

Challenges of Current and Retrospective Digital Archiving: A Step by Step Guide to Best Practices of ETDs

Tamal Kumar Guha

Lakshminath Bezbaroa Central Library, IIT Guwahati, Guwahati, Assam, India

Introduction

Today, creating digital archives has become customary for any library servicing an academic institute. In fact, like developing books and journal collections, creating repositories of their own digital collection is now standard practice for any higher education libraries. Hence, explaining how to create a repository for this august gathering would be redundant. Indeed, the technical components, like which operating software is to be used or which software is to be used for hosting or the bandwidth requirements for hosting an archive, are well known to all professional colleagues. However, while creating a repository, a librarian's most formidable challenge is not technical but primarily administrative, legal and workflow-related challenges. While some might be encountering the problem of retro-converting the existing collection, some face challenges keeping track of the in and out flow of the collection.

The primary purpose of this write-up is to share our experiences with professional colleagues and explain what we are doing and why we are doing what we are doing. Without hesitance, many might have come up with their own solutions, which might be suitable within their working framework, and this discourse does not intend to doubt their efficacies. While building up an Institutional Repository as such could have several challenges, the subsequent part of this discussion would remain restricted to creating a digital repository of electronic theses and dissertations.

Initiation of workflow:

At the outset, it is worth emphasising that while setting up a digital repository, the utmost care should be taken about the copyright issues of the contents to be hosted on the platform. It is essential to ensure that the researcher grants a non-exclusive, worldwide, irrevocable, royalty-free license to use, reproduce, translate without changing the content, publish, archive, preserve, communicate and distribute, in paper form, in microform, electronically by telecommunication or on the internet, and/or any other formats as may be adopted for such use from time to time to the educational Institute where s/he belongs.

Further, researchers often face challenges related to similarity reports encountered by publishers of their in-progress publications. To be more specific, researchers often prepare their dissertations and submit articles for publications borne out of the same research work simultaneously. While submitting and getting a degree awarded takes only a few months, it often takes about 18 months to publish such articles due to the processes involved in the publication industry. Suppose the full-text theses are made available publicly through institutional repositories before the completion of the editorial processes of publishers. In that

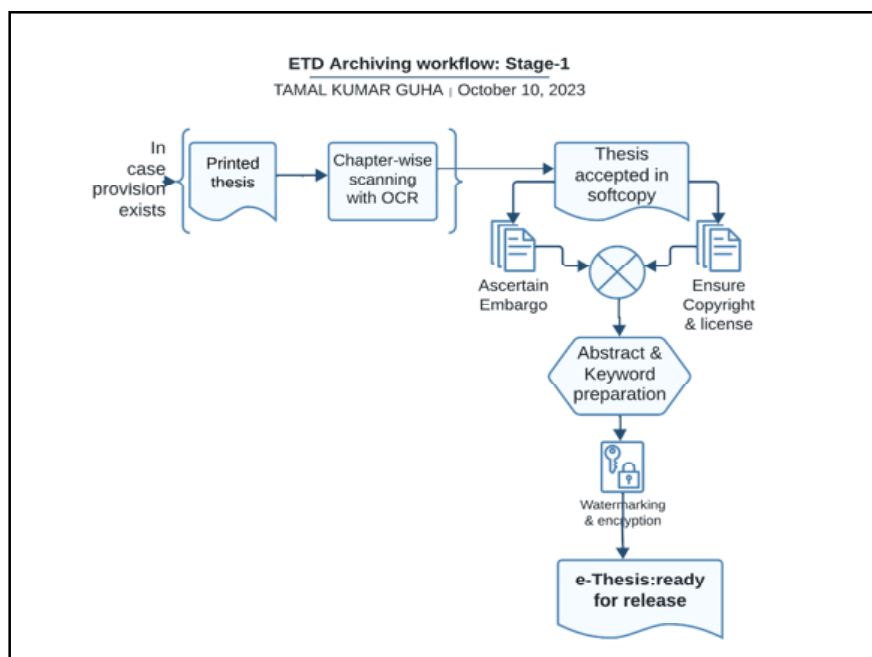
Corresponding Author: Dr. Tamal Kumar Guha, E-mail:librarianiitg.ac.in

case, it leads to a high similarity index at the end of the editorial board. As a result, publications often get denials from the editorial boards. Similar issues have been observed when a specific portion of a research work is filed for patenting, which usually creates more complexity due to several legal reasons. Hence, constant and consistent workflow monitoring is of utmost importance while uploading and making dissertations publicly visible.

As all of us know, the misuse of intellectual property is rampant, especially after affordable reach to the internet to most of us. We often observe that several portions of text and other intellectual property are reused by many without proper acknowledgement, and to prevent such losses, practices of this nature of society need to be addressed technically.

We at IIT Guwahati are fortunate enough that almost all of our theses and dissertations are digitally born due to very appropriate measures adopted by the Institute as early as 1999. In fact, from the beginning, the Institute adopted measures so that it was mandatory for the researchers to submit their theses both in print and electronic format (at that time, the softcopy theses had to be submitted in CD-ROM). Though the purpose of such an electronic copy of a thesis was to make a backup copy available for eventual loss due to unforeseen reasons, eventually, by 2014, when we started our digital repository and archiving those theses, we started with the policy of ensuring copyrights, an embargo for a limited time (if needed) and also protecting the contents from possible misuses.

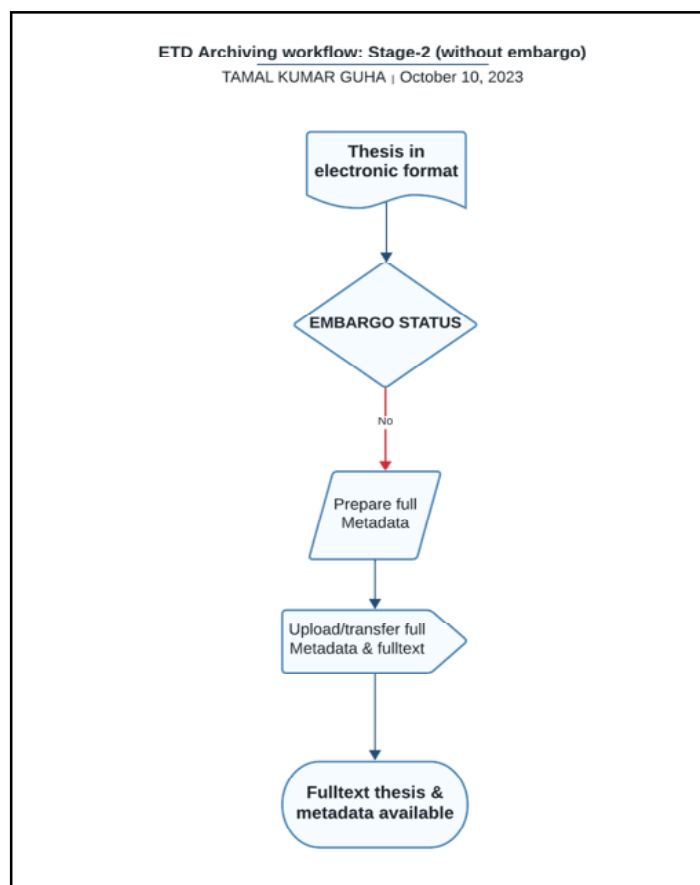
Following is an illustration of the workflow we adopted at the initial stage and has been followed since then while accepting and preparing any thesis before uploading to the archive:



Though we did not have any printed collection of theses that needed to be scanned (barring a couple of cases where CDROMs were damaged), the scanning loop has been kept deliberately to address the issues, where theses and dissertation inflow are in printed versions only.

Workflow of theses without embargo:

The process of releasing theses without embargo is usually straightforward; we need to prepare the full metadata, ensure that full-text theses are available for uploading in acceptable file format and then upload/transfer the entire metadata set along with the softcopy files. The following workflow narrates the steps which can be used while curating the metadata and uploading theses:

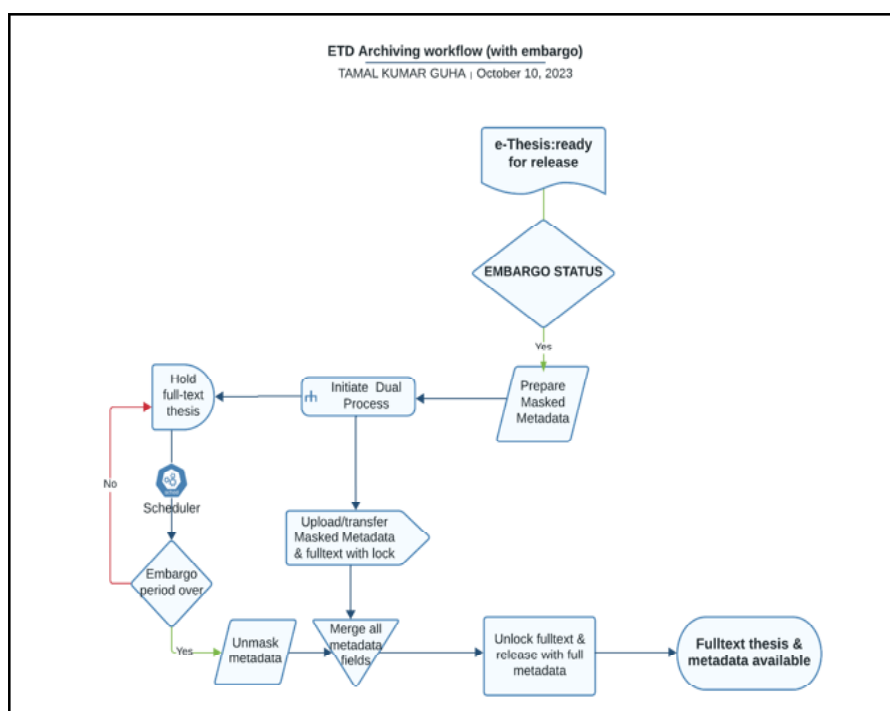


The only challenge one might encounter is when too many theses are received in one instance and must be released in one go. Keeping track of all theses and appropriate metadata mapping outflow needs proper attention and due diligence. To mitigate such challenges, over time, librarians can gain experience and devise their monitoring mechanism. This process is primarily cumbersome when large-scale retro-conversion is essentially, especially with the libraries with extensive backlog collections to bring on achieving platforms.

Workflow for theses with embargo:

Over time, it has been observed that workflow for this set of dissertations is usually more complex and involves iterating processes. Once it is earmarked as ‘under embargo’, a thesis needs to be scheduled for its release. Upon deciding on the release date, two sets of metadata shall be created, where the initial set only discloses a limited number of fields, like the author, supervisor(s), subject headings, etc. While the initial set can be disclosed without delay, the full-text thesis, along with the exhaustive metadata, consisting of title, abstract, etc. shall be kept under embargo till the date of release.

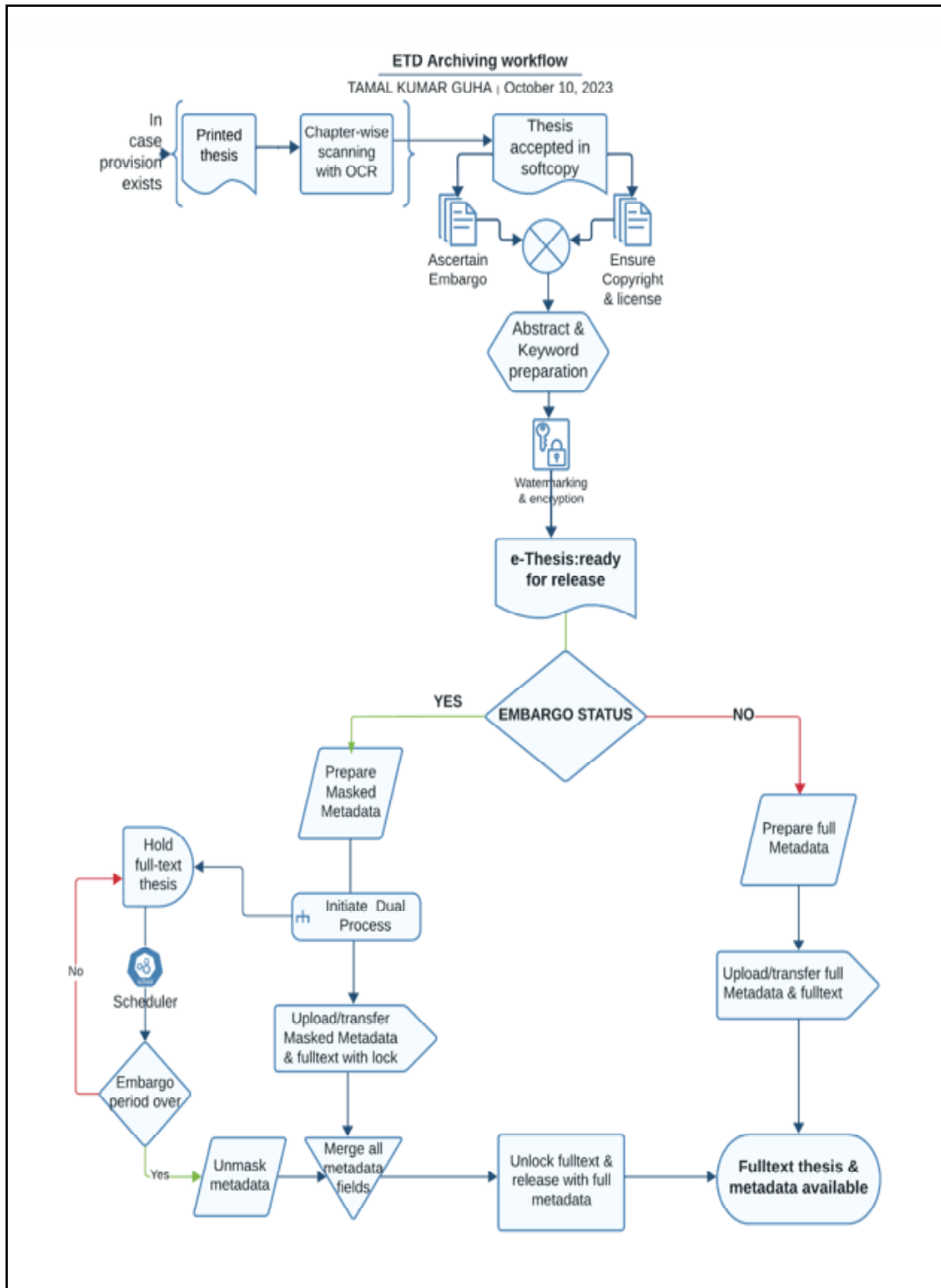
The entire workflow process for such set of theses has been summarised in the following flow chart:



The second workflow for this set of documents would start right after the embargo period has ended. At that stage, we need to release the full-text content, full abstract, full title, and any other field(s) suppressed initially. The workflow might also need an action to merge the initial metadata with the full metadata and follow the action process engaged in case of theses without-embargo.

Segregating theses without embargo and with embargo at the same workflow process, especially when inflow is large, seems to be a critical component. A meticulous process of iteration needs to be followed so that the researchers do not feel disabled due to unwanted and unexpected early release of content, causing serious concerns about high similarity reports from the editorial board, where they might have submitted article(s) for publication(s). For the convenience of the readers, a summed-up flowchart has been prepared and presented below, which brings all the loops in one place.

CHALLENGES OF CURRENT AND RETROSPECTIVE DIGITAL ARCHIVING: A STEP BY STEP GUIDE TO BEST PRACTICES OF ETDS



Conclusion

It might seem daunting, especially when a library attempts to retro-convert its existing collection and bring forward the current literature on the achieving platform. However, it becomes more manageable if planned meticulously, monitored consistently and followed all the subroutines in the workflow. In the libraries, where the provision of the embargo (withholding) does not exist, building up an ETD collection should be a manageable task, provided dedicated efforts are put forward. However, considering that a substantial number of researchers are publishing their ongoing research reports, it is prudent to provide appropriate measures so that the archive grows smoothly and the researchers do not suffer due to the complexity involved in the publication industry.

As this is not a research paper but a narration of the workflow being followed in my institute library, this write-up does not have any reference or bibliography. I certify that I am the sole author of the above write-up and it does not infringe any rights of others.