OPENSTARTS: A "LEAN" APPROACH TO ETD PUBLISHING

Stefania Arabito – Sistema Bibliotecario di Ateneo - Università degli studi di Trieste – arabito@units.it
Fabio Asnicar - Sistema Bibliotecario di Ateneo - Università degli studi di Trieste – fabio.asnicar@poste.it

Background

Italian Universities are now undergoing significant changes owing to the reform triggered by the Bologna declaration.[i] Unfortunately ever shrinking budgets are undermining the possibility of achieving greater standards of quality and of doing long-term planning. As the President of the Rectors' Conference keeps saying, Italy's investments for Universities and Research are well below the European average and in the lowest ranks in Europe. The soaring costs of e-journals subscriptions are jeopardizing scholarly communication. It is high time for Universities to take a systemic approach to the management of resources.

At the University of Trieste there has never been real interplay between libraries, administrative services, Faculties and Departments. There is no feeling that the overall mission of the institution is the same and that cooperation is a prerequisite for effectiveness. There are plenty of projects pertaining to the cycle of information, but more often than not different sections work independently of each other, not even knowing the mutual competences and interests. Often Department sites replicate the information contained in the Library System site and often sites are not updated. Only recently has the Rector started an action aimed at rationalizing information flows, but the old habits of fragmentation are difficult to overcome.

The legal status of theses and dissertations in Italy

Students are entitled to the intellectual property of theses, which are to be considered both administrative documents (i.e. a prerequisite to graduation) and intellectual works. In this case the law protects the form and expression of ideas not ideas themselves (tutors may have contributed to the ideas but not to their form and expression). Libraries need to be authorized to manage the circulation of theses. Authors (i.e. students) can limit access to their work; even tables of contents are copyrighted. Changing formats makes no difference. All forms of plagiarism are penal offences and are prosecuted accordingly.

The reform has changed the length and structure of University courses. After a three-year cycle the student is awarded a first degree which no longer requires the production of a degree thesis, even though at the University of Trieste virtually all faculties do. The quality of such essays is variable; only some of them are worth while archiving.

After a further two-year cycle the student is asked to write an extended thesis, which is supposed to be more scientific and original. Graduates will therefore be more sensitive to disseminating their work.

University libraries have a long-standing tradition in setting up suitable systems for the management of the archival, print copies of theses. They are therefore entitled to be key agents in the project for the electronic publishing of theses.

Electronic theses and dissertations (ETDs)

"*Digital archiving*" can be defined as the "*long-term storage, preservation and access to information that is born digital*"[ii].

Times are ripe for libraries to unlock their resources and assist students with publishing their works electronically. Just a very brief selection of initiatives concerning electronic theses and dissertations in the Anglo-Saxon countries and in Italy is listed below.

UNESCO is concerned with electronic theses and dissertations. The Internet is considered as a tool for the dissemination of scientific knowledge. In 2002 the "UNESCO Guide to Electronic Theses and Dissertations"[iii] was published and in 2003 pilot projects started in Africa, Europe and Latin-America.

The Networked Digital Library of Theses and Dissertations[iv] is an international federation of digital theses archives which started in 1996 at Virginia Tech[v], formerly under a grant from the U.S. Department of Education's Fund for the Improvement of Secondary Education .

The Theses Alive![vi] Project at Edinburgh University under a JISC (Joint Information System Committee) fund evolved into the Edinburgh Research Archive[vii], which hosts various kinds of documents.

The DAEDALUS[viii] project at Glasgow University Library, again JISC funded, started by making a case for the creation of institutional repositories dealing with theses and other types of grey literature and now hosts also published and peer reviewed papers, albeit in a different repository.

In Italy, Thesis 99[ix] has evolved from Titulus 97, a project concerning digital archiving. Apparently the project has stopped. The University of Pisa has joined the NDLTD and applies its model to the submission of theses. Several Universities have by now started institutional repositories which also host or will host ETDs (namely Padova, Milan, various institutions in Rome, Parma, Firenze, Trento and Trieste among others).

Best practices for digital archiving are developing.[x] The technology that is going to help us is already in place.

Open access and the Open Archives Initiative

The present crisis in scholarly communication is introducing new models and paradigms. Interoperability is the first step and a set of standard metadata has been created, based on the Dublin Core[xi]. Self-archiving is the second step. The focus is more and more on the application and exploitation of the Open Archives Initiative Protocol for Metadata Harvesting[xii], i.e. on the creation of institutional repositories or subject repositories.

This policy aims at eliminating barriers to dissemination and at providing easier access to information. Repositories can be harvested and are cross-searchable, authors are more "visible" and their works have greater impact. The way to open access has been paved by the Santa Fe convention (1999)[xiii], the Berlin declaration (2003)[xiv] and the subsequent Messina declaration (2004)[xv] whereby almost all the Italian institutions of higher education agreed on supporting open access as the only sustainable way to achieve broader and faster dissemination of scientific knowledge.

The Library System of the University of Trieste

At the University of Trieste the Library System is taking advantage of ICT to offer better services to users. Its Information Retrieval systems are integrated with its Library Management System and all functions are fully automated to avoid duplication of tasks and to offer more timely and exhaustive information to patrons. Its policy is focussed on access and user-centred. In the past few years the Library System has been forced to cut expenses owing to lower allocation of funds, a predicament which is concerning the whole University context to varying degrees.

The present state is that of a hybrid library which however has not a fully recognized status as its visibility within the parent organization is inadequate owing to a lack of communication and of promotion.

The University produces research and one of the factors that determine the allocation of funds is the amount of research that is carried out. Much of this work is currently undervalued and has a low impact. Starting a limited project on a limited sector of this local production could pave the way for the creation of an institutional repository.

The legacy
The General Library has always been entrusted with archiving, storing and preserving theses and dissertations on behalf of the Registrar's Department.
Prior to 1994 there was a separate card catalogue ordered by author and a topographic catalogue ordered by faculty in the back offices.
After the first library automation, theses were catalogued following the ISBD in the Library System and therefore accessed through both the local OPAC and the national OPAC. They were assigned a special code and were retrievable by author, title and general keywords. Access was restricted and they did not belong to the circulating collection.
The second automation enabled us to catalogue them in the local system so that they were visible only in the local OPAC. This was considered more suitable given their special circulation status.
Before submitting their theses to the Registrar's Department, students are asked whether they give full or restricted access to their works. A librarian at the General Library then processes and catalogues them. Another stores them in a faraway deposit – space problems are becoming dramatic – and collects them once a week on request. Patrons have to enter the library at least twice if they want to access them. The first time to fill the request form, then, after the thesis has been collected, to look at it in the reference room. Tables of contents and bibliographies are photocopied by staff on demand (if the author has given consent). The interlibrary loan and document delivery services also send copies of the tables of contents and bibliographies on request. Theses are assimilated to monographs and do not have particular visibility in the OPAC.
In 1997 the University of Trieste joined a national project called Alma Laurea[xvi] whose aim is to create a national database of theses which can be accessed on demand and for a fee by potential employers seeking graduates who have special qualifications. Therefore presently theses are also archived and indexed by the Registrar's Department. The General Library and the Registrar's Department databases are not interoperable and the staff belong to different divisions.

OpenstarTs
OpenstarTs is an action included in the project "University and work opportunities in Friuli Venezia Giulia", under a grant from the Italian Department of Education and Higher Education. The transition from education to the world of work is facilitated if theses and dissertations are more widely and effectively disseminated. The University of Trieste is engaged in many ways in supporting graduates and postgraduates in pursuing their professional/vocational careers. Moreover, a "lean" approach to theses management facilitates both the submission and the retrieval of such items. Tutoring students on these issues is also an additional asset for their portfolio of skills and competencies.
The project aims at overcoming the present weaknesses which can be summarized as follows.
In terms of input: slow cataloguing (a thesis is processed roughly one year after the student's presentation and defence) and duplication of tasks (the Library and Registrar's Department).
In terms of storage: dramatic lack of space and unfit deposits (mould, misplaced material…), plus theses are continously moved from the Registrar's Department, to the General Library, to the deposit and round and round.

In terms of retrieval: information is undervalued and has a very low impact, it is neither timely nor complete, and is confined to one location (the General Library).

The threats involved are that such a project is a joint venture involving not only librarians but also IT staff, academic administrators and faculty. The outcomes are in terms of enhanced visibility, credibility and sense of identity. As a rule, however, there is an unjustified feeling of competitive rivalry between library, computing and administrative staff and no true cooperation with the academic staff. Strategies to make all these people feel involved and committed to the project and to counteract defensive patterns have to be put in place.

On the other hand, the expertise needed for the electronic publishing of theses is already there. On the legal and archival side, Italian laws allow and even encourage and promote the digital preservation of administrative documents. Theses are not specifically mentioned, but are logically included.

An open archive for theses is the first step towards an open archive for educational material and research works at an institutional level, stopping the fragmentation of information in departmental servers.
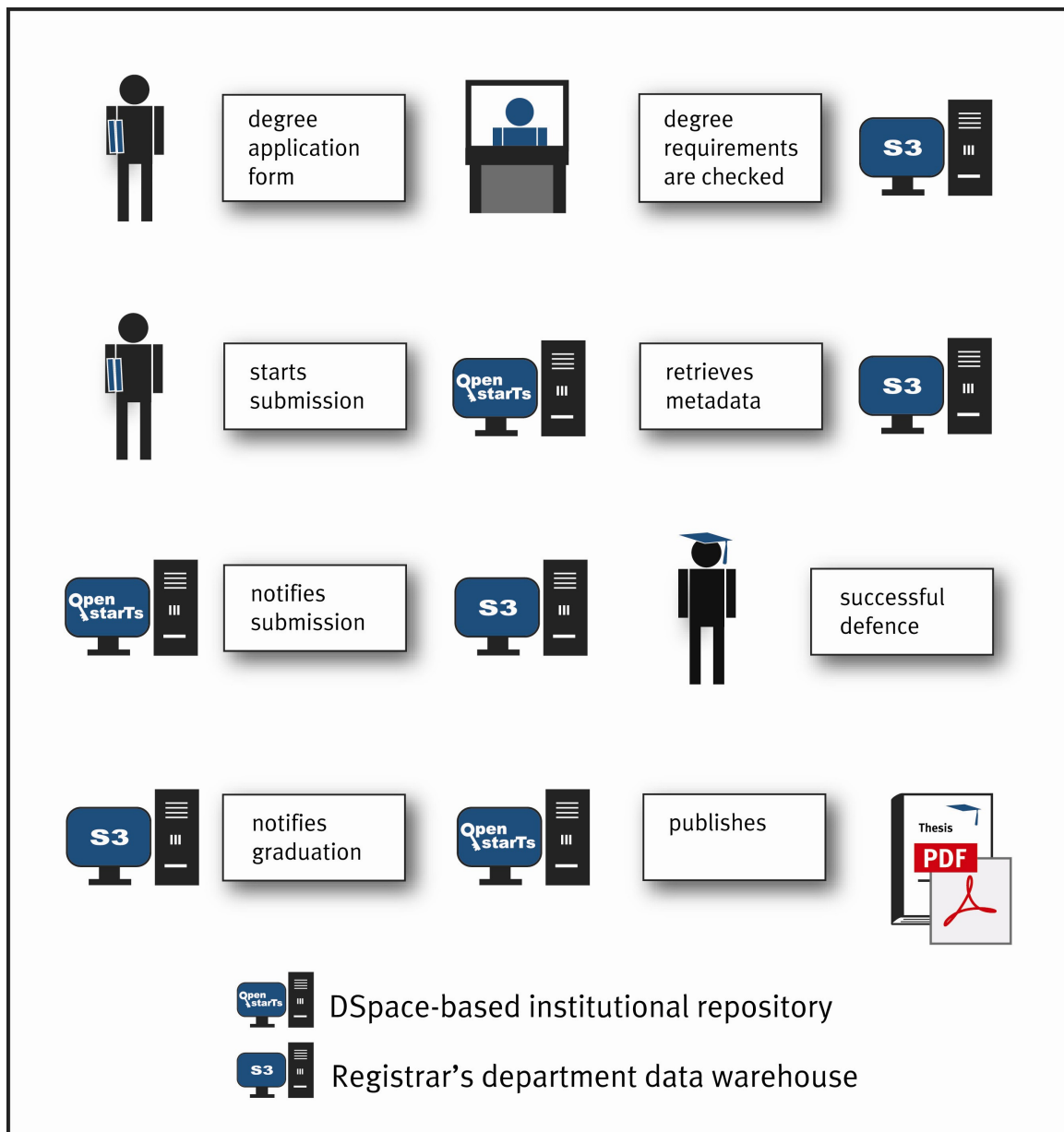
A "lean" approach

The whole workflow of the submission of theses needs to be redesigned by integrating the schedules already in place while avoiding duplications. It is strategic to fine tune the existing automatized processes while avoiding major alterations, lest resistance to change might jeopardize the success of the initiative. On the contrary, it is vital to exploit the present procedures to their full potential.

Trieste University already has a single sign on authentication system. Students, as well as academics and the administrative staff, are recognized through their registration number via LDAP. The Registrar's department already has a dataware house which is fully automated and which holds all the relevant information on the students' academic careers. These data are certified and have a legal value, as awards are automatically produced by the system. The Registrar's department data warehouse is also connected to the Staff data warehouse for the relevant information on supervisors and tutors.

The starting point is digital archiving through an OAI compliant system, as dissemination is enhanced by interoperability and compliance to the metadata harvesting protocol. The most "lively" open source software for open archives is presently DSpace[xvii], which is the result of a joint venture between MIT (Massachussetts Institute of Technology) and HP (Hewlett Packard). DSpace relies on a proactive international federation which meets regularly and keeps in touch through a "general" and a "technical" mailing list and a wiki[xviii] . Dspace is very flexible and can be customized to suit the different needs of institutions. It is organized around communities and collection, and it is apt to capture, store, index, preserve and distribute all kinds of digital materials.

After an accurate evaluation of the metadata relevant to electronic theses, also in view of efficient harvesting from service providers, a "usable" translation of the interface has been made by an expert colleague[xix]. Much attention has then been devoted to an thorough analysis of the existing procedures and on devising a viable and efficient model for ETD management.

In the OpenstarTs model the workflow of electronic theses is based on a full integration between LDAP, S3 (the Registrar's Department data warehouse) and OpenstarTs itself (a Dspace application). The table below summarizes the interaction between S3 and OpenstarTs:

1. The ETD workflow

degree application form

degree requirements are checked

S3

starts submission

Open starTs

retrieves metadata

S3

Open starTs

notifies submission

S3

successful defence

S3

notifies graduation

Open starTs

publishes

Thesis
PDF

Open starTs DSpace-based institutional repository

S3 Registrar's department data warehouse

source: the OpenstarTs project (www.openstarts.units.it)

Self-archiving procedures are simplified to a huge extent, and students find an extremely friendly and usable submission interface. They only have to type their registration number and the system retrieves the metadata related to their theses, such as faculty, course, supervisor and tutor names, academic year, and provisional title (which is modifiable). The error probability fall downs to zero. Students can then focus on providing accurate "semantic" metadata: keywords, abstracts and tables of contents. They can add their curricula vitae if they wish.

The last step is to attach as many files as the chapters of their theses and to specify the permissions on each file. As students are entitled to the intellectual property of theses, the theses submitted can have a mixed availability status, according to the will of their authors.

As soon as the students submit their theses, OpenstarTs accordingly notifies it to the Registrar's department data warehouse, and students can relax and prepare until the defence day comes. No need to print out and bind expensive paper copies, let alone queuing up to hand them in! After the successful presentation and defence of theses, metadata are validated and theses enter the repository to all purposes.

Validation procedures are consequently enormously facilitated and the whole workflow becomes extremely "lean". Life is made easier for students, librarians, and the administrative staff.

It is to be noticed that S3 is a leading software in Italian Universities, so the functionalities described above are applicable to other contexts.

Where we are now
A prototype is already in place. It is going to be tested and validated by a group of users at the beginning of May 2006. The outcomes of the questionnaires and of the observations will provide valuable clues for fine-tuning the procedures and making the interface even more usable.

Acknowledgements
The real strength of the project is its ability to catalyze the joint efforts of several components of the University, namely the Library System (and the Library System IT staff), the central computing services, the Registrar's department and Faculties. It is thanks to all these willing and cooperative people that we will manage to save the time of our students.

---

i http://europa.eu.int/comm/education/policies/educ/bologna/bologna.pdf
ii http://www.dlib.org/dlib/january00/01hodge.html
iii http://www.etdguide.org/
iv http://www.ndltd.org/browse.en.html
v http://scholar.lib.vt.edu/theses/
vi http://www.thesesalive.ac.uk/
vii http://www.era.lib.ed.ac.uk/index.jsp
viii http://www.lib.gla.ac.uk/daedalus/
ix http://www.unipd.it/archivio/th_idx.htm
x http://www.minervaeurope.org/guidelines.htm
xi http://www.dublincore.org/
xii http://www.openarchives.org/OAI/openarchivesprotocol.html
xiii http://www.openarchives.org/sfc/sfc_entry.htm
xiv http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html
xv http://www.aepic.it/conf/viewappendix.php?id=49&ap=1&cf=1
xvi http://www.almalaurea.it/
xvii http://www.dspace.org/
xviii http://wiki.dspace.org/
xix Roberto Sturman, Università degli studi di Trieste, roberto.sturman@infn.units.it

Bibliography

De Robbio, Antonella (2002) *Le tesi nel diritto d'autore : un argomento complesso.* Available at: http://eprints.rclis.org/archive/00000091/02/TESIe_DDA.pdf

Jones, Richard, Andrew, Theo & MacColl, John (2006) *The institutional repository.* Oxford: Chandos publishing

MacColl, John (2002) "Electronic theses and dissertations: a strategy for the UK", *Ariadne, 32.* Available at: http://www.ariadne.ac.uk/issue32/theses-dissertations/intro.html

MIT's DSpace experience: a case study. Available at: http://www.dspace.org/implement/case-study.pdf

Pistelli, Zanetta (2002) *La catalogazione delle tesi.* Roma: AIB

Smith, M. (2003) "DSpace: An open source institutional repository for digital material." *D-Lib Magazine* **8**(10) Available at: doi:10.1045/october2002-inbrief#SMITH