

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Winter 1-12-2014

An Analysis of Subject Coverage and Worldwide Involvement of E-LIS: the International Repository for Library and Information Science.

Nestor L. Osorio

Northern Illinois University, nosorio@niu.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Osorio, Nestor L., "An Analysis of Subject Coverage and Worldwide Involvement of E-LIS: the International Repository for Library and Information Science." (2014). *Library Philosophy and Practice (e-journal)*. 1067. <https://digitalcommons.unl.edu/libphilprac/1067>

An Analysis of Subject Coverage and Worldwide Involvement of E-LIS: the International Repository for Library and Information Science.

Nestor L. Osorio

Northern Illinois University
University Libraries
DeKalb, IL 60115, USA
E-mail: nosorio@niu.edu

Abstract: E-LIS, E-prints for Library and Information Science, is a subject-oriented repository initiated in January 2003 by a group of European information specialists. E-LIS is an open-access repository run by experts and editors from many different countries and with holdings originating in 110 countries. The first purpose of this study is to provide a description of E-LIS with special attention to the types of documents archived and the geographical distribution of its contributors. The second purpose is to determine the subject coverage, which is done by using several well-known bibliometric techniques. Using co-occurrences of subject terms, a cluster analysis is performed, producing four major clusters; a correspondence analysis of keywords and subject terms produces eight groups of association.

Keywords: Library and information science; Documentation; Digital repositories; Content analysis; E-LIS; e-prints.

Introduction and Literature Review

The creation and development of E-LIS, E-prints for Library and Information Science (LIS) <<http://eprints.rclis.org/>>, has been described and discussed by several of its prominent leaders and editors (De Robbio, 2003; DeRobbio and Subirats, 2005; Medeiros, 2004; De Robbio, et. al., 2011). Initiated in January 2003 by a group of European information specialists, it has the characteristics of being a truly open access, international endeavor, and is operated by a group of volunteers. Two of E-LIS' creators and administrators are Imma Subirats Coll (Food and Agriculture Organization (FAO) of the United Nations) and Antonella De Robbio (University of Padua, Italy); it runs with the assistance of an Administrative Board, an Executive Board, and an Editorial Board. The Editorial Board currently includes fifty-six members from twenty-eight countries and five continents.

E-LIS is not the only repository exclusively collecting LIS documents. De Volter (2011) has reported the existence of four others, two in France, one from India, and one in the United States. These are of limited scope in terms of types of documents collected, and their coverage is national rather than international. They are:

France

- Casic, Archive Ouverte en Sciences de l'Information et de la Communication <<http://archivesic.ccsd.cnrs.fr/>>, number of items posted: 1571 (December 20, 2013)
- mémSIC, Mémoires de master en Sciences de l'information et de la communication <<http://memsic.ccsd.cnrs.fr/>>, number of items posted: 362 (December 20, 2013)

India

- LDL - Librarians' Digital Library <<http://drtc.isibang.ac.in:8080/handle/1849/1>>, number of items posted: 490 (December 17, 2013)

USA

- DLIST, Digital Library of Information Science & Technology <<http://dlist.sir.arizona.edu/>>, number of items posted: 1562 (December 20, 2013)

Many other repositories include materials related to the LIS field, but not necessarily exclusively; a search in the Directory of Open Access <<http://www.opendoar.org/find.php>> using the Browse function for Library and Information Science produced 119 sites.

The E-LIS support and coverage in several countries has been discussed by some editors of the project. Villegas and Angel (2009) made a historical review of repositories in Mexico, institutions involved in open access projects and their contribution to education. The first documents from Mexico were loaded in 2004; they discussed the presence of documents in the Spanish language in E-LIS (37 percent); the value to the LIS community of E-LIS in that country, and future expansion of the project. Dimitri (2009) made a quantitative analysis of the contributions to E-LIS by Argentinean authors: authors, principal cited researchers, type of documents loaded, languages used in the cited documents, and publications cited. He also identified the 13 most-used keywords assigned to the documents from Argentina. Ferreira and Neves (2013) discussed the participation of Portugal's LIS community in the E-LIS project, which started in 2006, and presented the state of research of the LIS field in Portugal. They also indicated that the number of documents deposited is very small compared to the national production.

Adamick and Reznik-Zellen (2010) in their survey article about the literature of subject repositories, included E-LIS as one of the top ten subject repositories sharing space with well-known sites such as PubMed Central (medicine), CiteSeer (computer science), RePEc (economics), ArXiv (physics), and SSRN (social sciences). Although, their survey revealed *"a lack of literature on the ten identified subject repositories"* and also that they found *"difficult to understand the full impact that subject repositories have had on their respective fields, as well as the impact subject repositories have on larger scholarly communication and digital library issues as distinct from institutional repositories."* A gap in understanding the function of subject repositories appears to exist; therefore, the content analysis presented here can supply additional information about this topic.

The first purpose of this study is to present a general description of E-LIS after the first ten years of its existence; second, using several well-known bibliometric techniques, we analyzed the subject coverage of the repository. This second purpose is interesting since, as an open-access repository, the technical infrastructure already in place allows for authors from all corners of the world to upload their documents, while the regional editors' role is to maintain the established standards and assuring that the documents loaded are LIS-related. In this wide-open environment in which authors volunteer their publications with the main objective of making them available, the main questions are: what LIS topics are covered in E-LIS?; beyond the simple observation of data from the JITA classification system, is it possible to identify an structure of topics covered?

General Description

In this section, a brief description of ELIS is provided, particularly as it is related to the types of documents archived and the geographical distribution of its contributors. This study covers all the documents loaded in the repository since its inception in 2003, which includes publications dated from 1965 to 2013.

Type of Documents

E-LIS allows for the archiving of many types of documents; listed in order of occurrence are: journal articles; conference papers; presentations; book chapters; preprints; theses; reports; other types of documents; conference posters; reviews; newspaper/magazine articles; books; guides/manuals; conference proceedings; bibliographies; project/business plans; datasets; and technical reports. Journal articles and conference papers account for nearly 70 percent of the repository.

Geographical Distribution

It is important to emphasize the international coverage of this service. E-LIS' geographical representation of countries is significant; there are documents originating from 110 countries, including twenty-one countries from Africa; twelve from North and Central America; ten from South America; twenty-six from Asia; thirty-seven from Europe; and four from Oceania. Figure 1 shows the distribution of documents by continents. As it was mentioned before, this is one of the characteristics that set E-LIS apart from other LIS repositories. In spite of efforts to reach out to publications from all over the world, Figure 1 shows that the highest number of contributors is from Europe, followed by North America, South America, Asia, Africa, and Oceania. Russia is considered part of the European countries.

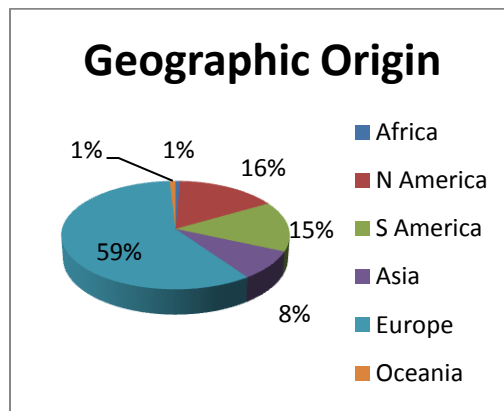


Figure 1. Geographic distribution of documents by continents

There are twenty-eight countries that have achieved the milestone of having at least 100 documents in the repository, see Table 2; leading the group are Spain and Italy which make up 34 percent of the database.

Table 2 - Countries with 100 or more entries

Country	Items	Country	Items
Spain	3675	Poland	407
Italy	1548	Colombia	316
Argentina	924	Serbia and Montenegro	243
United States	772	Chile	187
India	759	Peru	145
Brazil	633	Portugal	145
Mexico	614	Iran	144
Cuba	581	China, People's Republic	126
Germany	558	Switzerland	122
Austria	539	Indonesia	116
United Kingdom	488	France	115
Greece	476	Croatia	113
Turkey	437	Australia	100
Canada	418	Czech Republic	100

Methods and Data

In this section, the methods used and the data collection procedures are presented. Data were collected directly from the E-LIS website < <http://eprints.rclis.org/>>; they were obtained by utilizing the Browse and Search functions of the repository. Information about the distribution of countries and continents was obtained from the Browse facility, as well as data for the JITA distribution of subjects. This is a straight-forward procedure because the data is automatically stored and maintained by the producer. This study covers all the documents loaded in the repository since its inception in 2003; publication dates of the documents are from 1965 to 2013. The data was collected in September of 2013.

Data for the cluster analysis was obtained by performing co-occurrence searches of the JITA subject terms; all the 12 major classification areas (A-L) were included. For example, Class A was searched using the AND operator with Class B and so forth until all the classes were exhausted. This produced a matrix of 12 by 12 observations. Data for the correspondence analysis was obtained in two steps. First, a selected number of 13 keywords were identified, second, co-occurrence searches of keywords against the 12 major JITA classification terms were obtained; the result was a 13 by 12 matrix.

The method used in this paper for the use of cluster analysis mirrored the work done to visualize the structure of medical informatics (Morris, 2001) and the work done by Osorio and Osorio (2004) to study the literature of disasters from the human factors point of view. In these two projects, term co-occurrence analysis was used and cluster trees were produced.

Finally, correspondence analysis is mainly a descriptive tool; it is a graphical technique to represent associations of terms (Agresti, 2002). In this case, we explore relationships between selected keywords and JITA classification terms. Keywords assigned by authors are a common practice in vocabulary control, they are used to supplement the index terms or subject terms and allow the authors to contribute their subject expertise in the metadata aspect of the paper. The keyword searching technique is widely accepted. Nawaz et. al., (2001) built a keyword-base scheme and used it to retrieve the literature of engineering education research with the object of defining the field. The work of Hu (2013) is also critical to mention in our research for these reasons: by selecting keywords from major journals, they used co-word techniques and performed several multivariate statistical analyses to study the structure of LIS in China. The keywords element is important to reference because it is used by these two authors as a valid research tool for performing content and cluster analysis. In our project, keywords were selected by observing the terms and phases of the highest ranked secondary terms in the JITA system; 13 keywords were selected.

Results - Subject Coverage of E-LIS

In E-LIS, the author of a document must assign to a document being loaded two types of subject-related information. Keywords, which are selected by the author in order to freely describe the content of the document, and classification terms selected from the JITA Classification System. JITA is a two-level hierarchical-enumerative classification system which allows the classification of objects from main thematic areas to more specific topics. It is the creation of four information specialists (José Manuel Barrueco Cruz, Imma Subirats Coll, Thomas Krichel and Antonella De Robbio) and the JITA acronym represents its creators' first initials. JITA is based on two previously developed classification systems: the NewsAgentTopic Classification Scheme and the Review of Information Science (RIS) Classification Scheme (Bassi et. al., 2005; and Dal Porto and Marchitelli, 2006).

Therefore, documents indexed in the system allow for subject searching (JITA) and keyword searching and a combination of both. The search engine also allows for browsing the JITA subject terms

as well as authors, year of publication and countries. Others types of sear fields are also available. For this paper, JITA subject terms are important because they represent the main thematic areas of the repository.

Table 3 shows the twelve main thematic areas which are listed alphabetically from A to L. Each main thematic area is further expanded into more specific subject terms; at that level, a total about 160 subject terms are available.

Table 3. JITA Classification System of Library and Information Science

	Main thematic areas	Main area	Second level	Total
A	Theoretical and general aspects of libraries and information	1617	1044	2661
B	Information use and sociology of information	3906	4304	8210
C	Users, literacy and reading	1397	1434	2831
D	Libraries as physical collections	2442	2438	4880
E	Publishing and legal issues	1368	1350	2718
F	Management	1264	997	2261
G	Industry, profession and education.	1311	1452	2763
H	Information sources, supports, channels	3991	4431	8422
I	Information treatment for information services	2402	2339	4741
J	Technical services in libraries, archives, museum	1241	1137	2378
K	Housing technologies	137	168	305
L	Information technology and library technology	2814	2531	5345
		23890	23625	47515

Table 3 also shows the frequency of the assigned JITA subjects terms found in the database using the subject Browse capability. In total, 47,515 subject terms have been assigned to 14,317 documents, given an average of 3.3 primary or secondary subject terms assigned to an E-LIS document. The table shows in the third column the frequency of hits for primary thematic areas; in column four, the frequency of hits for secondary subject terms; and column five, the total of both. We have found some minor inconsistencies in the number of terms counted when using the Browse capability of E-LIS as compared to the results obtained from the Search engine. This anomaly cannot be explained; in most cases the percentage of discrepancy is relatively low, and therefore, dimensionally it is considered that the data obtained is a good representation of the actual data. The data from the Browse mode was chosen because is more consistent.

Based on Table 3, it can be determined that from a simple JITA classification point of view, the most common subject areas in the repository are Class B: Information use and sociology of information (8210, 17.28%) and Class H: Information sources, supports, channels (8422, 17.72%).

Co-occurrence

Table 4 shows the number of items retrieved when each major classification term of JITA was searched against each other (co-occurrences); the process systematically determined closed relationships between the terms used (Osorio and Osorio, 2004). These closed relationships would be the base for finding the subject clusters of the repository. The data in Table 4 was presented to SAS Pro Factor, following a similar a procedure used by Morris (2001), and the Ward's hierarchical cluster method was selected.

Table 4. Co-Occurrence values of JITA main subject terms

	A	B	C	D	E	F	G	H	I	J	K	L
A	0	473	136	241	107	152	195	193	184	82	29	211
B	473	0	450	553	452	243	308	950	414	155	29	551
C	136	450	0	418	67	109	163	260	124	61	20	197
D	241	553	418	0	132	423	191	596	292	317	71	285
E	107	452	67	132	0	63	103	517	108	97	10	158
F	152	243	109	423	63	0	176	233	162	146	35	820
G	195	308	163	191	103	176	0	273	115	66	15	179
H	193	950	260	596	517	233	273	0	729	472	35	820
I	184	414	124	292	108	162	115	729	0	239	20	619
J	82	155	61	317	97	146	66	472	239	0	33	307
K	29	29	20	71	10	35	15	35	20	33	0	26
L	211	551	197	285	158	820	179	820	619	307	26	0

Data obtained for the correspondence analysis is shown in Table 5. As mentioned in the previous section, keywords from the most used secondary terms from the JITA classification scheme were selected; these terms are represented in Table 5 with the code X1 to X13. The selected keywords are: X1: archives or repositories; X2: bibliometrics; X3: cataloging or bibliographic control; X4: education or training; X5: e-resources or e-journals; X6: information services; X7: information systems; X8: information technology; X9: intellectual property or copyright; X10: libraries; X11: networks; X12: society; X13: use studies or surveys.

Table 5. Co-occurrences of keywords X1-X13 vs JITA subject terms (A-L) (1-9-2013)

	A	B	C	D	E	F	G	H	I	J	K	L
X1	83	284	41	269	220	85	61	845	170	254	17	304
X2	42	435	6	20	16	5	6	68	23	4	.01	23
X3	26	95	20	65	32	31	18	178	311	158	3	112
X4	107	211	274	193	28	89	362	137	64	22	1	113
X5	133	531	111	221	297	13	87	916	233	184	12	287
X6	36	101	49	81	17	62	46	110	104	35	2	99
X7	64	122	23	68	19	49	40	103	110	36	4	138
X8	36	92	28	32	12	23	36	38	44	11	1	112
X9	28	125	15	39	315	14	14	101	24	37	2	46
X10	330	566	418	1128	193	439	214	655	306	297	54	549
X11	52	146	35	89	23	77	29	132	87	19	6	172
X12	67	220	30	45	24	18	47	41	19	8	.01	54
X13	64	199	185	105	37	66	58	125	50	28	2	65

Cluster analysis

The twelve principal subject classes of the JITA classification systems are not isolated islands, this is observable by browsing the metadata of the documents; usually a document has more than one of these classes listed in the record. A cluster routine took advantage of those relationships in order to create a model that represents them. The factor analysis, in this case, proposed a solution of four factors which can be interpreted as the existence of four clusters. Based on this, and by observing the possible solutions

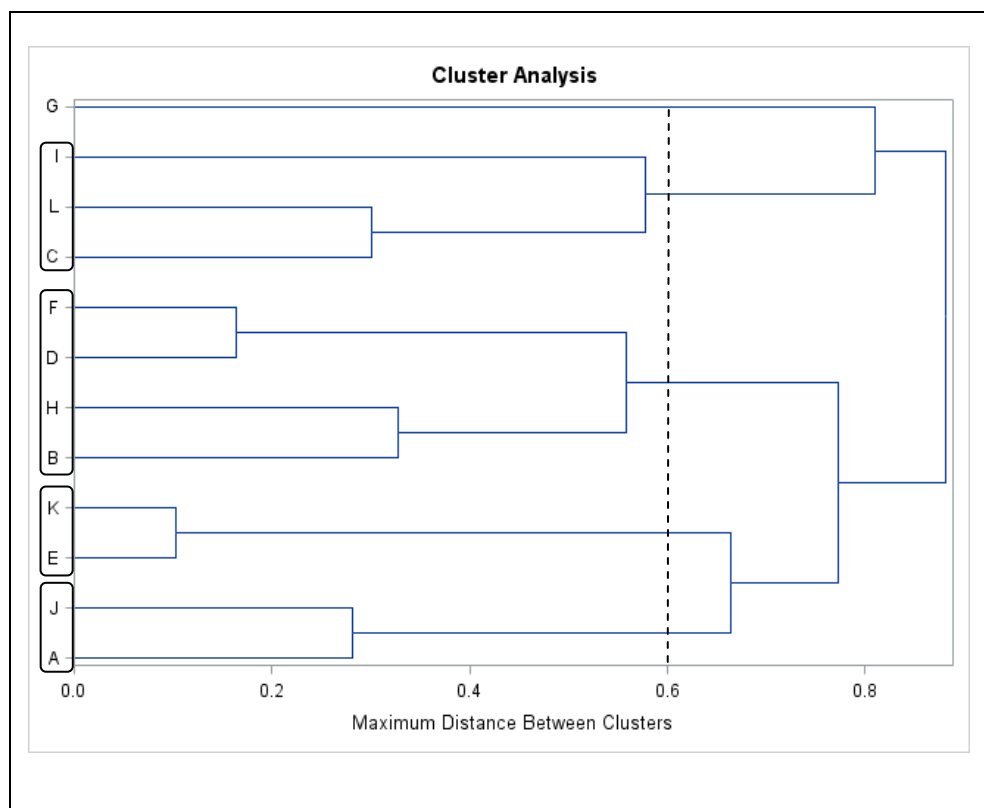


Figure 2. E-LIS four clusters model

Table 6. E-LIS thematic clusters

cluster	Class letters	Definition phrase
1	I, L, C	Information treatment for information services; Information technology and library technology; Users, literacy and reading
2	F, D, H, B	Management; Libraries as physical collections; Information sources, supports, channels; Information use and sociology of information
3	K, E	Housing technologies; Publishing and legal issues
4	J, A	Information treatment for information services; Theoretical and general aspects of libraries and information
5	G	Industry, profession and education

offered in the dendrogram (Figure 2), we have concluded that a solution of four clusters is adequate for this case. Table 6 shows the definition of these four clusters by using the phrases of the classes according to JITA. In addition, there is an isolated class (G). Several of the clusters discovered have a significant level of logical understanding. For example, in cluster one, elements of information services, information technology and user participation are incorporated, and in cluster 4, elements of information treatment and the theoretical aspects of information are present.

Correspondence analysis

In this correspondence analysis, associations of keywords and JITA are explored. The data collected in Table 5 was presented to the SAS Pro Corresp procedure. Figure 3 shows the results.

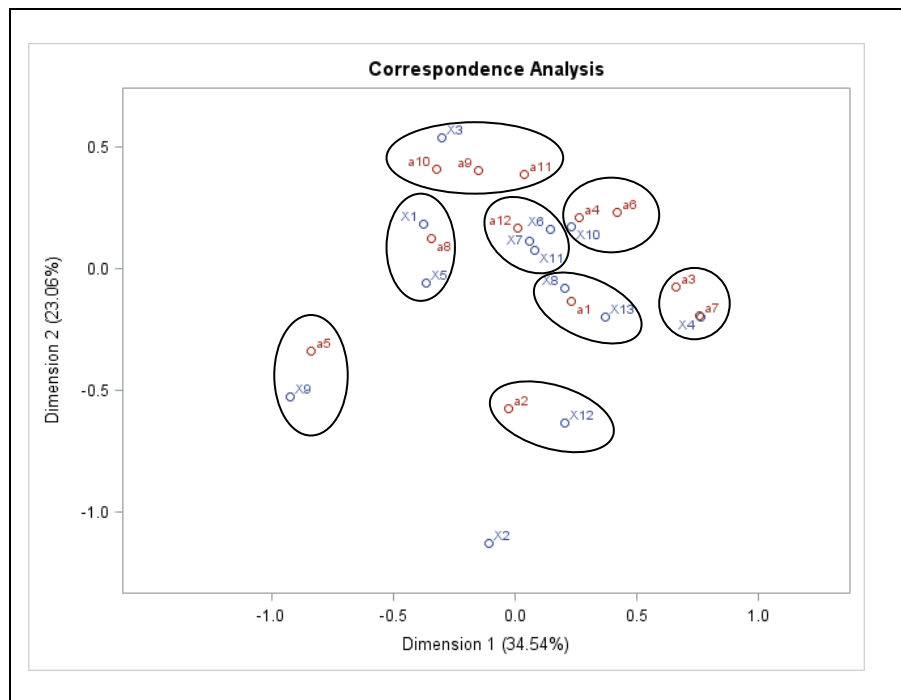


Figure 3. Correspondence analysis groups

Table 7. Correspondence analysis associated groups

Group 1	X9-A5	X9: Intellectual property or copyright A5 (E): Publishing and legal issues
Group 2	X12-A2	X12: society A2 (B): Information use and sociology of information
Group 3	X1-A8- X5	X1: archives or repositories A8 (H): Information sources, supports, channels X5: e-resources or e-journals
Group 4	X3-A10- A9-A11	X3: cataloging or bibliographic control A9 (I): Information treatment for information services A10 (K): Technical services in libraries, archives, museum A11(K): Housing technologies
Group 5	A12-X7- X6-X11	A12 (L): Information technology and library technology X6: information services X7: information systems X11: networks
Group 6	X8-A1- X13	X8: Information technology A1 (A): Theoretical and general aspects of libraries and information X13: Use studies or surveys
Group 7	X4-A3- A7	X4: education or training A3 (C): Users, literacy and reading A7 (G): Industry, profession and education.
Group 8	X10-A4- A6	X10: libraries A4 (D): Libraries as physical collections A6 (F): Management
	X2	X2: Bibliometrics

The eight associated groups identified are shown graphically in Figure 3, a description of them is included in Table 7. This description is based on the phrases used to define JITA's mail classification classes, and the words used to define the selected keywords. We have also found one isolated term: bibliometrics. There are some significant associations in the groups. For example, intellectual property and publishing in Group 1; archives, information sources and e-resources in Group 3; information technology, information services, information systems, and networks in Group 5; and libraries and management in Group 8. Overall, both the cluster analysis and the correspondence analysis have produced meaningful outputs.

Conclusions

This paper presents the results of studying term relationships using JITA classifications classes and selected keywords with the purpose of exploring the structure of a subject-related open repository. The results obtained show that the subject coverage of the E-LIS repository can be described in four subject clusters; the correspondence analysis has produced eight associated groups. These are indications that the E-LIS repository, in its tenth year, has achieved the status of a mature knowledge base archive for the international community of Library and Information Science.

Acknowledgments

The author would like to recognize the support provided to this project by Thomas F. Rhoden and Gabriel E. Osorio, both graduate students at Northern Illinois University (NIU) as well as to the NIU University Libraries.

References

1. Adamick, J., and Reznik-Zellen, R. (2010). Representation and recognition of subject repositories. *D-Lib Magazine*, 16(9/10). Available online: <http://www.dlib.org/dlib/september10/adamick/09adamick.html> (consulted December 10, 2013).
2. Agresti, A. (2002). *Categorical data analysis*, 2d ed., New York: Wiley. (Chapter 9: Building and extending loglinear / logic models).
3. Bassi, M.-C., Subirats-Coll, I., De Robbio, A., and Krichel, T. (2005). E-LIS: E-prints in Library and Information Science. In IFLA Council and General Conference (71st: 2005 : Oslo), Oslo (Norway), 14th-18th August, conference poster. Available online: http://eprints.rclis.org/6647/1/poster_IFLA2005.jpg (consulted December 21, 2013).
4. Dal Porto, S., Marchitelli, A. (2006). Functionality and flexibility of traditional classification schemes applied to a Content Management System (CMS): facets, DDC, JITA. *Knowledge Organization*, 33(1), 35-44.
5. De Robbio, A. (2003). E-LIS: and open archive in Library and Information Science. *Bibliotime*, VI (I). Available online: <http://eprints.rclis.org/4113/1/E-LIS-bibtime.pdf> (consulted December 10, 2013).
6. De Robbio, A., Mornati, S., Peset, F. Subirats, I., and Tsakonas, G. Developing methods for extending repository services to increase the accessibility of OA documents in Library and Information Science., 2011. In LIBER 40th Annual Conference, Barcelona, Spain, 29 June - 2 July 2011. Conference poster. Available online: <http://eprints.rclis.org/15666/6/ELIS-liber2011-poster.pdf> (consulted December 10, 2013).

7. De Robbio, A; Subirats Coll, I. (2005). E-LIS: an international open archive towards building open digital libraries, *HEP Libraries Webzine*, 11. Available online: <http://eprints.rclis.org/6634/1/e-lis.pdf> (consulted December 10, 2013).
8. De Volder, C. (2011). El ejercicio del autoarchivo en el repositorio temático E-Lis por parte de los bibliotecarios argentinos. Tesis de grado Licenciatura en Bibliotecología y Documentación, Universidad Nacional de Mar del Plata Facultad de Humanidades, Departamento de Documentación, Mar del Plata, 2011. Available online: http://eprints.rclis.org/15476/1/tesis_DeVolder.pdf (consulted May 20, 2013).
9. Dimitri, P.-J. (2009). Informe sobre E-Lis Argentina., 2009. In RIBDA 2009, La innovación en el acceso abierto a la información agrícola y de medio ambiente, Lima, Perú, 2009 October 27-29. Available online: http://eprints.rclis.org/14235/1/INFORME_SOBRE_E.pdf (consulted December 20, 2013).
10. Ferreira, C., and Neves, B. (2013). E-Lis: análise da participação dos investigadores portugueses no maior repositório para a Ciência da Informação., 2013. In VI Encontro Ibérico EDICIC, Porto, 4-6 November 2013. Available online: <http://eprints.rclis.org/20638/1/e-LIS%20an%C3%A1lise%20da%20participa%C3%A7%C3%A3o%20dos%20investigadores%20portugueses.pdf> (consulted December 20, 2013).
11. Hu, C.P., Hu, J.M., Deng, S.L., and Liu, Y. (2013). A co-word analysis of library and information science in China. *Scientometrics* 97(2), 369-382.
12. Medeiros, N. (2004). A repository of our own: the E-LIS e-prints archive", *OCLC Systems & Services*, 20(2), 58 - 60.
13. Morris, T. (2001). Visualizing the structure of Medical Informatics using term co-occurrence analysis: II: INSPEC perspective. In: Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology, 2001, Elizabeth Aversa and Cynthia Manley, editors, Medford, NJ: Information Today, 38, 489-497.
14. Nawaz, S., Rajan, P., Yu, J.H., Yi Luo, Choi, J.H., Radcliffe, D.F., and Strobel, J. (2011). A keyword based scheme to define engineering education research as a field and its members. In: Global Engineering Education Conference (EDUCON), 2011 IEEE, 4-6 April 2011, 201-209.
15. Osorio, N.L., and Osorio, G.E. (2004). Disasters in the literature of Human Factors: An analysis of subject and application Terms. In Proceedings of the ASEE Illinois-Indiana Section Annual Convention, East Peoria, Illinois, USA, 26-27. Available online http://eprints.rclis.org/6300/1/ASEE_IL-IN-Osorio-II-A-3.pdf (consulted December 15 2013).
16. Villegas, S., and Angel, M. (2006). Iniciativas de acceso abierto y perspectivas de E-LIS en México. In Proceedings Congreso Internacional de Información INFO 2006, La habana, Cuba. 2006. Available online: http://eprints.rclis.org/7528/1/E-LIS_M%C3%A9xico.pdf (consulted December 15, 2013).