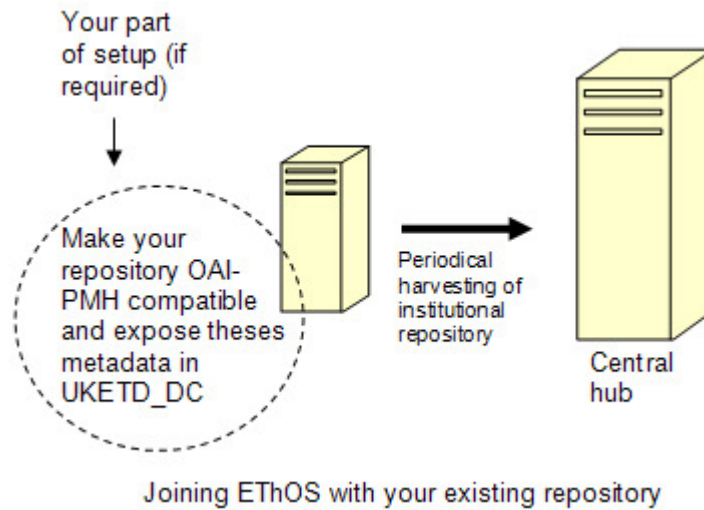


<u>Guide to using an existing repository</u>	2
Your repository should be able to handle OAI-PMH requests	2
Have a single point of collection for all your theses metadata	3
Technical Factors	4
<u>Guide to setting up a new repository</u>	5
Project Preferences.....	5
DSpace	6
Eprints	7
Fedora	7
Technical Factors	7
<u>Guide to using the central hub repository</u>	8
Physical loading via CD or DVD.....	8
Using central hub group submission interface developed by EThOS	9
<u>Training and Guidance</u>	9
<u>Training methods</u>	10
Personal, individual training	10
Informal workshops	10
Formal workshops.....	10
On-line help screens.....	10
Presentations and publications.....	11
<u>Training content</u>	11
Training for students	11
Training for supervisors	11
Training for Library, I.T. and administrative staff.....	11

Guide to using an existing repository



In order to enable the central hub to harvest the theses content, your repository should meet following three requirements:

Your repository should be able to handle OAI-PMH requests

The harvesting protocol of the central hub harvester is based on version 2 of Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and therefore this is an essential criterion.

Most of the well known repository software, such as Dspace, Eprints and Fedora, can readily handle the OAI-PMH request (see [OAI website](#) for more).

Your repository should expose, at least, the following metadata fields for harvesting

- Author
- Title
- Awarding institution
- Qualification level (the EThOS system is for doctoral theses)
- Date of award

However, EThOS recommends following list of metadata to be exposed for harvesting.

- [List of metadata to be exposed in UKETD_DC format for harvesting](#)

In order to make things easier, EThOS have developed plug-ins for two of the well known repository software: DSpace and GNU EPrints. Installing the plug-in will allow these repositories to automatically expose the theses metadata in UKETD_DC format.

Plug-in for DSpace: You can download the DSpace OAI Crosswalk plug-in together with full installation instruction by following the link below. You should also install the DSpace custom input submission form for theses, which allows you to enter the metadata into DSpace repository in the format required by the EThOS central hub.

- [DSpace OAI Crosswalk](#) (for DSpace versions 1.2.2 - 1.4.2)
- [DSpace custom input submission form for theses](#) (for DSpace version 1.2.2. or above)

Note: as of DSpace version 1.5.2, the UKETD_DC crosswalk is already included in the distribution files and can be activated by uncommenting the line:

[Crosswalks.uketd_dc=org.dspace.app.oai.UKETDDCCrosswalk](#) in `oai.cat.properties`

Plug-in for EPrints: You can download the EPrints OAI Plugin together with full installation instruction by following the link below.

- [EPrints OAI Plugin](#) (for Eprints v2)

NB: Eprints 3 users - the OAI interface to expose UKETD_DC is already bundled with the Eprints 3 distribution. However, it is still necessary to configure your installation to allow entry of metadata conforming to EThOS requirements. To this end you may find the [EPrints v3 Configuration Notes](#) helpful.

Have a single point of collection for all your theses metadata

A theses collection point is a URL, where EThOS can send the OAI request and harvest the metadata of your theses. In order to prevent any confusion and extra administrative cost involved in maintaining multiple collection points for a single institution, EThOS requires each institution to provide a single collection point for all their theses content.

This is not an issue for institutions with a repository which contains only theses. In this case the institution only needs to provide the base URL for their OAI interface.

This is also not an issue for the institutions which, although having a mixed repository with a multiple theses collection, have also mapped all of their theses metadata to an alternative single collection. This, especially, applies to the DSpace repositories where a mapping option has been made available. In this case providing the base URL with the set identifier for the alternative theses collection is sufficient.

The problem arises when:

- An institution maintains only multiple theses collections
- An institution has multiple repositories containing theses

For an institution that maintains only multiple theses collections and their repository is based on GNU EPrints 2, installing the above mentioned [EPrints OAI Plugin](#) will resolve this problem (and the plug-in is pre-installed for Eprints 3 users). In this case providing the base URL for your OAI interface is sufficient.

For an institution that maintains only multiple theses collections and has a repository based on DSpace, EThOS provides a 'UKETD_DC OAI modification'. Installing this modification will force DSpace repositories to only expose the theses metadata that are in UKETD_DC. In order for UKETD_DC OAI modification to work, it is necessary to first install the above mentioned DSpace OAI Crosswalk. Once the 'UKETD_DC OAI modification' is installed, the institution only needs to provide the base URL for their OAI interface. Follow the appropriate link below to download the UKETD_DC OAI modification along with full installation instructions.

- [UKETD_DC OAI modification for DSpace 1.3.2](#)
- [UKETD_DC OAI modification for DSpace 1.4.x](#)
- [UKETD_DC OAI modification for DSpace 1.5.2](#)
- [UKETD_DC OAI modification for DSpace 1.6](#)

For an institution that runs multiple repositories containing theses, the above mentioned plug-in will not provide a solution. In practice, it is assumed that it is highly unlikely that many institutions will maintain more than one repository which holds theses. There are several alternatives that can be adopted to overcome such a situation.

1. One alternative is to gather all the theses and metadata into a single collection point through physical loading by CD or DVD. More information on how to do this can be found in the [Guide to using the central hub repository](#).

2. Another alternative is to install a harvester locally, and harvest metadata from their multiple repositories into one place and then re-expose the theses metadata from there. The most complete and configurable harvesting solutions currently available are the following packages:

- ARC - Arc is released under the NCSA Open Source License. Arc is a federated search service based on OAI-PMH. It includes a harvester, a search engine together with a simple search interface, and an OAI-PMH layer over harvested metadata. Arc can be configured for a specific community, and enhancements and customisations by the community are encouraged. Arc is based on Java Servlet technology and requires JDK1.4, Tomcat 4.0x, and a RDBMS server (tested with Oracle and MySQL). (<http://sourceforge.net/projects/oaiarc/>)
- Celestial 1.1.1 - Celestial (formerly OAI) is an OAI 2.0-compliant OAI aggregation/caching/mirroring tool. Using MySQL and Perl, Celestial allows the administrator to mirror 1.0, 1.1, and 2.0 repositories in their own 2.0-compliant repository. (<http://celestial.eprints.org/>)

Technical Factors

Character sets: For compatibility with the Central Hub, you should ensure that you expose your metadata using the UTF-8 charset.

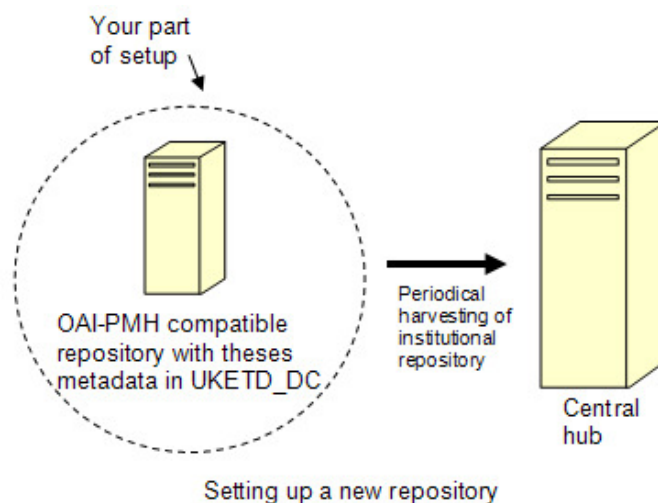
Apart from above requirements, the following general technical factors should also be considered while maintaining an institutional repository

Traffic: what is the expected or ongoing traffic on your repository? (This helps you to decide whether the specification of your current server machine is sufficient enough for your needs).

Backup: what backup arrangements are in place?

Security: Are security issues such as virus protection, unauthorised access, etc. are properly addressed?

Guide to setting up a new repository



If you decide to set up your own repository, then you have to install repository software of your choice and make it harvestable by the EThOS central hub.

There are several key issues to be taken into consideration when selecting a software solution for setting up a repository.

They are:

- **Availability:** A range of potentially suitable systems are available and there is evidence that developmental work is still ongoing. The previous e-theses projects have given consideration to systems that have been tested and adopted by a number of respected institutions: DSpace, EPrints, Fedora, Virginia Tech ETD-DB, Greenstone and several other software packages identified from the Open Society Institute '[Guide to Institutional Repositories](#)'.
- **Interoperability:** To ensure cross institutional access and also to allow BL to harvest your metadata, the software system selected should comply with the most up to date version of the '[Open Archives Protocol for Metadata Harvesting \(OAI-PMH\)](#)'.
- **Suitability:** The software selected should be relatively simple to install across a range of hardware types and operating systems. The software should meet the needs of individual institutional policies with regard to the integration of ETDs with other material in electronic repositories. The software should also offer an intuitive and appealing user interface.
- **Sustainability:** The creators or existing users of the software should demonstrate a commitment to the future support and development of the product.

Project Preferences

Recent JISC funded projects associated with the establishment of institutional repositories that contain e-theses have successfully made use of **DSpace**, **EPrints** and **Fedora** software.

You can choose any of the above or any other software that meets the data transfer standards identified by EThOS; the choice will depend upon your hardware infrastructure and available technical support. Described below is the guidance on

how to setup each of these repository softwares and make them compatible to EThOS.

DSpace

[DSpace](#) was designed in 2002 by [MIT Libraries](#) in collaboration with the [Hewlett-Packard](#) Company. Its focus in the submission, storage, access and preservation of research material in digital format makes it well suited to the needs of an institutional repository. Follow the links below for more information on DSpace and Installation guidelines.

- [DSpace Introduction](#)
- [Download latest version of DSpace](#)
- [Hardware/Software requirements and Installation](#)
- [FAQs on how to install](#)
- [FAQs on how to use](#)

In addition to the above general setup of DSpace, you need to perform some additional setups, as required by EThOS, which will allow the EThOS central hub to harvest your DSpace repository successfully:

EThOS has developed guidelines on how to customise your DSpace metadata-entry pages for these submission (for version 1.2.2 or later). This will help you to make sure that all required fields from UKETD_DC are entered in the correct format. Instructions for implementing customised e-thesis submission forms conforming to EThOS requirements, a predefined XML fragment for cutting and pasting, and an example of a working customised input submission form can be downloaded by following the link below.

- [DSpace custom input submission form](#) for theses (for version 1.2.2. or above)

EThOS requires UKETD_DC metadata of your theses to be exposed via the OAI interface of your repository. A DSpace OAI Crosswalk plug-in has been developed for this purpose. You need to install this plug-in to expose your metadata in UKETD_DC format. You can download the plug-in with full instructions on how to install it by following the link below.

- [DSpace OAI Crosswalk](#) (for versions 1.2.2 - 1.4.2)

Note: as of DSpace version 1.5.2, the UKETD_DC crosswalk is already included in the distribution files and can be activated by uncommenting the line:

`Crosswalks.uketd_dc=org.dspace.app.oai.UKETDCCrosswalk` in `oai.properties`

In order to minimise the administration cost of maintaining the multiple URLs for a single institution, EThOS has adopted the policy of only maintaining a single point of collection for all your theses. For those DSpace repositories which contain multiple collections of theses, EThOS has developed a modification to OAI interface that forces DSpace to only expose metadata for theses only in UKETD_DC. In this case EThOS central hub can use just the base URL of your modified OAI interface to harvest all your theses content. You can download the modifications along with full installation instructions for DSpace 1.3.2 and for DSpace 1.4 by following the links below.

- [UKETD_DC OAI modification for DSpace 1.3.2](#)
- [UKETD_DC OAI modification for DSpace 1.4.x](#)
- [UKETD_DC OAI modification for DSpace 1.5.2](#)
- [UKETD_DC OAI modification for DSpace 1.6](#)

Eprints

[EPrints](#) was developed at the University of Southampton and first made available in 2000. Supported by an active user base, and with the capability of being modified to meet local requirements, EPrints is well suited for an institutional repository. The EThOS central hub repository maintained by BL is also based on EPrints. Follow the links below for more information on EPrints and the installation guidelines.

- [Eprints Introduction](#)
- [Download latest version of Eprints](#)
- [Hardware/Software requirements and Installation](#)
 - [FAQs](#)

EThOS requires UKETD_DC metadata of your theses to be exposed via the OAI interface of your repository and also requires you to provide a single point of collection (i.e. URL) for all your theses contents. So, in addition to the above general setup of EPrints, you need to perform some additional setups, as required by EThOS, which will allow the EThOS central hub to harvest your EPrints repository successfully:

Eprints v2. EThOS has developed a EPrints OAI Plugin for the institutions with an EPrints v2 repository. It allows the entry of metadata and theses conforming to EThOS requirements and their subsequent exposure in UKETD_DC via the OAI interface. You can download the EPrints OAI Plugin together with full installation instructions by following the link below.

[EPrints OAI Plugin](#)

Eprints v3. the OAI interface to expose UKETD_DC is already bundled with the EPrints 3 distribution. However, it is still necessary to configure your installation to allow entry of metadata conforming to EThOS requirements. To this end you may find the [EPrints v3 Configuration Notes](#) helpful.

Fedora

[Fedora](#) is a general purpose digital repository software developed jointly by Cornell University Information Science and the University of Virginia Library. Its development started in 2001 and several versions have been released to date. At its core is a powerful digital object model that supports multiple views of each digital object and the relationships among digital objects. This makes it a highly configurable and customisable system which is well suited for an institutional repository. Follow the links below for more information on Fedora and the installation guidelines.

- [Fedora Homepage](#)
- [Download latest version of Fedora](#)
- [Hardware/Software requirements and Installation](#)
 - [FAQs](#)

EThOS requires your repository to expose UKETD_DC metadata of your theses via the OAI interface and also requires you to provide a single point of collection (i.e. URL) for all your theses contents. In order to meet these requirements, you may need to perform some additional setups to your Fedora repository in addition to the general setups.

Technical Factors

When setting up a repository:

Character sets: For compatibility with the Central Hub, you should ensure that you expose your metadata using the UTF-8 charset.

And, the following general technical factors should be considered:

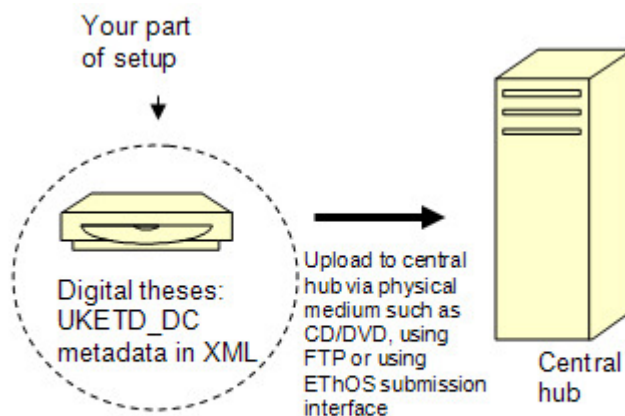
Traffic: What is the volume of your theses and expected traffic to your repository?

Server type: Depending upon the technical infrastructure available to your institution, you may wish to use one of your existing servers or may wish to set up a separate server. Most repository software is open source and is intended for UNIX/LINUX based servers. However, you may install your repository in a MS Windows based server if the repository software permits.

Backup: In case of server breakdown proper backup arrangements should be in place.

Security: Security issues such as virus protection, unauthorised access, etc. should be addressed.

Guide to using the central hub repository



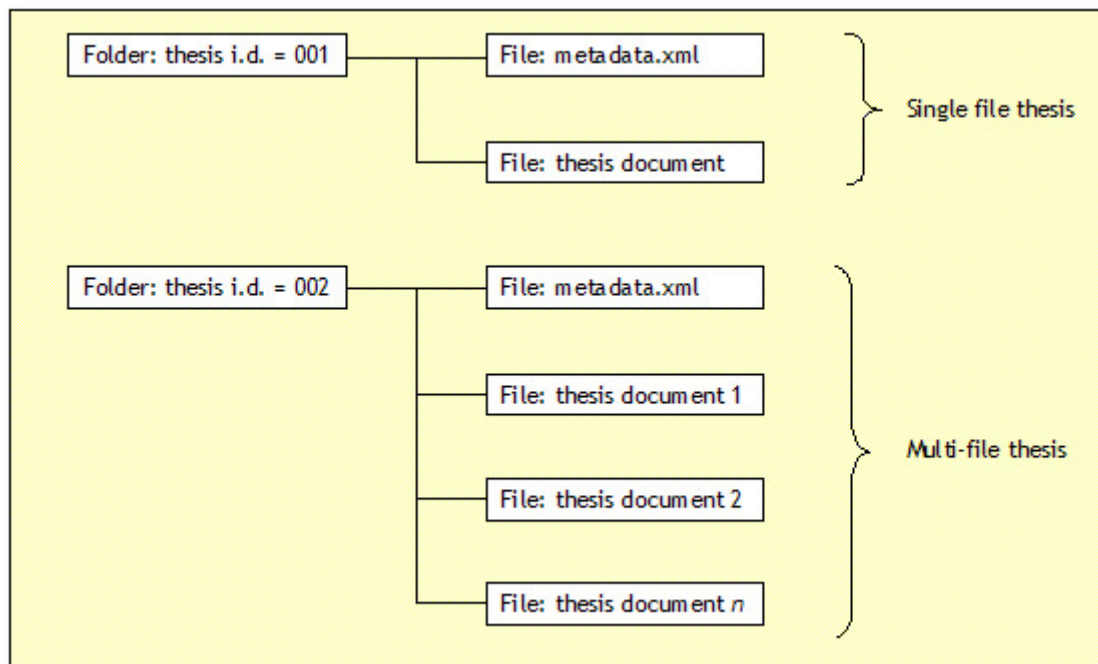
Using the Central Hub repository

If you decide not to set up your own repository, you can still use the central hub as an e-theses repository.

In this case, instead of EThOS harvesting your repository, you will have to submit your e-theses to EThOS as follows:

Physical loading via CD or DVD

You can supply your theses to the EThOS central hub by putting your theses content in the CD/DVD and physically send them to EThOS. If this method is to be used, supplying institutions must present the metadata and theses in a standard manner. The folder and file structures for supply of content to the EThOS central hub administrator should be as follows:



The 'metadata.xml' file should contain metadata expressed as UKETD_DC and the format of the theses documents should be in keeping with the policies of the EThOS central hub. Click in link below to view a sample metadata.xml file.

- Sample [metadata.xml](#)

If you wish EThOS to harvest your metadata (from a location/server specified by you), your metadata file also requires the "OAI wrapper" to be added on top of general UKETD_DC fields. Click in link below to view a sample metadata.xml file with required OAI wrapper.

- Sample [metadata.xml with OAI wrapper](#)

Using central hub group submission interface developed by EThOS

EThOS has developed a "group submission interface" that can be installed at the institutional end. Group submission interface is a client application which connects to a server at the EThOS central hub. It includes a theses submission interface and allows you to enter the metadata and the content for the theses which can then be directly submitted to the central hub.

You can download the EThOS group submission interface together with full installation instruction by following the link below.

- [EThOS Group submission interface](#)

Training and Guidance

This section focuses on training for library and IT staff, academics and students; offers advice on the creation and submission of e-theses and how to provide support to those involved in the process

When changing procedures to encourage or require the submission of theses in electronic format, it is necessary to ensure that adequate training is provided for Library, I.T. and administrative staff as well as for students and supervisors.

It is also necessary to provide guidance to those who will be responsible for establishing and maintaining an institutional repository and to those who will be responsible for publicising the availability of the institution's theses on the Web.

Training methods

Training and guidance may be provided in a number of ways and different approaches may be adopted for different groups of stakeholders.

Personal, individual training

This approach benefits the recipient of the training significantly but it is time consuming and therefore costly to provide. A session can be arranged to focus directly on the needs of the individual. It can be pitched at the right level and adapted in response to feedback obtained at the time. This service may prove most worthwhile when targeting individuals who are not convinced about the changes proposed, or when encouraging individuals who are concerned about potential additional workloads or about lacking the skills and expertise to fulfil the new requirements.

Informal workshops

Workshops may be arranged at intervals throughout the year and advertised to students (or other specific groups involved in the process) as available on a bookable or drop-in basis. This approach has the advantage of allowing students to choose to attend training sessions at the time most suitable to them and they are therefore likely to be receptive to the ideas and advice imparted. The disadvantage is that students may not appreciate how early in their studies they would benefit from participating in a workshop and they may only attend at the 'writing up' stage. Many students may fail to attend the sessions, even if they are well publicised, if the workshops are offered on a voluntary basis. Informal workshops may be of most value where there is a 'voluntary submission' arrangement for e-theses and where those students who are enthusiastic about having their work made more accessible, or producing their theses in non-traditional formats, can take advantage of an opportunity to receive 'face-to-face' training.

Formal workshops

The provision of formal, timetabled workshops may ensure a high level of attendance, but students may not receive the advice they need at the most appropriate time. This approach to training may be most positive where there is a significant amount of support from supervisors and university management and where time is allowed on students' schedules for participation in a major training event or series of sessions. This approach may be most appropriate where the submission of theses in electronic format is compulsory and where it is necessary to demonstrate that an effort has been made to provide training to every relevant student.

On-line help screens

On-line help screens are likely to be of value in addition to any of the above training routes. Basic information can be imparted and specific tutorials can be provided (e.g. for converting 'Word' documents into PDF). Links can be provided to Web pages that contain information about university regulations, etc and students may find it useful to have a single point of contact for information about e-theses creation and submission.

Presentations and publications

Staff and students who have participated in organised training sessions will need to be updated about new e-theses developments at intervals. If your institution has an annual programme for training research supervisors and a research methods training course for students it may be worth ensuring that time is included for a presentation about e-Theses. Papers may also be submitted to relevant university committees to ensure that key individuals are aware of recent developments, and articles can be written for inclusion in in-house newsletters, etc.

Training content

The topics to be covered in training sessions will vary according to the target audience.

Training for students

Training for students may need to include basic tips on using word processing software correctly to produce an e-thesis: inserting page breaks, page numbering, indenting paragraphs and aligning the text in 'Contents' pages are amongst the commonest areas where mistakes are made. Consideration also needs to be given to how to scan drawings and photographs for inclusion, how to create navigation aids, whether to divide the document into multiple files and how to name the files, etc. Students may need advice about how to convert their documents into PDF and how to submit the finished product. General tips about saving work, allowing time for dealing with presentational aspects of the thesis, and identifying key words may be included in the training schedule. A key element of any training programme must be IPR and copyright issues, particularly the provision of information about how to deal with third-party copyright and the types of situations where it is important to obtain permission from rights holders.

Training for supervisors

Much of the training provided for students is also relevant to supervisors since many students will approach them rather than Library or I.T. staff for guidance. It is also important to engage supervisors in discussions about the extent to which they are prepared to accept multimedia based theses - and to consider the training needs associated with this.

Training for Library, I.T. and administrative staff

Those responsible for training students and supervisors, for creating and maintaining an institutional repository, and for organising the submission procedure and liaising with EThOS will need to keep abreast of developments relating to e-theses nationally and internationally. The [NDLTD Web pages](#) provide a good global overview of activities relating to electronic theses and dissertations and the [annual ETD Symposia](#) provide an insight into potential forthcoming developments. It may be helpful to identify a small team of staff, possibly those responsible for the institutional repository where this exists, who will communicate the latest developments to the relevant stakeholders and co-ordinate the introduction of changes as these appear necessary or beneficial. This team could also be responsible for ensuring that the existence of the institution's e-theses repository is well-publicised externally and for monitoring and publicising the extent to which the theses are being accessed in electronic format.