

## **Un archivo abierto en ciencias de la documentación e información**

Por Imma Subirats Coll y José Manuel Barrueco

- Diciembre 2003 -

**Resumen:** *E-LIS (E-prints in Library and Information Science) es un archivo abierto dedicado a las ciencias de la documentación e información. Se trata de una biblioteca digital a nivel internacional en funcionamiento desde enero de 2003 que tiene el objetivo de difundir la filosofía open access en nuestra disciplina. Se describen brevemente sus características y funcionalidades principales así como el contexto donde se enmarca y el porqué de su creación. Se explica cómo el uso del software GNU Eprints ha sido determinante en su actual estructura. Finalmente se ofrece a la comunidad de bibliotecarios, documentalistas y archiveros el uso de este archivo no sólo como fuente de información donde pueden recuperar los documentos a texto completo sino como archivo abierto desarrollado con el ánimo de difundir dentro de nuestra profesión los cambios que se están llevando a cabo en el sistema de comunicación científica.*

**Palabras clave:** *Bibliotecas digitales, Protocolos de internet, Metadatos, Eprints, Archivos abiertos, Acceso abierto*

**Abstract:** *E-LIS (E-prints in library and information science) is an open archive devoted to library and information sciences. It is an international digital library that has functioned since January 2003, with the goal of spreading the open access philosophy within our community. Its characteristics and functionality are described briefly, as are the background and reasons for its creation. The article explains how the use of the GNU Eprints software was a determining factor in the archive's current structure. In concluding, E-LIS is promoted not only as a source of information for retrieving full text documents, but also as an example for the LIS community of open archives and of the changes that are affecting all scholarly communication.*

**Keywords:** *Digital libraries, Internet protocols, Metadata, Eprints, Open archives, Open access*

### **Introducción**

El trabajo que presentamos tiene como objetivo divulgar entre los profesionales de nuestro país la existencia de una biblioteca digital en ciencias de la documentación e información: *E-LIS (E-prints in library and information science)*. Se trata del primer archivo abierto desarrollado en España y, aunque lleva tiempo en funcionamiento, aún no ha sido ampliamente utilizado por los bibliotecarios de nuestro país. El desconocimiento de las iniciativas a nivel internacional que se están llevando a cabo entorno al áclo de la comunicación científica es quizás una de las causas de por qué el uso de este archivo abierto no ha sido tan extendido como en el caso de otros países. <http://eprints.rclis.org>.

Nuestra intención es proponer a todos los profesionales que trabajan en ciencias de la documentación e información la utilización y explotación de este archivo. Para ello la

sección 1 dará una visión del contexto en que se enmarca *E-LIS* dentro de los cambios que se están produciendo en el sistema de comunicación científica. En el punto 2 se sitúa el proyecto en el ámbito de las ciencias de la documentación y se justifica la necesidad de crear un archivo en esta disciplina. Con posterioridad se describen cuáles son sus características así como su funcionamiento. La sección 4 propone una mayor intervención y uso del archivo como una de las mejores maneras para conocer cuáles son los cambios que se han producido, se están produciendo y se producirán en nuestro entorno como profesionales dedicados a ofrecer información a otras disciplinas.

## 1. El contexto

Desde la aparición de internet han sido muchos los proyectos surgidos a nivel internacional con el objetivo de mejorar el sistema tradicional de comunicación científica. Tal mejora puede concretarse en tres puntos: aumentar la rapidez de distribución de los trabajos por medios electrónicos, reducir los costes de las publicaciones científicas y aumentar la visibilidad de los trabajos publicados.

Estas iniciativas se podrían categorizar en tres grupos:

- Creación de repositorios de documentos. El primero y más importante fue el creado en 1991 en *Los Alamos National Laboratory* (EUA) por **Paul Ginsparg** para recoger las prepublicaciones de física. Actualmente proporciona acceso gratuito a cerca de 250.000 documentos (tanto prepublicaciones como trabajos publicados en revistas) y se ha convertido en la principal fuente de información para físicos de todo el mundo. Su éxito llevó a la aparición, a lo largo de la década de los 90, de iniciativas similares en otras disciplinas: *CogPrints* en psicología, *Ncstrl* en informática, *RePEc* en economía, etc.

- Creación de revistas electrónicas. En los 90 surgieron también las revistas científicas electrónicas editadas por centros de investigación gracias a las facilidades de distribución por medios informáticos y con costes muy reducidos que les brindaba internet. Nunca antes había sido posible llegar a una audiencia tan grande con unos recursos tan pequeños y prescindiendo de unos intermediarios necesarios en el mundo impreso: los editores comerciales. La mayor parte de estas nuevas revistas se distribuyen gratuitamente ya que son editadas por y para la comunidad científica sin intereses

económicos y con el único objetivo de facilitar el acceso a los resultados de investigación por el mayor número de lectores posible. El *Directory of open access journals* recoge 350 publicaciones de este tipo a nivel internacional.

- Acceso abierto a la literatura científica. Este movimiento propugna maximizar la visibilidad de los documentos publicados mediante la eliminación de las barreras económicas que entorpecen el acceso a los mismos. Con el sistema actual ninguna institución, ni siquiera en los países desarrollados, puede acceder a la totalidad de la literatura publicada debido a los elevados costes de las revistas científicas. Este problema denominado “crisis de las revistas” ha sido ampliamente estudiado. Su efecto es la limitación en la difusión de los trabajos yendo en contra de los intereses tanto de autores que esperan sus que trabajos sean leídos por el mayor número posible de colegas, como para los lectores ya que sólo aquellos en instituciones con recursos económicos podrán acceder a la literatura científica. Entre las iniciativas que propugnan un acceso libre a las publicaciones científicas está la *Budapest Open Access Initiative (BOAI)*<sup>1</sup>. Su objetivo es alcanzar el acceso abierto por dos vías: el aliciente a los autores para que coloquen sus documentos publicados en archivos abiertos y en segundo lugar mediante el incentivo a la creación de nuevas revistas de acceso abierto.

Todas estas iniciativas demuestran que nos encontramos en un momento de redefinición del sistema de comunicación de la ciencia, en el cual se están investigando alternativas a los esquemas tradicionales. Son muchos los proyectos de investigación que se están llevando a cabo en nuestro entorno en este sentido. Por ejemplo, en EUA, la *National Science Foundation* ha financiado el trabajo técnico de la *Open Archives Initiative* y ha puesto en marcha la *National Science Digital Library* con un número muy importante de recursos de investigación de acceso abierto<sup>2</sup>. En el Reino Unido el *Joint Information System Committee* ha financiado proyectos como *Romeo* para investigar los derechos de propiedad intelectual en un entorno abierto. La Unión Europea ha financiado el *Open Archives Forum* destinado a difundir y potenciar los archivos abiertos en Europa.

Nuestro país ha estado al margen de estas iniciativas hasta el momento, con una inexistencia casi absoluta de proyectos nacionales relacionados con el acceso abierto a la literatura científica y una ausencia total en las reuniones e instituciones internacionales. Creemos que esta situación debe cambiar y que debemos sumarnos al resto de países de

nuestro entorno en la búsqueda de alternativas al sistema tradicional de comunicación de la ciencia.

## 2. E-LIS: un archivo abierto dirigido a los propios bibliotecarios

E-LIS (figura 1), el primer archivo abierto internacional en ciencias de la documentación, se establece como servicio innovador con el fin de difundir la filosofía de acceso abierto (open access) en nuestra disciplina. Sin duda alguna, el colectivo bibliotecario tiene un rol fundamental en el desarrollo del cambio que se está generando en la comunicación científica y, por lo tanto, es necesario que los bibliotecarios dispongan de un archivo abierto para probar, difundir y mantener una base de datos relacionada con uno de los cambios más importantes que se están produciendo en el inicio de este siglo en nuestro entorno.

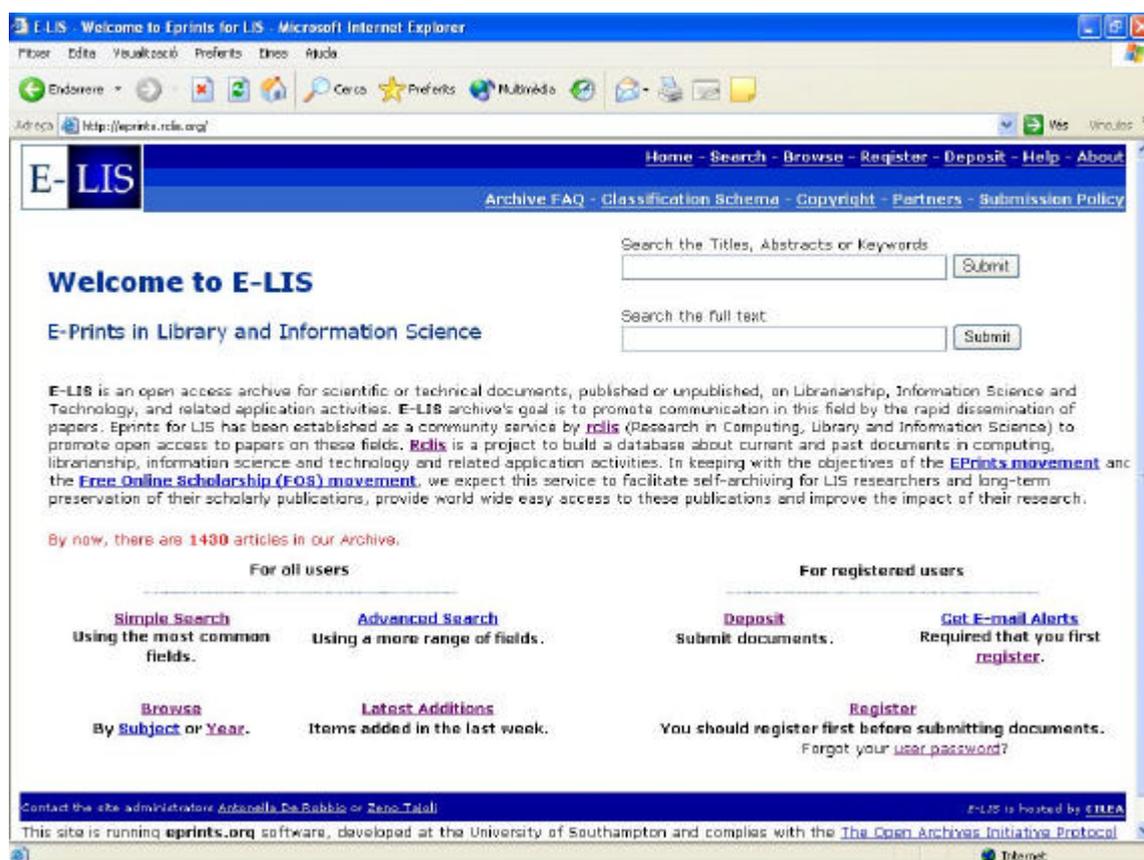


Figura 1. Página inicial de E-LIS

*E-LIS* es una evolución natural de *RcLIS* (*Research in Computing, Library and Information Science*) y *DoIS* (*Documents in Information Science*). *RcLIS* es un proyecto dirigido a construir una base de datos de documentos en informática, biblioteconomía y ciencias de la información así como disciplinas relacionadas. Está coordinado por **Thomas Krichel**, fundador de una de las más grandes redes de archivos abiertos en el mundo: *RePEC* (*Research Papers in Economics*).

*DoIS* es un servicio que lleva más de 3 años ofreciendo documentos (en estos momentos dispone de más de 16.000, 9.000 de los cuales son accesibles a texto completo) en ciencias de la documentación e información basándose también en un arquitectura distribuida en el que el trabajo de descripción se reparte entre un conjunto de instituciones. Estas organizaciones, habitualmente bibliotecas o investigadores a título personal, participan de varias formas, dependiendo si proporcionan datos o documentos (archivos) o si ofrecen servicios finales a los usuarios (servicios) o, finalmente, si ofrecen ambos. Esta estructura sigue unas reglas fijas definidas en el *Protocolo de Guildford*, con objeto de permitir a los servicios el fácil acceso a los datos.

En la actualidad *DoIS* cuenta con **Julio Alonso Arévalo** de la *Universidad de Salamanca* y **Antonella De Robbio** de la *Universidad de Padova* (Italia) como responsables del 95% de los registros que almacena. Aunque ambos están llevando a cabo un excelente trabajo, en cierta manera la filosofía de *RcLIS* no ha conseguido desarrollarse plenamente. El equipo de *DoIS* se ha encontrado a lo largo de estos años con serios problemas para establecer colaboraciones con instituciones y proveedores de metadatos. Si a esto le sumamos que la evolución de este tipo de repositorios se enlaza con el cambio en la información científica (open access) y el desarrollo de nuevas tecnologías como el protocolo *OAI-PMH* (*Open Archives Initiative Protocol for Metadata Harvesting*) o softwares como *GNU E-prints*, *E-LIS* surge como una posible solución ante la escasa participación de nuestras instituciones ofreciendo directamente a los autores la posibilidad de difundir sus publicaciones. La flexibilidad de su arquitectura les permite almacenar sus documentos (*self-archiving*)<sup>3</sup>, ofreciéndolos de forma más rápida a su comunidad científica.

Aunque la iniciativa nació en España mediante una subvención del *Ministerio de Educación, Cultura y Deportes* se trata de un trabajo conjunto dentro del marco de un

equipo internacional donde bibliotecarios de países como Italia, España, Croacia, Bosnia, India, Indonesia, Sudáfrica, Pakistán, Alemania, Reino Unido o USA entre otros están colaborando en el desarrollo de un único archivo en nuestra disciplina. La difusión y la concienciación de los autores en ciencias de la documentación es uno de los objetivos primordiales de todos aquellos que participan en *E-LIS*.

### **3. Funcionamiento de *E-LIS***

*E-LIS* ofrece el *self-archiving* a todos los profesionales en las ciencias de la documentación e información que quieran publicar o dar acceso a sus documentos en internet, publicados o no, con el objetivo de que cualquier usuario potencial pueda recuperar, acceder y usar los documentos a texto completo. Es absolutamente libre para todos, el único requisito es que los autores se registren, proceso mediante el cual se les requiere un nombre e identificador. De hecho, el funcionamiento básico de *E-LIS* se basa en este registro como condición indispensable para el depósito.

Dispone de uno o más editores en base a la procedencia de los autores de los documentos. El nombre de editores va creciendo a medida que se van adhiriendo más países. En estos momentos están trabajando bibliotecarios de España, Italia, Reino Unido, Alemania, Cuba, India, Croacia, Bosnia y Herzegovina, Serbia y Montenegro, Indonesia, Sudáfrica, Pakistán y Turquía. Son los responsables del control de los metadatos y del correcto estado de los archivos electrónicos depositados. A su vez, el coordinador/administrador es el responsable tanto de la calidad del contenido del repositorio como de la coordinación de todo el equipo así como el responsable de la calidad del archivo desde un punto de vista técnico. En estos momentos la coordinación y administración se está llevando a cabo conjuntamente desde España y Italia. Cabe destacar que ninguno de los miembros del equipo recibe remuneración alguna por el trabajo que realiza. Por parte de *CILEA*, *Consorzio Interuniversitario Lombardo Elaborazione Automatica*, se está llevando a cabo el mantenimiento del software y del servidor de forma altruista.

Aunque *OAI-PMH* permite la recuperación de los metadatos y documentos procedentes de diferentes archivos y, por tanto, la coexistencia de varios en una misma disciplina, se esta potenciando la idea de mantener un archivo internacional. Las razones son simples:

- Mantener una coherencia en la base de datos: mediante el uso de un mismo esquema clasificación (*JITA classification schema*) y unos mismos criterios de entrada de los metadatos. No debemos olvidar que son los autores quiénes introducen la descripción bibliográfica de sus documentos. Se debe remarcar el hecho que desarrollar un archivo abierto en ciencias de la documentación cuenta con la ventaja respecto el resto de archivos abiertos de que los autores son bibliotecarios, archiveros o documentalistas, es decir profesionales acostumbrados en su gran mayoría a catalogar y, por tanto, que son conscientes de la importancia de introducir los datos de forma correcta para su posterior recuperación.

- La presencia de diversas lenguas: sin duda, esta coexistencia es difícil en cualquier base de datos. No obstante, poder trabajar de forma coordinada para paliar los efectos de ello sobre un mismo archivo lo minimiza, manteniendo por ejemplo la obligatoriedad de una traducción al inglés del resumen entre otras decisiones.

- El software: sus posibles actualizaciones y mejoras en los servicios sólo se deben llevar a cabo en un archivo, por lo que se evita la posible coexistencia de diferentes versiones de la misma aplicación.

Para poder coordinar a todo el equipo se ha definido un protocolo, *Beijing Protocol*<sup>4</sup>, en el cual se basan los principios de *E-LIS*. En la actualidad, está en funcionamiento la versión **2.0.**, a medida que aparecen nuevas necesidades para el desarrollo del archivo se modifican y amplían sus directrices.

La estructura de *E-LIS* está particularmente enlazada al software que usa. Dentro del marco del *Open Citation Project* dirigido por **Stevan Harnard** de la *Universidad de Southampton* (Reino Unido) se ha elaborado *GNU Eprints* con el objetivo de facilitar el *self-archiving* y el desarrollo de archivos abiertos. Aunque en estos momentos no es el único, es uno de los más usados en el mundo con esta finalidad. Se trata de un software libre, por lo tanto sin restricciones de uso, modificable y redistribuible conjuntamente con su código fuente siguiendo *GNU General Public License*. Es fácil de implementar con numerosas funcionalidades y potencialidades, y uno de los aspectos más importantes es que permite personalizar parte de su configuración.

Esta estructura proporciona a *E-LIS* las siguientes características:

1. Cuatro funciones principales:

- la búsqueda (simple o avanzada),
  - el *browsing* (por materia o año),
  - la gestión de usuarios (el registro, la cancelación, el servicio de alertas...)
- y
- el depósito de documentos.

Próximas actualizaciones permitirán añadir otras funciones como por ejemplo el *browsing* por autor o por revistas. La actual versión en *E-LIS* es *EPrints* 2.3.3, y aunque la 2.3.6. ya está disponible, aún se encuentra en fase de pruebas<sup>5</sup>, por lo que no se ha implementado en este archivo aún.

2. Tres interfaces de acceso:

- La del autor,
- la de los editores y
- la del administrador.

Cada una de ellas dispone de diferentes posibilidades de búsqueda y una gama de funcionalidades diferenciadas según su rol.

3. Cuatro estados diferentes para los documentos:

- Archivo general. Donde se localizan los documentos aceptados y validados recuperables a través de la búsqueda y el *browsing*.
- *Buffer*. Espacio donde los documentos esperan ser aprobados por los editores.
- *Workspace*. Donde cada uno de los autores registrados puede dejar sus documentos, recuperarlos cuando les sea necesario y decidir cuándo quiere que se incluyan en la base de datos.
- *Deleted articles*. Aquí se sitúan los documentos que han sido borrados del archivo general.

### **3.1. E-prints.**

*E-LIS* acepta cualquier documento científico, publicado o inédito, en ciencias de la documentación e información y en cualquier idioma, sólo requiere que el resumen y las palabras clave se traduzcan al inglés. El criterio básico para su aceptación es que debe ser relevante y debe tener la forma de documento acabado, es decir apto para su lectura.

Siempre será admitido cuando algún componente del comité editorial considere su pertinencia adecuada. En el caso que no fuera aceptado sería devuelto a su autor con una notificación del porqué de ello.

Es requerido el texto completo de un documento, aunque es posible poner todas las urls que un autor considere necesarias como lugares alternativos donde acceder a él. A su vez, un mismo documento puede ser depositado en varios formatos, no hay ninguna restricción aunque alguno de ellos debe ser en .pdf o .html.

En la actualidad hay dos acuerdos para el depósito: el mismo autor es quien deposita un e-print en *E-LIS* o bien un trabajo es depositado por alguien que no es su autor. En cualquiera de los dos casos los materiales depositados en *E-LIS* mantienen los derechos de propiedad intelectual del autor. El que un documento no tenga ninguna restricción en su difusión electrónica es responsabilidad de los mismos autores. El acuerdo de depósito se basa en el hecho que un autor que decide archivar un documento y su material bibliográfico asociado concede a *E-LIS* el derecho de almacenarlos y de hacerlos permanentemente disponibles en internet. El autor acepta que este material es de su propiedad y que *E-LIS* no asume ninguna responsabilidad de ello. Cuando un documento es depositado por otra persona, ésta debe asegurar que se encuentra en el dominio público, y si no es así debe aceptar cualquier responsabilidad si su distribución conlleva problemas de propiedad intelectual.

### **3.2. JITA Classification Schema.**

Antes de depositar un documento es recomendable consultar este esquema de clasificación, accesible desde la página principal. Se trata de un solo nivel jerárquico con una descripción de cada uno de los elementos en base a una agrupación de palabras clave:

*[Theoretical and general]: general level.*

a. Theoretical and general aspects of libraries and information:

- Library and information science as a field. Information theory and library theory. Relationship of library and information science with other fields such as computing and communication science and cognitive science.

b. Information use and sociology of information:

- Use and impact of information. Bibliometric methods. Information in society. Information society. Information economics. Information policy. Information dissemination and diffusion. Information needs and information requirements analysis. User interfaces, usability.

*[User oriented, directional, and management functionalities]: intermediate level (socio-economical and legal issues included).*

c. Users, literacy and reading:

- Use studies. User studies. User categories: children, young people, social groups. User training, promotion, activities, education. Literacy. Reading and story telling.

d. Libraries as physical collections:

- World libraries. National libraries. Public libraries. Academic libraries. School libraries. Government libraries. Private libraries. Special libraries. Science libraries. Technical libraries. Health libraries, medical libraries. Archives. Museums.

e. Publishing and legal issues:

- Mass media. Printing, electronic publishing, broadcasting. Bookselling. Intellectual property: author's rights, ownership, copyright and copyleft. Intellectual freedom. Censorship.

f. Management:

- Co-operation. Marketing. Finance. Public relations. Personnel management. Funding. Local government. Reorganisation. Unitary authorities.

g. Industry, profession and education:

- Information industry. Software industry. Computer and telecommunication industry. Organisations. Staff. Biographies. Curricula aspects. Education. Training.

*[Objects, pragmatics and technicalities]: specific level.*

h. Information sources, supports, channels:

- Periodicals, newspapers. Gray literature. Archival materials. Rare books and manuscripts. Print materials. Microforms. Non-print materials. Audio-visual, multimedia. Electronic media. Cd-rom. Online hosts. Databases and database Networking. Opac's. E-journal. E-books. E-resources. Web pages. Portals. Repositories (OAI-compliant and not).

i. Information treatment for information services (information functions and techniques):

- Cataloguing, bibliographic control. Content analysis: abstracting, indexing, classification. Index languages, processes and schemes. Data and metadata structures. Knowledge representation. Information transfer: protocols, formats (xml), techniques. Information presentation: hypertext, hypermedia. Image systems. Filtering. Reference linking. Design, development, implementation and maintenance of information systems and services.

j. Technical services in libraries, archives and museums:

- Acquisitions. Serials management. Withdrawals. Stocktaking. Record keeping. Paper preservation. Digitization. Digital preservation. Circulation. Document delivery. Interlibrary loans.

k. Housing technologies:

- Resource centres. Library, archive and museum buildings. Furniture. Vehicles. Architecture. Planning, design, removal. Safety. Disaster planning.

l. Information technology and library technology:

- Telecommunications. Computer networking. Internet, including www. Computers. Scanners. Digital cameras. Photocopiers. Computer and network security. Authentication, and access control. Software. Software methodologies and engineering. Automated language processing. Automatic text retrieval. Data base management systems. Object-oriented DBMS. Intelligent agents. Library automation systems. Opac systems. Search engines.

Se trata de un esquema clasificatorio muy simple, fusión del *NewsAgentTopic Classification Scheme*, mantenido por **Mike Keen** desde 1998, y del *RIS Classification Scheme* del *Review of information science* originalmente ideado por **Donald Soergel**.

### **3.3. Cómo depositar un documento en E-LIS.**

Se trata de un proceso muy simple y tiene como objetivo facilitar al autor la introducción de los documentos:

a. En primer lugar debe registrarse. Cuando han sido validados el correo electrónico y el nombre de usuario, se deben completar los datos personales. A su vez puede suscribirse al servicio de alertas a través del que recibirá diariamente, semanalmente o mensualmente notificación de los últimos e-prints que se han ido depositando en el archivo.

b. Cuando un autor decide introducir un e-print debe rellenar el formulario de información bibliográfica, donde el nombre de campos que se deben completar dependerá de la tipología del documento (figura 2). Cuando un autor introduce un documento, lo está llevando a cabo desde su área de trabajo (*workspace*), lo que supone su espacio personal como usuario registrado. Si decide no enviar el e-print inmediatamente, éste puede estar disponible en su área.

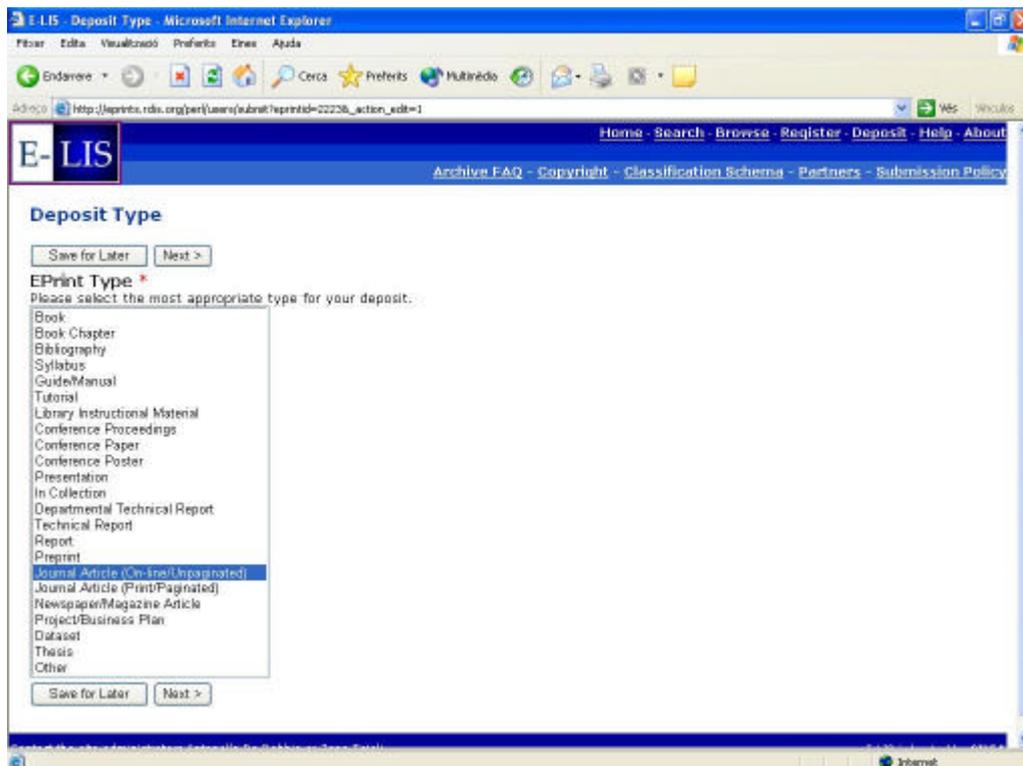


Figura 2. Tipología de documentos

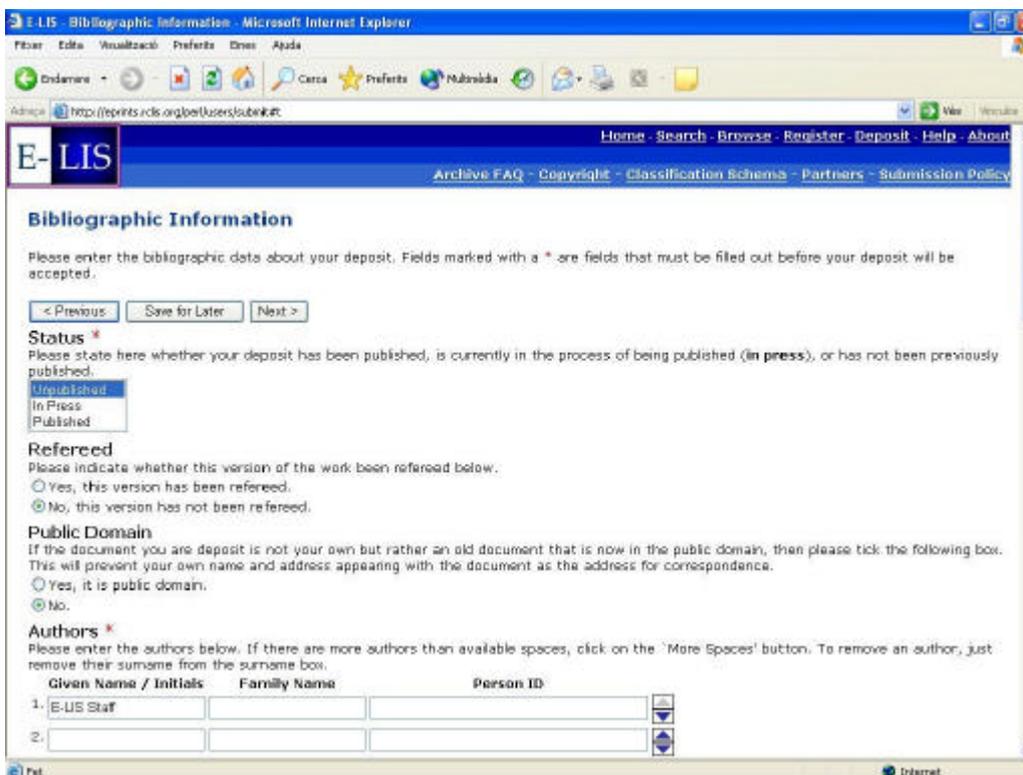


Figura 3. Formulario para la introducción de la información bibliográfica de un e-print

c. En el momento en que el autor decide enviar a publicar un documento, es transferido al buffer, donde uno de los editores procederá a la validación.

d. Cuando el editor lo ha aprobado se deposita en el archivo general donde ya es accesible para cualquiera.

#### **4. El futuro de E-LIS**

En estos momentos se está desarrollando una amplia difusión a nivel internacional. No sólo para que se incluyan documentos en la base de datos sino también con la intención de dar a conocer entre los profesionales de la documentación todo lo que envuelve a un archivo abierto y aquello que se está debatiendo acerca de posibles cambios en el sistema de comunicación científica. Se están llevando a cabo acuerdos con diferentes organismos, instituciones y organizaciones para difundir y explotar el repositorio. Se puede obtener más información al respecto en

*<http://eprints.rclis.org/credits.html>*

En estos momentos ELIS es el archivo abierto con más documentos en Ciencias de la información. Su evolución se puede observar en la figura 4. El gráfico se ha llevado a cabo mediante Celestial, elaborado por **Tim Brody** y accesible desde el *Institutional Archive Registry*.

El futuro de este archivo abierto está en manos de los propios bibliotecarios, no sólo como una herramienta útil donde podemos acceder al texto completo de todo su contenido, sino también para conocer más de cerca uno de los posibles cambios que se llevarán a cabo en el futuro.

#### **Notas**

1. *<http://www.soros.org/openaccess/>*

2. **Lagoze, Carl; Van de Sompel, Herbert**. "The making of the Open Archives Initiative Protocol for Metadata Harvesting". En: *Library hi tech*, 2003, v. 21, n. 2, pp. 118-128.

3. **Harnard, Stevan**. "The self-archiving initiative freeing the refereed research literature online". En: *Nature*, 2001, April, n. 410, pp. 1.024-1.025.

*<http://www.ecs.soton.ac.uk/%7Eharnad/Tp/nature4.htm>*

4. *<http://netec.wustl.edu/~subirats/Beijing/Protocol.htm>*

5. <http://software.eprints.org>

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

***Imma Subirats Coll, Dept. de Política Territorial i Obres Públiques, Generalitat de Catalunya.  
immas@correucatala.com ; wisubirats@gencat.net***

***José Manuel Barrueco, biblioteca de Ciències Socials, Universitat de València.  
jose.barrueco@uv.es***