



HEP Libraries Webzine

Issue 11 / August 2005

E-LIS : an International Open Archive Towards Building Open Digital Libraries

Antonella De Robbio(*) and Imma Subirats Coll(**)

13/06/2005

Abstract

Established in 2003, E-LIS is an international open access archive related to librarianship, information science and technology, and related disciplines. It uses the Open Archives Initiative (OAI) protocol and tools to facilitate interoperability between repository servers. To date E-LIS is the biggest repository in library and information science and after only two years contains over 2200 papers. E-LIS is the first international e-server in this area, is part of the RCLIS (Research in Computing, Library and Information Science) project and is organised, managed and maintained by an international team of librarians working on a voluntary basis. This paper describes the main characteristics (technical and organizational) of the archive and its configuration and customization, and discusses its policies, aims and mission. Its main focus, however, is on the E-LIS organizational model and on the strategic issues correlated with Open Access (OA). It also delineates some of the challenges and opportunities consequent on a global vision for the Library and Information Science (LIS) field which envisages papers coming from all over the world and which gives E-LIS the impetus and motivation to stimulate participation in the venture and to further develop international research activities. Finally, this paper also emphasises that the promotion of E-LIS further enhances the OA movement in general, so E-LIS can be regarded as a tool for the dissemination of the OA philosophy.

Introduction

One of the more recent developments in the field of Library and Information Science (LIS) is the trend towards digital libraries and self-archiving. Self-archiving can be defined as the deposit of a digital document in a public, free-access repository, for example, an e-print archive. An e-print archive is a collection of digital research documents such as articles, book chapters, conference papers and data sets. E-prints are the digital texts of peer-reviewed research articles, before and after refereeing. Before refereeing and publication, the draft is called a "preprint". The refereed, accepted, final draft is called a "postprint". The term e-prints include both preprints and postprints.

Established in 2003, E-LIS is an international, open access archive related to librarianship, information science and technology, and related disciplines, in keeping with the objectives of the e-prints movement [1] and the Free Online Scholarship (FOS) movement [2]. To date E-LIS is the biggest repository in LIS and after two years contains over 2200 papers. The E-LIS model is based on community standards which provide the mechanism for enforcement of proper attribution and responsible use of published works in line with the two above movements. On the technical level, it is based on the Open Archives Initiative (OAI) and shares its standards and protocols. E-LIS is the first international e-server in this area and is part of the RCLIS (Research in Computing, Library and Information Science) project [3]. The purpose of the E-LIS archive is to make full-text documents visible, accessible, harvestable, searchable and useable by any potential user with access to the Internet. It also aims to support individuals who wish to publish or make their papers available worldwide and it can be used by LIS communities in any country. Works can be deposited in any language and format; authors can self-archive, and a proxy service supports depositors. The archive is open to proposals from new partners and it has agreements with institutions and library associations of various countries. E-LIS is organised, managed and maintained by an international team of librarians working on a voluntary basis. The flexible architecture of the OAI is sub-divided into data provider and service provider. A data provider maintains one or more repositories (web servers) that support the OAI protocol as a means of exposing metadata. A service provider issues requests to data providers and uses the metadata as a basis for building value-added services. By self-archiving, authors can store a copy of their documents in a disciplinary or an institutional archive. As the OAI architecture enables documents to become swiftly available worldwide, self-archiving helps to maximize the visibility and accessibility of refereed research, and hence to maximize its usage by researchers and its impact on research.



Figure 1: The new layout of E-LIS, the international open archive for Library and Information Science, has recently become public <http://eprints.rclis.org/>. In the left-hand column there is a link (Information -> Brochures) to information and leaflets about E-LIS in different languages, <http://eprints.rclis.org/otherlang.html>

Self-archiving systems can be either centralised or distributed and are usually subject or institution-based. At present there are several subject-based archives, including ArXiv, CogPrints and E-LIS, which are centralised, and RePEC, which is distributed. There are also many institution-based archives but the advantage of subject-based archives, like E-LIS, is that they are specific to discipline needs and requirements. Many repositories use the OAI protocol and tools to facilitate interoperability between repository servers. The main characteristics of a centralised archive such as E-LIS are:

- Storage and distribution of data from a single location;
- Centralised access control over the supply and re-use of data;
- Checking, cleaning and processing of data according to standard criteria;
- Centralised support service for describing the contents of the data, defining the principles and practices governing the collection of the data, and other relevant properties of the data;
- Cataloguing of the technical and substantive properties of the data for information retrieval and,
- User support following the supply of data.

E-print repositories are complementary to, rather than a replacement for, scholarly journals. The peer-review process provided by journals is of critical importance to scholarship. However, the proliferation of institution- and discipline-based e-print repositories could accelerate changes that are currently taking place in the scholarly communication process which in turn could increase the number of journals moving from toll access (reader pays/subscription-based model) to open access (author-pays/free access model).

The evolving academic discourse surrounding the concept of Open Access (OA) provided the underlying motivation for the establishment of E-LIS. The library and information world is highly integrated with the areas of computing science and technology and it was felt that the LIS discipline should set an example to other communities by providing a state-of-the-art model for the OA movement and digital libraries, particularly in relation to the open archive model, within which E-LIS is a disciplinary repository. The extension of the OA concept to LIS works and the dissemination of material within the LIS community will contribute to the development of an international LIS network; E-LIS is mutually beneficial. For librarians, metadata creation is costly and the growing trend of authors to self-archive in the OAI framework is proving an effective way to reduce some of those costs. For librarians as authors, archiving their work in E-LIS gives them an increased understanding of the process of self-archiving and the E-LIS archive ensures data preservation and a wide data visibility in addition to facilitating active participation in the international librarian community.

E-LIS defines Open Access as a property of individual works and it adheres to the Bethesda Statement [4] which states that publication must meet two conditions:

1. *The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.*
2. *A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.*

The international context

There are similar initiatives within LIS, but they differ from E-LIS, in terms of organisation and vision, which in E-LIS is international. It is important to collaborate with these different but similar repositories in order to build a service provider for the LIS context. These repositories differ in their missions, aims and functions (see tables).

However, E-LIS differs from these models in that it is based on the voluntary work of people with different backgrounds and it has a non-commercial orientation (there is neither funding nor interest in receiving a revenue or profit). It is based on the spirit of the open source initiative in computing whereby people from all over the world, work together to build software for the public domain. E-LIS works to create a digital library which is freely available to the LIS community and is managed by an international team of volunteers comprising professionals from many different countries.

Table 1. NATIONAL REPOSITORIES

Australia

- *Australian Library and Information Association e-prints.*

<http://e-prints.alia.org.au/>

This is a new archive (2004) which contains 16 LIS papers from Australia.

France

- *CCSD: @rchiveSIC*
<http://archivesic.ccsd.cnrs.fr/>
- *CCSD: MémSIC*
<http://memsic.ccsd.cnrs.fr/>

@rchiveSIC (528 documents) and MémSIC (48 documents) are servers for author self-archiving - one for articles and working papers, the other for memories in the field of communication and information sciences (SIC). The majority of the articles are in French. Both the repositories were born from collaborations between many French institutions.

India

- *LDL, Librarians' Digital Library*
<http://drtc.isibang.ac.in/>

Located in the Documentation Research & Training Centre (DRTC), Indian Statistical Institute of Bangalore and based on the DSpace platform, it contains 172 documents.

Namibia

- *DSpace at the University of Namibia Library*
<https://dspace.unam.na:8443/dspace/>

It is not clear if this will be an institutional repository for the University of Namibia and managed by the Library, or if it will be set up for only LIS. At the moment only one collection has been created for holding library-authored documents. To date the server is empty.

Russia

- *DSpace at lib.usu.ru*
<http://dspace.lib.usu.ru>

Managed by and located at the Scientific Library of the Ural State University. It is based on the DSpace platform and contains 219 documents in Russian.

Table 2. REPOSITORIES OF LIS SCHOOLS

USA

- ***DLIST, Digital Library for Information Science and Technology.***

<http://dlist.sir.arizona.edu/>

DLIST is the Digital Library of Information Science and Technology, an OA, cross-institutional repository of full-text, electronic resources in the domains of LIS and Information Technology (IT). The server is implemented by the School of Information Resources and Library Science and the Arizona Health Sciences Library at the University of Arizona. It contains 354 documents and only accepts documents in English.

- ***SILS ETDs, University of North Carolina School of Information and Library Science (SILS).***

<http://etd.ils.unc.edu/>

Electronic Theses and Dissertations (ETDs) are digital representations of student Master's theses (or papers) and doctoral dissertations. SILS students are required to submit an electronic version as part of their degree. The repository contains 110 ETDs from the North Carolina School of Information and Library Science. It is based on the DSpace platform.

Table 3. LOCAL STAFF REPOSITORIES (reserved for the staff of certain projects)

Italy

- ***CNR Bologna Research Library EPrints Service.***

<http://biblio-eprints.bo.cnr.it/>

CNR Bologna Research Library Eprints Service is a repository for the self-archiving of research and technical papers in information science and computer science, and other disciplines pertinent to Library activities. The Eprints Service also hosts workshop proceedings and other technical material related to events organised by the CNR Bologna Research Library. It contains 10 documents. The repository relies on JITA, the E-LIS classification scheme, for the organisation of its documents.

USA

- ***CaltechLIB, Caltech Library System Papers and Publications.***

<http://caltechlib.library.caltech.edu/>

Caltech Library System Papers and Publications is an archive of papers and publications by the professional staff at the Caltech Library System. They have organized the papers according to the JITA classification scheme developed by E-LIS. To date, 29 documents have been deposited in the archive.

The E-LIS organizational model

Authors who contribute to an e-print archive are participating in a global effort by universities, researchers, libraries, publishers, editors, and readers, to redefine the mechanisms of scholarly communication. E-LIS will make LIS research more visible and accessible, which in turn will increase its status and public value.

E-LIS is divided into three sections: administrative, editorial and technical. A discussion list for each section provides the basis for action. The administrative section deals with strategic issues including the future direction of the initiative, its policies and their impact on the user community, and its connection with other scientific communities. The editorial section is devoted to metadata quality and guidelines. The technical section concentrates on the software - its implementation, enhancement, development, added-value functionality, and its operation within the OAI framework. These sections provide the over-arching structure within which the methods and procedures followed by E-LIS staff are established, maintained and developed.

Users perceive two main structures around which all the contents are organised; both structures have two levels. The first structure is the classification scheme for LIS (called JITA) divided into main subject divisions and subdivisions; the second is a geographical ordering based on continent and country. Users can navigate these structures to retrieve the required information. They can also view documents by browsing by author or year. Information can also be accessed via the search interface through a basic or an advanced search. In order to improve the search mechanisms, E-LIS has enhanced the full-text search furnished by the e-prints software it uses.

The submitter of a document is required to assign subjects and keywords from the E-LIS classification scheme: JITA [5]. JITA has recently opened a second-level, browse-by-country view which has over 120 sub-sections. This simple scheme, which has also been adopted by other repositories, is the result of the fusion and re-arrangement of the NewsAgentTopic classification scheme and the Review of Information Science (RIS) classification scheme. The JITA classification scheme is not intended to be a comprehensive classification scheme, but to facilitate document retrieval through the archive's browsing facility. It is divided into twelve blocks (categorised alphabetically from A-L) which have been created around the three following implicit (virtual) areas:

1. Theory and generalities (general level) This is divided into: theoretical and general aspects of libraries and information; information use and the sociology of information.
2. User-oriented, directional, and management functions (intermediate level) - socio-economic and legal issues are included here. This divides into: users, literacy and reading; libraries and information repositories; publishing and legal issues; management; industry, profession and education.
3. Objects, pragmatic issues and technicalities (on a specific level). This covers: information sources, supports and channels; information treatment for information services; technical services in libraries, archives and museums; housing technologies; information technology and library technology.

Since April it has been possible to browse by country. This gives a truly international aspect to the archive and is particularly aligned with the organisation of the editorial board whereby work is channelled through international staff on an individual country basis.

Strategic issues

The core of the organisational model is the administrative section which also has responsibility for determining the international and future vision of the archive, keeping in mind an understanding of national and international needs. In some countries, librarians want to create national archives for LIS instead of inserting papers into an established international archive, such as E-LIS. This is understandable because there are many technical difficulties associated with the use of different languages, alphabets, and non-alphabetic languages and with the consequent problems of input, output and the sorting of data and author names. E-LIS co-operates with each country individually to decide the best solution to the technical and non-technical barriers so that international visibility can be promoted whilst national interests are served. One idea, currently being debated, is to create a service provider with a harvester to gather national metadata from the national LIS archives and which could be part of the RCLIS infrastructure.

Thirty countries are currently involved with E-LIS and each of them works towards a common vision while maintaining their own cultural individuality. The editors are entrusted with the responsibility of promoting E-LIS within their own country and every editor has a different approach which is tailored to the specific needs of that country. The status of librarianship, which differs significantly from country to country, is the result of several factors such as the role of the professional association or of the LIS School, institutional assets, initiative on the digital library front, the impact of LIS in the social environment and how much OA has become a part of scholarly communication. However, all the editors share a common vision, and bring their own disciplinary and personal experience to bear on the position.

In E-LIS, papers can be viewed on an individual country basis, which highlights the internationality of the project, in addition to facilitating user access. In the LIS community most published papers traditionally came from the US, UK and Europe, but in E-LIS, where OA encourages a wider audience, it has been noted that high quality papers now come not only from the countries believed most innovative, but also from developing countries, such as India, and others in Africa, Asia, Eastern Europe, the Russian States and also Central and South America. This has been a very welcome, if unforeseen, development. The sense of this inclusiveness is highlighted by the recent collaboration with Cuban librarians, who have deposited papers in E-LIS from ACIMED, the Journal of Information Professionals in Health. The Open Access vision of E-LIS can be juxtaposed against the blockade imposed by the United States on Cuba, which is designed to prevent the publication of Cuban papers in the main U.S. journals. The Cuban case illustrates that E-LIS knows no borders, and negotiations are in progress regarding the involvement of Arabic countries, Israel and China. Collaboration with other countries has been very beneficial for E-LIS. In particular, Indian librarians have demonstrated a very high competency in bibliometric issues which has proved very stimulating for the E-LIS community as a whole.

E-LIS policies

E-LIS is driven and directed by its policies which determine its identity, quality and direction. It is not sufficient to create an archive merely by putting software on a machine, particularly in the case of E-LIS: an archive's organisational model is the sum of its policies and an archive without policies is like a library without a librarian. The principal policy concerns for E-LIS, which are discussed and democratically agreed upon by the editorial staff, are its:

1. Mission: its aims and objectives, what it is and where it is going, its target audience and the communities involved. Its international mission, (previously discussed) is clearly outlined on its website.
2. Submission policies: who can deposit material and how such deposits should be made (described in the following section).
3. Copyright policies (described below) are fundamental to any repository and are assigned due importance in E-LIS.
4. Organisational model (already described above), which is the core of any repository and determines its institutional or disciplinary nature.

Submission policy

Every librarian in the world can deposit papers in E-LIS as it promotes intellectual freedom and maximisation of impact in the LIS discipline. With E-LIS, as with arXiv in the physics field, experience has shown that this freedom has not led to the submission of low quality papers because authors are aware that their papers will ultimately be judged on a virtual basis by a very wide community of peers. The technical submission requirement is that authors who wish to submit a document must register in order to obtain a user I.D., which is also the basis for obtaining author-view browsing. Librarians, libraries, research institutes, organizations, and individual researchers involved in LIS and related fields are encouraged to make use of, and contribute to, the expansion of the archive and in turn to the critical mass of information available and useful for the building of digital libraries.

Figure 2: E-LIS submission interface

The E-LIS submission policy states that the archive accepts any scientific or technical document, published or unpublished, on librarianship, information science and technology or related activities. In this context, categories for different types of material have been created with respective sets of metadata. The criteria for acceptance are that the e-prints are relevant to research in LIS fields, and that they have the form of a finished document ready to be entered into a process of communication. Publications may include preprints, postprints, conference papers, conference posters, presentations, books, book chapters, technical reports/departmental working papers, theses, and newspaper and magazine articles.

Submitted documents are placed in the submission buffer for approval or rejection by E-LIS staff. Rejection is on the basis of pertinence to the archive. E-prints may also be returned to the author for metadata modification or if there are problems with the electronic file or format. Editors may make formal corrections but they do not make substantial modifications to the text or contents of the registered data. Documents in the submission buffer are manually reviewed to ensure they conform to E-LIS policy. E-LIS staff control the metadata quality of the document and are allowed to make changes if the metadata is incorrect. Generally, a paper becomes publicly accessible two working days after its deposit in E-LIS. The approval process is conducted by the editor from the country of the depositor. Once submitted, a document cannot be removed from the repository. Even when changing institution, Stevan Harnad argues [6] that "Wanting to remove one's work from the [repository of the] old Institution is [as] absurd as wanting to remove it from the shelves of one's old library - or any library". Indeed, removal would seriously hamper the communication process and users would be unable to chart the evolution of the idea or debate. Furthermore, because documents remain in the archive for posterity, the submission of inferior work is thus discouraged.

As an international open archive, E-LIS supports all languages; however, if a document is in a language other than English, it must include an English abstract and English keywords. If the English abstract is missing, the editor inserts it on behalf of the author. The following document formats are accepted: PDF, PostScript, Tex, LaTeX (DVI), HTML, XML, ASCII (text), PowerPoint, MS Word DOC and RTF. The use of HTML and PDF formats are strongly recommended.

Copyright policies

E-LIS does not want to infringe copyright. OA is encouraged where possible, but authors can restrict access to their papers if necessary. Access can also be restricted to the group of registered users of the E-LIS archive, which is a limited and known group of people. Furthermore, access can, in special cases, be restricted to only the depositor and archive staff (the archive administrator and any selected editors, evaluators, etc. in particular countries and organizations).

All work residing on the E-LIS server remains the property of the author. The author holds the copyright for the pre-refereed preprint and therefore, it can be self-archived without any other permission being sought. An author's works are that author's own intellectual property and they therefore own copyright and other proprietary rights until and if they grant otherwise. Authors submitting to the repository are responsible for ensuring the documents they archive do not have any restrictions on their electronic distribution imposed by a third party (such as a publisher). A pre-refereed preprint can be self-archived at a time when no copyright transfer agreement exists and so the author holds exclusive and full copyright; the author may no longer have the right to self-archive a refereed postprint if a copyright transfer agreement has been signed granting all rights to the publisher. In general, when an article is published in a journal, copyright is transferred to the publisher. Most journals permit self-archiving of the preprint and sometimes also the postprint, but it depends on the publisher's copyright policy. To avoid infringing any copyright, authors can deposit a postprint inside the archive with restricted access. Another way round this problem is for the author to request that the publisher allow them to retain certain rights, e.g. the right to deposit the postprint in an open access archive, or to place a copy on their homepage. Some publishers have stated that they grant these rights as a standard procedure. Alternatively, authors can replace the full text of the preprint with a link to the published version, if it is freely accessible. Copyright law gives to the creator of a work exclusive rights, which may be both segmented and transferred to others. Publishers have adopted various policies to facilitate author self-archiving. To help authors and editorial staff establish the copyright situation for deposit of a particular item, E-LIS uses the SHERPA [7] database of publisher copyright policies and self-archiving. SHERPA is a project investigating key issues in creating, populating and maintaining e-print collections, including specifically: Intellectual Property Rights (IPR), quality control, collection development policies, business models, scholarly communication cultures, and institutional strategies. In addition, E-LIS is aligned with the Rights Metadata for Open Archiving (RoMEO) project which cites the E-LIS policy on metadata [8]. E-LIS's metadata policy permits third parties to collect metadata from the archive via mechanisms that create end-user services to support the discovery and presentation of the archive content. The general policy is to allow harvesting of metadata but not the harvesting of full content.

Whatever the policy, the authoritative document for a published paper is the copyright agreement signed with the publisher. Fortunately, many publishers are adapting to the changing environment of electronic publishing. Nevertheless, some publishing policies contain ambiguities or have different approaches to self-archiving. For example, Elsevier, which publishes journals such as *International Information and Library Review* and *Library and Information Science Research*, has some restrictions regarding self-archiving. Its new policy appears friendly to self-archiving, but some important restrictions are evident in the details, for example, the conditions regarding deposits are as follows:

- Deposits must be on author's personal or author's institution's server;
- Published source must be acknowledged;
- Deposits must link to journal home page;
- Publisher's version/PDF cannot be used.

E-LIS can agree with the last three conditions, but operates in contradiction with the first condition which disallows deposit in a disciplinary repository such as E-LIS. According to Elsevier policy, therefore, authors can only deposit in two places:

a) on the author's personal web page. There is no co-ordination to facilitate the search and retrieval of scientific papers posted in the vast ocean of personal web pages.

b) on the institutional server. Institutional servers are not well developed or are non-existent in many universities throughout the world. The current reality is that there are very few institutional servers and

their archives are not well stocked.

Elsevier's decision to only allow self-archiving of preprints on personal or institutional servers deprives the author of the possibility of archiving in well-established, well-run, disciplinary repositories with good reputations. For many authors, a disciplinary archive is more relevant than an institutional one, particularly as many would prefer display their work and search for papers on a disciplinary, rather than an institutional, basis.

In an OA repository such as E-LIS, a user may find an exact copy of a paper from a journal which is normally only sold by a subscription. When a publisher's archiving policy is quite open it is possible to place a version (preprint or postprint) of a published paper in a repository. A user who knows how to use repositories can find high quality material in a free-access database, instead of acquiring it via a payment. Occasionally, a paper can be found in an aggregator site through a pay-per-view model (on average at \$30 per article), but in some cases papers are only found in electronic journals, some of which are "open access" and some "toll access". Some very high quality papers published in toll access LIS journals are available freely within an open access repository but it is imperative that archiving authors have a clear understanding of the different archiving policies adopted by publishers.

Editorial section

All work performed by the editorial section is developed by the editorial staff from discussion on a mailing list. Topics include metadata issues, guidelines for cataloguing, promotion of E-LIS and OA in general, and questions which arise from international co-operation. The principal editorial aim is to reflect the best practices of librarianship in each country by inviting the top scholars in the discipline to contribute to E-LIS. Therefore, the choice of editor for a particular country is crucial as they must be thoroughly conversant with the debates and personalities involved in the LIS disciplines in their country and they must also have the dynamism to promote E-LIS and the commitment, talent and patience for organising people, events and documents. Again, it must be stressed that this activity is currently done on an entirely voluntary basis, but it is evident that it has been of immense benefit to the LIS discipline as a whole. International co-operation can facilitate debate on current issues on many levels and provides the editors, on a personal level, with new professional experiences and expertise. Editorial tasks include:

- Creating and maintaining contact with university-based LIS academics and researchers and LIS-related course personnel;
- Approaching LIS publishers, with a view to obtaining permission to add articles from journals (both those already freely available on the web and those still using authentication systems) to E-LIS;
- Sending publicity e-mails to national mailing lists and submitting press releases to web-based publications;
- Writing literature reviews regarding Open Archive activities and writing journal articles on the E-LIS initiative.

Fundamental to obtaining high metadata quality are the repository's 'guidelines for cataloguing' which were created to facilitate editorial work and standardise metadata. The purpose of having the guidelines is to ensure that E-LIS adopts recognised, established practices for the creation of records. These guidelines assist the editors in assigning the correct value for each metadata field and were created taking into account the following points:

- Types of documents accepted
- Variety of contents accepted
- Who is submitting the material
- Metadata standards
- Advice concerning file formats and preservation of documents.

Each of the twenty-three document types has its own set of metadata which is checked in accordance with the editorial guidelines set by the international editorial committee. The twenty three document types are:

- Books;
- Book chapters;
- Preprints;
- Journal articles - online;
- Journal articles - print
- Conference proceedings;
- Conference papers;
- Conference posters;
- Presentations;
- Theses;
- Working papers - technical reports;
- Departmental technical reports;
- Reports;
- Project/business plans;
- Newspaper/magazine articles;
- Guides/manuals;
- Tutorials;
- Bibliographies;
- Library instructional material;
- Datasets;
- Syllabi;
- In collection;
- Other.

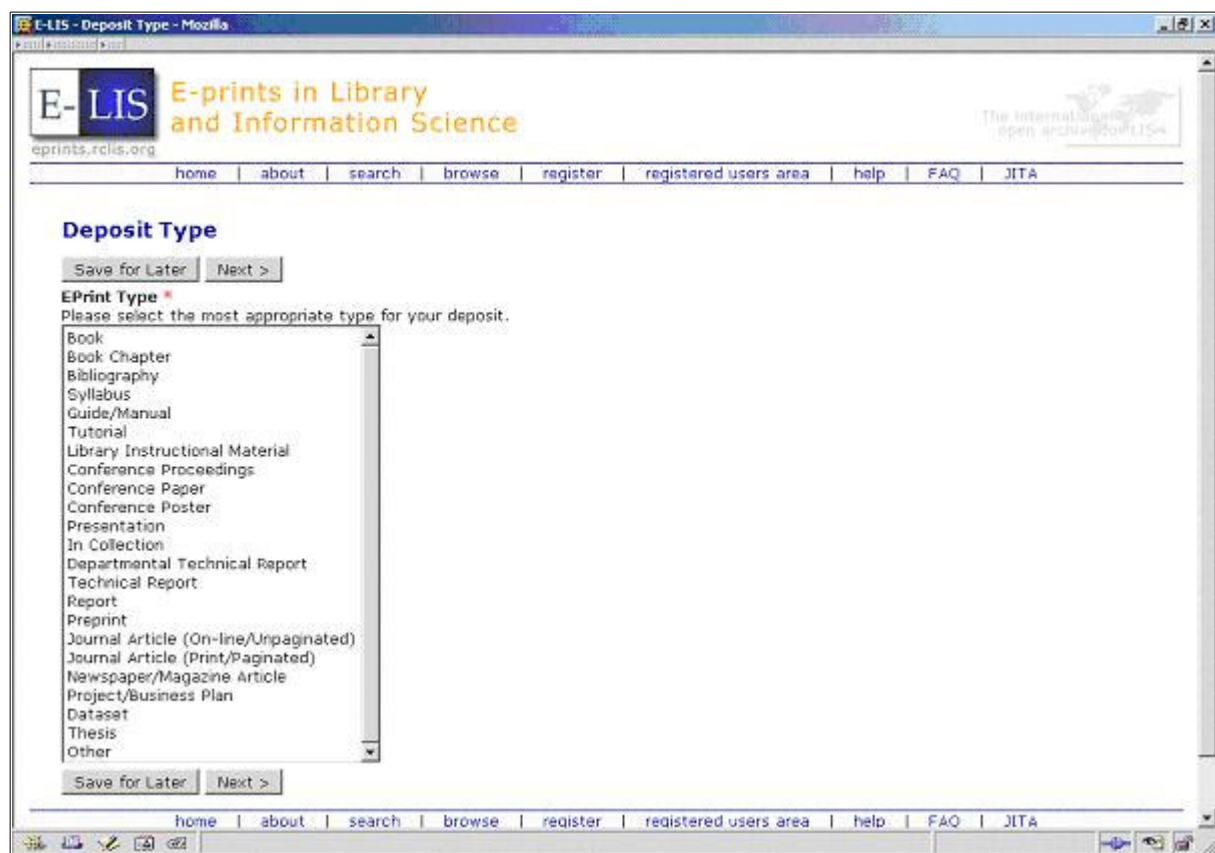


Figure 3: E-LIS interface showing the options for defining the type of document being deposited.

In addition to a common metadata core (Dublin Core) there is also a specific configuration of metadata for each category of these document types.

The technical section

Access to information is essential in a democratic society and the OAI framework helps to furnish tools and standards which give full access to information. Public health, the economy and public policy all depend on access to, and use of, information including copyrighted works. Technical solutions based on Open Source software, such as those adopted by E-LIS, are recommended to build open digital libraries. The OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The OAI Protocol for Metadata Harvesting (OAI-PMH) allows third-party services to gather standardized metadata from repositories and conduct searches amongst that assembled metadata to identify and ultimately retrieve documents. E-LIS is part of such a framework because the archive was built around OAI standards allowing the submitted records to be shared with, and contribute to, a growing global network of distributed, interoperable, institutional archives. Since the beginning of 2003, E-LIS has been validated within the OAI framework and has been registered in the official OAI register of data providers as an OA archive. All the metadata present in E-LIS is therefore exposed to harvesters, one of which is OAIster, the OAI service provider which harvests from all OAI archives.

The E-LIS server is hosted by the AEPIC team (<http://www.aepic.it>) of the Consorzio Interuniversitario Lombardo per L'Elaborazione Automatica (CILEA) in Segrate (Milano, Italy). CILEA is a non-profit organization consisting of nine universities in the Lombardy region of Italy. It provides information technology services on behalf of universities and related organizations, public organizations and enterprises. It also provides professional advice for both the planning and the dissemination of advanced technology in the fields of high performance computing, networking services and informatics. A full back-up is performed weekly and an incremental back-up is performed every day.

The archive is based on GNU EPrints archive software (version EPrints 2.3.7.1 (George) released on 12 March 2004). GNU EPrints generates archives that are compliant with the OAI-PMH (version 2.0). It originated from an idea of Stevan Harnad's, and was developed at the Electronics and Computer Science Department of Southampton University in the UK by Christopher Gutteridge, in collaboration with Mike Jewell (the design and development of version 1.0 was undertaken by Robert Tansley and dates back to 2001). A GNU EPrints server is an example of an OAI data provider, created for the self-archiving or mediated deposit of electronic resources.

The EPrints software is the most commonly installed and the most widely distributed of any of the repository software systems, according to the Open Society Institute [9]. From its origins in 1997 as being software designed to manage the Cogprints archive of papers on cognitive science, through to its current support for OAI-compliant institutional archives, development of EPrints software at ECS Southampton University has been largely funded by JISC. As a result, EPrints is free, GNU open-source software. E-LIS has a specific discussion list for technicians to discuss technical issues and try to find the best solutions for enhancing functionalities and developing new parts to the software that will then be put in the public domain.

The E-LIS service is accessible in two ways. First, a search engine can be used to search across the bibliographic descriptions (metadata). Three search screens are available for quick, simple or advanced searching. The simple and advanced searches can search within the full-text of documents. Second, views of the documents are offered for browsing. E-LIS users can browse by author/editor name, book/journal title, subject, year or country. Items deposited in the previous week can be viewed within the latest additions section. Having browsed or searched the archive, users can then view the metadata associated with a document, and download a copy if the access is not restricted.

Several added-value functions have been developed, for example, the layout of the home page and co-related pages has been re-styled and E-LIS now has its own logo. There is also a counter on the home page which displays the number of papers in the main archive in real time. This development was made possible by a Perl programme written by E-LIS. Furthermore, as E-LIS is a tool for librarians, a function - 'show all fields' - has been created which allows users to view an item's full list of metadata with field

labels.

The screenshot shows a web browser window with the title "E-LIS - Bibliographic Information - Mozilla". The page header includes the E-LIS logo, the text "E-prints in Library and Information Science", the URL "eprints.rciis.org", and a navigation menu with links: home, about, search, browse, register, registered users area, help, FAQ, and JITA. A banner on the right says "The international open archive for LIS".

The main content area displays the text: "This is complete contents of all the fields in this eprint:". Below this is a table of metadata fields:

ID Code:	3418
Type:	Journal Article (Print/Paginated)
Deposited By:	Brown, Kristina
Deposited On:	18 March 2005
Status:	Published
Refereed:	Yes
Public Domain:	Yes
Authors:	Defert-Wolf, Lidia and Dobrzynska-Lankosz, Ewa and Dziadkiewicz, Wanda and Gorny, Miroslaw and Gorska, Elzbieta and Gorski, Marek and Jazdon, Artur and Sokolowska, Anna and Widhardt, Teresa
Title:	Performance Indicators and Standards for Polish Research Libraries
Subjects:	G. Industry, profession and education. > G2. No one of these, but in this section. A. Theoretical and general aspects of libraries and information. > AA. Library and information science as a field.
Year:	2003
Abstract:	Performance Indicators and Standards for Polish Research Libraries Abstract: The article discusses proposed standards for Polish research libraries evaluation. At the beginning, the authors present the situation of research libraries in Poland. They write about the effects of the social-political transformation in the 90s, present selected statistical data, and describe the progress in computerisation. The following part of the article relates to the currently applied in Poland standards of library evaluation. Discussed are e.g. the presently applied tools for data collection and analysis. The last part includes proposed methods for the preparation of standards and assessments for Polish research libraries.
Publication:	Library and Information Research
Volume:	27
Number:	87

Figure 4: Display of an item using the link to 'show all fields' which displays all metadata fields and labels.

E-LIS uses ParaTools for reference linking, which is a set of Perl modules for parsing citations (the references at the end of papers). E-LIS has installed an enhanced version of ParaTools so that citations point directly to documents present on the Web and an indication is given if the cited document is itself in the E-LIS archive. ParaTools' main library uses a template-matching technique to extract metadata from citations, while secondary libraries provide support for OpenURLs and reference extraction from documents. Its main aim is to perform accurate citation parsing and there are specific modules within E-LIS to support this function. Another module can take parsed metadata and create/enrich OpenURLs and there is experimental support for extracting references from files (including PDF, PostScript and HTML). Other modules will be added as new ideas are proposed.



Figure 5: ParaTools is used to extract metadata from the citations and to make a link which will seek the item on the Web

Usage statistics

Since March 2004, monthly usage statistics for hits, visits, and downloads have been available on the statistics page. Older data from the previous hardware system (September 2003 - March 2004) are still available on another page.

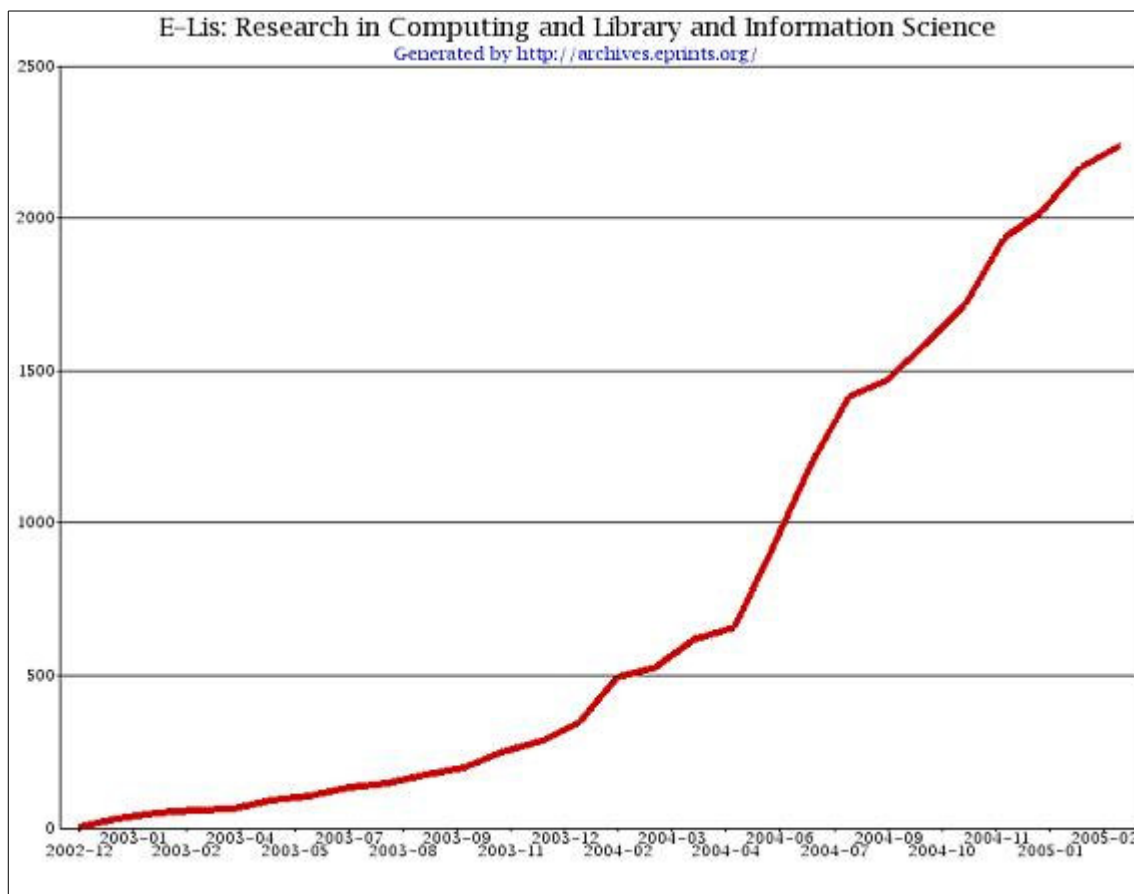


Figure 6: The size of E-LIS as measured by Celestial

The number of records in E-LIS can be measured by taking the date-stamp of records harvested from E-LIS by Celestial and is available at the Institutional Archive Registry [10]. The Institute Archive Registry tracks the number and size of open-access, e-print archives. The code behind this listing was written by Tim Brody, based on the GNU EPrints listing by Chris Gutteridge, with input from Steven Harnad, Les Carr and Steve Hitchcock and runs on Apache/PHP/Perl. The registry has two functions:

- to monitor overall growth in the number of e-print archives and
- to maintain a list of GNU EPrints sites (the software Southampton University has designed to facilitate self-archiving and used by E-LIS).

Conclusion

The challenges and opportunities created by a vision of a global collection of papers in the LIS field gives E-LIS the impetus and motivation to stimulate participation in the venture and to further develop international research activities. Furthermore, the promotion of E-LIS also enhances the OA movement in general, so E-LIS can be regarded as a tool for the dissemination of the OA philosophy, particularly in terms of strategic issues such as (explained further below):

- reporting statistics on authors and single papers;
- adding Creative Commons licences to metadata in the archive;
- involving other LIS partners by wide promotion in every country;
- funding and long term sustainability;
- enhancing technical functionalities;
- enriching metadata on the basis of the AMF (Academic Metadata Format) model;
- building of a LIS Service Provider for LIS.

E-LIS members believe that authors need to have appropriate tools and ways of maximising the impact of

their research papers in order to improve scholarly communication. The production of reliable and relevant statistics structured on the level of author output and single papers is an important and strategic issue. The development of tools to do this will play a leading role within national frameworks in respect of the evaluation of research channels, and on an international basis in order to obtain a common model where all intellectual output can find wide visibility and can be used and disseminated between different scholarly communities. Therefore, the maximisation of author visibility in the LIS field is a priority for E-LIS. E-LIS wants to focus its future efforts in this direction, and will endeavour to pursue agreements for collaboration so that funding for the development of new software can be obtained. The long term sustainability of E-LIS is another matter of great importance, strongly correlated with its partnership and future alliances.

As previously stated, E-LIS is engaged on several fronts. One future development will be the creation of a service provider tool. METALIS [\[11\]](#), a prototype service provider for the LIS field has been developed which currently harvests metadata from nine institutions (repositories) that offer full-text papers and documents about LIS. A possibility could be to combine the advantages of using centralised archiving with a distributed system composed of national repositories for different countries. The creation of a distributed network archive, within an RCLIS infrastructure, could be subject-based and would involve using several archives. In a model such as this, information would always be retrievable, no matter where it was deposited.

The possibilities offered by Creative Commons (CC) [\[12\]](#) are of great interest, particularly Science Commons - an exploratory project to apply the philosophies and activities of Creative Commons in the realm of science, and whose mission is to encourage scientific innovation by making it easier for scientists, universities, and industries to use literature, data, and other scientific intellectual property and to share their knowledge with others. Creative Commons works within current copyright and patent law to promote legal and technical mechanisms that remove barriers to sharing knowledge. Financial or human resources need to be found to develop software that can be added to the EPrints software, which will manage CC licences attached to the metadata submitted by authors during the self-archiving process. These licenses will not release work into the public domain, but they will encourage creative re-use of work in ways that full copyright protection does not. Licenses help authors retain their own copyright while allowing certain exceptions in certain conditions. The CC lawyers have worked to ensure that licences are enforceable in as many jurisdictions as possible. Furthermore, the practises of the RoMEO project regarding metadata were considered when developing the CC set of rights expressions to meet the requirements of academic research papers. The licences have three incarnations: a simple "human-readable" version, a "lawyer-readable" licence document, and machine-readable rights metadata.

As the RoMEO project decided on the use of CC licences to express rights over academics' open access research papers and metadata, E-LIS will follow suit. The joint OAI/RoMEO Technical Committee is working to create opportune guidelines and is also exploring a methodology for disclosure and harvesting the rights expressions inside metadata via the OAI-PMH. OAI/RoMEO suggests a five-pronged approach:

- Default, repository-wide rights expressions covering the metadata;
- Default, repository-wide rights expressions covering the resources;
- Optional, set-level rights expressions covering the resources;
- Rights expressions covering the individual metadata records;
- Rights expressions covering the individual resources.

E-LIS needs to analyse how it should adopt the correct metadata and CC licences, at which level of granularity the new metadata rights management procedure should be implemented, and at which step in the self-archiving process it should be introduced.

Last but not least, is the issue of enriching the metadata. The relationship between the OAI protocol and Dublin Core (DC) was discussed during a breakout session co-ordinated by Carl Lagoze during the CERN Workshop Series on Innovations in Scholarly Communication: Implementing the benefits of OAI (OAI3) [\[13\]](#). Since the first release of the OAI-PMH, unqualified Dublin Core has been the basic metadata format,

useful for interoperability. There was much discussion about the quality of the metadata being created by non-qualified people during the self-archiving process. With regards to the question of using other metadata formats, it was suggested that the mandatory link to DC be kept, but that the DC .type and .identifier should be used and then the record linked to the richer metadata format. It is imperative that any solution should not invalidate previous work, in particular because the Dublin Core is intended mainly for resource discovery of "document-like objects", and OAI-PMH is being applied in a number of other contexts. DC is useful not only for indexing, but also for listing and browsing results. RCLIS is working to provide information to LIS communities about evaluating and perhaps changing this requirement which may be interfering with a good application of the protocol. Therefore, while maintaining Dublin Core as mandatory, the use of richer formats must be considered in parallel within the RDF framework:

- MARC coming from traditional library environments;
- LOM for learning objects placed inside e-learning platforms or learning object repositories;
- OLAC the Open Language Archives Community metadata set based on qualified Dublin Core, but allowing the use of extensions to express community-specific qualifiers;
- AMF - Academic Metadata Format: a particular format to support scholarly communication developed by Thomas Krichel and Simeon Warner [14].

Another alternative for E-LIS could be to consider Dublin Core as non-preferred by default, and to use other rich metadata, such as the AMF, more tailored to scholarly communication needs. The AMF format is associated with a conceptual framework that comprises four classes of entities: resources, groups of resources, people and institutions.

The entirely voluntary character of E-LIS must again be stressed, as it gives a particular connotation to this support system which facilitates co-operation, promotes consensus, forms of expression and the identities of OAI group members, and nurtures relationships between different actors and countries in this wide, international community. E-LIS is confident of its progressive and comprehensive nature in technical matters and content, and particularly in the internationality of its vision and philosophy, and in its attempt to enhance and improve professionalism within the discipline. However, it is not lacking challenges, particularly concerning its budget and the resources needed for a more incisive performance, and as grants and other sources of funding are elusive, the success of E-LIS is largely based on the voluntary efforts of its members.

Announcement

We would like to take this opportunity to inform you that in October 2005 the first international conference of E-LIS will be held in the Geneva offices of CERN. Additional information is available on the conference website <http://www.aepic.it/conf/index.php?cf=4>.

Finally, we would like to inform those that are interested that during the next IFLA conference (in Oslo 14-18 August 2005) the new E-LIS poster, completed thanks to the sponsorship of AIDA (Italian Association for Advanced Documentation) <http://www.aidaweb.it/ingaida.html>, will be presented.

References

[1] Stevan Harnad. "The self-archiving initiative: freeing the refereed research literature online"
[Published in Nature 410 (26 April 2001): 1024 - 1025]

URL: <http://www.ecs.soton.ac.uk/~harnad/Tp/nature4.htm>

[2] Peter Suber. Open Access News (Formerly FOS News).

URL: <http://www.earlham.edu/~peters/fos/fosblog.html>

[3] RCLIS is a project dedicated to building a database of current and past documents in computing,

librarianship, information science and technology, and related application activities.

URL: <http://rclis.org/>

[4] "Definition of Open Access Publication" [Drafted during the Open access publishing conference April 11, 2003] Open Access Now : Bethesda principles.

URL: <http://www.biomedcentral.com/openaccess/bethseda/>

[5] JITA Classification Schema of Library and Information Science.

URL: <http://eprints.rclis.org/jita.html>

[6] Stevan Harnad. Email: "Re: Central versus institutional self-archiving" on *American Scientist Open Access Forum*. Posted Tue Nov 02 2004.

URL: <http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/4112.html>

[7] SHERPA is part of the JISC FAIR (Focus on Access to Institutional Resources) Programme which supports projects aiming to achieve the "disclosure of institutional assets" with the vision of setting up a "web of resources built by groups with a long term stake in the future of those resources, but made available through service providers to the whole community of learning."

URL: <http://www.sherpa.ac.uk/index.html>

[8] "E-LIS Copyright issues".

URL: <http://eprints.rclis.org/copyright.html>. Listed as "E-LIS Deposit Agreement" and linked from Web page: Project RoMEO - Licenses. URL:

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/Romeo%20Licences.htm>

[9] Raym Crow. "A Guide to Institutional Repository Software". Open Society Institute.

URL: <http://www.soros.org/openaccess/software/>

[10] Institute Archive Registry.

URL: <http://archives.eprints.org>

[11] METALIS is built by Zeno Tajoli of the E-LIS staff. Part of the code is from scripts developed by Simeon Warner, Alessandro Tugnoli and UKOLN.

URL: <http://metallic.cilea.it/>

[12] Creative Commons

URL: <http://creativecommons.org>

[13] CERN Workshop Series on Innovations in Scholarly Communication: Implementing the benefits of OAI (OAI3). 12-14 February 2004.

Agenda: <http://agenda.cern.ch/fullAgenda.php?ida=a035925> .

URL of workshop: <http://cern.ch/oai3/>

[14] Thomas Krichel and Simeon M. Warner. "A metadata framework to support scholarly communication" in International Conference on Dublin Core and Metadata Applications (DC-2001) October 22-26, 2001 (National Institute of Informatics, Tokyo, Japan, 2001).

URL of paper: <http://openlib.org/home/krichel/papers/kanda.html>

Further reading on E-LIS

1. Lewandowski, Dirk (2004) E-LIS : E-prints in Library and Information Science. In Proceedings Internationales Symposium fur Informationswissenschaft (ISI).
URL: <http://eprints.rclis.org/archive/00002532/>
2. Subirats Coll, Imma and Barrueco Cruz, José Manuel (2004) Un archivo abierto en ciencias de la documentacin e informacin. El Profesional de la Informacin 13(5): pp. 346-352

- URL: <http://eprints.rclis.org/archive/00002472/>
3. Medeiros, Norm (2004) A repository of our own: the E-LIS e-prints archive. OCLC Systems & Services 20 (2): pp.58-60.
URL: <http://eprints.rclis.org/archive/00001655/>
 4. Kumar, Anil and Kalyane, V. L. (2004) Bibliographics for the 983 eprints in the live archives of E-LIS : trends and status report up to 7th July 2004, based on author-self-archiving metadata.
URL: <http://eprints.rclis.org/archive/00001927/>
 5. De Robbio, Antonella (2003) E-LIS: un Open Archive in Library and Information Science. Bibliotime VI (1) (Italy)
URL: <http://eprints.rclis.org/archive/00000201/>
 6. De Robbio, Antonella (2003) E-LIS: un open archive per library and informations science. AIB Notizie 15, 2003(2): pp.12. (Italy)
URL: <http://eprints.rclis.org/archive/00002154/>
 7. Barrueco Cruz, Jose Manuel and Subirats Coll, Imma (2003) RCLIS: towards a digital library for Information Science. In Proceedings Libraries in Digital Age (LIDA), Dubrovnik and Mljet (Croatia).
URL: <http://eprints.rclis.org/archive/00000352/>
 8. Subirats Coll, Imma and Arencibia Jorge, Ricardo and De Robbio, Antonella (2004) E-prints for Library and Information Science (E-LIS) : la tecnología al servicio de la bibliotecología y las ciencias de la informacin. ACIMED 12(6). (Cuba)
URL: <http://eprints.rclis.org/archive/00002849/>

Author Details

Antonella De Robbio

E-LIS Manager

<http://eprints.rclis.org/>

CAB Centro di Ateneo per le Biblioteche

University of Padua, Via Anghinoni, 3 - 35121 PADOVA (Italy)

Tel: +39 049 8273654

Email: antonella.derobbio@unipd.it

URL: <http://www.math.unipd.it/~derobbio/home/schedaeng.htm>

Antonella De Robbio is the Coordinating Manager of the Library System of the University of Padova, the University Centre for Libraries (CAB). Since 2002 she has been responsible for the Projects Section and Digital Library: digitalization, eprints repositories (open access in general), bibliographical services to support e-learning. Until 2002 she was responsible for the Mathematics Library of the University of Padua, where in collaboration with other authors, she developed a set of tools for displaying and connecting classification schemes, applied to the MSC, ZDM, PACS and CCS, as well to DDC. She is the Italian coordinator of MAI the Italian MetaOpac (since 1999). In 2003 she implemented E-LIS of which she is the Manager.

Imma Subirats Coll

E-LIS Editorial staff coordinator

<http://eprints.rclis.org/>

Generalitat de Catalunya

Departament de Política Territorial i Obres Públiques

Av. Josep Tarradellas, 2-6 Barcelona (Spain)

Tel: +34 626 25 40 57

Email: imma.subirats@gmail.com

URL: <http://openlib.org/home/subirats/>

Imma Subirats Coll works at the Ministry of Territorial Policy and Public Works, Government of Catalonia. Since 2003, she has been working on E-LIS as a member of the coordinating editorial staff. Her research interests include the effects of open access on scholarly communication and the librarians' role in the development of open archives.

For citation purposes:

Antonella De Robbio and Imma Subirats Coll, "E-LIS : an international Open Archive towards building Open Digital Libraries ", *High Energy Physics Libraries Webzine*, issue 11, August 2005

URL: <http://library.cern.ch/HEPLW/11/papers/1/>

Reader Response

If you have any comments on this article, please contact the [Editorial Board](#)

[Top](#)

[Home](#)

[Editorial Board](#)

[Contents](#)

[Issue 11](#)

Maintained by: [HEPLW Team](#)

Last modified: 26 July 2005