

After DSpace Installation Tasks

hgibson@sun.ac.za / hilton.gibson@gmail.com
n.bimbe@ids.ac.uk / nasonbimbe@gmail.com

Tasks to complete immediately (1)

DAILY ADMIN

REBUILD DSPACE

RESTART DSPACE

REBUILD INDEXES

Tasks to complete immediately (2)

DAILY ADMIN CRONJOBS

Just after installation it is critically important that you enable daily automated tasks for your digital archive. In order to send out subscription emails, update search, browse, full-text indexes and do general daily housekeeping on the system, a regular maintenance script must be run automatically daily. On a Unix/Linux based system this is easy to accomplish with use of the "crontab" facility.

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Daily_Admin

<http://en.wikipedia.org/wiki/Cron>

Tasks to complete immediately (3)

REBUILD DSPACE SCRIPT

To apply customisations, DSpace usually needs to be rebuilt, to update the Java WAR's for re-deployment by the Tomcat webapp server. This wiki page helps you to create a customised rebuild script that you can use later any time you need it.

Also see: <http://wiki.lib.sun.ac.za/index.php/SUNScholar/Customisation>

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Rebuild_DSpace

<https://wiki.duraspace.org/display/DSPACE/Rebuild+DSpace>

Tasks to complete immediately (4)

RESTART DSPACE SCRIPT

This is the DSpace restart script setup. The restart can be used for the following:

Restart DSpace after a "dspace.cfg" change.

Restart after an updated or new customisation.

Restart manually after a system failure.

Restart automatically at a predetermined time using a root user "crontab".

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Restart_DSpace

Tasks to complete immediately (5)

REBUILD INDEXES SCRIPT

After modifying any index configuration and then rebuilding your webapps, you will need to rebuild your indexes. This script allows to rebuild indexes at any time later.

<http://wiki.lib.sun.ac.za/index.php/SUNScholar/Indexes/Rebuild>

Tasks to complete ASAP (1)

OPTIMISATIONS

HANDLE SERVER

INTERNET SECURITY

DISASTER RECOVERY

Tasks to complete ASAP (2)

OPTIMISATIONS

Several major optimisations of the system performed at Stellenbosch University in an attempt to create a truly production optimised version of DSpace.

[Tomcat Webapp Server](#)

[ANSI SQL Database](#)

<https://wiki.duraspace.org/display/DSDOC5x/Performance+Tuning+DSpace>

Tasks to complete ASAP (3)

HANDLE SERVER

The handle service allows us to apply a short URL, which is persistent, for the purposes of citation and discovery on the web.

<http://hdl.handle.net>

<http://www.handle.net/documentation.html>

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Handle_Server

Tasks to complete ASAP (4)

DISASTER RECOVERY

Since your archive is now the vehicle for maintaining the permanent digital academic research record of your institution, you will be concerned about its sustainability

!!! You will want to make sure it is backed up and monitored correctly !!!

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Disaster_Recovery

Tasks to complete at a later stage (1)

RESEARCHER AUTHORISATION

RESEARCHER IDENTIFICATION

CUSTOMISATION

SYSTEM ADMINISTRATION

Tasks to complete at a later stage (2)

RESEARCHER AUTHORISATION

Digital assets must only be managed by users authorised to do so. DSpace can authenticate using ePerson accounts or using LDAP server accounts. This wiki page describes the method of setting up DSpace to use an institutional LDAP server for user provisioning. After that is done, it is then up to the repository manager to define privileges for individuals on the repository.

See: <http://en.wikipedia.org/wiki/Ldap> for more info.

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher_Authorisation

Tasks to complete at a later stage (3)

RESEARCHER IDENTIFICATION

A unique researcher identification system (ORCID in our case) eliminates any author name ambiguity there may exist about who authored a paper. These machine-readable id's can automatically programmatically gather data from various sources and cluster research information around an individual researcher.

Also see: http://wiki.lib.sun.ac.za/index.php/SUNScholar/Authority_Control

http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher_Identification

Tasks to complete at a later stage (4)

CUSTOMISATION

This customisation relates to the technical details involved in the customisation of features available with DSpace, not the customisation required when performing operational management of the research archive (for example collection hierarchy).

Please read about advanced customisation before starting any customisation work, as it is the basis for most of the customisation tasks.

<http://wiki.lib.sun.ac.za/index.php/SUNScholar/Customisation>

Tasks to complete at a later stage (5)

SYSTEM ADMINISTRATION

After building the system, regular server maintenance must be carried out and involves the following activities:

1. Software Upgrades
2. Check server backups
3. Check server health

<http://wiki.lib.sun.ac.za/index.php/SUNScholar/Server>

Customisation: Making the DSpace your own

Areas to be covered: -

1. A look at the knowledge and tools needed to perform the tasks
2. Mirage theme customisation
3. Enabling OAI-PMH server
4. Implementing multilingual user interface
5. Improving discovery through search engines
6. Basic web traffic insights using Google Analytics (GA)

Knowledge Required (1)

XML, XSL and CSS

XML stands for EXtensible Markup Language and was designed to store and transport data. XML was designed to be both human- and machine-readable. Quick guide at <http://www.w3schools.com/xml/>

XSL stands for EXtensible Stylesheet Language, and is a style sheet language for XML documents. XSLT stands for XSL Transformations. Quick guide at <http://www.w3schools.com/xsl/>

CSS is a stylesheet language that describes the presentation of an HTML (or XML) document. CSS describes how elements must be rendered on screen, on paper, or in other media. Quick guide at <http://www.w3schools.com/css/>

Knowledge Required (2)

Web developer tools (found in web browsers) - used to test and debug web UIs.

Can be invoked on most web browsers by hitting F12

In Chrome through menu and then More tools->Developer tools or Ctrl+Shift+I or simply by right mouse click and then on the context menu select Inspect element.

Google webmaster services - has tools that let webmasters:-

Submit and check a sitemap.

Allow for GA to monitor your traffic for statistics purposes – you need to obtain GA UA id that you need to apply in the dspace.cfg for GA to be able to collect statistics.

Check and set the crawl rate, and view statistics about when Googlebot accesses a particular site. And many more.....Details at <https://www.google.com/intl/en/webmasters/>

Knowledge Required (3)

Concepts and use of 'overlays'

Overlays are a means of separating the core DSpace code from those areas (code) that are prone to customisations.

Overlays are best described in a slideshare presentation located at <http://www.slideshare.net/tDonohue/making-dspace-xmlui-your-own>.

Toolset

SSH Access using PUTTY - PuTTY is a free implementation of SSH and Telnet for Windows and Unix platforms, along with an xterm terminal emulator. It is written and maintained primarily by Simon Tatham. Details at <http://www.chiark.greenend.org.uk/~sgtatham/putty/>

WinSCP - Free SFTP, SCP and FTP client for Windows. Details at <https://winscp.net/eng/index.php>

PgAdmin - pgAdmin is the leading graphical Open Source management, development and administration tool for PostgreSQL, running on Windows, Linux, Solaris, etc. Details at <http://www.pgadmin.org/>
(Please NOTE, you will not need to look at SQL ever UNLESS you know what you are doing!)

Toolset (2)

Webmin - A web-based interface for system administration for Unix. Details at <http://www.webmin.com/>

Nano - A text editor for Unix-like computing systems or operating environments using a command line interface. Details at <http://www.nano-editor.org/>

Web browser Developer Tools - already explained and helps when 'skinning' the UI using CSS

Pixie - A utility made especially for webmasters and designers and it is a color picker with few extra goodies. Details at <http://www.nattyware.com/pixie.php>

Mirage Theme Customisation (1)

Create the Overlay folders to hold the customisation and copy/create the required files. When you build DSpace, all custom changes made in the overlay folders are merged with the core code.

[dspace-source]/dspace/modules/xmlui/src/main/webapp/i18n

Copy the messages.xml file that will be modified from [dspace-source]/dspace-xmlui/src/main/webapp/i18n into the new folder

Modify the file as appropriate. File contains all the text for the labels displayed in the DSpace user interface.

[dspace-source]/dspace/modules/xmlui/src/main/webapp/static

Mirage Theme Customisation (2)

Copy the robots.txt file that will be modified from [dspace-source]/dspace-xmlui/src/main/webapp/static into the new folder

Modify the file as appropriate. File contains information that tells web search engine crawlers what to index and what to ignore.

Mirage Theme Customisation (3)

[dspace-source]/dspace/modules/xmlui/src/main/webapp/themes/Mirage/images

Add the images needed into the new folder. This is where the logo is located.

[dspace-source]/dspace/modules/xmlui/src/main/webapp/themes/Mirage/lib/xsl/aspect/artifactbrowser

Copy the xsl files that will be modified from [dspace-source]/dspace-xmlui/src/main/webapp/lib/xsl/aspect/artifactbrowser into the new folder

The file that may really be needed to be modified is [item-view.xsl](#)

Mirage Theme Customisation (4)

This file contains XSLT statements that controls the item detail view. If you want extra metadata fields added for display on this page.

[dspace-source]/dspace/modules/xmlui/src/main/webapp/themes/Mirage/lib/xsl/core

Copy the xsl files that will be modified from [dspace-source]/dspace-xmlui/src/main/webapp/lib/xsl/core into the new folder

The file really needed to be modified is [page-structure.xsl](#)

This file contains XSLT statements that controls the general layout and structure of a DSpace HTML page. This is where the links in the footer are made. You can also use this file to add links to 'social sharing' services for instance.

Mirage Theme Customisation (5)

[dspace-source]/dspace/modules/xmlui/src/main/webapp/themes/Mirage/lib/css

Copy the css file that will be modified from [dspace-source]/dspace-xmlui/src/main/webapp/lib/css into the new folder

The file really needed to be modified is [style.css](#)

This file contains CSS statements that controls the 'look and feel' of a DSpace HTML page such as colours and branding. This is where changes to the logo file can be made. Use the Developer Tool play around with CSS settings before updating the CSS file. You can use Pixie to identifier the colours you need before adding them to the CSS settings.

Mirage Theme Customisation (6)

Modify the needed configuration files

Modify all the files in [dspace-source]/dspace/config/emails with your required settings. Contains email templates used by DSpace.

Modify default.license in [dspace-source]/dspace/config with your required settings. Contains text for submission Terms and Conditions (T&C)

Modify dspace.cfg in [dspace-source]/dspace/config with your required settings. Any other settings deemed necessary for your installation other than those already mentioned in the initial DSpace setup.

Modify input-forms.xml in [dspace-source]/dspace/config with your required settings. Contains statements that controls the input form fields, so use if you need to add/remove fields.

Modify news-xmlui.xml in [dspace-source]/dspace/config with your required settings. Contains opening text for your repository on the home page.

Enabling OAI-PMH server

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP.

DSpace can act both as a Data Provider and also as a Data Consumer.

Configurations for OAI-PMH server are held in OAI-PMH crosswalks and [dspace-source]/config/modules/oai.cfg. NOTE - unless you need to change the default behaviour, you may not need to touch them.

To enable DSpace's OAI-PMH server, just make sure the [dspace]/webapps/oai/ web application is available from your Servlet Container (usually Tomcat). You can test that it is working by sending a request to: [http://\[full-URL-to-OAI-PMH\]/request?verb=Identify](http://[full-URL-to-OAI-PMH]/request?verb=Identify)

If you are harvesting content (bitstreams and metadata) from an external DSpace installation via OAI-PMH & OAI-ORE, you first should verify that the external DSpace installation allows for OAI-ORE harvesting.

More Details are at <https://wiki.duraspace.org/display/DSDOC5x/OAI>

Implementing multilingual user interface

In order to deploy a multilingual version of DSpace UI you have to configure two parameters in `[dspace-source]/config/dspace.cfg`:

`default.locale`, e.g. `default.locale = en`

`webui.supported.locales`, e.g. `webui.supported.locales = en, de`

The Locales might have the form `country`, `country_language`, `country_language_variant`.

According to the languages you wish to support, you have to make sure that all the `i18n` related files are available. See the [Multilingual Support for XMLUI](#) in the configuration documentation.

You will find that most of popular languages have been already translated, see `[dspace-source]/dspace/modules/xmlui/src/main/webapp/i18n`. You therefore just need to copy the ones you are interested in into the appropriate overlay folder as already described and modify.

If there is no equivalent messages XML file for your language, you can create one by cloning the `messages.xml` into something like `messages_{your-language-identifier}.xml` and add your translations.

Improving discovery through search engines

Search engines use software programs that transverse the web gathering information for indexing.

These software programs are called crawlers or bots.

They are good at finding HTML files but since DSpace content is in a database system, you will need to tell the crawler how to get to that content.

First you need to make sure that you generate sitemaps. These can be done through the Cron job

```
# Generate sitemaps at 6:00 am local time each day
```

```
0 6 * * * [dspace]/bin/dspace generate-sitemaps
```

Improving discovery through search engines (2)

This will generate the sitemaps that are accessible at <http://{your-DSpace-URL}/sitemap> and <http://{your-DSpace-URL}/htmlmap>

Make sure also that the robots.txt contains directives to these paths. You will need the robots.txt placed in the root of your DSpace site. Make sure the robots.txt contains directives for what can and cannot be indexed

Details of robots.txt structure and instructions are at

<https://wiki.duraspace.org/display/DSDOC5x/Search+Engine+Optimization>

Basic web traffic insights using Google Analytics

DSpace has always provided a way to have the web traffic collected and reported on by GA.

GA uses a Javascript code that responds to 'some event' to capture the details of a 'web visit' such as IP Address, user-agent, date/time and so on of the visit.

Tracking downloads is not effective using this method as it operates on page load event while the download is triggered by another event such as onclick event.

Please note also now that DSpace can track downloads using the GA API as from DSpace 5 - but you will need to configure for this to happen, see

<https://wiki.duraspace.org/display/DSDOC5x/DSpace+Google+Analytics+Statistics>

Basic web traffic insights using Google Analytics (2)

To enable visits tracking, you will need to acquire a GA Key from Google - see Google Webmaster Tools already described.

Add the key to the DSpace configuration in `dspace.cfg` under the setting `jspui.google.analytics.key=UA-XXXXXX-X` (for JSPUI) or `xmlui.google.analytics.key=UA-XXXXXX-X` (for XMLUI) by replacing the `UA-XXXXXX-X` with the key you will be given by GA

The statistics can be viewed on the Google Analytics web application at <http://www.google.com/analytics/>.