

Theses e-submission tool at the National Technical University of Athens

Alexandros Koulouris^{1,3} and Angelos Anagnostopoulos^{2,3}

¹Department of Librarianship and Information Systems, Technological Educational Institute (TEI) of Athens, Agiou Spyridonos Str., 12210 Egaleo, Athens, Greece, koulouris.a@gmail.com, akoul@teiath.gr

²Converge ICT Solutions & Services SA, 74, Panormou Str., 11523 Athens, Greece, angelosanagnostopoulos@gmail.com

³Central Library, National Technical University of Athens, 9 Heroon Polytechniou Str., 15780 Polytechniupoli Zografou, Athens, Greece

Short biography

Dr. Alexandros Koulouris holds a Ph.D. and a B.Sc. in Library and Information Science (LIS) from Ionian University and the Technological Educational Institute (TEI) of Athens respectively, and a B.Sc. in International and European Studies from Panteion University. He works as a librarian at the Ministry of Economy and Finance, Hellenic Republic. He is research fellow of the Librarianship and Information Systems Department at the TEI of Athens and member of the Laboratory on Digital Libraries and Electronic Publishing at Ionian University. He has worked as developer and administrator of the National Technical University of Athens' digital library and as a head librarian at the Special Collections Department, Science and Technology Library, National Documentation Centre of Greece.

Angelos Anagnostopoulos is a graduate of the National Technical University of Athens and holds a degree in Electrical and Computer Engineering. He specializes in software engineering, using open source technologies. In the past he has worked both as an software developer for private enterprises and the academic sector, namely the NTUA's Central Library. Currently, he holds the position of senior software developer at Converge ICT Solutions & Services, a PRC Group member company.

Structured abstract

Research paper

Purpose

The added value that the theses e-submission system (tool) gave to the NTUA IR service is analyzed.

Design/methodology/approach

The tool improved the IR service, which can be certified by the continuing evaluation process.

Findings

This process produced the following main results: the automated batch importing feature should be integrated into the e-submission system; IR service tutorials should be given to the secretariats, faculty and students; in-house service marketing should be conducted.

Research limitations/implications

The present study provides a starting-point for further research in DL open source software.

Originality/value

The tool helped NTUA users in submitting their e-theses, and the IR staff in organizing efficiently the service workflow and the data integration into the DSpace system. The process weaknesses rectifications will enhance the service and allow it's broadness to all NTUA departments. This tool implementation may start a debate to the DSpace community in using proxies between DSpace platform and the end user, as an alternative.

Keywords: Institutional repositories, Evaluation, Theses e-submission system, DSpace platform, Decision making

Abstract

The added value that the theses e-submission system (tool) gave to the NTUA IR service is analyzed. The tool improved the IR service, which can be certified by the continuing evaluation process. This process produced the following main results: the automated batch importing feature should be integrated into the e-submission system; IR service tutorials should be given to the secretariats, faculty and students; in-house service marketing should be conducted. The present study provides a starting-point for further research in DL open source software. The tool helped NTUA users in submitting their e-theses, and the IR staff in organizing efficiently the service workflow and the data integration into the DSpace system. The process weaknesses rectifications will enhance the service and allow it's broadness to all NTUA departments. This tool implementation may start a debate to the DSpace community in using proxies between DSpace platform and the end user, as an alternative.

Introduction

The *National Technical University of Athens* (NTUA) is the oldest and most prestigious educational institution of Greece in the technological field. The NTUA Central Library, which was founded in 1916, serves the NTUA faculty, staff and students, as well as researchers, scientists and students from other universities. Since July 2007, the NTUA, through its Central Library, offers a new open access *Institutional Repository* (IR) (NTUA Central Library, 2007a) that contains *Electronic Theses and Dissertations* (ETDs) combined with an e-submission theses service (system, tool) (NTUA Central Library, 2007b). It is based on a DSpace installation, using *Dublin Core* (DC) descriptive metadata and *Open Access Initiative Protocol for Metadata Harvesting* (OAI-PMH) for interoperability reasons.

This paper focuses on how the e-submission tool (NTUA Central Library, 2007b) helped NTUA users in submitting their electronic theses and the IR staff in improving the service workflow and the data integration into the DSpace system. In addition, it emphasizes on the added value that the tool gave to the IR service. This can be certified by the continuing evaluation process results, which part of them are emphasized. Finally, conclusion and future thoughts on the e-submission process improvements are presented.

Initiative

The initiative started from the Central Library's vision in self-archiving and open access concept (movement). This vision was shared and accepted officially by NTUA administration, which with its Senate's decision (February 24, 2006). This decision, establishes the procedure of theses submission in electronic (digital) format and gives the right to the Central Library to collect and deposit the e-theses and the privilege to define the submission procedure specifications. In this context, Central Library has the opportunity to refine the necessary specifications, standards, and procedures for writing, submission, storage, deposit and access of e-theses.

The dilemma

The theses e-submission is an innovating procedure for the Greek Universities. NTUA is among the three universities (e.g. Aristotle University of Thessaloniki and Aegean University) that apply e-submission for e-theses. The Central Library is officially the NTUA department responsible to organize efficiently the procedure and to establish new routines for the academic community regarding e-submission and self-archiving.

The major dilemma that the IR staff faced concerned e-theses submission. There were two options: through DSpace, which is the system that support IR, or through a proxy between DSpace (the NTUA digital library software) and the end user. The IR staff decided, after a lot of research, that the NTUA digital library's DSpace v.1.3.1 system, is not suitable for content submission (i.e., e-theses) to non familiar users (e.g. students). Based on that decision the IR staff decided to develop from scratch a theses e-submission system (NTUA Central Library, 2007b), which acts as a proxy between DSpace and the end user. The theses e-submission tool, its technical characteristics, its initialization and its evaluation is analyzed below.

Theses submission procedure in printed and electronic form

At the time of writing, three (out of eight) university departments are participating in the IR pilot testing period. Students of these three departments are demanded to submit their e-theses through a web application (*the theses e-submission tool*) in PDF format. Subsequently, the associated personnel perform the necessary checks and corrections (e.g. spelling mistakes) and upload the metadata and the PDF file to the IR (DSpace). There are approximately 550 e-theses stored in the IR, the majority of them being diploma theses, providing full text worldwide access. The NTUA IR is indexed in the *Directory of Open Access Repositories* (DOAR) and in the *Registry of Open Access Repositories* (ROAR); additionally in the *Greek Digital Resources Index*.

The e-theses submission system comprises of a web application, a web-based form in which the users submit the metadata and the theses file in PDF format. The IR staff developed guidelines for the theses submission in printed and in electronic form. Especially for the theses e-submission via the web-based form, online help is provided in every step. The user interface, the guidelines, the form and the online help are provided, for the time being, in Greek and in English. The form fields have been selected to be compatible with the Dublin Core metadata element set that is used in the IR's DSpace application profile. This was decided, because the IR staff is planning, certain improvements, such as the transition from a semi-automated

metadata importing process into the main IR (DSpace), to a fully automated one (batch); this need arose from the results on the workflow evaluation.

The guidelines and the procedure for the printed and electronic theses submission is analyzed. The thesis should be completed before submitting the final electronic and printed copy. The thesis printed submission comes after the electronic one. NTUA Central Library does not accept any printed theses unless the e-submission, via the online e-submission form, has been completed.

In the first step of the thesis e-submission, users (students) define the desired access level (policy) based on the copyright and policy statement. NTUA Central Library abides by the following access policies that based on international practice and bibliography: a) worldwide (free) access – (*WWW access*), b) NTUA access (*onsite*) and c) no access. The *WWW access* has no restrictions and the e-thesis is available to all Web users. The *NTUA access* is restricted to NTUA community on and off-campus and is based on IP filtering. After three years it automatically reverts to *WWW access*. Finally, the *no access* applies in patent pending case and when intellectual property uncertainties are involved. In this option the e-thesis is not accessible for one year, when automatically it reverts to *NTUA access* and finally, after three years, to *WWW access*. The access policies are expressed in XML and the user may view them through the metadata (*dc.rights* field). The appropriate access policy documentation is available through the proxy (the theses e-submission tool) and the IR (DSpace system).

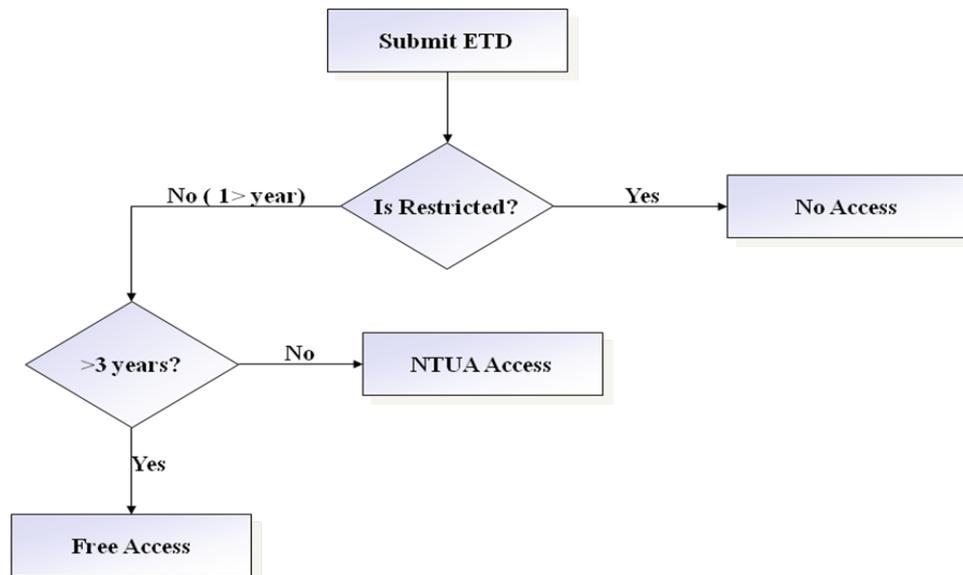


Fig. 1. Access policy workflow (Koulouris *et al.*, 2007)

The users' willingness to share their content was encouraging. The library promotes the open-access movement. It seems that users have recognized the advantages of knowledge sharing. The synergies that arise from e-theses free access policy increase the knowledge core competencies. This is certified from previous evaluation (Koulouris *et al.*, 2008) and from the current continuing one that operates in real-time.

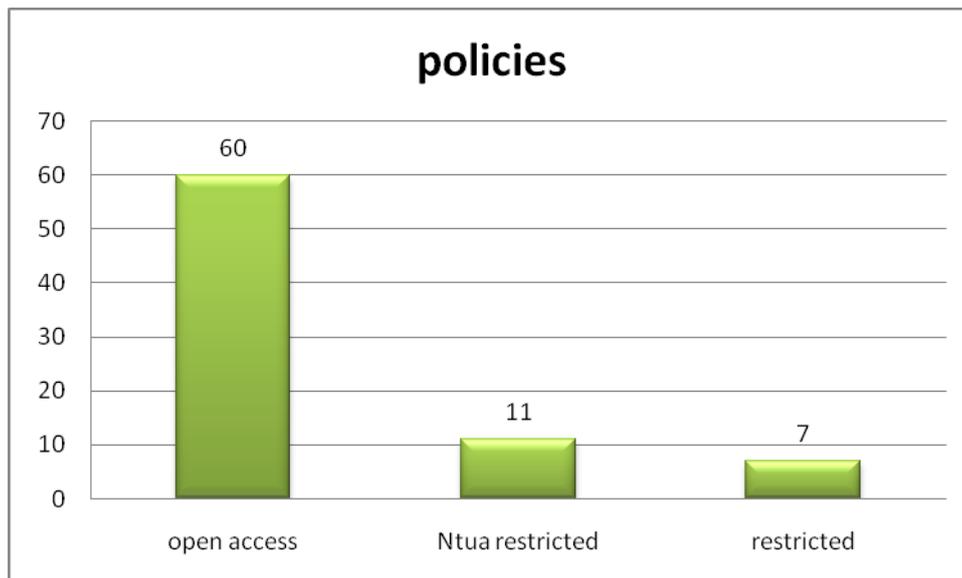


Fig. 2. Users' access policy selection for their e-theses

After selecting the access policy, users have to submit electronically the metadata (*title, author, summary, etc.*) as well as the thesis file only in *PDF* (that should not exceed 50 MB in size), via the online e-submission form. At each metadata input stage, online help is available. Students have to be very careful during the metadata e-submission stage. In particular, keywords given by the author (student) are not checked by librarians and do not follow standard prototypes (NLG Subject Headings, LC Subject Headings, etc.). It has been decided not to check for specific standardized subject headings as only the authors can provide keywords that define exactly the e-thesis content. It is very important for the authors (students) to pay attention when filling in these terms because the keywords provided make the thesis' search and retrieval easier.

When users (students) complete the thesis e-submission, they receive a verification code and a confirmation email indicating the successful thesis and metadata submission on the system. Once this phase is completed, they will have to demonstrate the same code, when physically depositing the print version and the CD/DVD with the electronic copy of the paper to the Library's IR Department.

Subsequently, the associated personnel perform the necessary checks and corrections (e.g. spelling mistakes) and upload the metadata and the PDF file to the IR (DSpace and the library's OPAC). Central Library has developed two ways of accessing ETDs, IR and OPAC, for several purposes: to intergrades the existing digital services and the online catalogue (OPAC) and also to secure the maintenance, preservation and viability of metadata to already existing formats (i.e. *UNIMARC*).

The theses e-submission tool

The theses e-submission system, developed from scratch by library staff members, acts as a proxy between DSpace (the NTUA digital library software) and the end user. The system is based entirely on open source technologies, namely the Postgres RDBMS acting as the persistence layer, Sun Microsystems Java at the business layer

and the JSF framework as the presentation layer. The e-theses submission system consists of a web-based form and it is fully interoperable with DSpace metadata element set; the application profile that the IR uses. In addition, the e-theses tool is fully interoperable with the open-source technology that the DSpace uses.

The DSpace platform was designed (back in 2002) using “older style” Java Servlets, combined with JSPs. This model is quite effective, but can increase the developer's workload exponentially when adding/modifying the application's business logic and/or presentation. Since the submission system that would be designed was going to be a new application, the development team chose to move to a more recent web application framework that would still be technically compatible with DSpace. Thus, JSF was chosen (which, in its core, still uses Servlets and JSPs, but provides the developer with a more component and event driven programming model).

A similar principle was applied when choosing the RDBMS. DSpace uses PostgreSQL, an open source, widespread, RDBMS. The development team decided to stay on the same “page”, and simply create a new, independent, database schema for the e-submission application, running on the same PostgreSQL server.

In summary, the decision to use similar technologies for both systems can be attributed to the following reasons:

- a) More efficient resource management: By using common hardware and software platforms, fewer resources are needed. The same development team can maintain/expand both applications. The same applies to the system administrators and, of course, the hardware resources themselves (e.g., servers, storage media).
- b) Interoperability: Data exchange between the two applications is much easier to implement, since both are based on Sun Java technologies. For instance, the automated batch import process (analyzed below), developed under the e-submission application, would directly insert data into DSpace, using the relevant Java API provided. Alternatively, data could be directly harvested from DSpace and exposed to the e-submission application, for manipulation and/or presentation.

The theses e-submission tool and IR continuing evaluation: from the user perspective

Trying to enhance the service, a process analysis was conducted, starting with the IR evaluation, in order to forecast, define and, finally, determine the process needs, both in technical and human resources terms. Quantitative and qualitative methods were employed for the process evaluation as an internal control mechanism to ensure that the service is efficiently and effectively used (Crawford, 2000).

The evaluation project was undertaken by implementing a web-based survey, hosted under the e-theses submission service website (NTUA Central Library, 2007b) and targeting on the e-theses submission users' population. Before officially starting the survey, an internal piloting was conducted, using six students working in the library as assisting staff. This procedure provided feedback and contributed in refining the questions.

The survey was conducted through a brief web questionnaire consisting of seven questions with pre-defined *Likert* values (Fowler, 1995). The questions are listed in the following screenshot:

Sex	<input type="radio"/> Male	<input type="radio"/> Female			
Age Group	<input type="radio"/> < 25	<input type="radio"/> 25 - 30	<input type="radio"/> > 30		
Rate according to the following scale:	Totally Disagree	Partially Disagree	Neither Agree or Disagree	Partially Agree	Totally Agree
The instructions received by the secretariat were clear:	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
The electronic submission web form help was clear:	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
I received clear instructions by the staff directly responsible for the electronic submission:	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Following your submission, do you plan to use the DSpace digital library to locate your thesis, or any other theses?	<input type="radio"/> Yes	<input type="radio"/> No			
Comments - Suggestions (Optional)	<input type="text"/>				

Fig. 3. Web questionnaire that is used for the e-theses and IR evaluation

Responses were gathered from the population that had submitted their theses up to January 31, 2008 (beginning July 1, 2008), at which point the survey closed. In total, 78 out of 128 users responded (61%) (Koulouris *et al.*, 2008). The response rate is considered satisfactory for extracting useful remarks; moreover, it seems all users show willingness in filling out the questionnaire. In addition, the evaluation is continual and the conclusions that are analyzed are derived from this ongoing evaluation process.

The data were imported into and processed by the *Statistical Package for Social Sciences* (SPSS) version 16.0 (Green and Salkind, 2004). The frequency and the cross tabulations tables, the correlations, and the *chi-square tests* (χ^2), derived from SPSS, whereas the diagrams and the table processing were carried out in MS Excel.

Descriptive statistical methods were used for variables frequencies: a) with frequencies tables and diagrams (vertical column graphs, circle graph) for ordinal variables, and b) average and standard deviation for qualitative variables.

The correlation co-efficient was calculated with: a) *chi-square tests* when referring to ratio or interval scale measurements, b) *t-test* or variance analysis when referring to qualitative variables compared to ratio or interval scale measurements, and c) *Pearson* correlation co-efficient in cases of qualitative data comparison. Finally, the

Cronbach's alpha was calculated in qualitative multidimensional variables (Gravetter and Wallnau, 2003).

A variety of statistics was created and the key results were exposed online, as part of the IR and e-theses submission system, and updated in real time. An extension to the existing e-submission web application was designed, which poses predefined queries to the submission database each time a user requests so. The query results are then converted into graphical charts, using open source Java based technologies, exported as *.png* images and embedded in a web page. The results are updated each time a user refreshes that page. This extension gives an evaluation snapshot in real-time and makes it an ongoing process.

The online document, aiming at helping users submit their content, is highly scored. The users' majority is fully satisfied by the web form help provided [Average 4, 5256 St Deviation 0, 6591]. Neutral scoring illustrates that the help was not used (according to user comments), because the users were very familiar with e-submission procedures.

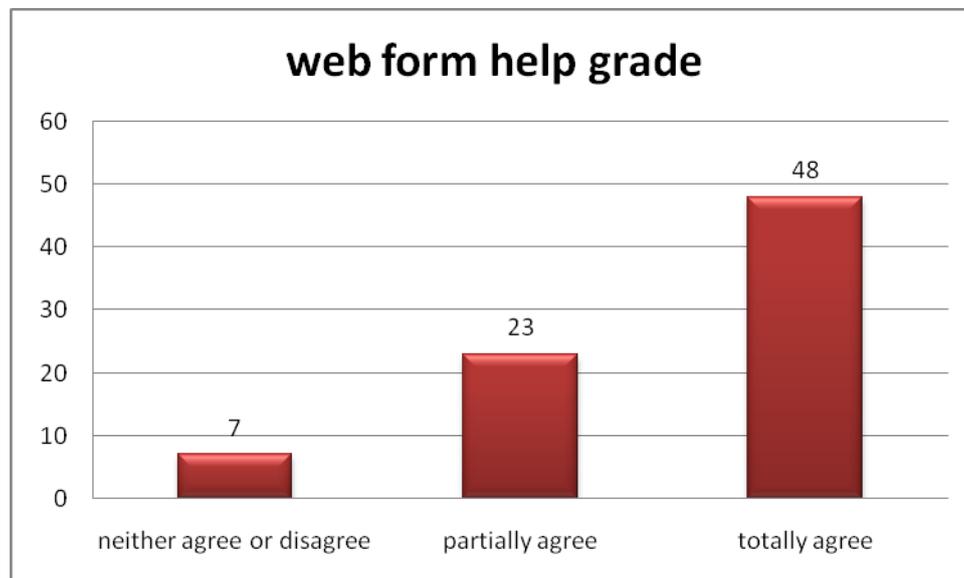


Fig. 4. Web form help grade

The e-theses submission procedures' acceptance rate can be explained. The users are engineers, familiar to IT procedures (e.g. web applications), and at the same time they are part of the most active and promising age group. This supports the effort of establishing new routines for academic publishing and self-archiving.

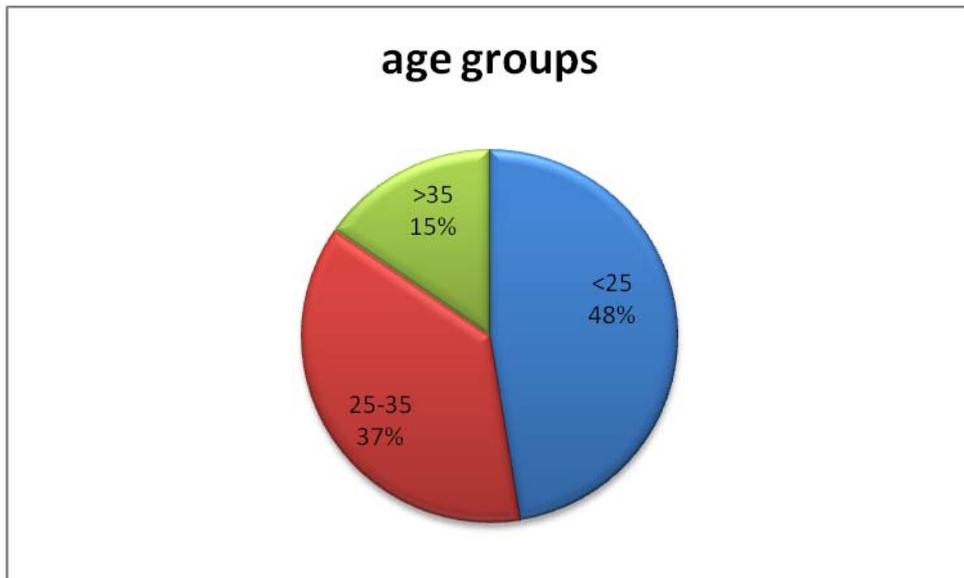


Fig. 5. E-theses tool acceptance rate according to age groups

All age groups scored similar ranks for the evaluation of the web form and its help, IR staff and secretariat. What should be of concern is the fact that the youngest age group was not as satisfied as other groups, so focus should be given to its comments and suggestions, in order to improve the service.

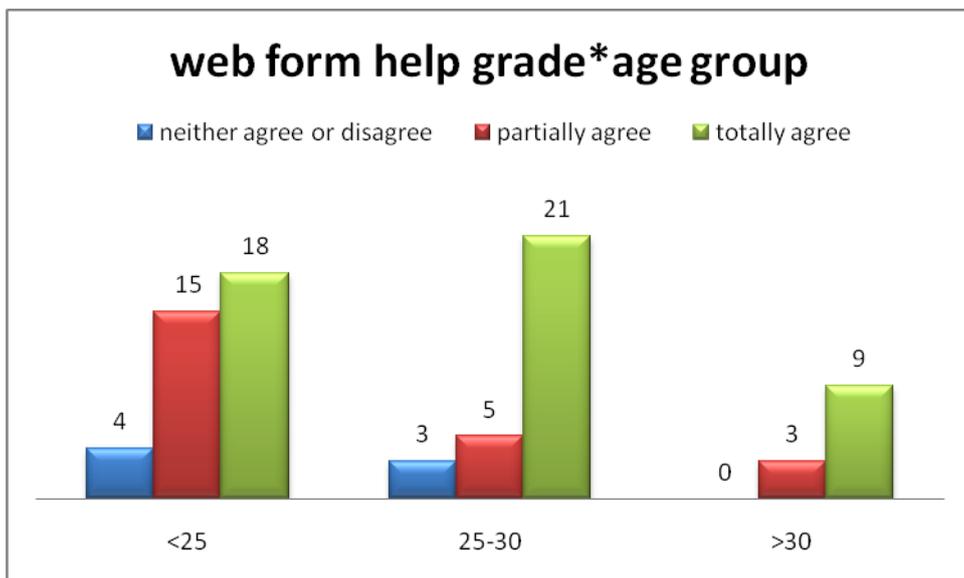


Fig. 6. Web form help grade and age group

The students' majority use the IR service after the thesis submission to certify the import of their e-thesis to the DSpace system, or to search other e-thesis in general.

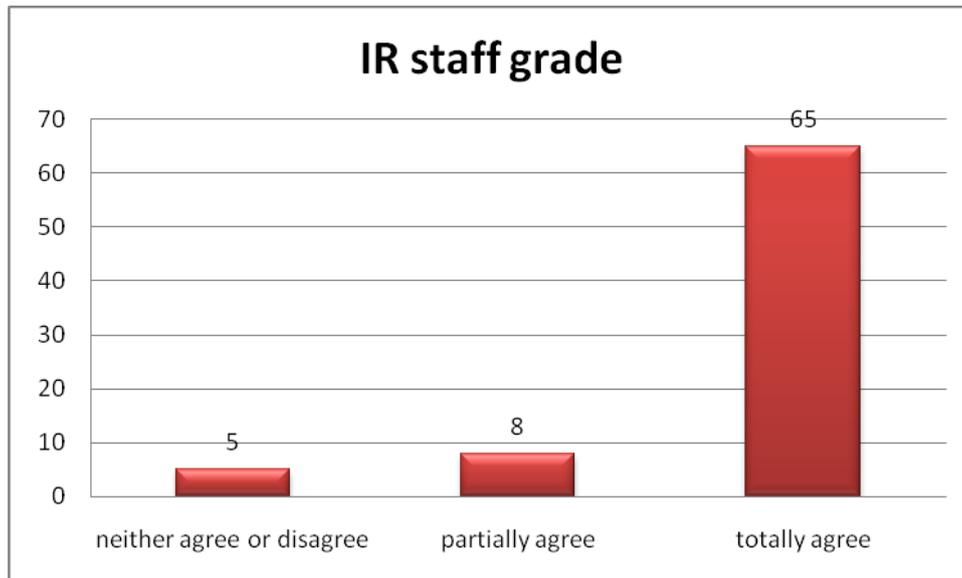


Fig. 7. IR staff grade

As far as the evaluation of the service provided to users by the IR staff is concerned, it is remarkable that the effort of the human resources supporting the whole procedure is highly scored by the majority of users [Average 4, 77 St Deviation wide range 0, 5565].

Finally, as a general remark, users find the e-theses submission tool very useful and according to their comments this gives to the IR service an added value.

The workflow evaluation from the user perspective and its improvement plan from the staff perspective

The main conclusion derived from the workflow evaluation, is that there are periods where the submission rate is very high, and others where the rate is lower. The discovery of the submission rate diversity is a strong motivation for doing the evaluation process continuous. In this context, the IR staff will forecast, define and, finally, determine the process needs, both in technical and human resources terms. This will help organize the procedure efficiently during “high-stress” working periods.

After observation of the entire procedure and testing of alternative serving methods (i.e. increasing the number of fully devoted employees, having in mind that the user should not wait in the queue for more than five minutes), it is concluded that the number of IR staff members currently involved, is unable to meet the demand posed by the users.

The statistical analysis results encourage the service broadness to all university departments, which is our first milestone. For that to happen, the transition from a semi-automated metadata importing process into the main IR, to a fully automated one (batch) is crucial. This will improve the service and will assist human resources administration. In addition, having in mind that less than the 30% of the total theses production is currently being archived, the staff needed for metadata importing will be

even more. This is the main reason that led to the decision to implement the automated batch importing system, which, at the time of writing, is tested and will be soon incorporated as a new functionality on the e-theses tool.

Establishing new routines for e-submission and self-archiving should be complemented by IR service use tutorials to the secretariats, faculty and students; this feedback was collected after tutorials were given to the library staff. Especially for the students, the tutorials may motivate the university departments to inject information literacy courses into the curriculum through e-learning platforms.

The service in-house marketing should be enhanced. The library staff, academic faculty and students should work together and contribute in changing the secretariats business culture, which is the main process weakness, illustrated by this survey. Additionally, aside from the library staff, professors should contribute in spreading the usefulness of the IR service to their students.

Conclusions and future work

The theses e-submission tool helped NTUA users in submitting their e-theses and the IR staff in organizing efficiently the service workflow and the data integration into the DSpace system. By integrating the automated batch importing feature into the e-submission system, will give added value to the IR service and will improve the workflow process even more. This critical point will enhance the service and will allow the service broadness to all university departments. The continuing evaluation that the IR staff has established helped in determining and rectifying the process weaknesses. Rectifications include giving IR service tutorials to the secretariats, faculty and students, conducting in-house service marketing, etc.

As the number of submitted theses grows, the volume of information increases as well. A simple keyword based search will soon be insufficient for the DSpace user searching for theses. Thus, data mining would significantly help categorize theses into various subjects, making the search a more 'e-shop' like experience (suggesting, for example, "related" theses to the ones found, or similar theses that other users viewed etc).

The main future obstacle faced is establishing a strategic preservation plan. NTUA Central Library, via its IR staff, has already participated in the Virginia Tech's ETD preservation survey (Virginia Tech, 2008). The participation in the *NDLTD DDP network*, which is part of the *MetaArchive* (Educopia Institute, 2007) Cooperative's preservation archives and employs the open-source *LOCKSS* (2008) software to harvest, cache, and validate files in a geographically distributed network, is under consideration.

References

Crawford, J. (2000), *Evaluation of Library and Information Services*, Aslib, London.

Educopia Institute (2007), "The MetaArchive cooperative", available at: www.metaarchive.org/ (accessed 12 August 2009).

Fowler, F. J. (1995), *Improving Survey Questions: Design and Evaluation (Applied Social Research Methods)*, Sage, London.

Gravetter, F. J., Wallnau, L. B. (2003), *Statistics for the Behavioral Sciences*, 6th ed., Wadsworth, London.

Green, S. and Salkind, N. J. (2004), *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, 5th ed., Prentice Hall, London.

Koulouris, A., Kokkinos, D., Anagnostopoulos, A. and Tanti, M. (2007), “The institutional repository of NTUA: challenges, concerns, pilot implementation and perspectives”, poster presented at the 10th International Symposium on Electronic Theses and Dissertations (ETD2007), Uppsala, Sweden, June 13-16, 2007, available at: epc.uu.se/etd2007/listofposters.html (accessed 12 August 2009).

Koulouris, A., Kokkinos, D., Anagnostopoulos, A. and Zidropoulos, S. (2008), “Evaluating the NTUA institutional repository”, paper presented at the 11th International Symposium on Electronic Theses and Dissertations (ETD 2008), Aberdeen, Scotland, United Kingdom, June 4-7, 2008, available at: www.rgu.ac.uk/etd/programme/page.cfm?pge=45695 (accessed 12 August 2009)

LOCKSS (2008), “LOCKSS (Lots of Copies Keep Stuff Safe)”, available at: www.lockss.org/ (accessed 12 August 2009).

NTUA Central Library (2007a), “DSpace@NTUA”, available at: dspace.lib.ntua.gr/ (accessed 12 August 2009).

NTUA Central Library (2007b), “ETD e-submission”, available at: dspace.lib.ntua.gr/ETDSubmit/ (accessed 12 August 2009).

Virginia Tech (2008), “Digital preservation of ETDs”, available at: lumiere.lib.vt.edu/surveys/ (accessed 12 August 2009).