Institutional Repositories

*a big picture*

Hussein Suleman
hussein@cs.uct.ac.za

University of Cape Town
Department of Computer Science
Digital Libraries Laboratory

May 2008
Overview

- **Past**
  - Digital Collections The Hard Way
    - 2 Case Studies

- **Present**
  - The Low-hanging Fruit
    - Institutional Repositories
    - ETDs

- **Future**
  - Ubiquitous Digital Collections

- Analysis and Implications
The W3C-WCA Repository

Past  Present  Future  Analysis

Search Repository

Search:  Submit

Browse Listing of Resources

All Publications

Complete Listing
(Publications, Tools and Traces)

All Tools

All Traces

Add New Resource

Read the procedure on submitting resources and then ...

Add New Resource

Log File Formats

Add New File Format

List File Formats

Download Validation Software

Conferences
Journals
Technical Reports
Theses/Dissertations
Drafts
Books
Other

General tools
Client logs
Server logs
Proxy logs
Network traffic
Logs of search engines
Miscellaneous logs
What were we storing?

- 66.249.67.89 - [10/Feb/2008:14:59:16 +0200] "GET /localdocs/openldap/rfc/rfc3712.txt HTTP/1.1" 200 62301 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

- 65.214.45.140 - [10/Feb/2008:15:01:07 +0200] "GET /localdocs/Sablot/jsdom-ref/apidocs/api-Node.html HTTP/1.0" 200 5451 "-" "Mozilla/5.0 (compatible; Ask Jeeves/Teoma; +http://about.ask.com/en/docs/about/webmasters.shtml)"

- 66.249.67.89 - [10/Feb/2008:15:02:01 +0200] "GET /localdocs/openldap/rfc/rfc2294.txt HTTP/1.1" 200 22059 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"


- 66.249.67.89 - [10/Feb/2008:15:04:55 +0200] "GET /localdocs/openldap/rfc/rfc2255.txt HTTP/1.1" 200 20686 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

- 66.249.67.89 - [10/Feb/2008:15:07:45 +0200] "GET /localdocs/openldap/rfc/rfc2253.txt HTTP/1.1" 200 18226 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

- 66.249.67.89 - [10/Feb/2008:15:08:59 +0200] "GET /workshops/lesotho HTTP/1.1" 200 47246 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

- 66.249.67.89 - [10/Feb/2008:15:10:55 +0200] "GET /localdocs/openldap/rfc/rfc3296.txt HTTP/1.1" 200 27389 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

- 66.249.67.89 - [10/Feb/2008:15:14:16 +0200] "GET /localdocs/libwmf/html/classes.html HTTP/1.1" 200 7529 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

Past

Present

Future

Analysis
Open Access?

- All documents and data (mostly log files from servers) were open to the world.

- This was the first OAI-compliant archive!
  - The repository was used as a testbed when the OAI protocols were developed.

- Data collections were contributed by:
  - AOL, Altavista, Boeing, etc.
Features

- Search and Browse interfaces.
- Users could submit documents.
- An editor always checked submissions.
- Data was carefully validated!
Validation

- Format of log files was defined using XML descriptions (long before XML schema).

  `<field>
      <quote/> <method/> <space/> <url/> <space/> <httpversion/> <quote/>
  </field>`

- Every dataset was tested against its formal definition using a special high-performance validation program.
  - Data items failing validation were excluded from the repository!
  - Datasets were certified (or not).
  - Certification and error report were added to metadata.
Scalability

- All datasets were stored on FTP servers.
  - FTP was (is?) more efficient than HTTP for large data transfers.
- Repository only handled metadata through Web UI.
- Datasets were very large and often distributed.
  - Magnetic tapes were sent by courier to exchange data among sites.
  - A tape robot storage system was used on the central server.
  - Subsets of data were kept online if complete datasets were too large.
Linking

- Repository initially only contained datasets, but expanded to include papers and tools.
- Metadata records for datasets and documents were stored in the same database.
- Search/browse operated over all metadata.
- Datasets were linked to documents in the metadata.
Datasets were contributed by their owners.

All sensitive data was first removed (by the submitters or archive managers).

Tools were developed to anonymise datasets (e.g., removing IP addresses).

The tools were themselves added to the repository!
Computer Science Teaching Ctr

Welcome!

The CSTC is a digital library of reviewed resources for teaching computer science. We invite you to submit a resource and browse our collection.

Submit a Resource

- Step 1: Create a CSTC account
- Step 2: Login to CSTC
- Step 3: Complete Submission Application

CSTC Newsletter

Our monthly newsletter keeps you informed about the latest additions to the library. To subscribe to the newsletter, simply create an account and put a check in the checkbox under the CSTC newsletter section. If you want to unsubscribe at any time, just go to the "account info" page and deselect the checkbox.

Volunteer as a Reviewer

- Step 1: Create a CSTC account

ACM JERIC

Resources submitted to CSTC may be considered for publication in the ACM Journal on Educational Resources in Computing (JERIC).
JERIC

Journal of Educational Resources in Computing (JERIC)

Online Submission and Review System

Welcome to the JERIC Submission and Review system.

From here you can submit papers and resources for consideration for publication in ACM JERIC, and if you are a reviewer or editor, you can submit reviews or manage the reviewing process respectively.

blah blah blah ...

blah blah blah

© 2002 ACM JERIC
CSTC / JERIC

- Developed in mid-1990s.
- Collection of teaching-related resources in Computer Science.
- Strong emphasis on peer review.
- Workflow of modern IR systems.
- JERIC prototyped an online journal review system.
  - Journal publication credit as an incentive for contributing resources to CSTC!
Open Archives Initiative

The Open Archives Initiative Protocol for Metadata Harvesting

Protocol Version 2.0 of 2002-06-14
Document Version 2004/10/12T15:31:00Z
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Previous protocol version: Protocol Version 1.1 of 2001-07-02
Instructions for migrating from Version 1.1 to 2.0
Implementation Guidelines

Editors
The OAI Executive:
Carl Lagoze <lagoze@cs.cornell.edu > -- Cornell University - Computer Science
Herbert Van de Sompel <herbertv@lanl.gov > -- Los Alamos National Laboratory - Research Library

From the OAI Technical Committee:
Michael Nelson <m.l.nelson@larc.nasa.gov > -- NASA - Langley Research Center
Simeon Warner <simeon@cs.cornell.edu > -- Cornell University - Computer Science

Table of Contents
1. Introduction
2. Definition
The Budapest Open Access Initiative arises from a small but lively meeting convened in Budapest by the Open Society Institute (OSI) on December 1-2, 2001. The purpose of the meeting was to accelerate progress in the international effort to make research articles in all academic fields freely available on the internet. The participants represented many points of view, many academic disciplines, and many nations, and had experience with many of the ongoing initiatives that make up the open access movement. In Budapest they explored how the separate initiatives could work together to achieve broader, deeper, and faster success. They explored the most effective and affordable strategies for serving the interests of research, researchers, and the institutions and societies that support research. Finally, they explored how OSI and other foundations could use their resources most productively to aid the transition to open access and to make open-access publishing economically self-sustaining. The result is the Budapest Open Access Initiative. It is at once a statement of principle, a statement of strategy, and a statement of commitment.

The initiative has been signed by the Budapest participants and a growing number of individuals and organizations from around the world who represent researchers, universities, laboratories, libraries, foundations, journals, publishers, learned societies, and kindred open-access initiatives. We invite the signatures, support, and participation of the entire world scientific and scholarly community.

E-mail: openaccess@scros.org

5210 signatures have been added to the initiative
Institutional Repositories

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 Julian Hofman (Department of Commercial Law) and Tobias Schonwetter (PhD candidate, Department of Commercial Law) and administered by Sue Wright.

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-oriented 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.

Collections in UCT Lawspace

Select to browse a Collection.

- Commercial Law
- Criminal Justice
- Private Law
- Public Law

South African Constitutional Assembly Database (Collection still under construction)
Institutional Repositories

- Collections of pre-prints and post-prints
- Research-centric
- Self-archiving
- Open Access
- Lots of software support: Dspace, Eprints, etc.
- Large international community
- Many established collections
- Clear value proposition
Welcome to ETD Individuals

This server supports submission of, and access to, electronic theses and dissertations (ETDs). We welcome works submitted by individuals whose degree granting institution does not have a program for ETDs. Works can be at the bachelor's, master's, doctoral, post-doctoral, and other levels. They typically will be a thesis or dissertation, but can be a project report or other resource prepared by a student with college or university supervision. Browsing and searching are anonymous. But if you wish to submit a work, you must first register.

If you need a bit more help getting started, the EPrints team has a commercial services arm which provides training, support, hosting and consultancy services.

Latest Additions
View items added to the repository in the past week.

Search Repository
Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

Browse Repository
Browse the items in the repository by subject.
<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ETD Collections</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Electronic Theses and Dissertations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submission/Review Workflow</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examination documents</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Similar to Pre/Post-print IRs, but different</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Reasons for existence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Legal status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Workflow</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Level of importance to institutions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Higher Level Services

Past

Present

Future

Analysis

Networked Digital Library of Theses and Dissertations

We found 382273 matching items.

Language

English [200213]
Portuguese [57171]
French [9202]
Swedish [8584]
German [4926]
Lithuanian [4181]

Table: Language

<table>
<thead>
<tr>
<th>Language</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>200213</td>
</tr>
<tr>
<td>Portuguese</td>
<td>57171</td>
</tr>
<tr>
<td>French</td>
<td>9202</td>
</tr>
<tr>
<td>Swedish</td>
<td>8584</td>
</tr>
<tr>
<td>German</td>
<td>4926</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>4181</td>
</tr>
</tbody>
</table>

Show More/Less

Continents

North America [221327]
Europe [47185]
South Pacific [36987]
Asia [16404]
Latin America [2687]
Africa [800]

Table: Continents

<table>
<thead>
<tr>
<th>Continent</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>221327</td>
</tr>
<tr>
<td>Europe</td>
<td>47185</td>
</tr>
<tr>
<td>South Pacific</td>
<td>36987</td>
</tr>
<tr>
<td>Asia</td>
<td>16404</td>
</tr>
<tr>
<td>Latin America</td>
<td>2687</td>
</tr>
<tr>
<td>Africa</td>
<td>800</td>
</tr>
</tbody>
</table>

Show More/Less

Country

USA [168321]
Canada [53006]
Australia [36356]
Sweden [18017]

Table: Country

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>168321</td>
</tr>
<tr>
<td>Canada</td>
<td>53006</td>
</tr>
<tr>
<td>Australia</td>
<td>36356</td>
</tr>
<tr>
<td>Sweden</td>
<td>18017</td>
</tr>
</tbody>
</table>

Show More/Less

Search for Items:

New Search

Search

Results 1 - 10 of 382273

1. Teoria de sistemas vibratorios aporticados não-lineares e não-ideais
   Author: Jorge Luís Palacios Felix
   Date/Publisher: 2011-09-13 by Biblioteca Digital da Unicamp.
   Language: Portuguese
   Description 1: Apresenta-se o estudo do comportamento dinâmico não lineal de um pórtico plano, excito por uma fonte não ideal - um motor elétrico de corrente contínua, desbalanceado e de potência limitada. Toma-se, um problema cujo modelo matemático representa um sistema simplificado (com característica do motor em regime estacionário) e completo (considera-se a equação eléctrica do motor). Adota-se a formulação Lagrangeana para gerar as equações de movimento, contendo termos não lineares até ordem cúbica. To Show More/Less

2. Saving, wealth, and retirement: evidence from the health and

UNIVERSITY OF CAPE TOWN

department of Computer Science
Higher Level Services

Search ETDs

Search

hussein suleman

ETDs

All of the scientific web

Search

Searched for: All of the words hussein suleman
Found: 3 results
Sort by: relevance | date

Open Digital Libraries

Suleman, Hussein. Nov 2002
Digital Libraries (DLs) are software systems specifically designed to assist users in information seeking activities. Stemming from the intersection of library sciences and computer networking, traditional DL systems impose library philosophies of ...

Open digital libraries

Suleman, Hussein. Jan 2002

Genetic Programming in Mathematica

Suleman, Mr Hussein. Jan 1997
GP has traditionally been implemented in LISP but there is a slow migration towards faster languages like C++. Any implementation language is dictated not only by the speed of the platform but also by the desirability of such an implementation. With a
Higher Level Services

Computing and Information Technology Interactive Digital Educational Library

Search CITIDEL

Advanced Search

Subject Search

Browse

Communities & Collections

Title

Authors

Subject

By Date

Sign on to:

Receive email updates

My CITIDEL authorized users

Edit Profile

Please use this identifier to cite or link to this item: http://hdl.handle.net/10117/28

Title: JPEG interactive Visualization

Keywords: Computing Methodologies sub_subject: Simulation and Modeling Interactive, visualization, JPEG

Issue Date: 12-Jun-2000

Abstract: Interactive visualization of the JPEG compression algorithm.

URI: http://www.citidel.org/handle/10117/28

Other Identifiers: 89

Appears in Collections: Computer Science Teaching Center

Files in This Item:
Higher Level Services

- Harvest / Search / Browse
- Uncertain
  - Costing
  - Ownership of data/metadata
  - Benefits
  - How to merge collections – what is relevant?
Open Educational Resources

Unlocking Knowledge, Empowering Minds.

Free lecture notes, exams, and videos from MIT. No registration required.

“OpenCourseWare is exactly the kind of thing that universities should be doing.”

Larry S. Barsamian
MIT Class of 1969 and OCW supporter
United States

FEATURED COURSE

6.032 User Interface Design and Implementation,
Fall 2004

Say the colors of each word aloud. Is it difficult? That's because of the Stroop effect.

OCW is grateful for the support of:

department of Computer Science
Open Educational Resources

THE CAPE TOWN OPEN EDUCATION DECLARATION

The Cape Town Open Education Declaration

The Cape Town Open Education Declaration arises from a small but lively meeting
convened in Cape Town in September 2007. The aim of this meeting was to accelerate
efforts to promote open resources, technology and teaching practices in education.

Convened by the Open Society Institute and the Shuttleworth Foundation, the meeting
gathered participants with many points of view from many nations. This group discussed
ways to broaden and deepen their open education efforts by working together.

The first concrete outcome of this meeting is the Cape Town Open Education
Declaration. It is at once a statement of principle, a statement of strategy and a
statement of commitment. It meant to spark dialogue, to inspire action and to help the
open education movement grow.

Open education is a living idea. As the movement grows, this idea will continue to
evolve. There will be other visions, initiatives and declarations beyond Cape Town. This
is exactly the point. The Cape Town signatories have committed to developing further
strategies, especially around open technology and teaching practices.

The Declaration has already been signed by hundreds of learners, educators, trainers,
authors, schools, colleges, universities, publishers, unions, professional societies,
policymakers, governments, foundations and other kindred open education initiatives
around the world. We encourage you to join us.

This work is licensed under a Creative Commons Attribution 3.0 License.
## Open Educational Resources

- Computer-Based Training
- Syllabi, Tests, Exams, Notes, Slides, etc.
- Course Packs
- Uncertain
  - Legal framework
  - Motivation/reward system
  - Responsibility
  - Linking into Learning Management Systems
THE DIGITAL BLEEK AND LLOYD

The Digital Bleek and Lloyd includes scans of every page of the 114 Lucy Lloyd /Xam notebooks, 13 Lloyd (mostly) /Xam notebooks and 28 Wilhelm Bleek /Xam notebooks. It also includes Jemima Bleek’s solitary Korana and /Xam notebook (but excludes three Korana notebooks in the collection of the Maingard Library at the University of South Africa) and all the drawings and watercolours made by Ikanqaso, Ditaqwa, Tamane, numa, Inani and Da. This represents the coming together of an archive dispersed across three institutions: the National Library, Iziko South African Museum and the University of Cape Town. With the scans is a 280 000-word searchable index, cross-referenced and including notes and summaries for each of the stories listed. Notes in italics are direct quotes from the reports of Bleek and Lloyd in which they detail the progress of their research.

KEYWORD SEARCH

Books
- Cover to cover
- Contributor
- Story
- Category
- Keyword

Drawings/Watercolours
- Contributor
- Category
- Keyword

Contributors
- /Xam
- /Xun
- Researchers
Cultural Heritage

- Collections of digitized objects related to heritage.
- Uncertain
  - Tools
  - Workflows
  - Sustainability
Personal Digital Collections

Past

Present

Future

Analysis
Personal Digital Collections

- Personal collections of documents/files.
- Unclear
  - Tools
  - Best practices
  - Archiving and preservation approaches
  - How to deal with scale
  - How to integrate with others
  - Metadata
Electronic Communication
Electronic Communication

Hussein Suleman

Hi Christopher,

some minor comments can be found at:

dont worry about the acknowledgements for now ... if the paper is accepted we can add that later.

ttfn,
-----hussein

--------------------------------------------------
hussein suleman ~ hussein@cs.uct.ac.za ~ http://www.husseinsspace.com
--------------------------------------------------

cparker@cs.uct.ac.za wrote:

Hi,

Sorry about the "spam" - I have included the sections in the text file and fixed some spelling and such after reading it through. One final mistake is on the last bit of text on page 2 - "of" must go. I forgot to ask you about acknowledgments for NRF. Shall I put these at the end after conclusion? Do you perhaps have the grant number for me?

Electronic Communication

- Personal/group collections of messages, conversations and collaborations.
- Network (sequence) of actions.
- Unclear
  - Public status
  - Tools
  - Best practices
  - Archiving and preservation approaches
  - How to deal with scale
Administrative Documents

Faculty handbooks (Handbooks 6 - 11)

Commerce
- Commerce - 2008 [download] (14 mb)
- Commerce, Part 1 - 2007 [download] (730 kb)
- Commerce, Part 2 - 2007 [download] (870 kb)
- Commerce - 2006 [download] (12 mb)
- Commerce - 2006 [download] (12 mb)
- Commerce - 2004 [download] (12 mb)

Engineering & the Built Environment - 2008 [download] (13 mb)

Health Sciences - 2008 [download] (2 mb)

Humanities
- Graduate School of the Humanities (postgraduate study) - 2008 [download] (18 mb)
- Performing and Creative Arts - 2008 [download] (740 kb)
- Undergraduate - 2008 [download] (1.2 mb)

Law - 2008 [download] (860 kb)

Science
## Administrative Documents

- Any set of documents related to administrative activities.
- Forms, handbooks, records, etc.
- Unclear
  - Tools
  - Best practices
  - Archiving and preservation approaches
  - How to deal with scale
  - Workflow
  - Legal requirements
ePortfolios

UCT Department of Computer Science
Honours Project Archive 2007

- An Extensible Software Testbed for Wireless Mesh Networks
  By: Paolo Pileggi
  File: pileggi.zip [View online] [Download]

- A System for Automated Geo-referencing
  By: Aziz Hendricks, Irwan Rakiep
  File: hendricks_rakiep.zip [View online] [Download]

- Cellphone Shopper
  By: Graham Hunter, Marc Pelteret, Tshifhiwa Ramuhaheli
  File: hunter_pelteret_ramuhaheli.zip [View online] [Download]

- Creating Usage Sensitive Knowledge Structures
  By: Sean Colledge, Ian Sander
  File: colledge_sander.tar.gz [View online] [Download]

- CyberTracker Mobile
  By: Liyaqat Mugienkar, Garth Whitaker
  File: mugienkar_whitaker.zip [View online] [Download]

- Developing Country E-Commerce Portal
  By: Khilali and Makave, Shikwakala
  File: kingdom_Makave_shikwakala.zip [View online] [Download]
### ePortfolios

- Persistent collections of student work.
- Either group- or student-based.
- Unclear
  - Complex object storage
  - Linking
  - Repository vs. Local stores
  - Ownership

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Big Picture

Higher Level Services

- Administrative Documents
- Cultural Heritage
- Educational Resources
- IR
- ETDs
- Data
- Communication
- ePortfolios

Past

Present

Future

Analysis

personal

external

institutional
Implications - Collections

- This is just the current known – other collection types may emerge.
- Common features in multiple digital collections.
- One repository does not fit all – it is not clear how to map collections to repositories.
- One software system (Dspace/Greenstone/XYZ) does not fit all.
## Implications - Institutions

- Digitization in all facets of university life.
- Need ETD/IR policies now.
- Need Digital Strategies in general.
- Need to move beyond donor funding to GOB.

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Implications - Staff

- Roles are not clear.
- Librarians and IT people have an increasing role to play.
- No one group owns digital domain.
- Common skills for different projects.
- Need for reskilling (end users, librarians, IT staff, etc.)
- Challenges or Opportunities?

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Conclusion ...

- Lots of common problems are being solved in general terms:
  - Multilingual issues
  - Usability of systems
  - Scalability of systems
  - Workflows, processes and roles
  - Legalities

- It is OK to tackle the low-hanging fruit ...
- But be aware of other digital efforts.
That's All Folks!

hussein@cs.uct.ac.za

Digital Libraries Laboratory
UCT Computer Science