: [EP-tech] Chicago citation style and e-prints

To show the family name first last you can do
@creators;order= fg@
or gf

gf= given family
fg= family, given

To change the separators to what you want add the following to your
local phrase file (not system-phrases)
<epp:phrase id= "lib/metafield:join_name">, </epp:phrase>
<epp:phrase id= "lib/metafield:join_name.last"> and </epp:phrase>

But that can't be set on a per-citation type level.



Repository Evils ?

□ Customisation requires programmer!

Date: Wed, 18 Oct 2006 10:24:45 +1300 (NZDT)

Subject: Re: [greenstone-users] Another question: multivalued fields

From: sw64@cs.waikato.ac.nz

Hello Ed,

You can use AZCompactList classifier, for which the "allvalues" parameter is the default. Also, you can use [sibling:Subject] or [sibling:Author] format statement to display all multiple values of Subject or Author in VList,

Regards
Shaoqun

- > Thank you for your help a few weeks ago. I did get a sample collection
- > working, and have just heard that the organisation is keen to continue – I
- > now need to get a couple of things working that I didn't finish before.
- > One of those is allowing for multiple values for Author and Subject. I have
- > created additional columns in the .cfg file and put some test data in.
- > Where is the "allvalues" parameter put and can it be done through the GLI
- > or do I have to edit the config file?



Repository Evils ?

□ Poor Scalability

- Most systems do not scale well beyond small collections.

EPrints

The database layout is not normalised and it uses the database in a non-traditional way, as is evident when it is indexing content. With our test data, it took roughly around 5 hours to index approximately 100,000 objects. To compound the problem the indexer will re-index the 100,000 objects each time it runs, as EPrints does not yet support incremental indexing.

DSpace

Preliminary evaluation of DSpace did not indicate any scalability problems. However after loading it with test data, there was obvious 'sluggishness' with

source: Technical Evaluation of selected Open Source Repository Systems, Catalyst IT



Repository Evils ?

- Identity not easily removable
 - DSpace is the name of the software, not the archive!

This letter was produced in Microsoft Word – please use Microsoft Word to view it properly.

Dear John

Using Microsoft Word, I regret to inform you that ...

Yours sincerely
John

|
file: Microsoft Word.letter to john.2006



Repository Evils ?

The screenshot shows the Cornell University DSpace Open Access Repository homepage. Several elements are circled in red:

- The browser address bar: <http://dspace.library.cornell.edu/index.jsp>
- The Cornell University Library logo.
- The main title: **DSpace Open Access Repository**
- The navigation menu: [Home](#) | [About Us](#) | [Browse](#) | [Search](#) | [Communities in DSpace](#) | [Collections in DSpace](#) | [Contact Us](#)
- The search bar: Search Repository: Go [Advanced Search](#)
- The navigation links: [Home](#), [About Us](#), [Help](#)
- The browse links: [Communities & Collections](#), [Titles](#), [Authors](#), [By Date](#)
- The sign-on links: [Receive email updates](#), [My DSpace authorized users](#), [Edit Profile](#)
- The main content area: **Welcome to the Cornell University Open Access Repository!**
- The search box: Search

The main content area contains the following text:

This Open Access Repository (OAR) is powered by [DSpace](#) and is open for anyone at Cornell University as a place to capture, store, index, preserve and redistribute Cornell faculty, staff, student or organizational research materials in digital formats.

Members of the Cornell Community, who wish to submit items to this OAR, must [register](#) with us.

If you would like to be notified when new items are submitted to a Community or Collection of interest to you, click on: [Receive email updates](#).

Some of our collections are restricted to specific members of the Cornell Community. If you wish access to any of these restricted resources, please contact us at dspace-admin-l@cornell.edu.

The Cornell University Library offers a wide variety of digital repositories and collections. To see what we have to offer go to our [Digital Initiatives](#) page.

Enter some text in the box below to search the Respository.

[Advanced Search](#)

For information on protecting your rights to your intellectual property, go to www.arl.org/sparc/author/addendum.html.

For information on copyright management at Cornell University, go to www.library.cornell.edu/scholarlycomm/copyright.



Repository Evils ?

□ Buy-In / Lock-In

- How easy is it to switch from DSpace to EPrints to Fedora to ... ?



Repository Evils

- Are these issues significant roadblocks?
- Are they being addressed at all?



Current Development Directions

Problem	Solution
Low-level interfaces	Greenstone 3 using Web Services
Scalability	Key concern in DSpace 3 architecture
Configurability	Tools are becoming more automated and flexible over time
OS integration	Will improve as tools become mature and development stabilises
Lock-in and Identity	Will improve when tools don't need to be "sold"



Current Research Directions

- Fedora
 - Generic interface to scalable repository
- Pathways / OAI-ORE
 - (Submit) Interface to repository components
- OCKHAM
 - Service registry for composition of systems from components
- DILIGENT
 - Scalability of repositories in a grid
- AJAX
 - Interactive Web-based interfaces

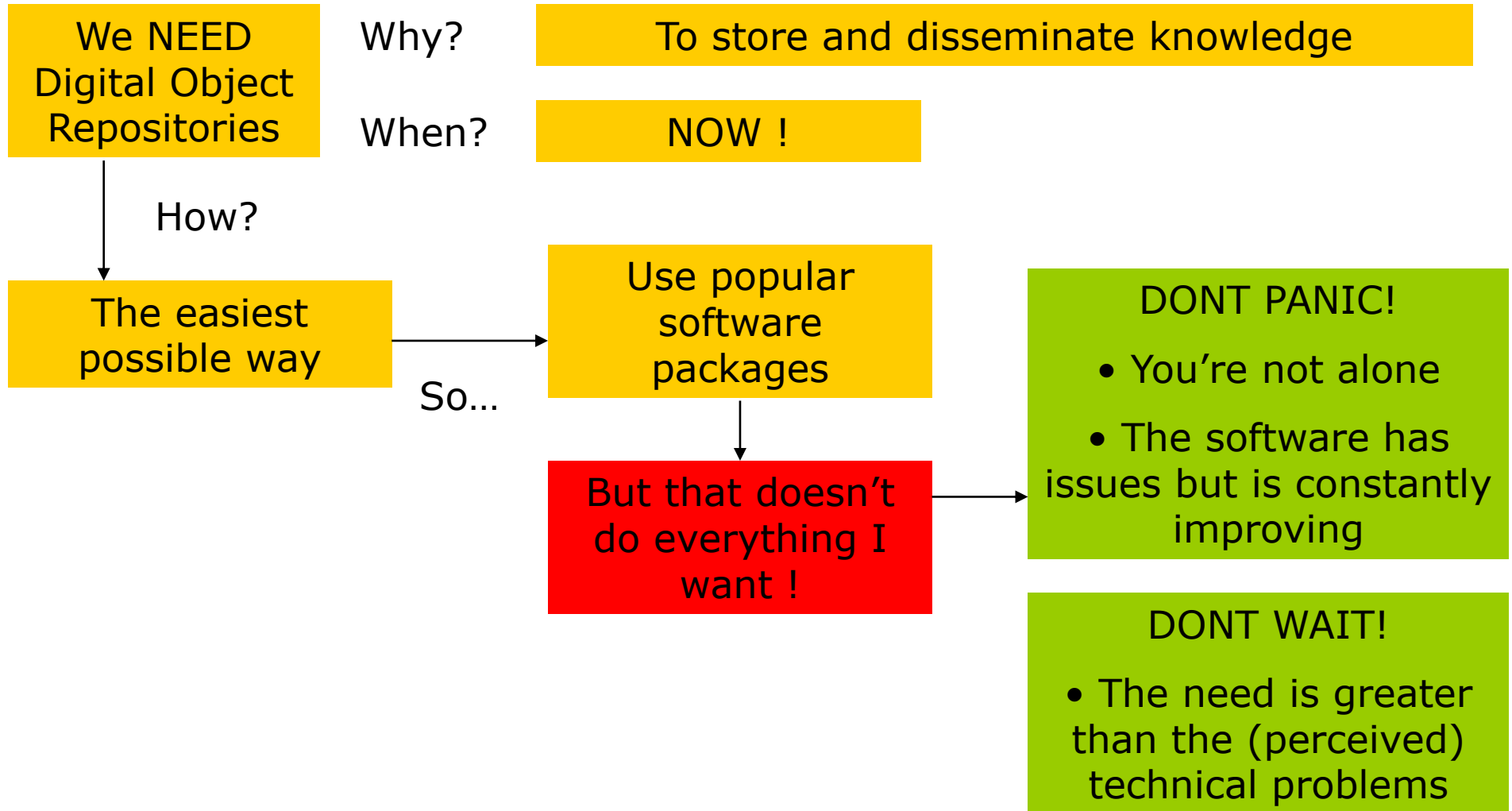


Future Im(perfect)

- Repository software gets a lot RIGHT,
- Repository software still has issues,
 - Software use should not require a programmer
 - Software is a means to an end, not an end
 - Software should be appropriately engineered for reuse on a large scale like other OSS tools
 - Software should easily integrated into other systems
- ... but most of these issues have been noted and/or are being actively addressed



Bottom Line



Open Access and Institutional Repositories NOW!



the end.

