Capturing knowledge in institutional repositories ... playing leapfrog with giraffes

Heila Pienaar (UP)

&

Martie van Deventer (CSIR)

KM Workshop: Durban 17 August 2007
Notes for clarification

• The giraffe must be one of Africa’s best known creatures.
• We are 100% sure that giraffes do not play leapfrog!
• They do, however, symbolise those giants upon whose shoulders we know we stand!
• Leapfrogging: …well you cannot keep on standing … action is required … get moving … go play!
Outline

• Definition & context
• Leapfrogging at the University of Pretoria
• UP Repository: capturing the University’s knowledge base
• CSIR: Learning from UP
• Experimenting … is it worth the effort
• Giraffe leap: our newly launched repository
• Knowledge CSIR ‘captures’
• Tips, tricks & traps
Definition

- In simplest terms, a digital repository is a database/catalogue where digital content, assets, are stored and can be searched and retrieved for later use.
- A repository supports mechanisms to import, export, identify, store and retrieve digital assets.
- Putting digital content into a repository enables staff and institutions to then manage and preserve it, and therefore derive maximum value from it.
- Digital repositories may include research outputs and journal articles, theses, e-learning objects and teaching materials or research data.

Definition cont.

- The SPARC organisation defines institutional repositories as follows:
  - Institutionally defined
  - Scholarly
  - Cumulative and perpetual
  - Open and interoperable

- Clifford Lynch: It is most essentially an **organisational commitment** to the stewardship of digital materials, including long-term preservation where appropriate, as well as organisation and access or distribution.
Impetus of the SARIS project: eResearch Support Service for SA

**Future eResearch activities**
- Research Portal (incl global searching)
- Data Transfer andSharing (processes and protocols, 3As, helpdesk)
- **Open Access** (Standards, common software, institutional repositories)
- The eResearch Librarian (Training and re-orientation)
- Digital Curation Services (Standards, software, marketing & training services)
- Portable & access constraints

**Immediately**
- SASLI+
- National Research and Education Network
- Centre for High Performance Computing

**Function**

**Activities**

**Innovative services move to Service Delivery**

**Governance & Management**

**eResearch Board**

**eResearch Development & Innovation**

**eResearch Service Delivery**

**Lead Users Forum**

*Usually sub-contracted to competent agents in the system*
Leapfrogging at the University of Pretoria because of existing **knowledge** on the Web & elsewhere

- 12 July 2004: Director gives assignment to Dr Heila Pienaar to establish a digital repository for the University
- 23 July 2004: University wide project team assembled
- Workshops on: 2 August, 12 August, 30 August, 20 September, 4 November
- **Demonstrations of open source repository software:** Greenstone, (Innovative), Fedora, e-Prints, DSpace, I-Tor etc
- **Study OSI’s “Guide to institutional repository software”**.
- Choose DSpace because of UP IT fit and decentralised architecture and workflow
- Create management team for implementation of DSpace on 30 September (managers and specialists)
- DSpace web page and project plan developed: use champions
- **Metadata standards document created**
- **Digitisation standards document created**
- **Digital objects identified**
- 2005: Software development to fit DSpace into University's portal (authentication)
- Jan 2006: UPSpace launch
WWW and other sources that made this quick implementation process possible

- Open source software available on the Internet
- Examples of the different software installations at different institutions on the Web
- Articles, presentations and discussions on the Web (open access)
- International courses e.g. DSpace course in India & digitisation course in the UK
DSpace is an open source dynamic digital repository.

**Submission**
- Submitter
- Web-based interface makes it easy for a submitter to create an archival item by depositing a set of related files. DSpace was designed to handle any format from simple text documents to datasets and digital video.

**Management**
- Data files, also called documents, are organized into bundles. Each bundle is linked to a collection format and its encoded information. A description of each item is preserved in metadata.
- An item is an "archival record" consisting of a group of related content and associated metadata. An item's metadata is indexed for browsing and searching. Items are organized into collections of logically-related material.

**Preservation**
- Preservation is achieved through a combination of bitstream format and current format, ensuring that the digital file is kept usable as technology changes.

**Community**
- A community is the highest level of the DSpace content hierarchy. They correspond to organizational bodies in the institution such as departments, labs, research centers, or schools.

**End-user**
- DSpace is committed to supporting browsing and searching the archives. Once an item is located, Web-ready versions can be displayed directly in a browser while others can be downloaded and opened using an application that can interpret the file.

UP digital repository

Specialist / expert roles:
- **Metadata**: Amelia Breytenbach
- **Digitisation**: Ria Groenewald
- **Platform manager**: Ina Smith
- **IT Specialist**: Dept IT / Psybergate
- **Consultant**: Theo Bothma

UP Digital Repository
Co-ordinator: Heila Pienaar

Champions

- **Special Collections**
  - Manager: Maureen du Pisanie

- **Scholarly Communication**
  - Manager: Monica Hames

- **Vet Science Faculty**
  - Manager: Erica vd Westhuizen

- **Dept of Architecture**
  - Manager: Hettie Groenewald

- **Dept of Chemistry**
  - Manager: Sunette Steynberg

- **Education Faculty**
  - Manager: Johann van Wyk
UPSpace today! Capturing the University’s knowledge base
Examples of UP knowledge captured

For more info: Tuesday 16:00-18:00 - “Adapt or die”: Energizing library staff and academics through the development of digital repositories
CSIR context

- The CSIR turned 60 in October 2005. This event marks major changes in our organisation.
- The CSIR was given the responsibility to drive South Africa’s Open Source initiative.
- The CSIR previously sold its publications but towards the end of 2005 that activity was stopped and the organisation started investigating Open Access.
- In October 2006 the Information Services took on the responsibility to establish an institutional repository for the organisation.
- UP agreed to help us learn and we thought we would have the repository set up in January 2007.
- Fortunately we have been managing the CSIR’s internal Technology Outputs database (which was established in the early 80’s) since 2004 and did not have to start from scratch to find content.
Fast tracking ... learning from UP

• Easy we thought: They (UP) had done the research, had gained the experience and were able and willing to train us ...

• For example we needed to know:
  – Which is the best repository software to use? UP shared their research results.
  – What items and under what circumstances may you place items in the repository? UP directed us to SHERPA.
  – Even ... why Dublin Core? Notes on standards.
  – What formats? Oops ... they showed us the help file!
  – What documentation needed to be prepared? Gave us their templates.

• The next set of questions related to the technology
  – Is vanilla version sufficient? Facilitated meeting with IT.
  – What is the server set-up? Our techie was introduced to their techie.
  – How does one manage security? More techie meetings!
  – How does the metadata-data ‘thing’ work? Demonstration!

• However, even with the help of our friends and with the best intensions in mind, the January launch did not become a reality.
Experimenting … would it be worth your while?

• We decided to slow down and to experiment … to establish if a repository would be suitable for our environment.
  – UP allowed us to build a collection inside their repository – so that we could gain experience in using the DSpace software. A set of 450 documents was identified and uploaded.
  – Simultaneously our IT department revived the CSIR’s eShop software … to test the repository concept in existing proprietary software with the same set of documents.

• We intended to compare effort.
• We did not communicate the fact that we had added open access content to the CSIR site … BUT … we had forgotten that the CSIR-site is indexed by Google.
• … then IT phoned to ask if we would like to see the activity associated with the ‘repository’.
By end April just < 6 000 documents had been downloaded.

Group Manager R&D Outcomes was informed and we took a serious look at the way in which we made the documents accessible. The additional default features of repository software was no longer an option.

By end June this figure has become >28 000 documents.
Favourites

February
22 Catastrophic failure of a raise boring machine during underground reaming operations
16 Impacts of air pollutants on vegetation in developing countries
15 Some failure analyses of South African Air Force aircraft engine and airframe components

March
50 Physio-climatic classification of South Africa’s woodland biome
43 Risk assessment as a management tool used to assess the effect of pesticide use in an irrigation system, situated in a semi-desert region
42 I think condoms are good but, aai, I hate those things: condom use among adolescents and young people in a Southern African township

April
93 Identification of atractyloside by LC-ESI-MS in alleged herbal poisoning
59 Physio-climatic classification of South Africa’s woodland biome
46 Costs and benefits of biological control of invasive alien plants: case studies from South Africa
46 Detective quantum efficiency of medical x-ray image intensifiers
46 Failure of rural schemes in South Africa to provide potable water

May
63 Costs and benefits of biological control of invasive alien plants: case studies from South Africa
65 Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series
67 Risk assessment as a management tool used to assess the effect of pesticide use in an irrigation system, situated in a semi-desert region

June
83 Catastrophic failure of a raise boring machine during underground reaming operations
71 Some failure analyses of South African Air Force aircraft engine and airframe components
54 Nitrate removal with reverse osmosis in a rural area in South Africa
54 Checking anion-cation charge balance of water quality analyses: Limitations of the traditional method for non-potable waters
Tracking individual items

- Realisation: the organisation could start tracking the impact of its research output much quicker and identify key expertise sooner.

- There was no turning back and as a team (together with IT and our Communications staff, and under the leadership of our Group Manager R&D Outcomes) we started redesigning the CSIR Internet site in such a way that the repository became an integral part of our knowledge dissemination strategy.
The Council for Scientific and Industrial Research (CSIR) in South Africa is one of the leading scientific and technology research, development and implementation organisations in Africa. It undertakes directed research and development for socio-economic growth.

Recent Studies

**Novel robot demonstrator for industrial use**

A novel robot, using parallel technology and designed according to the mechatronics engineering approach, is being developed by the CSIR. The technology demonstrator is intended for industrial use as either a welding or spray-painting robot.

**Natural fibre composites to be used in construction**

Researchers are investigating the application of natural fibre composites (NFCs) in construction, with the aim of producing construction materials that are non-harmful, recyclable and made from renewable materials.

**130 000-year-old fossil elephant found near Durban, South-Africa**

A modern African elephant tusk was discovered within a late Pliocene aeolianite at the base of an intertidal rock pool at Reunion Rocks, south of Durban. Dating indicates that the tusk is older than 112 kyr BP.
Also accessible from competence area

Joint effort … drawing upon the knowledge and expertise of our communications practitioners when new publications are posted in ResearchSpace
ResearchSpace: access to a subset of our explicit knowledge
Communities & collections

Same structure as throughout the Internet site
Knowledge we are ‘capturing’

- At this stage only published research articles – copyright cleared via SHERPA/Romeo.

- Selected research reports will be added next. Especially then, those reports that came about as a result of publicly funded research.

- The repository software makes provision for variety of formats. We’ll therefore also be adding data sets, source code, sound, video and graphic files when these become available.
The advantage/disadvantage of leapfrogging

- The standard time it takes to set up a successful repository is reported to be three years.
- UP, because of the knowledge they gained from the experience of others, (and with a lot of hard work) reduced that time to two years.
- The CSIR, in turn, was able to create its repository within 12 months.
- We have now added our own learning and are a point where we can start to reciprocate.
- On the flip side: do not expect to have the same in-depth understanding as the original investigator.
- Much patience/time is needed! Transfer of knowledge needs to be planned and manage or it could become all consuming!
Tips, trick & traps

- Commitment from organisation leadership is important … to ensure that there is focus as well as for the investment that will ensure sustainability.
- The library is now also a “publisher” of information and knowledge and not only a user and conduit.
- If you allow IT infrastructure and/or software to be stumbling blocks … they will be. The repository software is not the issue. The repository is.
- It is important to have the courage to experiment.
- It is lonely and frustrating to play by yourself. Make others aware that you would like to play and they’ll gladly help you leapfrog.
- Unfortunately: playing leapfrog is not the same as playing catch-up. The first step is to establish trust then be prepared to learn, to experiment, to work hard and to have fun.
Lessons & tips cont

- Do not plan the perfect repository. Get the basics under control and start experimenting!
- The Internet and especially open source and open access accelerate learning. Tips and techniques are available from a wide variety of sources!
- Personal knowledge transfer and personal networks remain key to ensuring successful learning partnerships.
- Be prepared to invest time and resources. CSIR had to invest in a new server and employ extra hands before we could get going and if that is what it takes to get your repository on-line make the commitment.
- Giraffes play … they do not travel barefoot over the Drakensberg!
- Africa’s knowledge needs to be set free … just imagine all those giraffes playing leapfrog!
- Be warned: Once the organisation KNOWS the benefits there is no back tracking … all systems go!
Questions?
Sources to consult … as a start

- Centre for research Libraries: [http://www.crl.edu](http://www.crl.edu)
- DSpace Project Implementing DSpace section of DSpace.org website: [http://dspace.org/implement/index.html](http://dspace.org/implement/index.html)
- Pathfinder Report on Web Based Repositories available: [http://www.palsgroup.org.uk](http://www.palsgroup.org.uk)
- SHERPA: [http://www.sherpa.ac.uk/documents](http://www.sherpa.ac.uk/documents)
- SPARC Europe: [http://www.sparceurope.org/index.html](http://www.sparceurope.org/index.html)