Institutional Repositories: Why they are important for African Research

Hussein Suleman
hussein@cs.uct.ac.za

University of Cape Town
Department of Computer Science

November 2007
WHAT
What Exactly is an Institutional Repository?

- Collection of research-related documents
- Author self-submission
- Checking of submissions
- Archive-everything!
- Hierarchical browsing
- Simple and fielded searching
- OAI-PMH compliance
WHY
because of Direct access from GOOGLE 1/3
Digital Libraries


Full text available as:
PDF - Requires Adobe Acrobat Reader or other PDF viewer.

EPrint Type: Book Chapter

H Information Systems: H.0 GENERAL
H Information Systems: H.1 MODELS AND PRINCIPLES

Subjects: H Information Systems: H.3 INFORMATION STORAGE AND RETRIEVAL
A General Literature: A.1 INTRODUCTORY AND SURVEY
H Information Systems: H.4 INFORMATION SYSTEMS APPLICATIONS

ID Code: 16
Deposited By: Suleman, Hussein
Deposited On: 11 July 2003
because of Direct access from GOOGLE 3/3

Digital Libraries
By Edward A. Fox, Hussein Suleman, Devika Madalli, and Lillian Cassel
Draft 5 / May 2, 2003

For: Practical Handbook of Internet Computing
(to be published by CRC Press)

1. Introduction
Definitions of “digital library” (DL) abound [1] [2], but a consistent characteristic across all definitions is an integration of technology and policy. This integration provides a framework for modern digital library systems to manage and provide mechanisms for access to information resources. This involves a degree of complexity that is evident whether considering: the collection of materials presented through a digital library; the services needed to address requirements of the user community; or the underlying systems needed to store and access the materials, provide the services, and meet the needs of patrons. Technologies that bolster digital library creation and maintenance have appeared over the last decade, yielding increased
because of the No of ACCESSES from Google

- 1734 hits directly from Google in March 2007.
- Example:
  - [http://www.google.com/search?q=questionnaire+system+UML](http://www.google.com/search?q=questionnaire+system+UML)
because academics can manage a research

OVERVIEW

2004

Conference Paper


Conference Poster


Departmental Technical Report


because WEB pages are not sufficient

Problems:
- Persistence – will the documents always be there?
- Authority – can we trust the authenticity of the website?
- Standards – are the formats and metadata the same across all websites?
- Discovery – how do we find these documents? Google?
because of REPORTING

- You can create a departmental or university listing of all research output.
- Where copyright does not allow, include just a citation – no files – for completeness.

NDLTD Union Catalog Project


Full text available as:
- **EPrint Type:** Book Chapter
- **Subjects:** H Information Systems: H 4 INFORMATION SYSTEMS APPLICATIONS
- **ID Code:** 14
- **Deposited By:** Suleman, Hussein
because of SHARING

- Most Institutional Repositories are compliant with Open Archives Initiative’s Protocol for Metadata Harvesting (OAI-PMH) v2.0.
- Metadata can be freely harvested by any service provider, e.g., search engines.
- Thus more people know what you have and can find it easily.
because of how many USERS we get

![Non-local Access Chart]

Legend:
- All hits
- PDF hits
because of no. of SEARCH ENGINES we get
because of how many RESOURCES get viewed
because of how OFTEN resources get viewed
because SUBJECT repositories failed

- In the 90’s subject archives were popular e.g., arXiv, RePEc, NCSTRL.

- Problems:
  - Sustainability – repositories run by organisations with limited funding.
  - Development skill – staff needed to develop and maintain all software needed to run the archive.
  - Should the archive be centralised or distributed?
  - Owner of data was not best provider of services.
because COPYRIGHT is changing in our favour

- Most society publishers will allow archiving on a website or IR e.g., ACM
- Most commercial publishers allow archiving on a website or IR after some time (typically 12-24 months), if not immediately.
- Newer commercial publisher agreements make greater allowance for IRs.
- You can always negotiate with a publisher!
because 400000 ETDs are already out there

<table>
<thead>
<tr>
<th>setSpec</th>
<th>setName</th>
<th>DC record count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADTP</td>
<td>Australian Digital Theses Program</td>
<td>30853</td>
</tr>
<tr>
<td>AUCKLAND</td>
<td>University of Auckland</td>
<td>375</td>
</tr>
<tr>
<td>BGYMYU</td>
<td>Brigham Young University Theses</td>
<td>910</td>
</tr>
<tr>
<td>BICBF</td>
<td>Bibliothèque interuniversitaire de la Communauté française de Belgique</td>
<td>880</td>
</tr>
<tr>
<td>CALTECH</td>
<td>California Institute of Technology</td>
<td>3776</td>
</tr>
<tr>
<td>CCSD</td>
<td>CCSD theses-EN-ligne, France</td>
<td>7936</td>
</tr>
<tr>
<td>CRANFIELD</td>
<td>Cranfield University</td>
<td>109</td>
</tr>
</tbody>
</table>
HOW
Set up Infrastructure

- Task for: ICT / Libraries.
- Acquire software (EPrints, DSpace, etc.).
- Test installation.
- Acquire hardware.
- Install and configure IR.
- Integrate into campus websites.
- Perform ongoing maintenance.
Convince Academics

- Leadership from senior management.
- Arm-twisting + Ego.
- Rewards.
- Payments for submission.
- Generate reports from IR.
- Link IR prominently.
- Work with younger, more dynamic, staff.
Discussion...

to find me, search on Google or Facebook for “hussein suleman”