

Regulations

UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005

To be Published in Gazette of India

NOTIFICATION

In exercise of powers conferred by clause (e) & (g) of sub-section (1) of Section 26 read with Section 14 of the University Grants Commission Act, 1956 (3 of 1956), the University Grants Commission hereby makes the **UGC (*Submission of Metadata and Full-text of Doctoral Theses in Electronic Format*) Regulations, 2005**

Preamble

Recognizing the need for a comprehensive database of doctoral theses for maintenance of quality of doctoral research in the country, these regulations are being issued to effect systematic creation, collection and compilation of cataloguing information, herein after called “metadata” of doctoral theses in a standardized format and to commence the process of submission of theses in electronic format, herein after called e-theses, in all universities. These regulations shall enable the UGC to:

- evolve a mechanism for creation of metadata of theses in a uniform format in all universities and its centralized collection, compilation and access through a database over Internet;
- commence the process of submission of e-theses in each university;
- establish OAI-complaint e-theses repositories at each university / regional electronic repositories and metadata harvesting services at the national level; and
- create, collect and compile metadata for all doctoral theses submitted to universities in India from retrospect. The resultant database would be made accessible on the Internet.

Regulations

1. Creation of Metadata for Current Theses

- 1.1. Each doctoral student shall create a record for his / her theses on the website of Indian National Theses Database. S/he would submit a print-out of bibliographic record created by him / her, a copy of title page of his / her thesis to a centralized agency to be named by the UGC;
- 1.2. The metadata submitted by the student should be duly checked by research supervisor / trained library staff for its completeness and accuracy. Corrections, if any should be duly made in the Indian National Theses Database;

- 1.3. Each university would assign Indian Standard Thesis Number (ISTN) recommended by the UGC once the final process of thesis adjudication is completed. The national committee would work-out proposition for generating unique ISTN number (s);
- 1.4. Each university would send the metadata information as above to a centralized agency to be named by the UGC; and
- 1.5. In addition to generating an online record, universities may send metadata of doctoral theses submitted to their universities in print format as well as in MS Excel Sheet / OpenOffice Calc (a column each for a field) or comma delimited file as soon as the Ph. D degree is notified.

2. Submission of E-Theses

- 2.1. Each university shall make the submission of electronic version of the doctoral thesis a mandatory requirement for all doctoral students. This may be introduced by modifying the existing Ph. D regulations of universities by including a clause on “*Submission of Theses*”;
- 2.2. The e-theses may be submitted in preferably in the PDF format. However, other formats such as MS-Word, RTF, PS, LaTeX, HTML or any other standard format may also be accepted;
- 2.3. Universities, through their libraries and computer centres, should provide facilities and assistance to researchers for converting their doctoral theses from word processing files to PDF;
- 2.4. The thesis should be formatted in a standard format as formally recommended by the university. The National Committee would evolve a model format along with mandatory fields.
- 2.5. Each university may develop naming convention for files being submitted. Submission of a thesis may be allowed in multiple file, one file each for a chapter and provision for separate files for graphics, animations and video clippings.
- 2.6. Bibliographic details of cited items should be given in a standardized format. The national committee would recommend a format based on international standards for citing printed works as well as for citing web-based electronic resources;
- 2.7. The thesis should include an abstract of not more than 200 word;
- 2.8. The printed as well as the e-thesis are to be submitted as a final version (including corrections / modifications, if any);

- 2.9. Each university will also get the Author Approval Form signed by the students for assigning non-exclusive rights to the university to archive and disseminate the thesis.
- 2.10. The metadata and e-thesis submitted by the student should be duly checked by research supervisor / trained staff members for its completeness and accuracy. Skills of Librarians may be utilized for the verification of rendering of metadata and choice of keywords assigned;
- 2.11. Each university has to evolve a copyrights policy relating to the doctoral thesis which may be included in the IPR policy of the university; and
- 2.12. The University should take necessary steps to ensure long-term presentation of e-theses. Necessary help may be obtained from the computer centre / outside experts in this regard.

3. Setting-up E-thesis Repositories

- 3.1. All universities should set-up an e-theses repository to facilitate e-submission, archiving, maintenance and access to these repository; at the University and also at a national repository at a website designated by the UGC; and
- 3.2. Universities may use OAI-complaint open source software to set-up their e-theses archives (e.g. D Space, ETD Software of VT, E-prints, etc.).

4. Retrospective Metadata of Ph.D Theses

- 4.1. With an objective to compile a comprehensive database of doctoral theses submitted to the universities in India, all universities are required to submit the metadata for all the Ph.D. theses submitted to their university from retrospect;
- 4.2. Meta data for retrospective records in a standardized format should including an abstract / summary. A format based on international standards would be adopted by the National Committee.
- 4.3. If the metadata for theses is already available with the universities, it is to be ensured that this confirms to the standard mentioned above, it is complete in all respect and is validated. The retrospective data in standardized format may be sent to a central agency to be designated by the UGC as a comma-delimited file / excel sheet.

5. Implementation

- 5.1. With the objective of facilitating and monitoring implementation of these regulations, a National Committee shall be constituted by the Chairman (UGC);
 - 5.2. The National Committee shall lay down data and technical standards, create tutorials and develop a guidebook for facilitating implementation of these regulations;
 - 5.3. The National Committee shall also develop a phased implementation plan and advise UGC and the universities on commitment of resources for successful implementation of these regulations; and
 - 5.4. The National Committee shall guide and advise the UGC and universities on all issues related or incidental to these regulations.
-

Part I

Background Information, Current Scenario, Major Issues & Data Standards

Introduction

1.1 Theses and dissertations are known to be the rich and unique source of information, often the only source for research work that does not find its way into various publication channels. Doctoral dissertations are manifestation of result of four to five years of intense work involving huge investment of resources, both mental and physical and infrastructure and other support from the universities. A thesis reflects quality of research work conducted by a student and the ability of an institution to lead and support original work of research in a given discipline. Research is characterized by originality, improvements and innovations. Scientific research, in particular, is cumulative in nature. The present research is built upon the past knowledge. The first and foremost step in research is, therefore “literature search” on the given topic. Scholarly communication system have evolved mechanisms to index past and current research publications through subject-specific bibliographic databases, such as MEDLARS in medical sciences, COMPENDEX in engineering and technology and INSPEC in physics, electrical and electronic engineering. Most of these bibliographic databases index research articles from national and international journals, conference proceedings and chapters in books. Most databases, baring one (Dissertation Abstracts International from Proquest) do not index theses and dissertations. Even the coverage of Dissertation Abstracts International is selective to those universities who have signed-up with Proquest. It does not contain bibliographic details of dissertations submitted to universities in India.

Why ETD?

1.2 The process of scrutiny, validation and approval of doctoral dissertations is confined to few experts (identified by the university on recommendation of theses supervisor). It is not open to the scientific community at large, and therefore, quality is sacrificed. The theses collection in most of the Indian libraries, are kept in closed access, making it difficult for other students to access them. It remains an un-tapped and under-utilized asset, leading to unnecessary duplication and repetition that, in effect, is the antitheses of research and wastage of huge resources, both human and financial. The UGC’s Regulatory Framework aims at evolving a mechanism to improve the quality, accessibility and availability of Indian theses and to implement uniform standards for creating metadata of doctoral theses and a system for collecting and collating this standardized data.

1.3 Electronic version of theses provides broader exposure to research students through greater accessibility. It offers opportunities to use new forms of creative scholarship through use of interactive elements, multimedia, hyperlinks, etc. It provides opportunities to research students for professional development as they learn the basic skills of scholarly publishing in electronic format. It prepares them for their future career and lead to the most expressive rendering of their discoveries and ideas. In other words, ETDs are new genre of documents that are being continuously redefined with evolution of e-publishing technology. It is expected that better type of ETDs may develop with development in e-publishing technology and their adaptation. While the simplest form of

ETDs can be thought of as “electronic paper” where the underlying authoring goal is to produce a paper form with diagrams and images in colour. Richer ETD may include links, hyperlinks from table of contents, tables, figures and indexes that are targeted to locations in body of the document. More evolved ETDs may include multimedia contents following international standards, and files including data and interactive or dynamic forms that may be harder to transform into print media.

1.4 Implementation of ETD would lead to streamline of workflow and save time and labour as checking of submissions and cataloguing of ETD would be faster, moving and handling of paper copy is eliminated and delay in binding is removed.

Present Scenario in India

1.5 There are a number of agencies that are involved in collection, compilation and presentation of metadata of theses in India. Some of the major initiatives are as follows:

Association of Indian Universities (AIU) publishes list of theses awarded in various universities in its weekly publication called “University News”. The AIU has also published a number of bibliographies on theses submitted to the Indian universities in various subject disciplines. However, coverage of University News and subject-specific bibliographies of theses published by the AIU is far from complete.

The INFLIBNET and DELNET host databases of bibliographic records of Ph.D. theses submitted to various universities in India consisting of 1,40,000 and 3,953 bibliographic records respectively. Vidyanidhi, a nation-wide effort on theses and dissertation, currently supported by the Ford Foundation and Microsoft India, hosts more than 500 theses in full-text and 85,000 bibliographic records of theses submitted to the universities in India. Vidyanidhi is a member of the Networked Digital Library of Theses and Dissertations (NDLTD), a global initiative with more than 174 members from different countries of the world. ETD@iisc is another initiative that hosts about 86 theses. It provides guidelines for submission, document conversion guidance, theses templates and sets the workflow for online theses submission. The National Social Science Documentation Centre (NASSDOC) a wing of the Indian Council of Social Science Research (ICSSR), hosts a library for the social scientists with Ph.D. theses in social sciences as its core collection. As a national institution, NASSDOC aims to build a truly representative collection of Ph.D. theses in each of the areas that make up the social sciences. The NASSDOC systematically acquires Ph.D. theses in social sciences submitted to the Indian universities. Currently, the NASSDOC has 4924 Ph.D. theses in social sciences in its collection.

1.6 In spite of a number of sporadic efforts mentioned above, India neither has a comprehensive and authentic source of information nor a mechanism to obtain information on all Ph.D. theses submitted to the universities in India. The situation calls

for a regulatory framework essentially to create a National Database of Theses and Dissertations for India as well as to initiate the process of electronic submission of theses and dissertations in universities.

International Scenario

Networked Digital Library of Theses and Dissertation (NDLTD)

1.7 The movement on electronic theses and dissertation is led world-wide by the Networked Digital Library of Theses and Dissertation (NDLTD) (<http://www.ndltd.org/>) initiative, taken-up by the Virginia Tech University, USA. It is an open federation consisting of currently 174 member universities and research institutions from all over the world. NDLTD's activities are focused on universities, libraries, faculties and research students in order to support authoring, indexing, archiving, dissemination and retrieval of electronic theses and dissertations worldwide.

1.8 NDLTD's vision is to increase the availability of student research documents for scholars and to preserve it electronically and to empower students to convey a richer message through the use of multimedia and hypermedia technologies. NDLTD encourages and supports universities to unlock their information resources and to advance digital library technology by sharing of experiences, tools, technology and knowledge.

UNESCO's Guide to ETD

1.9 The UNESCO has launched a project for development of an international framework for creation of electronic theses and dissertations (ETDs). The project aims to contribute to enhancing the production, access and archiving of scientific information by using the possibilities of new technologies. The specific objectives of the UNESCO project are:

- Establishing and disseminating guidelines, workflow models and best practices;
- Establishing a model training programme for project managers responsible for ETD programmes;
- Carrying out training courses and pilot projects

1.10 With the objectives mentioned above, the UNESCO has developed a “Guide to Electronic Theses and Dissertations (ETDs)”, a resource targeted to the research students who are writing theses or dissertations, for faculty who want to mentor ETD authors, for research administrators who want to initiate ETD programmes, and for IT administrators at universities. The Guide promotes the sharing of knowledge locked up in universities, and the collaboration of universities worldwide is designed specifically for academic researchers and their mentors.

1.11 Some of the major ETD initiatives taken-up the world over are given below:

Australian Digital Theses Programme (ADT) (<http://adt.caul.edu.au/>): ADT Programme is a collaborative initiative involving university libraries in Australia. The initiative is voluntary, any Australian university can join the programme. ADT software, available free to members in the programme, was designed to be transportable and flexible which can be installed at each member institution with minimum modification.

Cyber-theses (<http://www.cybertheses.org/>)

The Cybertheses started as a cooperative project amongst selected universities in France with an aim to publish and distribute electronic theses on the web. Cybertheses is now open to all institutions of higher learning in France. It allows theses to be indexed online using common metadata model. The Cybertheses database contain the metadata of theses from participating institutions. It provides an efficient indexing system and rapid searching, even while significantly increasing the visibility and the distribution of the theses.

Dissertation Online

In German, “Dissertation Online” project, funded by the German Research Foundation, is a multi-site, multidisciplinary project that involves several educational and research institutions in Germany. The project aims at bringing scholarly publications, such as dissertations, diploma and master theses online. The project is highly successful and works closely with the NDLTD. A Bureau of Coordination has been established in the German National Library (DDB) to coordinate all developments of “DissOnline.de”.

Other important ETD Projects

- MI (University Microfilms International) (<http://wwwlib.umi.com/>)
- National Library of Canada (<http://www.nic-bnc.ca/index-e.html>)
- TUG Electronic Theses Project (<http://www.lib.lutwaterloo.ca/TUG/ETD>)
- JETD Project (<http://www.fics.utoronto.ca/etd>)
- Digitale Dissertationen of Humboldt University at Berlin: (<http://www.edoc.hu-berlin.de>)
- University Lumiere Lyon 2 Digital Theses Project (<http://www.univ-lycon2.fr/search.html>)

Major Issues

Plagiarism

1.12 The risk of plagiarism is one of the important concerns that most students and faculty have. Although plagiarism cannot be ruled out even in print environment, it, however, cannot be denied that availability of documents in electronic format makes it easier for authors to copy. However, the risks of exposure of plagiarism is much larger in a scenario where theses are available in electronic format publicly given the fact that

most scholars and researchers still work in fields where a fairly small group of workers have detailed knowledge of their work. Moreover, the technology that has made ETD possible, also provides mechanism to detect plagiarized passages in electronic documents. Several software packages have now been developed that detect plagiarism. The software examines document files submitted for detection of plagiarism. It extracts the text portions from these documents and looks through them for matching words in phrases of a specified minimum length. When it finds matching files that share enough words in a number of phrases, a report is generated which contain the document text with the matching phrases underlined. Widespread use of such packages would increase risks of detection and, therefore, plagiarism.

1.13 Moreover, since ETDs are read more often than printed theses, there is a strong psychological pressure to discourage plagiarism. While on one hand, students would be more careful about consequences of plagiarism once detected, the associated faculty are likely to be more diligent than with paper works at the time of checking the validity and quality of results reported. In short, detection by machines or other users, and threat of severe penalties are likely to discourage students from considering plagiarism with regards to ETD.

Intellectual Property Right (IPR) and Copyright

1.14 The owner of copyright of a book or any other written document belongs to its creator or author, irrespective of media used for its presentations, i.e. paper or electronic. The author of an electronic thesis or dissertation is its copyright holder and thus owns the intellectual property contained in it. It is for author to decide how their works will be reproduced, modified, distributed, performed in public or displayed in public. However, an author may use another author's work with certain restrictions known as "fair use". The owner of ETD, i.e. a research student must agree in writing to host his / her thesis on the web with or without restrictions. Such declaration / undertaking is taken from the student at the time of submission of thesis on Student Approval Form (Annexure I).

1.15 Such a declaration gives a university or institution non-exclusive rights to archive and host an electronic theses on the institute ETD repository. It does not take away the right of an author to use intellectual contents of his / her theses for writing papers, books or taking patents, etc.

Metadata

1.16 Metadata is the term used to describe data about data. The primary function of metadata is to facilitate information access, search and retrieval. To achieve this goal, the metadata provides information known about the document, such as its title, creator (author), publisher, and date of publication, etc. in order to facilitate access, search and retrieval of document. It usually includes information about the intellectual content of the document (i.e. subject keywords or descriptors), digital representation data, and security or rights management information.

1.17 Besides providing access to intellectual contents of a document, a function analogous to bibliographic records, digital objects also require metadata about applications and formats used for creating a digital object. Such metadata is required to provide long-term access to a digital resource. In short, the following three types of metadata are associated with the digital objects:

- Descriptive Metadata: Include content or bibliographic description consisting of keywords and subject descriptors.
- Administrative or technical Metadata: Incorporates details on original source, date of creation, version of digital object, file format used, compression technology used, object relationship, etc. Administrative data may reside within or outside the digital object and is required for long-term collection management to ensure longevity of digital collection.
- Structural Metadata: Elements within digital objects that facilitate navigation, e.g. table of contents, index at issue level or volume level, page turning in an electronic book, etc.

1.18 Meta data support efficient and effective organization, access and retrieval of information contents in a digital library. Meta data is used in effective designing of browsing and search interfaces of a digital library. With attributes of a digital objects defined in the metadata, it is a simple task to organize digital objects into predefined categories specified in search / browsing interfaces.

Metadata Schemes

1.19 Institutions dealing with electronic theses and dissertations have either developed their own standards or adapted existing metadata standards. These metadata standards attempt to describe the author, the work, and the context in which the work was produced in a way that will be useful to the researcher as well as the librarians and / or technical staff maintaining the work in its electronic form.

1.20 There are quite a few metadata schemes. Some of these schemes are applicable to documents received in a library, others have broader scope. Some of the important metadata schemes are as follows:

- Machine Readable Catalogue (MARC)
Most traditional library systems exchange and store records using MARC format. MARC format has approximately 1,000 fields, several with repeatable sub-field. The use of this format allows very detailed description of the items. The MARC records for theses is not very robust, it often requires manipulation of fields to accommodate variations that are specific to theses as a type of document. Moreover, there are several fields in MARC record that are not applicable to theses and dissertations. Although, the MARC records have to be used for cataloguing theses and dissertations in a library, it is not designed for the ETDs.

- Dublin Core Metadata Elements Sets (DCMES)

Dublin Core is a set of 15 attributes divided into three groups, i.e. content, intellectual property and instantiation. Associate to Dublin core are Dublin Core qualifiers that enhance the identification of items. Annexure III provides qualified Dublin Core Metadata elements for an ETD.

Most of the institutional repositories use Unqualified Dublin Core (www.dublincore.org) metadata to ensure interoperability. Since OAI is based on the exchange of metadata, getting the metadata right is fundamentally important for a repository. The OAI compliant software automatically produce the necessary Dublin Core metadata for harvesting by service providers.

- ETD-MS

The ETD-MS standard is developed by the Networked Digital Library for Theses and Dissertations (NDLTD) which is used for submission of electronic theses and dissertations at Virginia Tech University. ETD-MS incorporates basic Dublin Core elements with the addition of several elements that further describe parts of the ETD process. The ETD-MS is the only metadata standard supported by the software. Annexure IV provides guidelines for information contents for each element of ETD-MS.

Standards for Metadata Harvesting

Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)

1.21 The Open Archives Initiative was originally proposed to enhance access to e-print / pre-print archives. Gradually, the scope of the initiative has broadened to cover any kind of digital content including images and videos. The OAI-PMH is a protocol devised to make machine-readable metadata widely available for use. The development of the OAI-PMH protocol took root in a meeting that was convened in late 1999 at Santa Fe, New Mexico to address problems of the e-print world. As disciplinary e-print servers became more common, it was difficult to support searching across multiple repositories. Repositories needed greater capabilities to automatically identify and access papers that had been deposited in other repositories. Need was felt to build a framework to bring about a kind of integration of these e-print/pre-print archives to solve these problems. The major work was to define an interface to permit e-print servers to expose their metadata for the papers it held, so that search services or other similar repositories could then harvest its metadata. These archives would then act as a federation of repositories by giving a single search platform for multiple collections.

Metadata standards and OAI-PMH

1.22 For the purpose of interoperability, the OAI Protocol for Metadata Harvesting specifies unqualified Dublin Core, encoded in XML, a mandatory metadata schema as the lowest common denominator. Almost any metadata scheme can be "downgraded" into unqualified Dublin Core. However, each server is also free to offer metadata in one or more schemas, and a harvester can request that metadata in any format in addition to the unqualified Dublin Core.

The OAI-PMH Framework

1.23 There are two classes of participants in the OAI-PMH framework:

- **Data Providers:** Data Providers, or repositories, administer systems that support the OAI-PMH as a means for exposing their metadata. All universities would serve as data providers once they have set-up their OAI-complaint ETD repositories.
- **Service Providers:** Service Providers, or harvesters, use metadata harvested via the OAI-PMH as a basis for building value-added services, such as building subject gateways, email alerts, etc. The central agency designated to maintain National Theses Database would work as service provider for ETD in India.

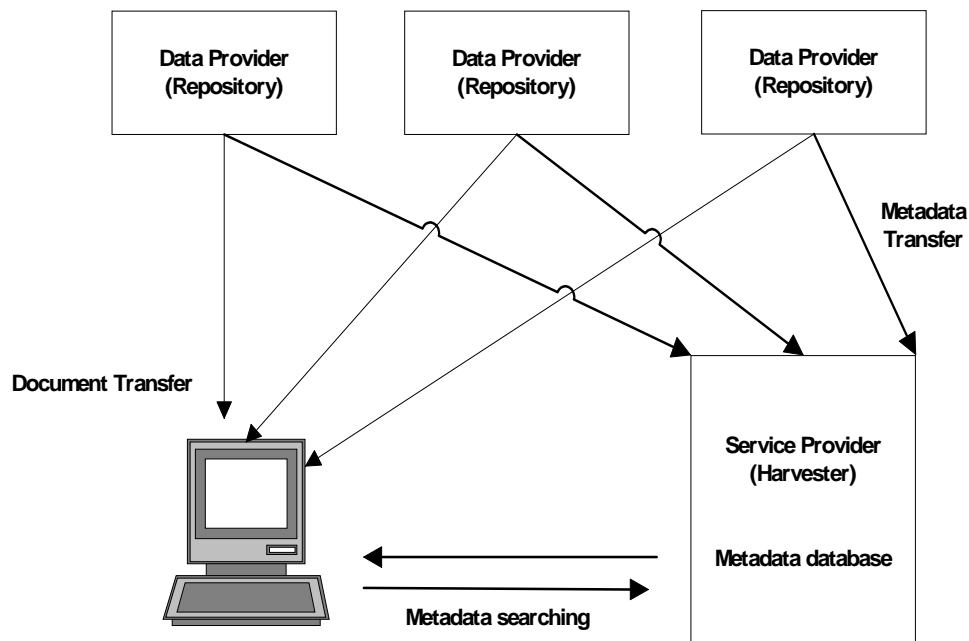


Figure 1: The OAI-PMH Architecture

1.24 The metadata stored in the data providers' database is transferred in bulk to the metadata database of the service providers. The transfer of metadata is done in a series of requests and responses between the data provider and the service provider/harvester. The

OAI-PMH Protocol depends upon the HTTP-transaction framework for communication between a harvester and a repository. Requests may be made using either the HTTP GET or POST methods. All successful replies are encoded in XML, and all exception and flow-control replies are indicated by HTTP status codes.

Standards for Data Format for Theses

1.25 The preparation of an ETD involves making of the electronic copy of the thesis / dissertation. There are many file formats available for text, image, sound and video. Basically, file formats can be proprietary or non-proprietary i.e. open file formats. MS-Word is an example of proprietary file format and OpenOffice.org is an example of non-proprietary file format.

1.26 Open standard file formats are strongly recommended to upload documents in the ETD. Using open standard file formats is very advantageous in the long run. The reason is proprietary standards require proprietary software which may not be accessible to everyone. Proprietary standards may not be backward compliant i.e. file formats created using an older version of the software is usually not readable by the newer versions. In the archival perspective it is feared that if the proprietary software become obsolete and go out of use there is no way one can retrieve the documents that have been written using these software. Being proprietary standards, nobody can even attempt to write programs to read these older versions of the file formats. Point in case is the WordStar document format of Microsoft. The software has become obsolete and is not available. And the current versions of MS-Office cannot read these files. Whereas, for open standards anyone can write a program to retrieve the document even if the software becomes obsolete. In cases where use of proprietary software is unavoidable, it should be urged to include the required software with the ETD file so that the user does not require to purchase additional software. Some of the open standard document formats are described below:

Open Text File Formats

Hypertext Mark-up Language (HTML)

1.27 HTML is the language with which Web pages are designed. This standard has been defined by the World Wide Web Consortium (W3C). HTML allows web documents to be created with ease. The primary objective of using HTML is to build a web page that communicates readily and effectively to make the document on the web most compelling to access and read. HTML is a plain text file and any text editor as simple as Notepad can be used to create HTML documents.

eXtensible Markup Language (XML)

1.28 XML provides a structured representation of data that can be implemented broadly and is easy to deploy. XML is a subset of SGML (Standard Generalized Markup Language), modified and optimized for delivery over the Web. This standard has been

defined by the World Wide Web Consortium (W3C). XML can be used to format and transfer data in an easy and consistent way. XML is more flexible, because one can define his/her own tags/elements. Hence it is possible to tailor the XML documents for different needs, and makes it possible to use XML to represent all kind of data for different purposes. XML is also a plain text format.

1.29 The advantages of using XML-based applications are that they can be implemented and used irrespective of the device and platform being used. XML is device independent and platform independent. The data encoded in XML can be accessed irrespective of the device i.e. on a wireless handset, a palmtop, a laptop, an airport kiosk, a projector in a conference room, or a desktop PC or operating system, i.e. Windows, Unix, Linux, Sun Solaris, etc.

Portable Document Format (PDF)

1.30 Invented by Adobe Systems, Adobe Portable Document Format (PDF) is a publicly available specification used by various standards bodies around the world for electronic document distribution and exchange. As an open file format specification, PDF is available to anyone who wants to develop tools to create, view, or manage PDF documents.

1.31 The most popular and preferred format is PDF (Portable Document Format). Adobe Acrobat's Portable Document Format (.pdf) is recommended, since it retains all format codes and graphic images, appearing as the original paper document and also because it is easily portable. In addition .pdf files can be indexed and searched by keywords. Apart from being an open standard, it maintains the integrity of the document. It can be converted to PostScript format, which can be used for electronic delivery and printed directly. Both Microsoft Word and WordPerfect files can be easily converted to .pdf files. Training and assistance in the conversion process to .pdf should be provided to the students.

TeX

1.32 TeX is a typesetting program designed for high-quality composition of material that contains a lot of mathematical and technical expression. (<http://www.tug.org/tex-pr-faq>) It has been adopted by many authors and publishers who generate technical books and papers. It was created by Professor Donald Knuth of Stanford University, originally for preparation of his book series "The Art of Computer Programming". TeX has been made freely available by Knuth in a generic form.

1.33 TeX implementations are governed by the principle that the same input should produce the same output, modulo font availability and output device resolution. All implementations of TeX must pass a "trip test" that assures adherence to these guidelines.

1.34 TeX has been tailored for and installed on almost every platform (computer + operating system), and is available as freeware, shareware and commercial

implementations. The TeX program is usually accompanied by other software to form a complete and usable system.

LaTeX (Lamport TeX)

1.35 LaTeX is a document preparation system for high-quality typesetting (<http://www.latex-project.org>). It is used for technical or scientific documents, but it can be used for almost any form of publishing. LaTeX is based on Donald E. Knuth's TeX typesetting language. LaTeX was first developed in 1985 by Leslie Lamport, and is now being maintained and developed by the LaTeX3 Project. LaTeX is available for free at <http://www.latex-project.org/ftp.html>.

Open Image File Formats

Portable Network Graphics (PNG)

1.36 PNG (pronounced 'ping'), the Portable Network Graphics file format, is an open raster image format. It is supported by the W3C and IETF and is expected to be released as ISO/IEC International Standard 15948. The latest version is PNG 1.2. It was developed in 1995 as a replacement for the GIF (GIF89a - Graphics Interchange Format) and a possible replacement for the TIFF (Tagged Image File Format). It is still not widely used and it has taken some time for Web browsers and image application software to support it. Now, PNG files have reasonable support among the leading browsers and can be created and manipulated within many image applications.

Joint Photographic Experts Group (JPEG)

1.37 The JPEG compression and its corresponding file format were developed in the late 1980s by independent members of the Joint Photographic Experts Group (JPEG). JPEG properly refers to the compression, with the file format officially termed JFIF (JPEG File Interchange Format). However, the format has become commonly known as JPEG and is usually given .jpg and .jpeg extensions. It is the most popular image format for the Web.

Poor Standard of Theses

1.38 Several researchers, research supervisors and education administrators would be reluctant to join ETD movement knowing the fact that research work being conducted in their institutions are of poor quality. We should take this as an opportunity to improve our research work. Considering the fact that ETDs are read more often than printed theses, there will be a strong psychological pressure on the research students as well as on the research supervisor to improve the quality of their research work. While on one hand, students would be more careful about the quality of their work, the associated faculty are likely to be more diligent than with paper works at the time of checking the validity and quality of results reported. Moreover, qualitative research work available on the Internet from other universities would give an idea to researchers as to how they can improve their work.

Part II

Implementation Process

Creation of Metadata for Current Theses

2.1 The metadata of theses and dissertations 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 students 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.

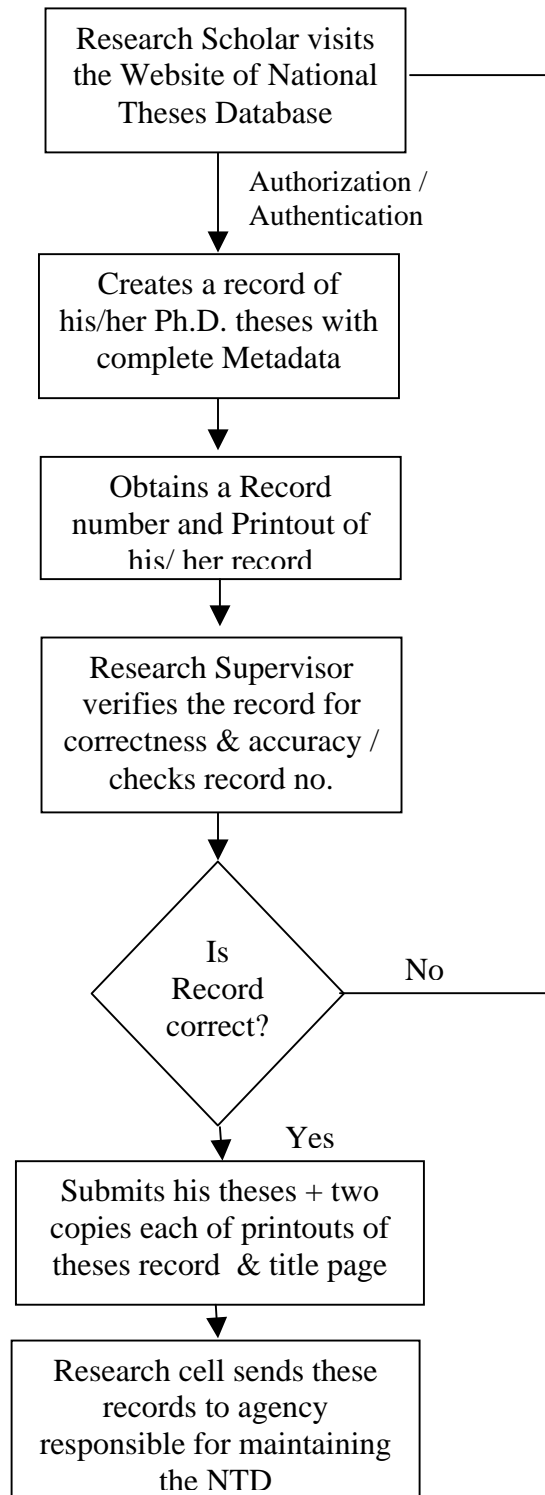


Figure 2: Flowchart for Creation of Metadata for Current Theses

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 or 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.

□

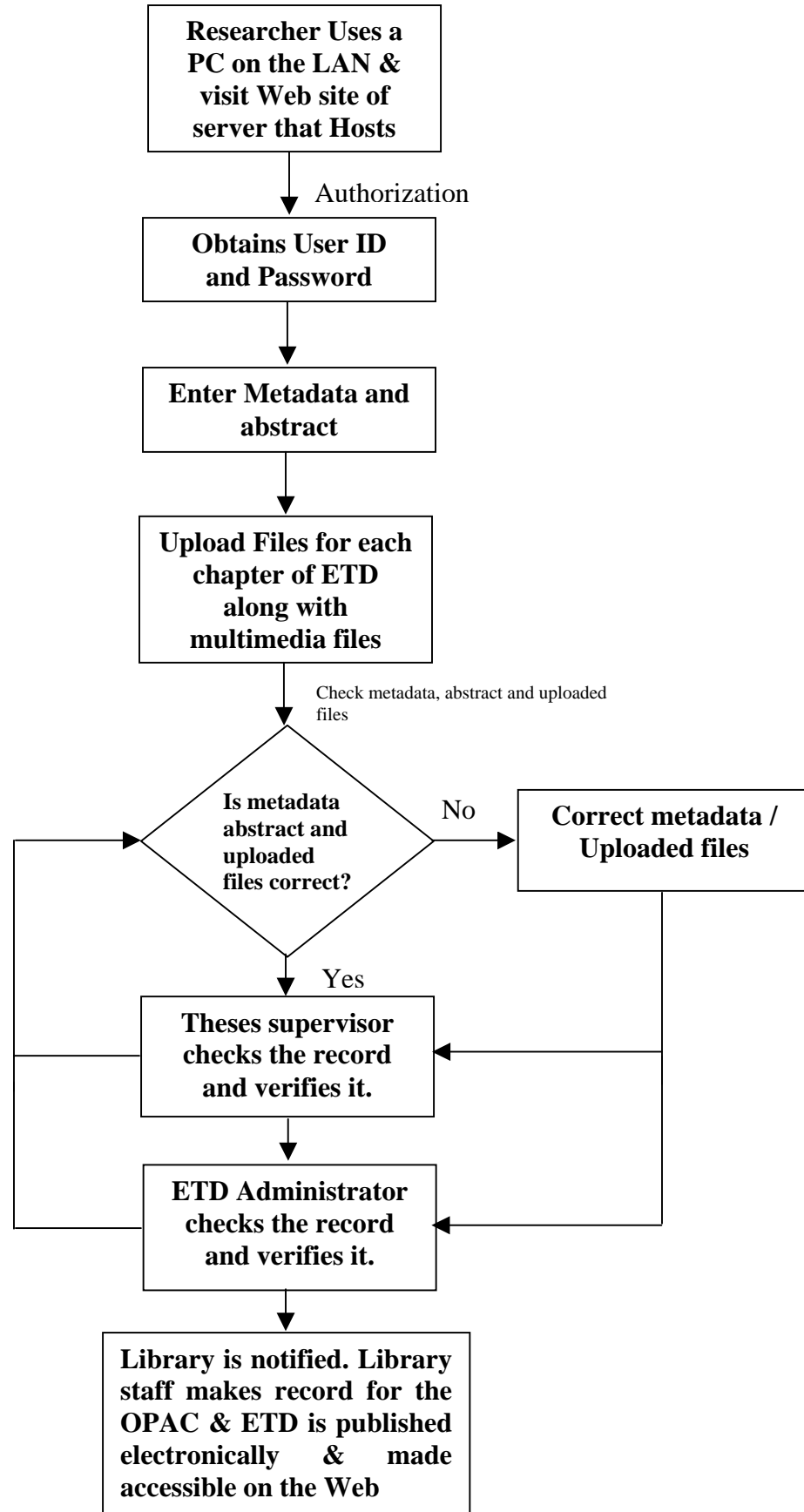


Figure 3: Electronic Submission of Theses and Dissertation

- **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 PostgreSQL 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/mailling.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 is 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 conforms 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
-