EIFL and IDS webinar on DSpace repositories

with Nason Bimbe (IDS), Andrew Mwesigwa (Makerere University), Hilton Gibson (Stellenbosch University) and Felix Rop (ADLC)
Challenges
Fine-tuning OA repositories

Interoperability with other institutional systems, e.g. integration with e-learning platforms;

Embedding new applications, e.g. mobile phone/tablet interfaces, research object level metrics, etc.;

OA repositories are more user-friendly for researchers and enable easier sharing of research output.
EIFL Plans for 2016-2017
“Next generation” OA repositories

Keeping up-to-date with the latest technology & policy developments; reinforcing local capacity in the management of OA repositories:

Develop check-lists on current requirements for OA repositories (infrastructure & policies).

Conduct an audit of current OA repository developments (infrastructure and content) & provide good practice recommendations.

Organize webinars & other knowledge sharing activities.
Mak IR

Welcome to Makerere University’s Institutional Repository (Mak IR). Mak IR was formerly called Uganda Scholarly Digital Library (USDL). Mak IR is a collection of scholarly output of by researchers of Makerere University, including scholarly articles and books, electronic theses and dissertations, conference proceedings, technical reports and digitised library collections. It is the official Institutional Repository (IR) of Makerere University.

Copyright Information:
For information about the publishers’ copyright policy on archiving your articles online or in an institutional repository, visit the Sherpa Site at http://www.sherpa.ac.uk/romeo.php. The site gives a summary of the permissions normally given as part of each publisher’s copyright transfer agreement. If you wish to publish your research findings in Mak IR, please contact Makerere University Library at makir@murlib.mak.ac.ug for details. Mak IR operates both open access and closed access models. Access to full text has been restricted in adherence to the Makerere University Intellectual Property Rights (IPR) and Copyrights policies.

Other Useful Resources
Africa Portal is an online repository of open access library collection with over 3,000 books, journals, and digital documents on African policy issues. This is an initiative by the Centre for International Governance Innovation (CIGI), Makerere University (MAK), and the South African Institute of International Affairs (SAIIA). Please visit the Africa Portal at http://www.africaportal.org/library.

More exciting news to appear here.

Communities in Mak IR

Select a community to browse its collections.

- College of Agricultural and Environmental Sciences (CAES)
- College of Business and Management Sciences (CoBAMS)
- College of Computing and Information Sciences (CoCIS)
- College of Education and External Studies (CEES)
- College of Engineering, Design, Art and Technology (CEDAT)
- College of Health Sciences (CHS)
- College of Humanities and Social Sciences (CHuSS)
- College of Natural Sciences (CoNAS)
- College of Veterinary Medicine, Animal Resources and BioSecurity (CoVAR)
Capacity Building For Sustainable Digital Information Services In An Academic Library

Hilton Gibson
The repository content and the repository infrastructure require support teams, to ensure that the repository is properly maintained, now and in the far future.
In the context of the library, the operational team provides information services, that are supported by the technical team using information systems, which are built and maintained in co-operation with the information technology department.


**Director**

*The director need not be a full time appointment, however someone at the director level should be made responsible.*

- **Excellent written and verbal communication skills.**
- **Doctorate in library or information science.**
- Academic journal publication and/or institutional policy development about "open scholarly communication and preservation".
- **Expert understanding of open access digital repository content management.**
- **Expert with metadata standards for academic publishing.**
- **Professional understanding of digitisation technologies.**
- Professional understanding of web 2.0 communication technologies.
- Familiarity with **repository software.**
- Familiarity with desktop computer operations.
- **Familiarity with open access digital repository systems management.**
- **Familiarity with the operations of the internet and how to use it securely.**
INFORMATION SERVICES

Manager

Repository

Subject Specialists

Journals

Research Office
Operational Manager

This is a full time appointment.

Core Skills

- Excellent written and verbal communication skills.
- Bachelors degree in library or information science.
- Major or Masters in "open scholarly communication and preservation".
- Expert understanding of open access digital repository content management.
- Expert with metadata standards for academic publishing.
- Expert with digital curation.
- Expert with digitisation technologies.
- Professional understanding of web 2.0 communication technologies.
- Professional understanding of repository software.

Specialist Skills

- Expert with the management of Open Research Repository Systems.
- Expert with the management of Open Journal Publishing Systems.
- Expert with the management of Open Conference Publishing Systems.
- Expert with the management of Open Data Management Systems.
- Expert with the management of Open Educational Resource Systems.

Familiarity with open access digital repository systems management.

Familiarity with desktop computer operations.

Familiarity with the operations of the internet and how to use it securely.
Repository and Journal Librarians - 2 each

These are full time appointments.

Core Skills

- Professional written and verbal communication skills.
- Technology certificate in library or information science.
- Major in "open scholarly communication and preservation".
- Expert with repository software.
- Expert with desktop computer operations.
- Expert with metadata standards for academic publishing.
- Expert with digital curation.
- Expert with digitisation technologies.
- Professional understanding of open access digital repository content management.
- Professional understanding of the operations of the internet and how to use it securely.
- Professional understanding of web 2.0 communication technologies.
- Familiarity with open access digital repository systems management.

Specialist Skills

- Expert with the operation of Open Research Repository Systems.
- Expert with the operation of Open Journal Publishing Systems.
- Expert with the operation of Open Conference Publishing Systems.
- Expert with the operation of Open Data Management Systems.
- Expert with the operation of Open Educational Resource Systems.
INFORMATION SYSTEMS

Manager

Software

Central IT Developers

Hardware

Central IT Infrastructure
**Technical Manager**

The technical manager should ideally be appointed internally from within the technology support division of the institution.

The technical manager is **NOT** currently technically skilled, since technology is evolving at a very rapid pace, but rather the manager is aware of the current technologies and how to effectively deploy them.

The technical manager should know how to lead an information technology team by making use of the specialised skills of the team and current technologies to achieve the information systems goals of the open scholarly communication office.

This person should have extensive experience in designing and implementing information technology infrastructure and software systems that fit the "business case" of the institution.

This position does not need to be full-time, rather the position can be shared with the central IT department, **however a person from top management must be made responsible for the long term maintenance and operation of the open scholarly communication office’s information systems needs.**

The person appointed should also be a **very good communicator** and be able to "bridge" the LIS and ICT worlds. In other words, the person can inform the library about ICT and vice versa, the person can inform ICT about the library, thereby creating a vital two-way communication "bridge".

In addition, though not strictly required, it would be very advantageous if the technical manager was personally supportive of **open systems** and **open access**. **Most importantly, this person can assist the library in the transition to a new digital era and can help the ICT department transition to the new digital information services era.** Also see: [http://wiki.lib.sun.ac.za/index.php/List_of_Repository_Software](http://wiki.lib.sun.ac.za/index.php/List_of_Repository_Software)
Web 2.0 Software Technologist or Web 2.0 Application Specialist or Java Webapp Developer These are current technology specialists whose expertise may be shared with the central IT department.

Core Skills

- Expert Java programming skills
- Expert XML programming skills
- Professional SQL programming skills
- Professional HTML programming skills
- Professional CSS programming skills
- Professional PHP programming skills
- Familiarity with the Tomcat java webapp server
- Familiarity with the LAMP stack webapp server

Specialist Skills

- Expert DSpace software programming skills.
- Expert EPrints software programming skills.
- Expert PKP software programming skills.
- Expert Open Data software programming skills.
- Expert Open Educational Resources software programming skills.
Web 2.0 Hardware Technologist or Web 2.0 Infrastructure Specialist or Ubuntu Linux System Administrator These are current technology specialists whose expertise may be shared with the central IT department.

Core Skills

- Expert with the Ubuntu Linux server operating system
- Expert with the Tomcat java webapp server
- Expert with the LAMP stack webapp server
- Expert with virtualization and cloud services
- Professional BASH programming skills
- Professional TCP/IP networking skills
- Familiarity with a major vendors server hardware
- Familiarity with a major vendors networking hardware

Specialist Skills

- Expert DSpace infrastructure support skills.
- Expert EPrints infrastructure support skills.
- Expert PKP infrastructure support skills.
- Expert Open Data infrastructure support skills.
- Expert Open Educational Resources infrastructure support skills.
What else can these teams do for the library and the institution?
These same teams, with their skills, are then also capable of enabling other open systems for the institution, for example:

- An open journal system
- An open conference system
- An open research data management system
- An open research collaboration system
- An open educational resources system
- An open bibliography system
- An open library management system
How do you build these teams?
References


http://staff.lib.sun.ac.za/~hgibson
IDS OpenDocs: A case study for the EIFL and IDS Webinar on DSpace

Nason Bimbe,
Information Systems Manager - BLDS,
Institute of Development Studies

n.bimbe@ids.ac.uk
@nbimbe
Skype. bimben

January 2016
Introduction

- COAR (Confederation of Open Access Repositories) discusses a number of profiles needed for sustainable practices for populating repositories and these are:
  - **Incentives**: promoting the benefits of repositories through advocacy and metrics, as well as the adoption of policies/mandates that require deposit
  - **Integration**: amalgamating repository services with other institutional services like research information systems and research biographies
  - **Mediation**: implementing tools, workflows, and agreements that ease and simplify the deposit process

- This webinar therefore I hope is a contribution in a practical way as a starting step to begin tackling some and if not all the areas highlighted in the profiles identified above.

The Institute of Development Studies (IDS) is a leading global institution for research, teaching and learning, and impact and communications, based at the University of Sussex.

British Library for Development Studies (BLDS), the IDS Library, which strives to improve the visibility, accessibility and ‘readability’ of research, in particular of that published in developing countries, both current and historical, which often remains unseen by an international audience, who could benefit from a wider range of perspectives. This is done through:

- Provision of core library services
- IDS Institutional Repository
- Digitisation and BLDS Digital Library
- Document Delivery service
About IDS OpenDocs

- OpenDocs is the IDS Digital Repository setup in 2010 and has two main communities – BLDS Digital Library and IDS Research Community.

- BLDS Digital Library - hosts collections of existing Southern research materials, available within BLDS and from partners that have agreed to be part of the DL, that were/are digitised, Creative Commons licensed, and are made freely available. It has over 4,800 items.

- IDS Research Community hosts publications by IDS Researchers and those from partner research centres and consortia. It has over 2,700 items.

- Together usage is over 250k downloads and over 150k views a month split 50/50 from both 'Southern' and 'Northern' based users.

- OpenDocs is now the default digital content store for all open access full text content produced by IDS and now including those from the IDS flagship journal 'IDS Bulletin' (on OJS) which is going Open Access this Year.

- The platform is based on DSpace, an open source digital repository software, from Duraspace.
OpenDocs: Architecture

- Runs on a virtual server managed by Microsoft Windows Virtualisation Server - Hyper-v
- Software stack is composed of:
  - Debian - a Linux flavour
  - PostgreSQL
  - OpenJDK
  - Tomcat
  - Apache Webserver - used as a (HTTP) Proxy using mod_jk proxy (mod_proxy is much easier to use though).
- We are on DSpace 4.2, soon upgrading to 5.4 in the next couple weeks
- Installation structure is based on separation of 'concerns' through the use of the Overlay facility, which allows separating the core DSpace code from your own - helps with customisation and upgrades.

TIPS

- DSpace is well tested in a Linux, PostgreSQL environment – unless otherwise, use this environment.
- Always run your repository with the latest software versions.
- We found putting Apache Webserver in front of Tomcat works very well and makes the platform stable and possibly scalable.
- Always use 'Overlays' to add your own code to change the behaviour of your DSpace.
- If possible use version control to manage your changes.
- Use profiles in the top-level POM for easy management between Development and Production systems.
Business Continuity

- OpenDocs runs on a virtual server therefore, a copy of the whole virtual machine is taken every night. This copy can be used for restore purposes if there are any problems by simply rebuilding the virtual machine from the copy.

- OpenDocs as one of the many ICT services in IDS, is included in the overall IDS’s ICT Business Continuity Plans managed by central ICT team – CATS (Computer And Technical Services)

TIPS
- It is important to have a Business Continuity Plan which would cover: -
  - Backups/Restore procedures
  - Policies for disaster recovery
Persistent Identifiers

Persistent identifiers (PIs) are simply maintainable identifiers that allow us to refer to a digital object. Persistent identifiers are also persistently actionable (that is, you can "click" them); however, unlike a simple hyperlink, persistent identifiers are supposed to continue to provide access to the resource, even when it moves to other servers or even to other organisations.

http://www.ariadne.ac.uk/issue56/tonkin#sthash.2eyvo7yu.dpuf

With OpenDocs, we are not subscribed to the Handle Service hence we use the service’s base URL together with the handle attributes for this.

We will soon be using Digital Object Identifiers (DOIs) to compliment our handle system setup.

---

**TIPS**

- It is important to make sure that Handle functionality is well configured.
- If you not subscribing to the Handle Service, make sure that handle setting in the `dspace.cfg` or `build.properties` is set as thus:
  ```
  handle.canonical.prefix = ${dspace.url}/handle/ // uncomment
  #handle.canonical.prefix = http://hdl.handle.net/ //comment
  ```
- Otherwise, your content URIs will not be resolving correctly – i.e. they resolve to a dead end. This is very important because content harvested through OAI-PMH uses the URI as the Resource Locator Address.
“Each individual repository is of limited value for research: the real power of Open Access lies in the possibility of connecting and tying together repositories, which is why we need interoperability. In order to create a seamless layer of content through connected repositories from around the world, Open Access relies on interoperability, the ability for systems to communicate with each other and pass information back and forth in a usable format. Interoperability allows us to exploit today’s computational power so that we can aggregate, data mine, create new tools and services, and generate new knowledge from repository content.”

https://www.coar-repositories.org/activities/repository-interoperability/coar-interoperability-project/a-case-for-interoperability/

- Out of the box, DSpace provides the following facilities for interoperability purposes: -
  - Open Access Initiative Protocol for Metadata Harvest (OAI-PMH)
  - RSS (Rich Site Summary) Feeds
  - (REST) Application Programming Interface (RESTITAPI)
  - Semantic Web, based on the Resource Definition Framework (RDF)
  - SWORD (Simple Web-service Offering Repository Deposit) which allows digital repositories to accept the deposit of content from multiple sources in different formats
  - Digital Object Identifiers (DOIs) and ORCID (Open Researcher and Contributor ID) can compliment these protocols.

- In IDS, we use OAI-PMH to enable OpenDocs interoperate with the corporate website (FarcryCMS) and also the Resource Discovery Tool (VuFind).
- We will also be adding RIOXX metadata application profile. RIOXX from JISC was developed with input from HEFCE (UK) to help repositories comply with the open access policy as mandated by RCUK.

**TIP**

Make sure you enable the appropriate webapps for interoperability more especially OAI-PMH. OpenDOAR relies on it.
Local Customisation

- Make the DSpace your own through local customisation
  - The 'Look and Feel' through modifications to the CSS
  - Extend metadata fields for your institution's specifics
  - Enhance the display fields for items
  - Add extra search facets
  - Add your own Item Types, Languages, Subjects, Licenses, etc.
  - Make your DSpace UI multi-lingua if needed
  - Add a responsive UI for example Mirage 2

- We have done all for OpenDocs except for the last two.
- We have a bi-directional link with the LMS (Koha) for items that are also listed in the catalogue.
- After the upgrade, we will enable Mirage 2.
- We will also be adding the ability to ‘stream’ the BitStream in addition to downloading, which will be helpful especially for our users with limited bandwidth. This functionality will be based on the module from @mire.

**TIPS**

- Always customise your Dspace installation to suit your institution branding and way of working.
- Using mobile access (small screen form factor) is prevalent in developing countries, therefore it is important that your DSpace has a Responsive User Interface like those based on Twitter’s Bootstrap Framework such as Mirage 2.
- Assess all **default configuration** settings to see if they meet you needs.
OpenDocs: Screen shots

**Home page**

**Item detailed page**
Discoverability

- It is important to make sure that your DSpace and its content is easily discoverable on the web.
- There are various ways and tools available to help achieve this.
  - Use sitemaps so that your DSpace can be indexed by Google and other search engines
  - As of DSpace 1.8, integration with Google Scholar comes ‘out of the box’
    - But you need to register with Google to have your content in Google Scholar
  - Use ‘Repository of Repositories’ such as OpenDOAR, the CORE at the Open University in the UK, etc.
    - Metadata (including full text for some) will be harvested and made searchable within these type of repositories
  - Use your other web presence channels and/or your collaborator’s web presence channels
    - Could be other websites, social platforms, etc.

- At IDS, we use all the above for OpenDocs.

**TIPS**

- Setup your sitemaps generation as a Cron job.
- Make sure that filter media (‘breaking words in documents’) is setup as a Cron job.
- Make sure you setup a robots.txt file and placed in the ROOT for your DSpace to aid search engine web crawlers.
- If you can, register your DSpace with Google Scholar.
- Always register your DSpace with all known ‘reputable’ aggregators – Repository of Repositories.
Knowing how much ‘stuff’ is in your DSpace and how also it is used, and from where is crucial to help in securing the sustainability of your platform.

Therefore having good and 'accurate' statistics of your repository is good for advocacy, buy-in, business case building.

- At best you should know: - Number of downloads and views; Geo-location of where these are coming from; and Content analysis such how many items are in the repository (by various dimensions i.e. by item type, subject, language, etc.)

DSpace provides a number statistics such as number of downloads, number of items, number of failed logins, etc. (all the data is available in the SOLR indexes – these can be queried directly as well.

Use Google Analytics (registration required) or Piwik (an open source web analytics) to collect information about site visits. Standard GA (JavaScript based) has limitation in capturing downloads.

As of DSpace 5, there is integration with Google Analytics via GA API, statistics are therefore accessible in DSpace.

Third party tools you could consider: -

- Content Usage Analysis (CUA) module - a DSpace add-on from @Mire (not open source)
- MyDash from Harvard University - a standalone Open Source web application ([https://github.com/oscharvard/mydash](https://github.com/oscharvard/mydash))
- Altmetrics such as those from Altmetrics.com or PlumX.

At IDS, We use mostly CUA coupled with Standard Google Analytics as a compliment.

**TIPS**

Assess what sort of insights you want from your Dspace and setup/acquire the appropriate visualisation and analysis tools.
Policies

- Policies are crucial in the success of your Repository in terms of content archiving, copyright, usage (what is acceptable use of the content), take downs, preservation and so on.
- The next slides will consider:
  - Archiving
  - Copyright
  - Types of OA Policies
- For these you can read the slides yourselves at your own time.

- At IDS, we have an archiving policy (not formal though), copyright, usage and take down which are easily accessible by anyone.

**TIPS**

- Make sure appropriate policies are setup.
- Setup a mechanism of updating them as you learn more and also to take in the changes happening in the environment (be at local, national, global levels).
- Make sure the policies you setup are not creating 'pain points' in your processes and objectives but instead are enablers.
- Plan on how they will be advocated for, disseminated, implemented and monitored.
Policies: To consider - Archiving

- **Mandatory self archiving/deposits**
  - requests or encourages researchers to self-archive their publications in an open access institutional repository.

- **Voluntary self archiving/deposits**
  - requires that researchers self-archive their publications in an open access institutional repository.

Note: Data from Australia obtained by Arthur Sale shows that only mandatory policies result in high levels of self-archiving.
Traditionally, in most cases authors are required to relinquish copyright to journal publishers in return for having an article published.

Not incompatible with an institutional self-archiving policy, as the majority of publishers, even if they demand copyright to be transferred, will allow a copy to be deposited in an Open Access repository.

Where publishers do not allow this, the article can be deposited in the repository under restricted access.

A growing number of authors are retaining copyright, and are often encouraged to do so by their institution where the author usually assigns the publisher a “licence to publish”.

Normally done through the addition of an "author addendum" to a publisher’s copyright transfer agreement (CTA), which the publisher requires the author to sign (and which normally transfers copyright to the publisher).

Example: Harvard, Stanford and MIT all require authors to retain their own copyright.

Consider embargo period.

Even where publishers permit self-archiving of the author’s final version of an article, may still impose an embargo period during which Open Access is not permitted.

Varies considerably from publisher to publisher and discipline to discipline.

may be as short as 3 months for some science journals or very long.

up to 24 months in the case of certain journals in the fields of humanities and social sciences.
Policies: Basic Types for OA

- Type 1: Immediate deposit with immediate Open Access
- Type 2: Later deposit after the embargo period
- Type 3: Immediate deposit with optional access
Type 1: Immediate deposit with immediate Open Access

- Requires authors to deposit their articles upon acceptance for publication, once the final corrections have been made, and to make their articles openly available immediately through the repository.

- This policy results in immediate Open Access.

- However, because some journals do not permit immediate Open Access, has the disadvantage that it restricts the choice of journals in which an author can publish.
Type 3: Immediate deposit with optional access

- Requires immediate deposit, but if the article is submitted to a journal with an embargo, then the policy permits access to be open only at the end of the embargo period.
- During that period, however, article’s metadata (title, authors, affiliation, abstract, references) should be openly available at all times (publisher embargoes cannot be applied to metadata, which are not subject to copyright).
- Complies with publisher embargoes but at the same time ensures that all the required research outputs are compiled in the repository at the earliest opportunity that is.
- If a publisher embargo must be accommodated, the deposit process allows the author to enter the date on which the embargo will be lifted and the full text of the paper can be made openly available.
- During the embargo period, the metadata (title, author names, affiliation, and abstract) of the article are Open Access (only the full text is closed) so that indexing services like Google can find and index the article.
  - Therefore, it is known to exist, even if the full text is not available Open Access until the end of the embargo period.
  - Potential readers can email the author for a copy of the article or, if they have access to the journal through a subscription, they can find the published version. Either way, the article is being used and its impact is increasing.
- DSpace repository software have a "request a copy" button for articles contained in the repository, so that when the full-text is under an embargo, would-be readers simply click on this button and an email is sent to the corresponding author with a request for a copy of the article by email.
  - This is a "fair use" or "fair dealing" fulfilment of a personal request from an interested would-be user.
Thank you

Nason Bimbe,
Information Systems Manager - BLDS,
Institute of Development Studies

n.bimbe@ids.ac.uk
@nbimbe
Skype. bimben

January 2016
Objectives

• Brief History
  (ADLSN - ADLC)
• ADLC Mandate
• ADLC Today
• Aspirations
ADLSN: A Living Regional Community
ADLSN Brief History

Initial network: South African Greenstone Support Network
To promote and provide training to use this popular open source digital library software to set up digital collections and repositories

US funding + eIFL FOSS support led to a series of projects:
To pilot and implement the network extension in collaboration with champion institutions & professionals

“Train-the-trainer” approach
Spreads to other Southern, then Eastern and Western African countries
First idea: Greenstone Support Organisation for Africa
A community of African practitioners and other interested actors with a common goal of supporting the preservation and dissemination of local content in digital form.
ADLC is one product of ADLSN
ADLC MANDATE

1. Facilitate Access
2. Organize training and provide resources
3. Provide community support
Collection Visibility Project

African Digital Library Central Hosting platform

- Increase global access for poorly visible collections, due to:
  - Insufficient and/or unreliable technical infrastructure
  - Lack of expertise, awareness, strategy
Today....... 

• A web platform ('in the cloud') set up to host various digital libraries and repositories technologies: Greenstone, Dspace, Omeka...
  • Remote management for collection's administrator of the institution
  • Seamless access for institution's end users
  • Measurement of impact with analytical tool to track

... and we plan...

• To encourage and assist the development of institutional and national sustainable hosting solutions
A digital collection of full documents, created by an institution, is hosted by the ADLCentral platform in the cloud, with one of the DL management tools available.
Demonstration
http://ds.adlc.info/
Final words
Recommendations

Reinforce infrastructural and hosting capabilities to unlock the African digital knowledge

Set up a Working Group to create an African-wide policy for data centers to take up the hosting challenge of local content

ADLSN's Collection Visibility project can support with contribution of experiences and lessons learned
Conclusion

A growing and lively African community is now more than ever getting together and generating more digital content.

Local and distributed initiatives – like ADLSN – are to continue to enhance the efforts of this movement.

And a pan-African coalition for pan-African challenges like hosting infrastructure could act to assure the world wide success of this community.
Thank you

ADLSN management  info@adlsn.org

Twitter  @infoADLSN
Facebook  http://www.facebook.com/infoADLSN
Website  http://adlsn.org
LinkedIn Group  http://linkd.in/11KXWhl