

## Evaluating the NTUA institutional repository

Alexandros Koulouris, Dionysis Kokkinos, Angelos Anagnostopoulos, Spilios Zidropoulos  
Central Library, National Technical University of Athens  
koulouris.a@gmail.com, dennis@central.ntua.gr, angelosanagnostopoulos@gmail.com,  
ilizidro@gmail.com

The *National Technical University of Athens* (NTUA), through its Central Library, offers an *Institutional Repository* (IR) and ETD submission service, currently operating in a pilot testing period. The main objectives of this paper can be summarized into three major points. Primarily, to evaluate the IR service pilot period, focusing on the ETD submission process. Secondly, to refine and improve the above mentioned process and, finally, to promote the concept of self-archiving and open access.

The project was undertaken by implementing a web-based survey, targeting on the ETD submission users' population. The data were imported into and processed by statistical analysis software. The key results were exposed online, as part of the IR system, updated in real time, since the survey is an ongoing procedure.

The statistical analysis produced useful results regarding various aspects of the IR service. The major descriptive statistics focused on user groups, administrative staff and procedure grading, as well as access policy selection. Moreover, cross tabulations and correlations were created between all variables, for example "university department" and/or "ETD type" associated with "access policy". A positive user attitude towards the procedure was noted, which motivates us to further enhance and expand the service.

Our first milestone is to broaden the service to incorporate all the university departments. For that to happen, the statistical results will be used to forecast, define and, finally, determine the process needs, both in technical and human resources terms. Process weaknesses detected will be rectified, wherever possible, whereas process strengths will be used to market the service. At the same time, certain improvements, such as the transition from a semi-automated metadata importing process into the main IR (DSpace), to a fully automated one (batch), are already in development.

## 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) with *Electronic Theses and Dissertations* (ETDs) combined with an ETD submission service, currently operating in a pilot testing period. 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's objectives focus on the IR service and the ETD submission process evaluation, the improvement of this process and the promotion of the self-archiving and open access concept. Three (out of eight) university departments are participating in the IR pilot testing period. Students are demanded to submit their ETDs through a web application (*ETD submission service*) 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). At the time of writing, there are approximately 250 ETDs 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*.

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).

## Methodology

The project was undertaken by implementing a web-based survey, hosted under the ETD submission service website (NTUA Central Library, 2007b) and targeting on the ETD 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:

<b>Sex</b>	<input type="radio"/> Male	<input type="radio"/> Female			
<b>Age Group</b>	<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
<b>The instructions received by the secretariat were clear:</b>	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
<b>The electronic submission web form help was clear:</b>	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
<b>I received clear instructions by the staff directly responsible for the electronic submission:</b>	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
<b>Following your submission, do you plan to use the DSpace digital library to locate your thesis, or any other theses?</b>	<input type="radio"/> Yes	<input type="radio"/> No			
<b>Comments - Suggestions (Optional)</b>					

**Fig. 1**

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%). The response rate is considered satisfactory for extracting useful remarks; moreover, it seems all users show willingness in filling out the questionnaire.

In order to gather sufficient answers, the survey was promoted using several ways. The users were motivated to participate in the survey. For instance, users were informed that the survey results would be dynamically published, not only on the website, but also in relevant conferences. The campaign was based on persuading the users that an open-access IR is primarily beneficial to them. Open ETDs mean knowledge sharing, research promotion, author's citation and reputation, etc. Furthermore, it was made clear to the users that the survey success assists the IR service viability. The fact that this survey is innovative for the Greek universities also constitutes motivation for the users.

The survey was conducted by the *IR staff* (administrative staff directly involved with the ETD submission and IR systems) for the service evaluation and for non-commercial purposes. It was clarified that the survey results will be used only for scientific reasons. Procedure security was taken under consideration. Users' privacy and anonymity were guaranteed. However, users were informed that their responses are combined with their submission data (e.g. access policy, department).

### Statistical analysis

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* ( $\chi^2$ ), derived from SPSS, whereas the diagrams and the table processing were carried out in MS Excel.

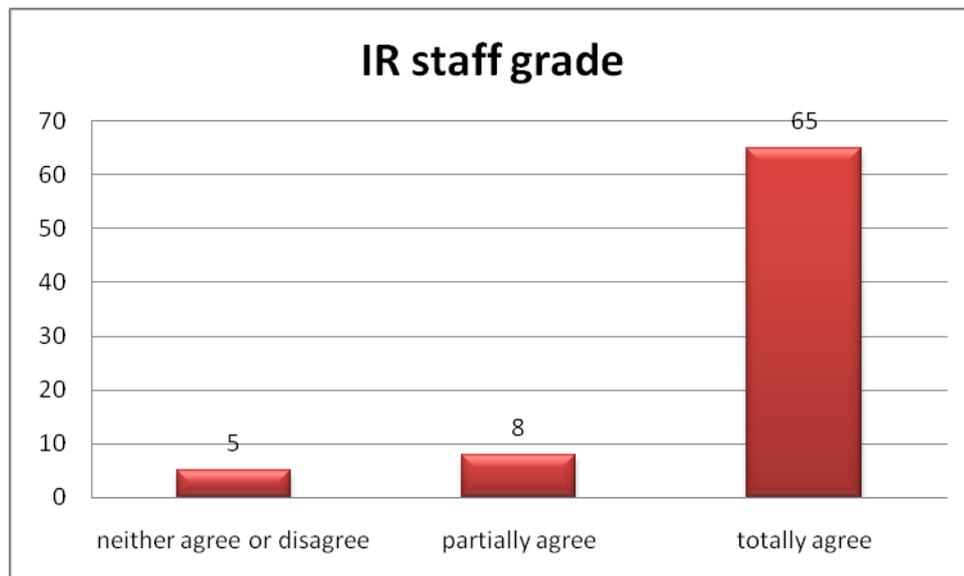
Descriptive statistical methods were used for variables frequencies: a) with frequencies tables and diagrams (vertical column graphs, circle graph) for ordinal variables, and b) average and standard deviation for qualitative variables.

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 ETD 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.

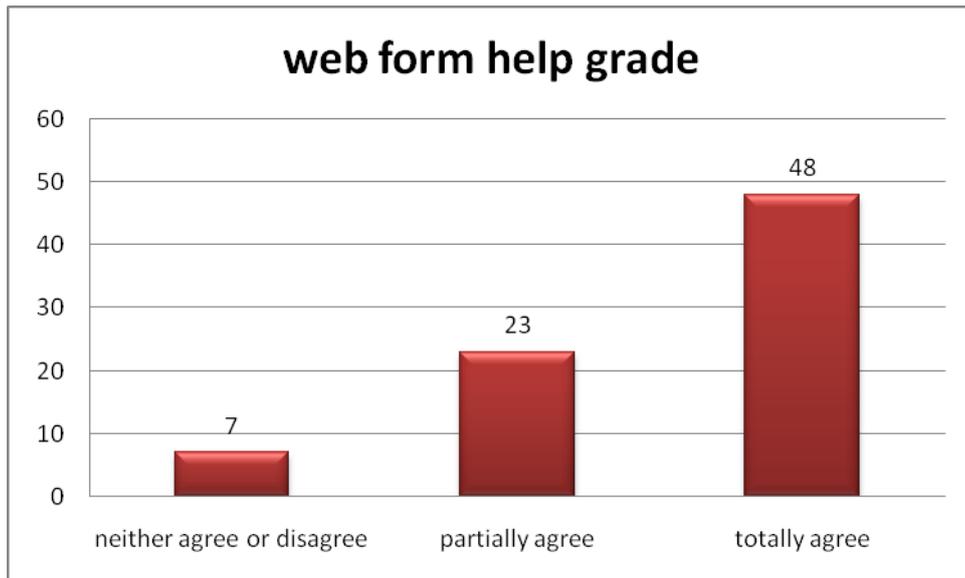
### ***Descriptive results***

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



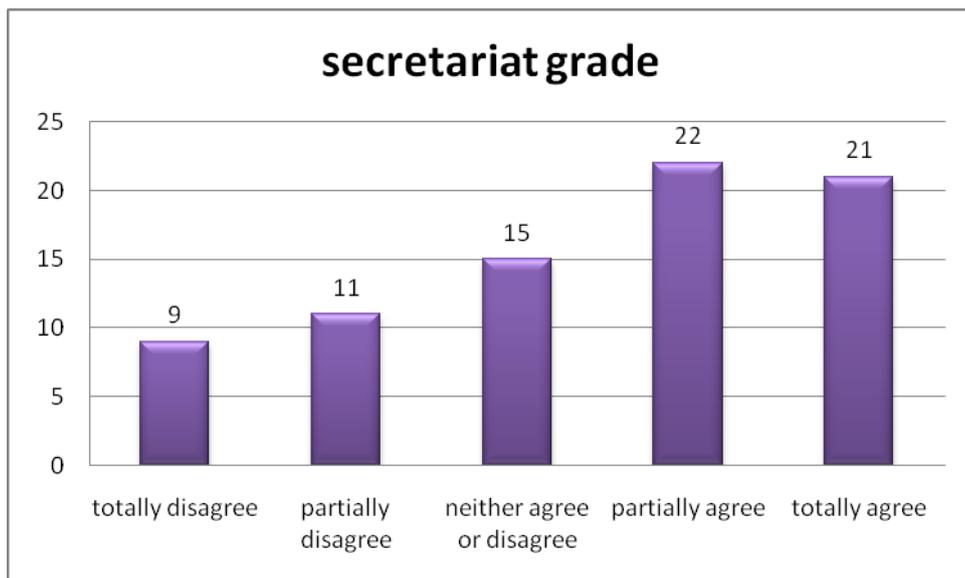
**Fig. 2**

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].



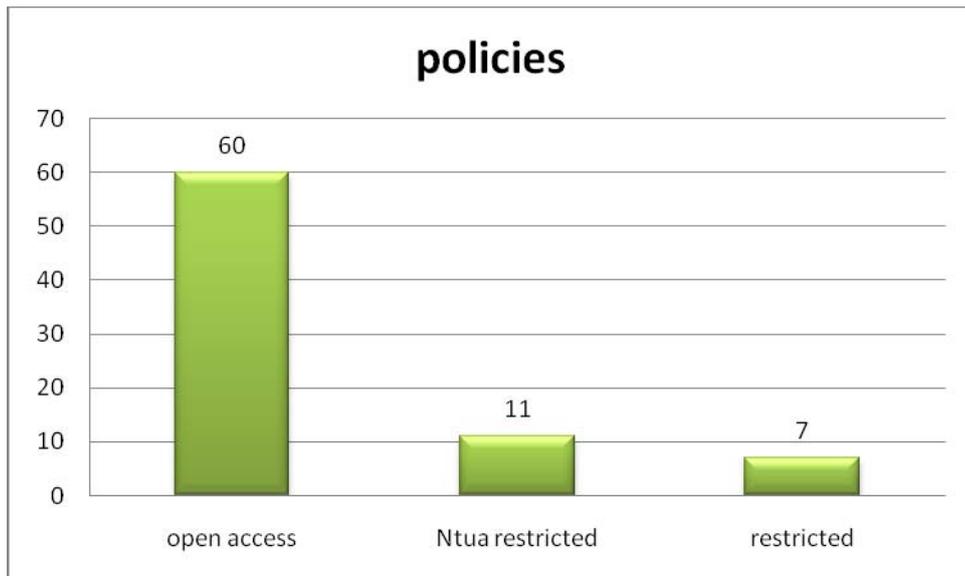
**Fig. 3**

The online document, aiming at helping users submit their content, scored even higher. 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.



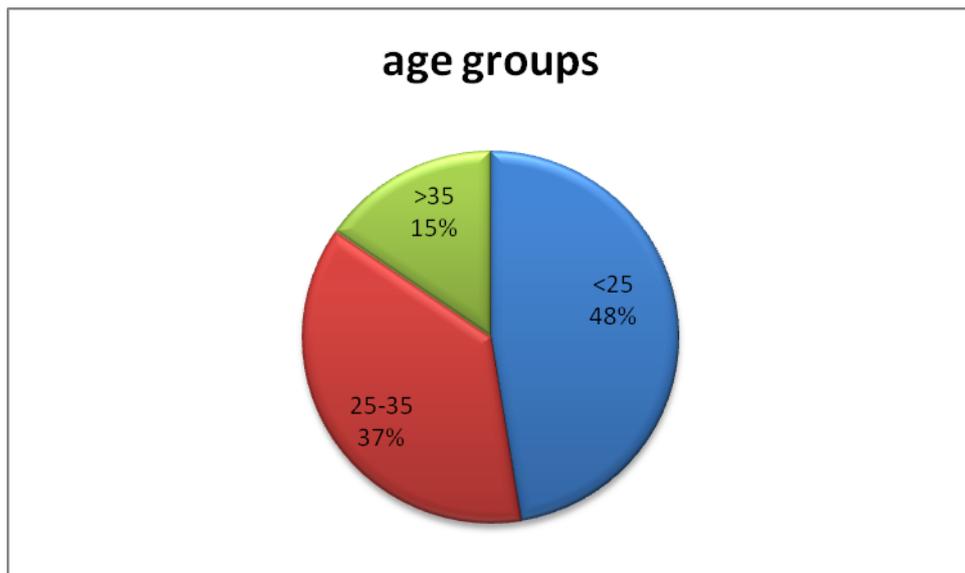
**Fig. 4**

Users' attitude against the secretariats' service, was weakly strong, as cross tabulations showed. Bureaucratic procedures are known to cause extra overhead to any business process and secretariats are no exception to this rule. The secretariats should be further assisted in promoting the whole service. Meanwhile, plans have been scheduled for conducting secretariat staff tutorials regarding the service. These tutorials will be implemented when the ETD submission procedure opens for all the university departments [Average 3, 4487 St Deviation 1, 3353].



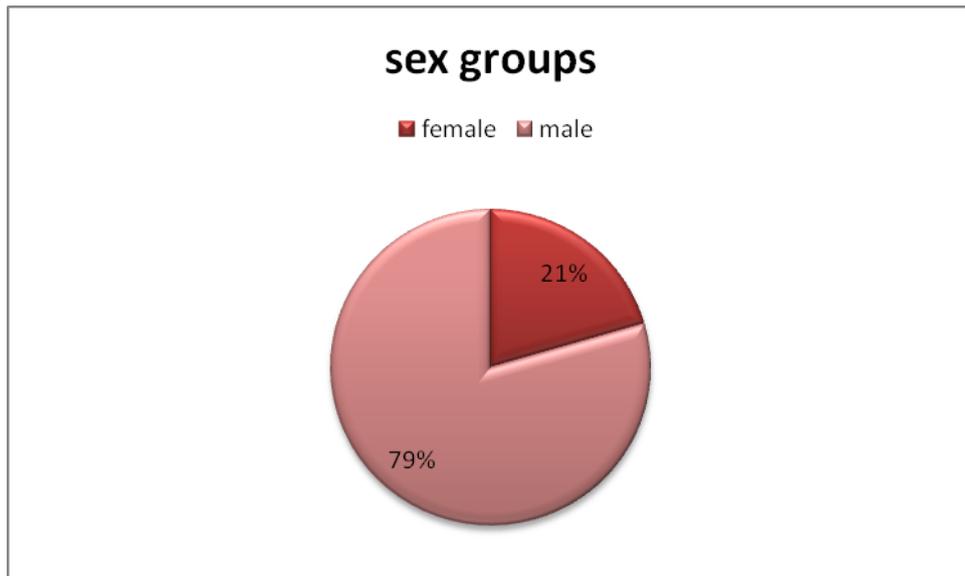
**Fig. 5**

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 ETDs free access policy increase the knowledge core competencies.



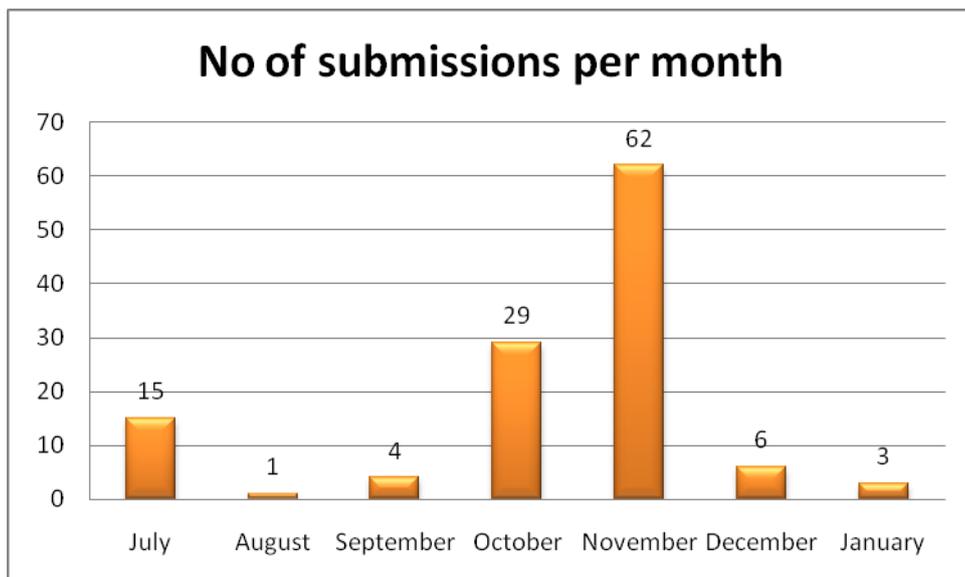
**Fig. 6**

The ETD 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. 7**

Men show higher preference in the technical and engineering departments that NTUA consists of. This explains the fact that males submit more than females.



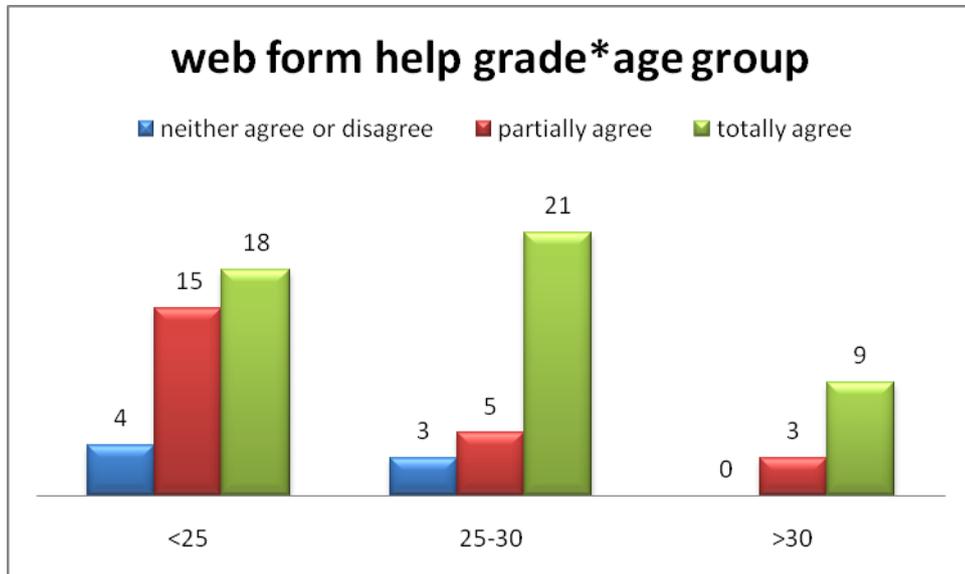
**Fig. 8**

The academic year of 2006-2007 was not a normal one, due to a national reformation in the field of education. Thus, a shift of all year's submissions was noted, towards October and especially November 2007. Additionally, the IR was established early in July 2007, so due to in-house bureaucracy, the information distribution was slow at first. A campaign (e.g. letters to secretariats, website posting) was waged, to officially inform the participants, students, academic and faculty staff. In parallel, the university departments have specific periods for accomplishing and examining their produced theses. As a result, 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 this research, in order to 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.

For example, lacking an automated batch importing system, it is clear that at least two employees should be occupied with the process of importing data to the DSpace system. Having in mind that less than the 30% of the total theses production is currently being

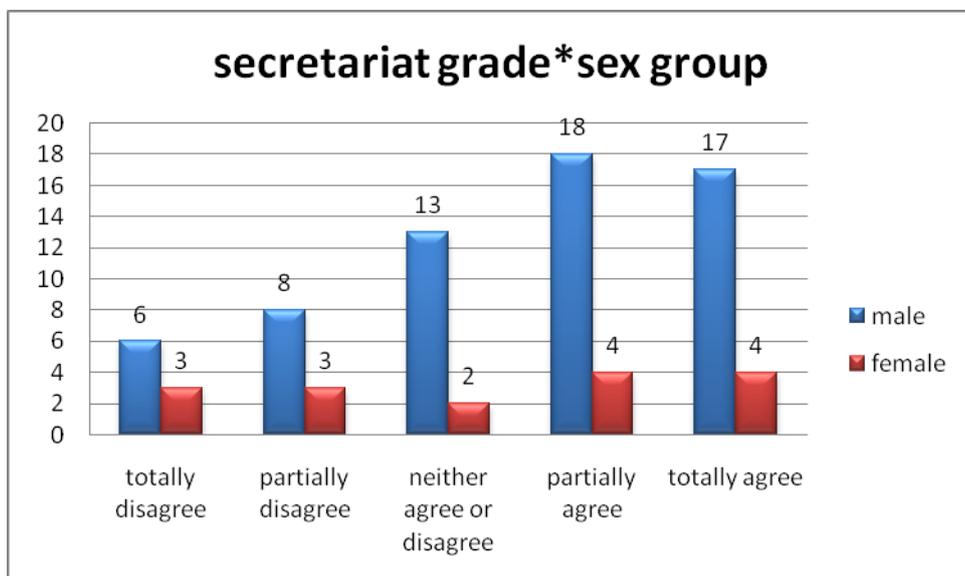
archived, the staff needed for metadata importing will be even more. This is the main reason that led to the decision to implement the automated batch importing system, which, at the time of writing, is under development.

**Cross tabulations**



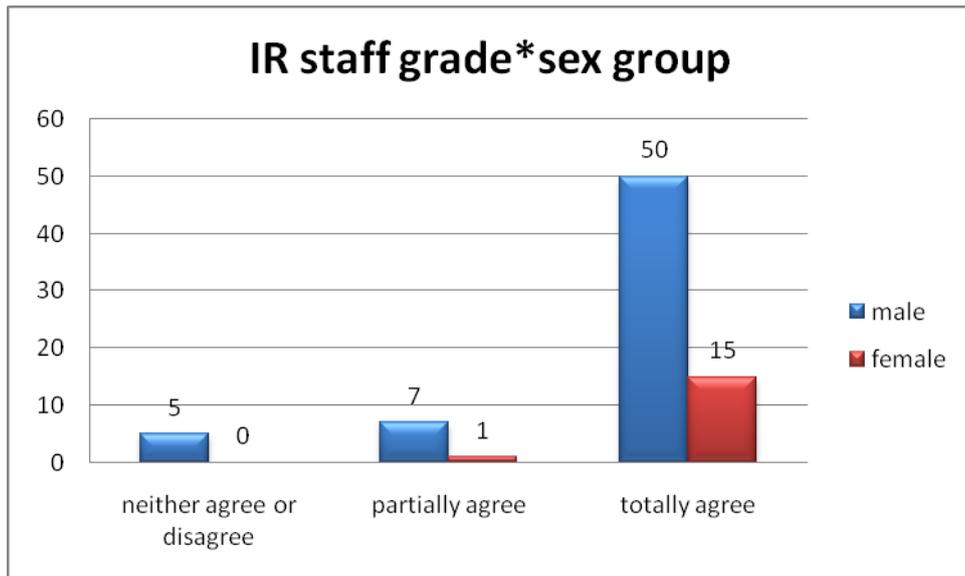
**Fig. 9**

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.

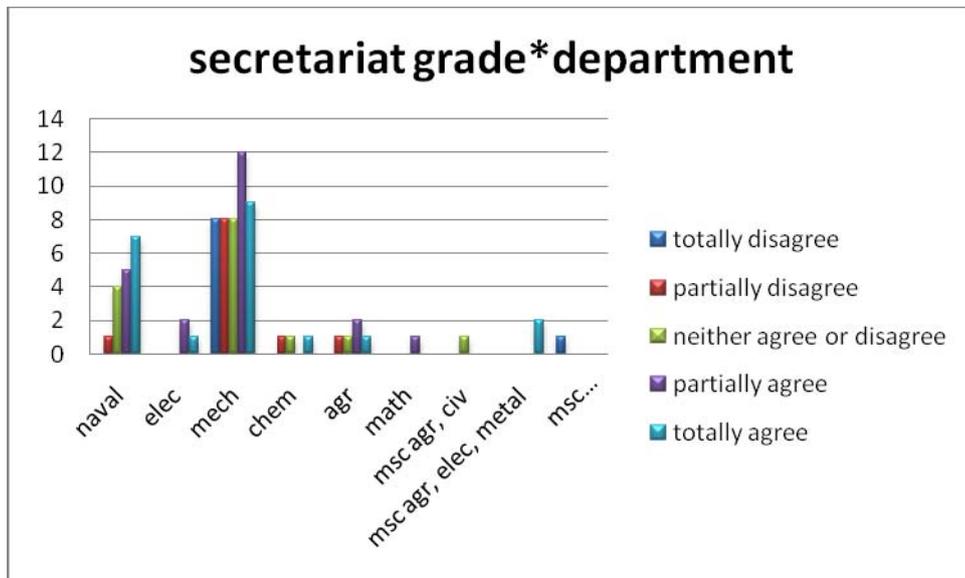


**Fig. 10**

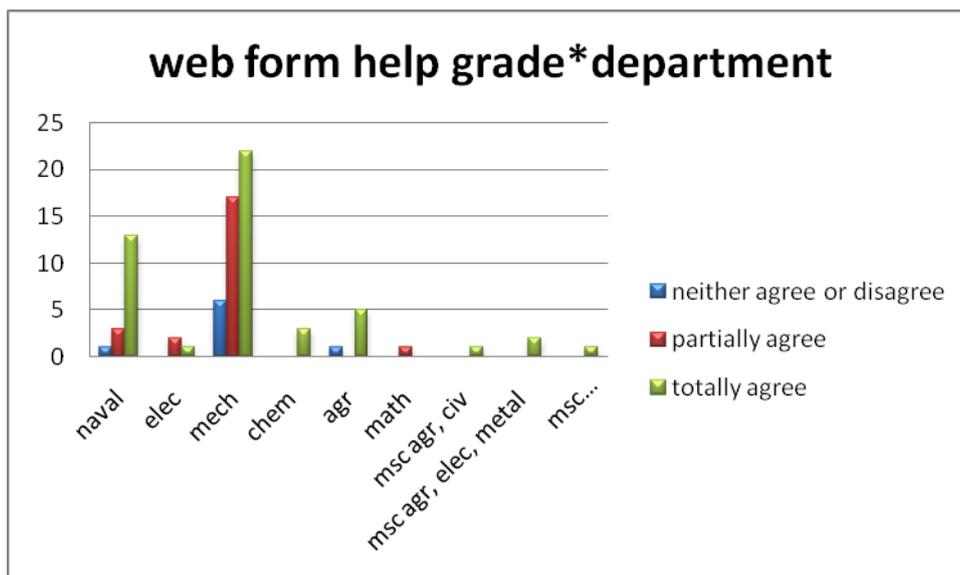
It is notable, that male users were more satisfied than female concerning the secretariat evaluation.



**Fig. 11**  
The women find the IR staff's service more reliable than the men do.



**Fig. 12**  
Regarding the cross tabulations, naval engineers rated the instructions they received from their secretariat quite highly (compared to other users). Moreover, the naval engineers' secretariat on graduate level, efficiently direct the users for the e-submission. One of the users certified this, by commenting on the positive attitude and the high level of work done by the specific secretariat. On the other hand, the naval engineers' secretariat on post-graduate level, does not seem as willing to help.



**Fig. 13**

The uniformity on web form grading regarding the users form various departments, can be justified by the fact that all users are engineers familiar with IT procedures.

### ***Correlation***

There is a weak correlation between the following:

IR staff grade and web help grade (0, 12257),

IR staff grade and secretariat grade (0, 12368), and, finally,

web help grade and secretariat grade (0, 171213)

[< 0, 30 weak linear association; 0, 31-0, 60 moderate linear association; >0, 60 strong linear association].

### **Conclusions**

What is rather expected is the fact that the younger the users, the more willing they are to use the IR. The department a user belongs to has also no relevance to what level of access one will allow to his document.

The thesis type (diploma, master, Ph.D.) is irrelevant to the access policy selection. However, more diploma theses are open than the other theses types. This can be attributed to the fact that master and Ph.D. theses deal with more specialized subjects and probably contain innovative and/or patented content.

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 was 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.

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.

## Future plans

Any strategic plan incorporates ambitious steps and projects. Plans that have been described in previous works (Koulouris, et al., 2007), i.e. federated search, are already in development. The statistical results will be used to forecast, define and, finally, determine the process needs, both in technical and human resources terms.

Process weaknesses detected will be rectified, wherever possible, whereas process strengths will be used to market the service. For example, the maximum allowed uploaded file size and the refinement of specific directions given to the users have already been modified. Development of a full text searching capability is also under consideration. Additionally, design and development of the automated batch importing system (mentioned above) is already underway.

It will be useful and crucial to conduct a new survey with all the university departments participating in the IR. The current evaluation results will be used as a methodology tool for constant observation of the procedure.

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, should be reviewed.

## References

- Crawford, J. (2000). *Evaluation of Library and Information Services*. London: Aslib.
- Educopia Institute (2007). *The MetaArchive Cooperative*. <http://www.metaarchive.org/> [Accessed 07-05-2008].
- Fowler, F. J. (1995). *Improving Survey Questions: Design and Evaluation (Applied Social Research Methods)*. London: Sage.
- Gravetter, F. J., Wallnau, L. B. (2003). *Statistics for the Behavioral Sciences*, 6<sup>th</sup> edition. London: Wadsworth.
- Green, S., Salkind, N. J. (2004). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, 5<sup>th</sup> edition. London: Prentice Hall.
- Koulouris, A., Kokkinos, D., Anagnostopoulos, A. and Tanti, M. (2007). "The institutional repository of NTUA: challenges, concerns, pilot implementation and perspectives". *Proceedings of the 10th International Symposium on Electronic Theses and Dissertations (ETD2007)*. Uppsala, Sweden, June 13-16, 2007.
- LOCKSS (Lots of Copies Keep Stuff Safe)* (2008). <http://www.lockss.org/> [Accessed 07-05-2008].
- NTUA Central Library (2007a). *DSpace@NTUA*. <http://dspace.lib.ntua.gr/> [Accessed 05-05-2008].
- NTUA Central Library (2007b). *ETD e-submission*. <http://dspace.lib.ntua.gr/ETDSubmit/> [Accessed 05-05-2008].
- Virginia Tech (2008). *Digital Preservation of ETDs*. <http://lumiere.lib.vt.edu/surveys/> [Accessed 07-05-2008].
- The project is co-funded by European Social Fund and National Resources.*