Comparison of Digital Archives from the Point of Sustainability

Sümeyye Akça: Hacettepe University Department of Information Management, 06800 Ankara, Turkey, sumeyyeakca@hacettepe.edu.tr

Zehra Taşkın: Hacettepe University Department of Information Management, 06800 Ankara, Turkey, ztaskin@hacettepe.edu.tr

Monika Berać: University of Zagreb Faculty of Humanities and Social Sciences, Croatia, monika.berac@gmail.com

Abstract: *Digital repositories applications have been proliferated recently throughout the world as a result of open access policies. In order to gain more efficient these attempts, it really important to meet requirements for creating well-designed archives. One of the most significant requirements of digital repositories is sustainability. The main aim of this study is to assess three selected digital archives which are e-prints in library & information Science (ELIS), Digital Archive of Hacettepe University Department of Information Management* **(***PDAHU) and Faculty of Humanities and Social Sciences Institutional Repository of University of Zagreb (DARHIV) in terms of sustainability by using the framework of Research Libraries Group (Research Libraries Group, 2002). We analyzed these three digital libraries under the six titles which are important issues for persistence of institutional repositories. Findings reveal that E-LIS is well structured digital repository rather than PDAHU and DARHIV. At the same time, all of these three digital repositories should increase their deposit activity in order to gain more efficiency on sustainability.*

Keywords: Digital libraries, sustainability, open access, E-LIS, DARHIV, Publication repository of Hacettepe University Department of Information Management.

Introduction

Digital archives require persistence and currency to gain their missions (Wright, 2006). It is really important to meet these requirements to create well-designed archives. One of the requirements of digital archives is sustainability. The term "sustainability", which is defined as "causing little or no damage to the environment and therefore able to continue for a long time" (Cambridge Online Dictionary, 2013), has become significant for digital archives recently. It is accepted as crucial to understanding and evaluating of digital archives by some studies (Dale, 2006). The main aim of this study is to assess three selected digital archives in terms of sustainability by using the framework of Research Libraries Group (Research Libraries Group, 2002).

This paper deals with the questions of sustainable information practices that are connected with the digital libraries and digital repositories. Specifying those keywords is a bit challenging task. Although digital libraries do not have long history there was a remarkable development of ideas, performances and theoretical descriptions of concepts. Some of the definitions have the broader meaning: Digital Libraries basically store materials in electronic format and manipulate large collections of those materials effectively. Research into digital libraries is research into network information systems, concentrating on how to develop the necessary infrastructure to effectively mass-manipulate the information on the Net (The National Science Foundation, 1999). More specific definition is by Arms (2000): "a digital library is a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network". Digital libraries contain diverse collections of information for use by many different users. The unifying theme is that information is organized on computers and available over a network, with procedures to select the material in the collections, to organize it, to make it available to users,

and to archive it. Also Arms (2000) finds difference between digital repositories and digital libraries by defining digital repositories as any computer system whose primary function is to store digital material for use in a library. Repositories are the book shelves of digital libraries.

Sustainability is also not unambiguous term. Applied to digital libraries, sustainability is a broad term, referring to everything from technical issues about the digital preservation of materials, to the social questions surrounding the long-term accessibility of resources to the public at large (McArthur, Giersch & Burrows, 2003). Sustainability issues in digital librarianship can be divided into two groups: 1) sustainability as the preservation of information by digitizing and storing in the digital repositories, 2) financial sustainability and cost effectiveness.

Publications' Digital Archive of Hacettepe University Department of Information Management (PDAHU - http://bbyeprints.hacettepe.edu.tr/), Faculty of Humanities and Social Sciences Institutional Repository of University of Zagreb (DARHIV - http://darhiv.ffzg.hr/) and E-LIS (http://eprints.rclis.org/) are the three digital collections that
are most often described as digital repositories and this paper will show the sustainability aspect of them.

Methodology

PDAHU (http://bbyeprints.hacettepe.edu.tr/), DARHIV (http://darhiv.ffzg.hr/) and E-LIS (http://eprints.rclis.org/), which is one of the well-known archives of LIS field, were chosen with the aim of evaluating digital archives from the point of sustainability. The common point of all these archives is to use ePrints software. In addition to this, the archives of Hacettepe University and University of Zagreb are brand new and open for improvements. E-LIS can be good example to develop these archives positively. The recommendations of Research Libraries Group for organizations to create trusted digital libraries were used to compare the systems' quality with each other (Research Libraries Group, 2002, p. i). And also, the systems are compared from various aspects which are listed below;

- Compliance with the open access systems such as ROAR and OpenDOAR
- Administrative structure
- Organizational and financial sustainability
- Information and technological infrastructure
- System security
- Policy formulation

This study is important to present the basic requirements to create sustainable digital archives. Besides, the findings of this study can be used by creators and policy makers of selected digital archives to improve the features of systems from the point of sustainability.

Findings

Compliance with the open access systems such as ROAR and OpenDOAR

Definition of open-access is literature that digital, online, free of charge for public with an internet connection and free of most copyright and licensing. It could be described as "barriers-free" access (Suber, 2013, p. 4). When an institution wants to create a repository, it must be chosen IR software which is the most important for accessibility it. Especially, as items stored in using ePrints that is one of the most usable software for open access repositories is so important to be accessible in various scholarly search engines such as Google Scholar, Scirus, Registry of Open Access Repositories (ROAR) and OpenDOAR (Adewumi & Ikhu-Omoregbe, 2010, p. 4).

ROAR is an indexing system which was created by using ePrints software at University of Southampton in 2003. The aim of Registry of Open Access Repositories (ROAR) is at promoting the development of open access by providing detailed information about the growth and status of repositories, accessibility to the literature throughout the world. Open access to research maximizes research access and thereby also research impact, making research more productive and effective (Roar.eprints.org, 2013).

Directory of Open Access Repositories (OpenDOAR) is an authoritative directory of academic open access repositories that provide in-depth search on the repository list, contents and give some statistical charts about this information (Opendoar.org, 2013).

A repository remain usable and current with evaluating what is they can harvest from each repository - and what rights they have in the use of the harvested information and what features should a repository support in order for its content to be fully harvested, and re-used by the latest innovative service (Opendoar.org, 2013).

In this study, we evaluate these three institutional repositories from the point of compliance with ROAR and OpenDOAR. PDAHU was registered as an institutional repository in ROAR. We see the register date and information about repository type, organization, using software et al. And it is quite open that there is no movement on deposit activity by the reason of being a new repository in the area.

When we search for information in the ROAR about DARHIV repository, we can access the records count, digital objects in the repository and daily deposit activity. Although DARHIV created in 2006, deposit activity has increased especially on 2012-2013.

E-LIS is one of the well-known archives of LIS field. It was registered in ROAR. As we look at the record details in the ROAR, it is obvious that daily deposit activity is so low and density appeared on the 2012 and 2013. When we search OpenROAR for E-LIS on 8 December 2013, It has already have policy and fill all requested information about the repository. DARHIV has registration in OpenDOAR, however there is some missing information in the web site about this repository. PDAHU hasn't been registered OpenDOAR yet.

Administrative Structure

Administrative structure is the essential part of a digital library. It would be almost obvious that any administrative system within an institutional digital library must harmonize with the other part and work in concert with them to achieve and sustain purpose of the IR (Jelinek and Burstein, 1982, p. 242). An institutional repository should provide that it has a fundamental commitment to implementing the standards and best practices that affect its operations - especially users that directly influence its viability and sustainability (Research Libraries Group, 2002, p.13). It should also be written all parameters which process of acquisition, maintenance, access, and withdrawal clearly on the web site. All the IR platforms analyzed allow an IR administrator to arrange import and export items and workflows (Adewumi and Ikhu-Omoregbe, 2010, p. 3).

E-LIS has a team of volunteer editors and authors are invited to contribute in the growth of repository, by sending their study, works to the E-LIS. E-LIS team for biennium 2013-2014 has comprised of administrative board, executive board and editorial board. There have persons responsible for quality assurance, standardization of authors' names, editorial board

coordinator, policy on copyright and rights management coordinator, JITA Classification Scheme in the executive board. Editorial Board consists of regional editors and editors. There are two regional editors in charge of North and Central America and South America. Editors responsible for their own countries to evaluate works for taking place in the repository. 31 countries have national editors (eprints.rclis.org, 2013).

PDAHU (http://bbyeprints.hacettepe.edu.tr/) has been created as a project by postgraduate students of Digital Libraries class. The publications can be loaded by administration or authors if register in the system. Yet there is no team for administration structure. Foundation history of DARHIV hasn't been presented in the website. There is only e-mail address presented to contact responsible people.

Organizational and Financial Sustainability

An institutional repository is also dependent on many issues for maintain their sustainability like organizational and financial requirements. Organizations should become trusted that digital repositories will establish themselves in ways which demonstrate their viability. They should give commitment clearly for the long-term retention, management of, and access to digital cultural assets on behalf of depositors and users (Research Libraries Group, 2002, p. 14). An institutional repository should be able to prove its financial sustainability over time (Research Libraries Group, 2002, p. 15). The effective utilization of financial resources is part of the digital library sustainability issues for all institutions. Organizational and financial fitness should be checked annual for sustainable encouraged to repository.

E-LIS which is so crucial for the LIS field has been encouraged by sponsors for financial and organizational sustainability. There are four financial sponsors for E-LIS. PDAHU has been supported by Hacettepe University. DARHIV has been supported by University of Zagreb. Two of the repositories which we choose to evaluate for sustainability are supported by State Universities.

Information and Technological Infrastructure

Technology infrastructure is the name generally given to the equipment, control and operating sub-systems, network connections and the structure of the total technological systems (Cooper, Demmers and Graham, 2004, p. 1). It is stated out in the literature that advanced information and technological infrastructures of digital libraries should be created from the point of following issues (Ioannidis, Maier, Abiteboul & et al., 2005, p. 2);

- Unitary and comprehensive theories and frameworks
- Interoperable multimodal and multilingual services for anyone.

All three digital libraries use ePrints Software and its technological infrastructure. However, ePrints is open-source software and it is recommended that each institution have to spare special technical programmer effort to make these libraries sustainable with regard to following technological developments. Hosting for digital libraries is generally made by institutions which are the owners of the libraries. Therefore, it is important to provide reliable infrastructures for them. E-LIS digital library has some information about its hosting and technology in the page of "About us". Similarly, PDAHU also has small quantity of information about technological features of installed ePrints, like version. Operating system information is also given the "About us" page of PDAHU. In the repository of DARHIV, there is small information in the about page, however, the links for detailed information is not working during the evaluation.

In the other perspective, it is vital to serve unitary, comprehensive and interoperable frameworks for these libraries. The criteria can be about used languages, "file not found" errors, search tips and etc. It is confirmed in this study that E-LIS is created flawlessly in the sense of frameworks. All the pages are designed in a standard structure. PDAHU uses the Turkish Language and all the pages translated into Turkish in a standard pattern. Besides, there is no Turkish language pack on the ePrints Bazaar (http://bazaar.eprints.org/). It means designers haven't made their language pack public yet. Designers should make their language pack apparent to help other Turkish repositories. Only DARHIV has some "file not found" errors and non-standard language. The designers of this repository should consider these issues to improve quality and to make their repository sustainable. There are two language selection options in this archive, but it is not working effectively.

System Security

Security is really important issue for digital libraries. However, security is not only about the infrastructure of the libraries, but also about contents, users and intellectual property etc. In the DELOS Digital Library Reference Model, there are 6 main elements are identified for digital libraries (Candela, 2010). These are content, user, functionality, architecture, quality and policy. Each of these elements has some security issues on it (Fox and ElSherbiny, 2011).

Most of the security issues for the evaluated repositories mentioned in the repository policy pages. To clarify the security principles of the repositories, content security of them are provided by people who upload papers on them. Content of DARHIV and PDAHU are collected from the departments, so the owners of these publications have rights for self-archiving. Similarly, papers placed in E-LIS are uploaded by authors or editors. All uploaded papers by users of the repositories are evaluated and validated before moving to repository by repository admin or editors. In addition, user security control provided by using *role-based access control model* for all three databases. Admin of repositories have an authority to give roles for users. Users can be under control by the help of role-based access control model.

Policy

Policy is defined in the literature as "governing how a digital library is instantiated and run" (Innocenti, Smith, Ashley & et al., 2011, p. 111). Policies are important to identify repositories' perspective for metadata, data, content, submission and preservation. Some tools to constitute repository policies are developed to create standard policy texts. Standard policy texts are important to create interoperable policies for digital repositories. Repository owners can form their own policies by using OpenDOAR's *Policies* tool (http://www.opendoar.org/tools/en/policies.php).

All three repositories have detailed repository policy texts in their websites. E-LIS has the well-designed and instructive policy page for users and authors. Authors' rights are explained deeply in the E-LIS Policy. Other two repositories also have structured interoperable policy texts.

Results and Discussion

Digital repositories have been an important part of scholarly communication recently from the point of accessing scholarly articles in an open access platform. Institutions create their own repositories with the help of open-source software like ePrints or DSpace. Developing and making these repositories sustainable is the duty of the owner institutions. They need to consider compliance with the open access systems, administrative structures, organizational

and financial sustainability, information and technological infrastructure, system security and policy issues while developing their repositories. We have analyzed three digital repositories which are E-LIS, PDAHU and DARHIV with regard to these issues. The common point of these repositories is usage of ePrints software. This makes evaluation process consistent.

All the analyzed repositories were registered to ROAR, and also E-LIS and DARHIV were registered to the OpenDOAR. Being indexed in important open access repository directories makes repositories more visible in the World Wide Web. As a suggestion, PDAHU repository should be registered to the OpenDOAR and the deposit activities of all the repositories should be increased. There is a really small quantity of publications placed in PDAHU. The others have more publications than PDAHU, however, their deposit activities in ROAR is really low. Users should encourage submitting their papers to these systems. It is generally accepted that an administrative structure has key role to maintain the organization. Presenting administrative structures in the repositories is important to make users informed anything about administrative bodies. However, only E-LIS has the detailed administrative information on its website. Others should develop web pages about their administrative body.

Organizational and financial sustainability of PDAHU and DARHIV has been provided by their owner institutions. E-LIS has four sponsors for its financial sustainability. Organizational and financial annual controls are vital for repositories to provide continuance. All three repositories should pay attention to organizational and financial issues.

Technological infrastructures of three systems are like each other in consequence of using the same software. However, E-LIS's interface is more unitary, comprehensive and well-designed than others. PDAHU and DARHIVs' technical structures should be improved to procure technological sustainability.

System security is the issue belongs to the policy of the repositories. If the security issues are represented in the repository policy clearly, there will be no problem about system security. All three repositories have similar policies and protection about their contents. It is a good point for future of these repositories.

When we evaluate findings about these three digital repositories, we generally realized that PDAHU and DARHIV should develop their policy, structure and deposit activity for more efficient using. Although E-LIS is well structured rather than two digital repositories, it also should be increased their deposit activity[1] in order to gain more impact on the field. In addition, the findings of this study can be useful for other digital libraries to provide their sustainability
for the future.

References

Adewumi, A.O., & Ikhu-Omoregbe, N.A. (2011). Institutional repositories: Features, architecture, design and implementation technologies. *Journal of Computing*, 2(8).

Arms, W.Y. (2000). Digital libraries. Cambridge, MA: MIT Press.

Cambridge Dictionaries Online. (2013). *Sustainability*. Retrieved 2 October, 2013 from

[1] "Deposit activity measures the number of days in the last year that had "low" (1-9), "medium" (10-99) or "high" (100+) numbers of records deposited" (Roar.eprints..org, 2013).

http://dictionary.cambridge.org/dictionary/british/sustainable?q=sustainability

Candela, L. (2010). DL.*org: Coordination action on digital library interoperability, best practices and modelling foundations.* Project Funded under the Seventh Framework Programme, ICT Programme - Project No: 231551. Retrieved December 6, 2013 from http://www.dlorg.eu/uploads/DL%20Reference%20Models/The%20Digital%20Library%20Reference%20Model_v1.0.pdf

Cooper, L., Demmers, L. & Graham, C. (2004). Technology infrastructure design for libraries. California: Libris Design Project.

Dale, R. L. (2006). Developing metrics to evaluate digital archives. *Portico Participants' Meeting*. ALA, New Orleans. Retrieved 2 October, 2013 from http://www.portico.org/-digital-preservation/wp-content/uploads/2010/01/DaleALA062406.pdf

Eprints.org. Retrieved 8 December, 2013 from http://roar.eprints.org/.

Eprints.rclis.org. Retrieved 8 December, 2013 from http://eprints.rclis.org/information.html

Fox, E. & ElSherbiny, N. (2011). Security and digital libraries. In *Digital Libraries - Methods and Applications* (pp. 151-160). Rijeka: InTech Open Access.

Innocenti, P., Smith, M., Ashley, K. & et al. (2011). Towards a holistic approach to policy interoperability in digital libraries and digital repositories. *The International Journal of Digital Curation*, 1(6), 111-124.

Ioannidis, Y., Maier, D, Abiteboul, S. & et al. (2005). Digital library information-technology infrastructures. *International Journalon Digital Libraries*, 5(4), 266-274.

Jelinek, M. & Burstein, M.C. (1982). The production administrative structure: A paradigm for strategic fit. *The Academy of Management Review,* 7(2), 242-252.

McArthur, D.J., Giersch, S. & Burrows, H. (2003). Sustainability issues for the NSDL. In *Proceedings of the 2003 Joint Conference on Digital Libraries (JCDL'03)* (pp. 395). Retrieved December 8, 2013 from http://csdl.computer.org/comp/proceedings/-jcdl/2003/1939/00/19390395.pdf

National Science Foundation. (1999). *Digital Libraries Initiative: Available Research*. US Federal Government

Opendoar.org. Retrieved 8 December, 2013 from http://www.opendoar.org/

Research Libraries Group. (2002). Trusted digital repositories: Attributes and responsibilities. *An RLG-OCLC Report*. Retrieved 3 October, 2013 from http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

Wright, R. (2006). What archives want: The requirements of digital technology. *EBU Technical Review*. Retrieved 2 October, 2013 from http://tech.ebu.ch/docs/techreview/trev_308-archives.pdf

Suber, P. (2013). *Open access*. Cambridge: The MIT Press.