Part II

Implementation Process

Creation of Metadata for Current Theses

2.1   The metadata of theses and dissertation submitted to a university / institution is created when a copy of thesis is given to its library. Like other documents, the library catalogue these theses using tools and techniques that are used for cataloguing other types of documents, predominantly books, received in the library. In this process, the libraries ignore some of the bibliographic details, such as names of the research supervisor, name of department and faculty, degree awarded, etc. that are peculiar to theses and dissertations. The metadata for theses and dissertations, therefore, cannot be sourced through the library.
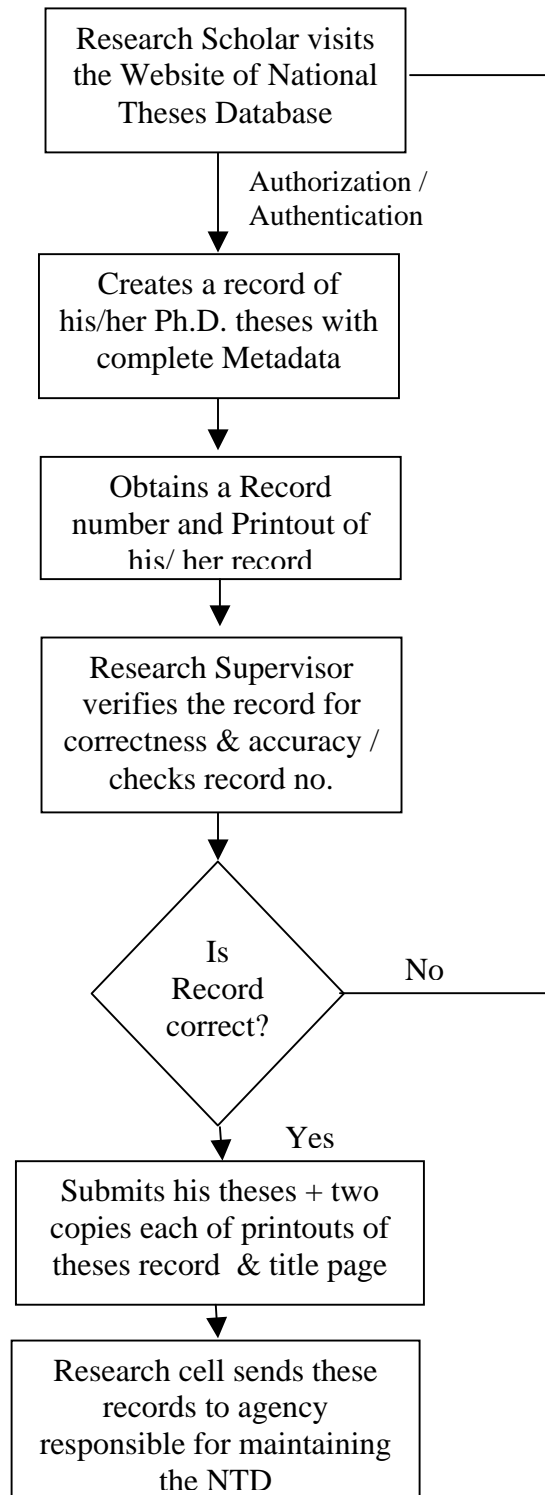
2.2   Good and reliable metadata for theses and dissertations can easily be created at the time of submission of theses. Most universities have Research Cells (or other similar units) where student submit their theses. The Research Cells can be made responsible for coordinating the activities related to creation and submission of metadata for theses and dissertations.

Methodology

2.3   Step-wise procedure to be followed by a student for creation of metadata shall be as follows:

i)   The research student logs-on to the web site of "Indian National Theses Database" (INTED) and creates a record online. S/he receives a "Record Number" as an acknowledgement. S/he is also prompted to take a print-out of bibliographic record created by him / her in this process. Universities/central agency would develop registration based authentication/authorization process for students to log on to the INTED

ii) Research student submits a copy of print-out of bibliographic record created by him / her in the step (i). Alternatively, s/he fills-in a form (Annexure III) to provide metadata of his / her thesis being submitted along with an abstract. The "Record Number" obtained in step (i) is also mentioned in this form;

iii) The research supervisor verifies the correctness and accuracy of record created in the step (i). S/he also verifies that the student has also created a record online and has obtained the record number;

iv) The INTED would also facilitate subsequent editing of record, if required.

v)   Two copies of the record, duly verified by the Research Supervisor in step (iii) is submitted to the Research Cell of the university along with a copy of title page of the thesis. A template for title page for a thesis along with its information content is given as Annexure V;

vi) The research cell collects these records and sends them to the agency made responsible for maintaining the "Indian National Theses Database" in a batch of five records; and

vii) The agency made responsible for maintaining "Indian National Theses Database" would verify records entered in the theses database, generate and send quarterly reports giving details of theses submitted by various universities for their verification.

```
        ┌─────────────────────────┐
        │  Research Scholar visits │────────────┐
        │  the Website of National │            │
        │    Theses Database       │            │
        └─────────────────────────┘            │
                     │ Authorization /          │
                     │ Authentication           │
                     ▼                           │
        ┌─────────────────────────┐            │
        │     Creates a record of  │            │
        │  his/her Ph.D. theses with│           │
        │     complete Metadata    │            │
        └─────────────────────────┘            │
                     │                           │
                     ▼                           │
        ┌─────────────────────────┐            │
        │     Obtains a Record     │            │
        │   number and Printout of │            │
        │      his/ her record     │            │
        └─────────────────────────┘            │
                     │                           │
                     ▼                           │
        ┌─────────────────────────┐            │
        │    Research Supervisor   │            │
        │    verifies the record for│           │
        │  correctness & accuracy /│            │
        │     checks record no.    │            │
        └─────────────────────────┘            │
                     │                           │
                     ▼                           │
                   ◇ Is                 No       │
                   ◇ Record ◇───────────────────┘
                   ◇ correct?
                     │ Yes
                     ▼
        ┌─────────────────────────┐
        │    Submits his theses + two│
        │  copies each of printouts of│
        │   theses record  & title page│
        └─────────────────────────┘
                     │
                     ▼
        ┌─────────────────────────┐
        │  Research cell sends these│
        │     records to agency    │
        │  responsible for maintaining│
        │         the NTD          │
        └─────────────────────────┘
                     20
```

Figure 2: Flowchart for Creation of Metadata for Current Theses

Major Issues

2.4 (i) Development of Web-based Interface for Indian National Theses Database: The agency designated to maintain the Indian National Theses Database would be required to develop a web-based interface that would enable its users to search and browse the database. Besides, it would also facilitate research scholars to create bibliographic records online for their theses submitted to their respective university.

(ii)Indian Language handling: The Indian National Theses Database, to begin with, would consist of bibliographic records in English language. Later capability for handling other languages would also be incorporated.

By-product

2.5   While the online version of "Indian National Theses Database" can be made accessible free of cost. Its CD ROM version can be made available at a price.

Incentives

2.6   The universities, which successfully regulate the process mentioned above, would qualify for additional points in the process of accreditation.

Individual universities may be given an output on CD for metadata of theses by the concerned University.

Submission of Ph.D. Theses in Electronic Format

2.7   Doctoral dissertations submitted to a university / institution are originally created in digital format using one of the word processing software packages like MSWord, LaTex, Word Perfect, Word Pro, etc. or one of the desktop publishing packages like Page Maker, Ventura, etc. These documents are undisputedly highly valuable collection especially in digital format that qualify to be an important component of a digital library.

2.8   The documents composed in word processing / desktop publishing packages can be easily converted into PDF, Post Script or marked-up in XML using appropriate software tools so as to host them on the web.

Implementation

University's Responsibility

2.9   Universities can do the following for implementation of this initiative:

i)   Mandate Submission of E-theses: All universities should mandate the submission of electronic version of theses. The existing sets of Ph.D. regulations may be suitably modified to include a clause for submission of an e-version of Ph.D. theses along with printed theses.

ii)   Student Approval Form: Each university should obtain declaration / consent from the students assigning non-exclusive rights to archive and distribute their doctoral work on the Internet in full-text with our without conditions. (Student Approval Form- Annexure I).

iii)   Submission Format: Students may be allowed to submit their theses preferably in PDF (Portable Document Format). Alternatively, they may also be allowed to make submission in MS-Word, Tex, LaTex, or other standard formats approved by the university. Facility for converting various formats into PDF may be provided in the Research Cell or in the Library.

iv)   Training to Students: Universities may organize formal training programme to educate students on various aspects of electronic submission of theses and dissertations including its advantages, submission procedure, conversion from native format into PDF, etc.

v)   Institutional Repository: All universities should set-up and maintain an institutional repository of e-theses. See 7.3 for further details in this regard. All the theses submitted to a University should be archived and maintained in the institutional repository set-up for this purpose.

vi)   Alternate Submissions: In case a university is not been able to set-up its institutional repository, it may still accept e-version of theses on CD ROM. The CD ROMs may be properly labelled and stored till the institutional repository is set-up. Once the institutional repository is set-up, the theses available on the CD ROM may be submitted to the e-repositories.

vii)   E-theses Web Site: The web site of each university should provide information to the students regarding policy and procedures for electronic submission of theses and dissertation.

viii) Develop File Naming Convention: Universities may define a formal convention for file name so as to have unique name for each file that are meaningful and convey the contents of files being uploaded. Besides, a formal convention should also avoid chances of having same files names for submissions made by different students. Students may be instructed to store and upload each chapter in a separate file using an agreed file naming convention. All files should have a distinct name with an appropriate standard suffix according to the convention for its type (e.g. ".doc" for MS Word files, ".gif" for GIFs, ".jpg" for JPEGs, ".txt" for text files. Initial pages of thesis (i.e. acknowledgement, contents, certificate, etc. may be clubbed into one

single file which can be named as "intro.pdf". For example, each chapter in a theses may be named as follows:

InitialsLast name of author_chapter1
jarora_intro.pdf: Introductory chapters submitted by Jagdish Arora
arora_chapter1.pdf : Chapter one of theses submitted by Jagdish Arora
jarora_chapter2.pdf : Chapter two of theses submitted by Jagdish Arora

Suitable procedures may be followed to avoid anomalies, for example duplicate file names.

The standard mentioned above will facilitate users to easily identify the relevant parts of a thesis via a file structure that is common across all theses. It will also allow users to quickly look at the table of contents and other introductory information without having to wait for the entire thesis to download.

ix) IPR Policy: Each University should evolve copyright policy relating to the doctoral theses, which may be included in the Intellectual Property Right (IPR) policy of the university. A detailed note on IPR and electronic thesis is enclosed as Annexure V.

x) The centralised agency identified for co-ordinating ETDs will also take up the responsibility of tracking international developments in ETD hosting and maintenance and make appropriate suggestions from time to time.

Students'/Research Scholars' Responsibility

2.10 A student/Research scholar would be responsible for creation of a bibliographic record of his/her thesis and subsequent submission. Instructions to research scholars would vary from university to university depending upon the software used for e-submission. Steps generic in nature are given below:

i) Log-on to ETD Server: A researcher logs-on to the Web Site of his / her university that accepts electronic submission of theses. S /he creates a new account in the ETD database so as to submit his / her thesis.

ii) User ID: The research is assigned a unique session ID and password to register in the database that can be used for submission as well as for corrections of submitted records and files in future.

iii) Enter metadata or bibliographic information: The researcher is prompted to create a record for their ETD based on basic bibliographic information. Follow the online instructions and complete the form. Also enter an abstract for your theses.

iv) Processing and Change Bibliographic Information: Once a researcher has filled-in all bibliographic information, s/he can process it by clicking the submission button. The system displays a bibliographic record of his / her theses created by him / her and prompts him / her to edit the record, if required.

v) Uploading ETD Files: Once bibliographic information for your thesis has been created successfully, the system prompts you to upload files containing full-text of your thesis. You will generally receive confirmation that files have been successfully uploaded, but if there are problems with the filenames you have used, this system will also inform you and instruct you to make necessary changes and upload files with corrected names.

vi) File Names: A formal convention for file name is desirable so as to have unique name for each file that are meaningful and convey the contents of files being uploaded. For further details, please refer to 7.2.1.1-viii: File Naming convention.

vii) Use Corrected Version: The electronic version of theses should incorporate all corrections / modifications in it.

Types of Access

2.11   A university has an option to open-up its theses world-wide to the entire research community or restrict it to a few privileged users. The types of access that can be given are as follows:

▪ Worldwide Access: This access level allows immediate worldwide access to all ETDs, as soon as they are submitted. Onsite and offsite users can read and download the ETDs.

▪ Campus-only Access: The access in this category is restricted only to a particular group of users on-campus. The access is controlled either by IP addresses, a valid ID, or through a proxy server.

▪ Temporary Restricted Access: The ETD could be restricted for one, two or three years and then it is released for worldwide access, unless the University / Department authorities give an approval for delayed release.

▪ Mixed Access: The student has the opportunity to release parts of the ETD but not the entire thesis or to publish two versions of a chapter, each with different access levels. For example, one version of an ETD chapter, which might contain information to be published in a journal, might be restricted to on-campus use only, the second version might be available for everyone in an unrestricted access level but might not contain the data that is going to be published later.
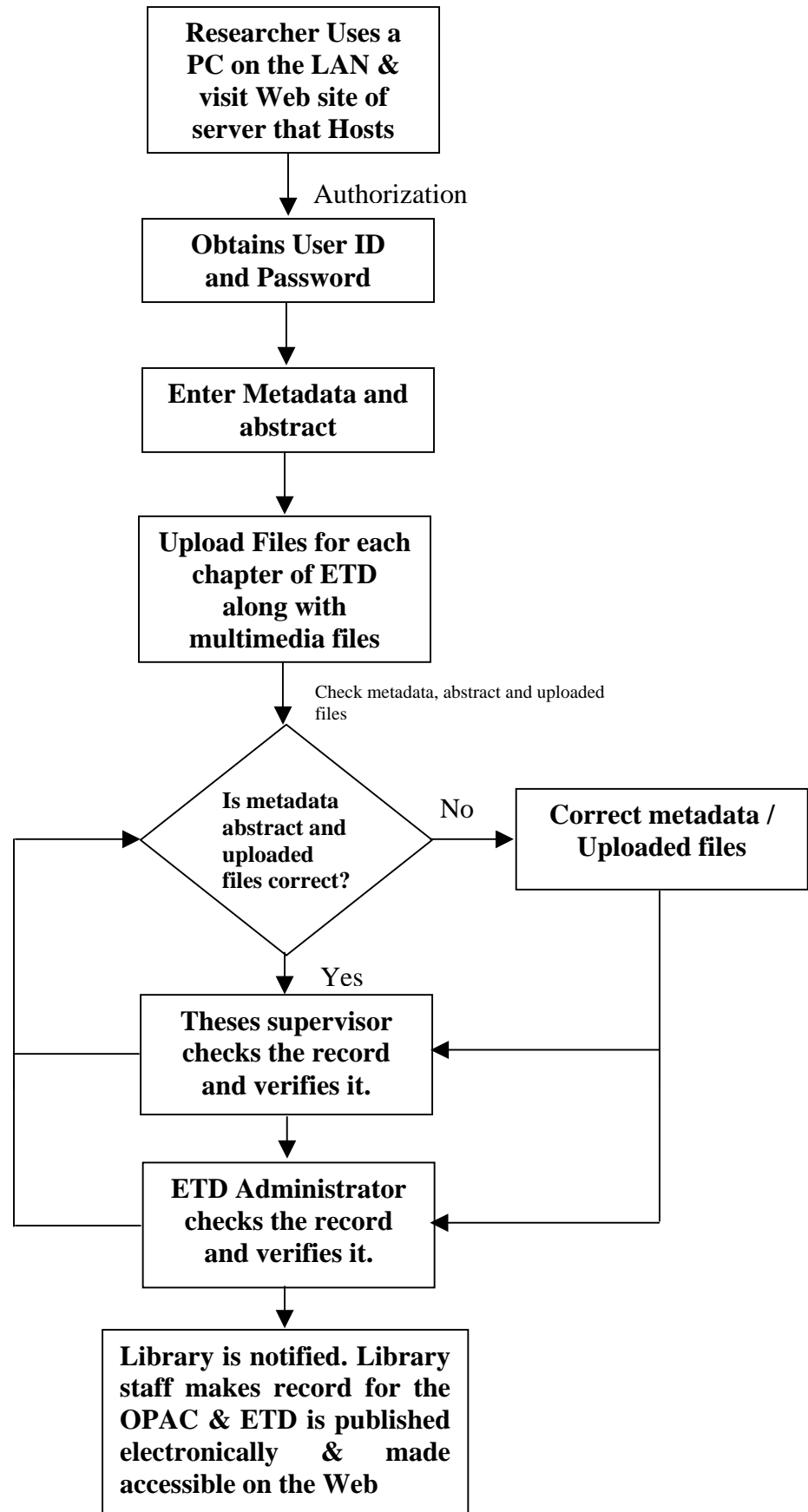
```
┌─────────────────────┐
│  Researcher Uses a  │
│  PC on the LAN &    │
│  visit Web site of  │
│  server that Hosts  │
└─────────────────────┘
           │
           │  Authorization
           ▼
┌─────────────────────┐
│   Obtains User ID   │
│    and Password     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Enter Metadata and │
│      abstract       │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Upload Files for   │
│   each chapter of   │
│    ETD along with   │
│   multimedia files  │
└─────────────────────┘
           │
           │  Check metadata, abstract and uploaded
           │  files
           ▼
        ╱╲
       ╱  ╲        No    ┌─────────────────────┐
      ╱ Is ╲──────────── │  Correct metadata / │
     ╱metadata╲          │   Uploaded files    │
     ╲abstract╱          └─────────────────────┘
      ╲ and  ╱                      │
       ╲uploaded                    │
        ╲files correct?╱            │
         ╲╱                         │
           │  Yes                   │
           ▼                        │
┌─────────────────────┐             │
│  Theses supervisor  │◄────────────┘
│  checks the record  │
│   and verifies it.  │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  ETD Administrator  │◄────────────
│   checks the record │
│   and verifies it.  │
└─────────────────────┘
           │
           ▼
┌──────────────────────────┐
│ Library is notified. Library │
│ staff makes record for the   │
│ OPAC & ETD is published      │
│ electronically    &    made  │
│ accessible on the Web        │
└──────────────────────────┘
```

Figure 3: Electronic Submission of Theses and Dissertation

- Withheld Access: The access is restricted for all. In these cases the university is able to give access only to the bibliographic details/metadata of the thesis/dissertation along with a summary / abstract.

The Student Approval Form can have options for researchers to choose the level of access s/he would like to offer.

Incentives

The universities, which successfully regulate the process mentioned above, would qualify for additional points in the process of accreditation. Such universities may also apply for a research project to establishing repository for ETD. The UGC would keep provision for funding such projects.

Setting-up Institutional Repositories for ETD

2.12   The regulation aims at setting-up OAI-complaint institutional repositories using standardized metadata scheme(s) that is to be used uniformly in all universities for hosting e-theses submitted in the process mentioned above.

2.13   An institutional repository (IR) is a digital archive of an academic institution's intellectual output including ETD. IRs adhere to an open access model, by centralizing and preserving the knowledge of an academic institution and making it accessible to anyone with Internet access. Institutional repositories are not discipline-specific, instead it aims to archive the entire range of a university's intellectual output. IRs also form part of a larger global system of repositories, which are indexed in a standardized way, and are searchable using one interface, providing the foundation for a new model of scholarly publishing. It is assumed that universities will be implementing their IR using one of the OAI-complaint IR software such as:

➢ D Space (MIT, Cambridge, USA) (http://www.dspace.org/)

➢ E-prints (University of Southampton, UK)   (http://www.eprints.org/)

➢ ETD-db (Virginia Tech University, USA)   (http://scholar.lib.vt.edu/ETD-db/)

➢ Fedora, University of Virginia Library (http://www.fedora.info/)

➢ CDSWare, CERN, Geneva, Switzerland (http://cdsware.cern.ch/)

The basic aim of using OAI-compliant IRs software is to ensure interoperability contents and metadata with each other as well as with other OAI-compliant archives.

Steps for Setting-up Institutional Repositories

2.14   The steps presented here are based on documentation produced by the Open Archives Initiative (www.openarchives.org), Eprints (http://software.eprints.org), Dspace (http://www.dspace.org/) and institutional repositories designed by other universities.

Pre-implementation Steps

2.15   Setting up an institutional repository is not a trivial task. However, the biggest issue is actually deciding our needs and archiving policies. The following is a list of the issues that should be addressed to prior to software implementation, since they will effect the software configuration:

Content
2.16   The first step for creating an institutional repository is to establish some content guidelines. A review of a number of existing repositories shows that institutional policies regarding content vary substantially. The contents, in this case, are restricted to the ETDs submitted to a given university and, therefore, the metadata and information fields may be customized (or qualified) to accommodate metadata specifically for ETDs. Copyright, archiving and distribution guidelines in relation to ETD may be clearly detailed. Annexure VI provides a detailed write-up on "Intellectual Property and Electronic Theses."

Metadata

2.17   Institutional repositories must incorporate, indexes and search options from diverse collections deposited in diverse formats. It needs to deal with standard vocabularies from diverse fields of study, and include metadata pertaining digital object structure, administration, and content. Unqualified Dublin Core (www.dublincore.org) is the minimum metadata required for OAI interoperability, however, depending on the type of content a repository has, you may include other metadata sets (such as name thesis supervisor, degree awarded, etc).

2.18   Since OAI is based on the exchange of metadata, getting the metadata right is fundamentally important for a repository. The OAI compliant software would produce the necessary Dublin Core metadata for harvesting by service providers.

Document Type

2.19   OAI-complaint institute repositories software packages are designed to accommodate different document types such as e-prints, books, posters and conference papers. Institutional repositories may, however, include additional document types such as journal article, preprint or thesis, depending on the content guidelines of the repository. In this particular case, the document type would be theses.

Subject Headings

2.20   Identifying a useful set subject headings is one of the major challenges for repository implementers. Broad subject headings may be appropriate for a single institutional repository, however, as access to institutional repositories becomes federated (across institutional repositories installed in various universities), it becomes more problematic. It is impossible for a user to browse papers from a variety of repositories that use different subject terminologies. It would therefore, be better to provide a set of descriptors to each record that are both useful to experts in the field and those who are not.

Format

2.21   The default formats accepted by most IR software are PDF, Postscript, ASCII, and HTML. Individual installations may want to add or forego some of these formats. Possible additions may be specialized formats for data sets, or other common formats, such as Rich Text Format. There are open-source utility programs available to convert from non-supported to supported formats. Consideration may also be given as to whether any of these default formats should be switched off. For example, HTML is very fluid format that is difficult to validate easily and some may choose not to accept documents in this format.

2.22   The most IR software enables implementers to design their own subject hierarchy and load it into the database fairly easily. However, it is far more complex to alter this once you have started to upload documents. So it is important to get this right before uploading too many papers. To view some examples of additional content types, formats and subject headings you can visit the advanced search pages of some existing archives: http://eprints.anu.edu.au/perl/advsearch,   http://eprints.ncsi.iisc.ernet.in   and http://eprints.lib.gla.ac.uk/perl/advsearch.

Implementation

2.23   It is suggested that, as a first step, a demonstration version of an IR using the default configuration is set-up. Previous experience has shown that the initial configuration can be quite a bit of work if you have complex requirements, and rather easy if the only part of the default configuration is required to be change.

Installation of IR Software

2.24   D Space, E-prints, Fedora and CDS Ware are some of the OAI-complaint software that are available free of charge from the Web sites mentioned above. All these software are OAI-compliant and once installed, they are automatically ready to generate metadata in a form that can be picked up by OAI harvesters. Installation of these software takes approximately one to two days (or more depending upon the extend of customisation). The installation and customisation of these software requires in-depth knowledge of operating system, RDBMS system being used as back-end database (PostgreSQL for Dspace MySql for e-prints) and other pieces of software that are harmonized with IR software (for example Dspace uses Tomcat for Java and Eprints uses

Perl programming language). Detailed installation instructions for most the IR software are available on their web sites. Section 7.3.2 provides details on software options for institutional repositories.

Technical Requirements

2.25   The hardware and software requirements IR software are as follows:

- Any computer capable of running GNU/Linux or similar operating system. Obviously, the faster, the better, but any Intel Pentium IV processor or Zion processor will give good performance.

- A GNU operating system. GNU/Linux (www.gnu.org or www.linux.org) a very advanced and free UNIX-like operating system works just fine.

- The Apache WWW server (www.apache.org/httpd.html) another professional-quality free software product, often included with GNU/Linux distributions, such as that produced by RedHat (www.redhat.org).

- The Perl programming language (www.perl.com) also included with most GNU/Linux distributions for Eprints and Tomcat for Dspace.

- The mod perl (http://perl.apache.org) module for Apache (www.apache.org/httpd.html), which significantly increases the performance of Perl scripts.

- The MySQL Database (www.mysql.com), a free database system, for Eprints and PostgresSql for Dspace.

- The IR software (Dspace / E-prints)

Potential Problems

2.26   Most IR software are considered ready for use. There are still some issues and bugs, but not too many or too major. A list of known issues are listed on the Web sites of IR software. All IR software host discussion list (http://software.eprints.org/tech.php/ in case of Eprints and (http://www.dspace.org/feedback/mailing.html) in case of DSpace). Discussion lists and their archives are generally a good place to find solutions to problems that may encounter during implementation.

Post Implementation

Interface Design

2.27   Most institutional repository software provides a web interface for managing, submitting, searching, browsing and downloading documents. It may take a few days and some expertise to customize the web interface for the repository.

Quality Control

2.28   Most institutional repository software have a submission buffer, in which all content are stored before they are cleared by a mediator for making them publicly available. The system administrator can accept, edit or reject a submission at this stage. This allows the administrator to approve deposited material before it goes live, ensuring a certain level of quality control over metadata, formatting and in some cases, content of the deposited material. If there is a problem with a thesis that has been deposited, it can be returned to the submitter's "workspace", and the researcher is sent an e-mail explaining the problem. Repositories may outline policies, design template regarding quality control of submissions.

Documentation

2.29   ETD repositories are required to provide detailed documentation to assist the researchers, both in submission and preparation of their theses. Most IR software web site provides sample instructions for submitters that can be used as the basis for help documentation by an ETD repository.

Registering the Repository

2.30   Once the Institution Repository software has been installed, the server needs to be registered with the Open Archives Initiative (http://www.openarchives.org/data/ registerasprovider.html). The OAI maintains a list of OAI-compliant archives for OAI Service Providers to be able to visit. Before registering the archive, the OAI will perform a set of conformance tests on the repository, to ensure integrity of the registry. When this is completed, they will confirm by email and the archive will added to the public list of OAI compliant data providers (http://www.openarchives.org/Register/BrowseSites.pl) In the case that the repository fails to complete the tests, the repository will be removed from the registry and an email will be sent containing an explanation of why the repository did not conform. The OAI periodically retests repositories for their conformance.

Promotion and Advocacy

2.31   Setting an archive up is one thing, but getting users to participate in its ongoing development is quite another. One of the most difficult tasks in setting up the archive is getting the content. This would be applicable in case of ETDs if submission of e-version of theses in made voluntary for researchers. Since the UGC regulation would make the ETDs mandatory, every thesis that is submitted to a university would be submitted electronically. In case submission of thesis is voluntary for researcher, their participation would be critical and will go through two important phases: first, the goal will be to just

get enough content in place to set up a demonstration. Once the demonstrator is in place, the second step will be to get a critical mass of content in order to provide a useful service and a repository more representative of all of the departments in the institution.

2.32   Institutions may publicize and promote the ETD repository through university magazines, and the Library user newsletter. Literature about the value of ETD repositories may be distributed to the potential users.

Software Options for ETDs

2.33   A number of software solutions are currently available for building digital publishing systems that can support different document types including electronic theses and dissertation. Several software are open-source, OAI-complaint solutions that allow its users to build an open-standards compliant interoperable ETD repository. A brief account of four of these software is given below.

ETD-db

2.34   The ETD database (ETD-db) was developed at Virginia Tech as a joint project between the Graduate School at Virginia Tech, the Digital Library and Archives (a division of the University Libraries), and the National Digital Library of Theses and Dissertations. The software was developed by Anthony Atkins and currently maintained by Carolyn Kletnieks, and is available free of charge. It is used for ETD submission, archive and search tool. The latest version is 1.8 and is available for download at http://scholar.lib.vt.edu/ETD-db/developer/download/

2.35   The ETD database is a series of web pages and Perl scripts that interact with a MySQL database. These scripts provide a standard interface for web users and researchers, ETD authors, graduate school personnel, and library personnel to enter and manage the files and metadata related to a collection electronic theses and dissertations.

2.26   Installation: ETD-db depends upon the Perl programming language and the MySQL open source database system. Perl is native to most Linux and Unix installs. In addition to the standard Perl installation, it is also necessary to install additional 'Perl Modules' which enhance the functionality of the language. It requires a reasonably experienced systems administrator to do the prerequisite installation. The ETD-db website (http://scholar.lib.vt.edu/ETD-db/) provides instructions for installing the software using the Apache web server.

2.37   Metadata: ETD-db supports the ETD-MS standard set by NDLTD. ETD-MS is the basic Dublin Core elements with the addition of several elements that further describe parts of the ETD process. This is the only metadata standard supported by the software.

2.38   Comments: The ETD-db system is designed specifically for ETDs. Furthermore, it is designed to support only the ETD-MS standard for ETD metadata. However, it is an open system and is modifiable. Modifications to the software is possible.

DSpace

2.39   DSpace (www.dspace.org) has been developed in partnership between Hewlett-Packard (HP) and MIT (Massachusetts Institute of Technology). Development work is still in progress, but Dspace, as institutional repository software, is making its mark with an increasing number of institutions around the globe installing, evaluating and using the package. The latest stable version is 1.2 and is available for download at http://sourceforge.net/projects/dspace/.

2.40   The original developers undertake most of the core development, but a growing technical user base is generating suggestions for future releases as well as producing some add-on modules. In addition, the DSpace Federation is guiding the transition to a more community-wide open-source development model.

DSpace captures, stores, indexes, preserves, and redistributes the intellectual output of a university's research faculty in digital formats. DSpace accepts all forms of digital materials including text, images, video, and audio files. Possible content includes: articles and preprints, technical reports, working papers, conference papers, e-theses, datasets (statistical, geospatial, matlab, etc.), images (visual, scientific, etc.), audio files, video files, learning objects and reformatted digital library collections.

2.41   Installation: The DSpace software is written in Java and requires the use of the Apache Tomcat server and the PostgreSQL database server. Installation requires access to the Tomcat applications directory and administrative privileges for the PostgreSQL server. Some knowledge of system administration is required to configure Tomcat, PostgreSQL, and Apache (use of Apache is optional). Directions for installation is provided on the DSpace website (http://dspace.org/technology/system-docs/). It requires an experienced systems administrator to do the prerequisite installation.

2.42   Metadata: DSpace uses a set of qualified Dublin Core elements loosely based on Library Application Profile set of elements and qualifiers. This set is modifiable but is used for all objects in the DSpace collection. This presents a problem when multiple document types are stored in the same DSpace installation. One solution is to create a set of metadata that can describe any object in the repository, while only providing the appropriate fields for each object when they are submitted.

2.43   Theses Alive Plug-in for Institutional Repositories (TAPIR): Tapir stands for Theses Alive Plug-in for Institutional Repositories. Tapir is being developed at Edinburgh University Library (EUL) (http://www.thesesalive .ac.uk/dsp_home.shtml). Tapir is an add on for DSpace. The current version is 0.3. Tapir is being developed with funding from JISC, with the intention of making it a open source, freely available addition to the DSpace.

2.44   Tapir provides the ability within DSpace to operate a supervised authoring facility, allowing Thesis and Dissertation Advisor/Guide/Supervisors to observe the ongoing work by their student on their project, to comment and to even make changes. This comes with an addition to the DSpace administration area to manage the supervising groups and their access policies to the student's work. It is envisaged that although developed specifically with ETDs in mind, that this software may also find other applications.

2.45   In addition, two submission interfaces (one for E-prints and other documents, and one for E-theses) are now supported, with the option to choose between them. Each of these submission interfaces provides custom metadata collection and licensing options for submissions.

2.46   Comments: DSpace is a very flexible digital collections system. The software is under active development and many features that would be beneficial to an ETD repository are being considered. Both a DSpace Thesis SIG and an ETD DSpace committee have been formed, and are coordinating activities.

E-Prints

2.47   GNU E-Prints 2.x is an open source digital library software package designed primarily to create institutional repositories (*http://software.eprints.org/)*. The default configuration creates a research papers archive. With its origins in the Scholarly Communication movement, E-prints default configuration is geared to research papers but it can be adapted for other purposes and content. It was developed in the Intelligence, Agents, Multimedia Group at the Electronics and Computer Science Department of the University of Southampton. GNU E-Prints is freely distributable and subject to the GNU General Public License. The latest version is 2.3 and is available for download at http://software.eprints.org/download.php.

2.48   Installation: Installing the E-prints software is relatively easy, and being made easier with each successive release of the software. Knowledge of MySql (used as backend database), apache WWW server and Perl programming language would be helpful. Mod_perl module for Apache significantly increases the performance of Perl scripts. Complete documentation for installation of E-prints is available on their web site (http://software.eprints.org/docs/eprints-docs.pdf).

2.49   Metadata: E-prints uses a set of unqualified Dublin Core elements. This presents a problem specially to accommodate fields that are specific to theses and dissertations.

2.50   Comments: E-prints is quite similar to Dspace. However, Dspace is a very flexible digital collections system and much more capable.

Flexible Extensible Digital Object Repository Architecture (FEDORA)

2.51   The Fedora repository system (http://www.fedora.info) is an open source, digital object repository system developed jointly by University of Virginia Library and Cornell University. The Fedora project is devoted to the goal of providing open-source repository software that can serve as the foundation for many types of information management systems. The software demonstrates how distributed digital information management can be deployed using web-based technologies, including XML and web services.

2.52   Features

- XML submission and storage: Digital objects are stored as XML-encoded files that conform to an extension of the Metadata Encoding and Transmission Standard (METS) schema.

- Parameterized disseminators: Behaviors defined for an object support user-supplied options that are handled at dissemination time.

- Access Control and Authentication: Although Advanced Access Control and Authentication are not scheduled until Phase II of the project, a simple form of access control has been added in Phase I of the project to provide access restrictions based on IP address. IP range restriction is supported in both the Management and Access APIs. In addition, the Management API is protected by HTTP Basic Authentication.

- Default Disseminator: The Default Disseminator is a built-in internal disseminator on every object that provides a system-defined behavior mechanism for disseminating the basic contents of an object.

- Searching: Selected system metadata fields are indexed along with the primary Dublin Core record for each object. The Fedora repository system provides a search interface for both full text and field-specific queries across these metadata fields.

- OAI Metadata Harvesting: The OAI Protocol for Metadata Harvesting is a standard for sharing metadata across repositories. Every Fedora digital object has a primary Dublin Core record that conforms to the schema. This metadata is accessible using the OAI Protocol for Metadata Harvesting, v2.0.

- Batch Utility: The Fedora repository system includes a Batch Utility as part of the Management client that enables the mass creation and loading of data objects.

Incentives

2.53   The universities, which successfully set-up the ETD repository in their universities, would qualify for additional points in the process of accreditation. Such universities may apply for research projects to carry out further research,

development and training on e-repositories for ETD. The UGC would keep provision for funding such projects.

Universities may approach UGC to fund them for metadata creation and digitisation of old Theses and Dissertations. The UGC in turn may selectively fund such projects to institutions of higher learning.

Creation of Metadata for Old Theses

2.54   A comprehensive database of doctoral theses submitted to the universities in India would require submission of metadata for all the Ph.D. theses submitted to their university from retrospect. If the metadata (bibliographic record) for theses is already available with the universities, it may be ensured that it confirms to the standard given in Annexure I, it is complete in all respect and is validated. The retrospective data in standardized format may be sent to the agency assigned to develop and maintain National Theses Database as a comma-delimited file / excel sheet,/OpenOffice's Calc, etc. The agency assigned to develop and maintain National Theses Database may also develop online interface that would facilitate uploading of data for theses from retrospect.

Concluding Remarks

2.55   This handbook provides basic guidelines for implementation of "UGC Regulation Framework on UGC Regulations for Submission of Metadata and Full-text of Electronic Form of Doctoral Theses". Individual universities are required to develop detailed guidelines for students, research supervisors and research administrators for actual implementation based on current practices, software being implemented and other local factors. The universities may develop following documentation / guidelines:

- Guidelines for metadata creation (For students)
- Guidelines for validation of metadata (For librarians)
- Guidelines for reviewing submitted theses & dissertations (for guides)
- Guidelines for subject descriptors (for metadata validation)
- Guidelines for submission of Theses and Dissertations (for students)
- Handbook of Software used for ETD (e.g, Dspace, ETD-dp etc.) for Digital Repository administrators
- Guidelines for Digitisation of hardcopy theses
- Copyright Declaration