Exploring usefulness and usability in the evaluation of open access digital libraries

Giannis Tsakonas*, Christos Papatheodorou
Laboratory on Digital Libraries and Electronic Publishing, Department of Archive and Library Sciences, Ionian University, Palaia Anaktora, Plateia Eleftherias, 49100, Corfu, Greece

Abstract: Advances in the publishing world have emerged new models of digital library development. Open access schemes are expanding their presence and realize the idea of digital library in various means. While user-centered evaluation of digital libraries has drawn considerable attention during the last years, these systems are currently view from the publishing, economic and scientometric perspectives. The present study explores the concepts of usefulness and usability in the evaluation of an e-print archive. The results demonstrate that several attributes of usefulness and usability, as well as functionalities commonly met in these systems, affect user interaction and satisfaction.

Keywords: digital libraries evaluation, open access, e-prints, interaction, usefulness, usability

1. Introduction

Evaluation of digital libraries has drawn considerable attention during the last years, not only because of the money spent for the acquisition of access to them, but also because of the major investments in their development. The term digital library is vast, covers many and different applications and has been used interchangeably for systems, like digitized collections, e-journals platforms, network databases, library websites etc. Moreover the current electronic publishing business models enrich the DLs technology aiming to provide powerful information access options to the users. However despite the polymorphous nature of DLs, one feature is stable: information provision through networked systems.

The advance of open access systems increases the complexity and diversity of DLs because the growth of their collections requires the involvement of new actors; e.g. authors submit their own original material, produce the corresponding metadata, etc. Although a significant amount of research is noted on the evaluation of commercial or research DLs, the equivalent work is not dedicated to the freely accessible DLs. However the evaluation of such systems is imperative, due to their worldwide acceptability and usage, but their special characteristics liquefy the objects under evaluation.

In general usefulness and usability are two major research issues in the field of DL evaluation. Originating from the fields of information behavior and human computer interaction, they both hold significant role in pursue of user satisfaction and system usage, as they study users' interaction with representations of information systems and information objects. Järvelin & Ingwersen have based the formation and analysis of user's cognitive processes and information seeking behaviour on the interactivity between the user and both the information system and the information items, that stand behind an interface (2004). Toms uses this tri-polar structure as a platform, where users' information seeking behaviour takes place, and upon this platform proposes ways of extracting meaningful design suggestions for information architecture (Toms, 2002). According to the DeLone/McLean model (1992), one of the eminent models in IS success, system and information quality affect user satisfaction, system use and influence organizational impact in later stages. In a previous study, it was shown that usefulness and usability are related properties of user interaction (Tsakonas & Papatheodorou, 2006).

In the present paper we investigate which DL content and system features affect most significantly the overall DL usefulness and usability. In other words we try to explore the most significant DL features that affect the users information seeking and work completion processes. In particular we present a user-centred evaluation approach and we apply it to an open access system, E-LIS (*http://eprints.rclis.org*), which is an e-print archive in the field of library and information science. The aim of the paper is twofold; firstly to apply a DL evaluation framework -with a special interest on usability and usefulness- on the E-LIS archive and to explore its appropriateness for open access systems, and secondly to provide an insight about the E-LIS usefulness and usability. Since the E-LIS features and operations are commonly met in institutional repositories and open access systems, the derived results might empower the design of such systems informing about the users' opinion for their strong and weak features.

*Corresponding author. Tel.: +30 2610 969625; fax: +30 2610 969672.
E-mail addresses: *john@lis.upatras.gr* (Giannis Tsakonas), *papatheodor@ionio.gr* (Christos Papatheodorou).

Date: 30/01/07

Section 2 resumes the scientific literature in regard to evaluation of DLs and open access systems. Section 3 presents the research setting in terms of problem, aims and models employed. Section 4 presents the methodology of the current study, while the following section, number 5, presents the results. Section 6 is dedicated to the discussion of the main findings and section 7 summarizes and suggests further research steps.

2. Literature Review

There are various DL evaluation initiatives that use different protocols and models; some of them continuously tested, other created or adapted to fit to specific systems and needs. Among the various concepts under investigation, two concepts have gained significant attention in user-centered evaluation, usefulness and usability. Technology Acceptance Model (TAM) is widely used in the field of information systems success and it has been used in prior studies of predictive usage of digital libraries (Thong, Hong & Tam, 2002; Hong et al., 2002), search engines (Liaw & Huang 2003) and portals (Yang et al., 2004). TAM exploits users' perceptions of usefulness and ease of use of an information system, relates them with external variables of users' profile and combines them to predict attitudes and actual usage. DigiQUAL (Kyrillidou & Giersch, 2005) originates from the field of service quality in the library sector and aims to be extended to the digital space of service provision, covering areas such as design features, accessibility/navigability, interoperability, DL as community for users, developers and reviewers, collection building, role of federations, copyright, resource use, evaluating collections and DL sustainability.

The interest on the concepts of usefulness and usability is reinforced by the users themselves. Xie (2006) collected user identified DL evaluation criteria, analyzed and classified them in five categories, such as usability, collection quality, service quality, system performance efficiency and users opinions. The principal finding was that *"users like to apply the least effort principle to finding useful information to solve their problems"*. Similar results by Kani-Zabihi, Ghinea and Chen (2006), who examined the users' preferences on system features, designate that users prefer learnable and reliable DLs, than aesthetically pleasant and supportive ones.

Concerning the DL evaluation methodology Bertot et al. (2006) report an iterative campaign based on the system-driven concepts of functionality, usability and accessibility. One of the main suggestions by the authors is that evaluation activities should be based on a number of constructs (evaluation concepts), rather than to focus on one dimension, in order to assist the formation of a holistic picture for user interactivity and to extract easier a set of implications on system design.

Since the present paper focuses on the open access systems we present some representative studies in this particular field. Although open access DLs and electronic information services have undergone various evaluations of their impact, their technological infrastructure (Wyles, Maxwell & Yamog, 2006) and integration with library resources (CRL, 2006), there are a few studies concerning the interaction of the users with such systems. An attractive research issue is the acceptability of these systems and the corresponding services by the scientific community. Nicholas, Huntington and Rowlands (2006) gathered the opinions of a world wide sample of scientific authors to estimate the rate of adoption of open archive publishing schemas and to compare it with the dominant publishing habitus. Kurata et al. (2006) showed that the usage of e-prints servers in Japan is limited to specific disciplines, such as physicists, and that electronic journals remain the principle vehicle of scientific communication. Finally, Correira and Castro Neto (2002) reported that e-prints servers' usage is mainly for retrieval purposes, rather than publishing.

Hitchcock et al. (2002) evaluated the usefulness and usability of the open-access bibliographic service Citebase. Although this formative evaluation campaign demonstrated a useful and usable system, problems with the coverage and the navigation ability were reported. Despite the fact that Citebase is not a DL, but a bibliographic service that in parallel provides access to full text, where available, the results showed problems concerning the features of content and system. Silva, Laender and Gonçalves (2005) examined several features of the BDBComp self archiving system. In their usability study, among others, counted the average submission time. It was shown that differences were attributed to users experience and expertise, as well as to content types.

In conclusion, user-centered evaluation should be multi-construct, capable to capture a panoramic view of users' opinions, taking into account their characteristics and needs and grounded on their perceptions and goals.

3. Research setting

For the aims of our research we applied Interaction Triptych Framework (ITF) (Tsakonas, Kapidakis & Papatheodorou, 2004; Fuhr et al. 2006). ITF is a theoretical model, which attempts to integrate knowledge from the fields of information behaviour and human computer interaction for the benefit of DL stakeholders. ITF is based on the concept of DL components interaction. Each DL is consisted by three main components, namely system, content and user. Previous theoretical models (Fuhr et al. 2002) have acknowledged these three components as key factors in the DL

lifecycle, which are able to predict DL usage. Each component interacts the other two and each interaction defines an evaluation axis, which is considered as the resultant of a set of descriptive attributes of that interaction. In particular the interactions between the three components (i.e. the evaluation axes) are defined as usability, usefulness and performance and each of them is a resultant of a number of attributes, which are considered as evaluation variables (Fig. 1). The following subsections present analytically the evaluation axes and their components attributes.

3.1. Usefulness

The concept of usefulness defines whether DLs constitute valuable tools for the completion of users' tasks. Usefulness answers the questions if DLs support users' information needs and work completion.

Users' work tasks are formed by their social and organizational context and responds to needs like research, authorship etc. Users seek in DLs the appropriate resources for their work tasks, both relevant, in terms of subject proximity, and "integrateable", in terms of content morphology. On the other hand information tasks include all actions related to need formulation, need expression, querying, relevance assessment, all combined and executed in an iterative manner. During all those phases of information seeking activity there are common attributes that are expressed either as requirements or criteria, depending on the stage of the information seeking procedure.

Specifically users assess the applicability of the documents to their work tasks by the relevance of the source, the reliability, the level of the information, the format of the document and the coverage of the deposited documents. Hong et al. (2002) report that system features have effect on both users' perceptions of ease of use and usefulness. However relevance, which is considered as a system feature, was related only to perceived usefulness. The authors interpret this relation as the association of relevance to the content of the DL. Liu (2004) signifies the importance of information reliability and credibility for the selection of the appropriate resources, while Vakari and Hakala (2005) include in their relevance criteria the level of the provided information. Moreover users' information searching behaviour has demonstrated that despite retrieval of full text resources is significant, other levels of information, such as abstracts, are also preferred (Wolfram & Xie, 2002), probably due to the content overview they provide to users (Krottmaier, 2002). Current evaluation practices involve recording and analysis of usage statistics based on users' format preferences, such as .pdf and .html file formats (Mercer, 2000).

3.2. Usability

Usability stands on the user-system axis, focuses on the effective, efficient and satisfactory task accomplishment and aims to support a normal and uninterrupted interaction between the user and the system. DL community has shown an increasing interest in usability and through the research activities a set of attributes have been identified, such as ease of use, terminology, navigation, aesthetic appearance, learnability. Easiness of use is considered as a crucial attribute of DL interaction, especially in advanced systems, like aggregated search interfaces (Park, 2000). Previous usability studies (Ebenezer, 2003; McMullen, 2001) have shown that terminology raises important barriers in user's understanding of principal functions and contribute to negative changes in their affective state. Navigation easiness is related to the improvement of users' performance, as they are able to trace their place in the DL and to direct to previous or next destinations (Hartson, Shivakumar & Pérez-Quiñones, 2004). Additionally the aesthetic appearance and layout has a crucial role to the overall satisfaction rate (Jackson 2001). Van House et al. (1996) suggested a simplified interface that would reduce users' efforts and recent usability studies have concentrated on the effect of inappropriate visual layout to user interaction (Allen, 2002; Cockrell & Jayne, 2002, Fuller & Hinegardner, 2001). Finally some researchers have emphasized on the ability of a DL to be easy to learn in order to improve user familiarity and performance (Ferreira & Pithan, 2005; Jeng, 2005; Sutcliffe, Ennis & Hu, 2000).

3.3. Performance

Performance evaluation is often a system-centered process, based on quantitative data and doesn't involve real users. Precision and recall are the main metrics, which are transferred from the field of Information Retrieval. With the advance of networked systems other metrics are introduced, such as response time (Kobayashi & Takeda, 2000). Response time is considered as a metric sensitive to the subjective judgement of users and is often reported by end users as a criterion for the adoption of a DL.

4. Methodology

4.1. Evaluated System - E-LIS

According to the repositories typology proposed by Heery and Anderson (2005), E-LIS is an international e-print archive that holds full text pre-print and post-print documents and their

metadata and provides enhanced access to researchers in the domain of field of librarianship and information science. E-LIS offers free access to content through simple and advanced search interfaces and various indices, personalized access for document deposit and several other tools, such as document interlinking. E-LIS is an effort that covers 80 countries and hosts more than 4 500 documents in many different languages. The main requirements for the inclusion of archives in E-LIS are domain relevance and document integrity for the promotion of scientific communication. E-LIS is grounded on a solid and vivid international community, which has structured communication channels and organized dissemination events for the achievement of internal consistency. In each country-member there are one or more editors who are responsible for the promotion of E-LIS, the support of native users and the control of the deposited archives.

4.2. Procedure

Data for this survey were gathered by an online questionnaire. A call for participation was distributed to several national and international mailing lists by the respective national E-LIS editors. Editors were also invited to take part in the survey, as well as the administrative personnel. The call was also visible on the main E-LIS website and the Greek website for a period of one month (May to June 2006).

4.3. Instrument

The questionnaire consisted of thirty four (34) questions, structured in three main parts. The first part (six questions) consisted of questions that asked users about their familiarity with E-LIS, the perceived usage, the attributed significance of E-LIS in their information seeking activity and the potential dedication of time and effort to retrieve the desired information. The second part (sixteen questions) investigated the role of the attributes of the usefulness, usability and performance axes. In detail it consists of three subsets of questions each one corresponding to each axis attributes. Each subset includes also an ending question that measures the overall satisfaction with the subset (i.e. the evaluation axis). The last part of the questionnaire (twelve questions) focused on the E-LIS usefulness and usability. The aim of this part was to examine the relative influence of the system features on usefulness and usability. In all the questionnaire parts, participants were invited to deposit their degree of agreement to a number of statements through a five-point Likert scale (from "Disagree" to "Agree").

Reliability of the research instrument was measured. Cronbach alpha tests were conducted separately for each group of items and rates were high for every group and every separate item respectively (table 1), superseding the 0.70 threshold proposed by Nunnaly and Bernstein (1994). There was an exception in the question 3.2 (Response Time), which scored higher than the value of its group, but the difference was not considered that affected drastically the group's value.

4.4. Participants

A total of 131 of valid questionnaires were collected for analysis, after the removal of duplicate and test submissions. Table 2 presents descriptive statistics concerning respondents' relation with the archive, such as role, geographic area (continent-wide) and E-LIS' introduction channel. The results show that almost the half (50%) of the participants were unregistered users, revealing that E-LIS has an equal readership degree compared to its active members. The results also show that E-LIS has a plenitude of dissemination channels, which introduce effectively the service to the potential users. It is remarkable however that almost half of the respondents learned about E-LIS by informal means, like their own searches or personal communication.

Table 3 presents the descriptive statistics of the questions reflecting users' characteristics. Participants reported their usage frequency, appointed significance and intention to spent time and effort to complete successfully their work tasks through E-LIS. Although the mean usage is below the medium value, there is a normal distribution, indicating that there are several usage types, either intensive or not. Accordingly participants appoint mediocre importance to E-LIS in their regular information seeking activity, while they are more positive towards the statements of spending significant amounts of time and effort to complete their information tasks. This shows a commitment to expressions of productive information seeking activity.

Cross-tabulation of the role in E-LIS and the usage revealed that usage is consistently increased when the type of respondent indicates an active role, such as editor (Agree n=10) or Registered User (Agree n=11). Likewise E-LIS significance and intention to spent the needed time and effort for the retrieval of the desired information increase as the participants report more intense usage (Agree n=10 in all cases). Analysis of Variance reported that the factor of the role shows considerable differences at the questions of usage (F=9.016, p>.001) and appointed significance (F=3.466, p>.010). However there are non-significant differences at the questions for time (F=0.984, p>.419) and effort spent (F=1.059, p>.380). The same pattern was observed at the reverse cross-tabulation between the usage and the other variables, where role (F=13.913, p>.001) and appointed significance (F=40.399, p>.001) showed important differences, but time (F=1.099, p>.360) and effort (F=0.437, p>.782) did not show.

5. Results

5.1. Descriptive statistics

Table 4 presents descriptive statistics of the participant answers for all features. A first observation from this view is the participant's positive attitude towards the evaluation of each attribute. The rest of this section presents and analyzes the results in terms of the appointed role and usage. Usage is a critical factor for the creation of user profiles and user classification, because it implies experience in effective and efficient use of a DL. The factor of the appointed role is introduced by dictation of the system nature and may suggest some sort of experience with the use of the particular DL.

5.1.1. Usefulness questions

All of the usefulness features (1.1.-1.5.) scored high rates, indicating a very positive attitude towards the usefulness attributes of the DL. However the only feature that failed to satisfy users clearly is Coverage. It seems that the participants were not satisfied entirely with the content coverage of E-LIS. The comparison of responses between different roles demonstrated that participants with more active role in the system are more favourable to E-LIS coverage (Editors M=3.88, S.D.=1.11 and Registered Users M=3.75, S.D.=0.94), while Unregistered Users are not completely satisfied by the content quantity found in it (M=3.09, S.D.=1.09). Those who reported very rare use disagreed with the statement of satisfaction of coverage (M=2.67, S.D.=1.17), while those who reported very often use stated also their agreement with the statement (M=4.12, S.D.=1.05). ANOVA showed that the factors appointed role and usage differentiate significantly the behaviour of the sub-samples ($p<.001$ for all features in the usefulness category).

5.1.2. Usability questions

In the usability category (2.1.-2.5.), Ease of Use and Learnability scored very high mean rates (4.06 and 4.08 respectively). Once again users with active presence in the archive (such as editors and registered users) consider E-LIS as a very easy to use DL (Editors M.=4.00, S.D.=0.79, Registered Users (M.=4.41, S.D.=0.76), but those with little interference with formal roles do not share the same opinion to the same extend (Unregistered Users M.=3.86, S.D.=1.18). Similar were the results in the Learnability attribute, where Editors (M.=4.12, S.D.=0.86) and Registered Users (M=4.36, S.D.=0.72) believe that E-LIS is a fairly learnable system, while the opinion of the Unregistered Users was the same, but to a lesser degree (M.=3.89, S.D.=1.10). Participants reporting very rare use deposited that they do not consider E-LIS as an easy of use DL (M.=3.17, S.D.=1.27) they were in contrast with those who reported very often use (M.=4.44, S.D.=0,65). Participants with very often use also believed that E-LIS is an easy to learn DL as well (M.=4.32, S.D.=0.75), but the participants with very rare use argued on its learnability (M.=3.08, S.D.=1.28)

A non-significant difference between the sub-groups representing different roles was traced in the case of Learnability (F=2.314, p>.05), while all other features showed significant differences at p<.001 for Aesthetic (F=7.410) and Terminology (F=5.558) and p<.05 for Ease of Use (F=2.781) and Navigation (F=3.023). Significant differences were also found between the sub-groups based on usage at the levels of p<.001 for Ease of Use (F=8.906), Aesthetic (F=7.806) and Navigation (F=5.820) and p<.05 for Terminology (F=4.619) and Learnability (F=3.646).

5.1.3. Performance questions

In the Performance category (3.1.-3.3.), the two dominating IR evaluation measures, Precision and Recall, didn't achieve to gather the preference of the participants. They did not supply a clear answer for their satisfaction with the Precision and the Recall of E-LIS. While Editors responses for Precision were above medium (M=3.65, S.D.=1.22) and were in accordance with the rates provided by Registered Users (M=3.61, S.D.=0.92), Unregistered Users' rates were below medium showing a disagreement about the ability of E-LIS to return precise results (M=2.85, S.D.=0.98). Similar were the results for Recall. Editors (M=3.65, S.D.=1.00) and Registered Users (M=3.48, S.D.=0.88) agreed that E-LIS could provide them the right number of results, but Unregistered Users demonstrated a negative attitude towards this feature (M=2.94, S.D.=0.96). Participants that reported very rare use believed that both Precision (M.=2.42, S.D.=1.18) and Recall (M.=2.46, S.D.=0.98) are not satisfactory. The participants with more frequent use of E-LIS declared their mediocre satisfaction with Precision M.=3.76, S.D.=1.05) and Recall (M.=3.84, S.D.=0.90).

Significant differences were indicated at the level of p<.001 for Precision (F=6.521) and Response Time (F=5.501), and p<.05 for Recall (F=4.446). Participants with very low frequency of usage showed their disagreement with the statements of satisfaction in Precision (M=2.42, S.D.=1.18) and Recall (M=2.46, S.D.=0.98). On contrast participants with high levels of usage agreed with the respective statements and provided scores above medium (M=3.76, S.D.=1.05 for Precision and M=3.84, S.D.=0.90 for Recall). Significant differences were reported for all types of reported usage (p<.001).

5.2. Inter-axes correlations

Although descriptive statistics show an image of users' opinions about the DLs features, we investigated the correlations that they appointed to them. Figure 2 presents the inter-category correlation values. All correlations have high rates above 0.70, with most notable the correlation between usefulness and performance. Although all correlations are significantly high, it can be inferred that participants are correlating particularly Usefulness with Performance due to their awareness of the criticality of these two concepts in the operation of a DL. It can be deduced that information scientists manage usefulness attributes and know how they are supporting users' work tasks, while they know the significance of IR evaluation criteria for the successful completion of their information searching activities.

Remarkable is the correlation between usefulness and usability verifying that user interaction quality depends strongly on both axes. Moreover it recommends that the DL evaluation process should consist of the examination of criteria that represent the characteristics of both usefulness and usability.

Further correlation analysis was performed to find significant associations between the model variables. Table 5 presents the inter-attribute Pearson correlations, all of which were found to be significant, ranging from r=0.35 to r=0.79. An overview of the table demonstrates significant correlations between Reliability and Format, Reliability and Level, and Coverage and Level for the usefulness category, Ease of Use and Navigation, Ease of Use and Learnability, Navigation and Aesthetics and Terminology and Learnability for the usability category and Recall and Precision for the performance category.

5.3. Predictive strength of factors

Stepwise multiple regression analysis was performed for each category of constructs in order to estimate their predictive strength. The attributes of each category were entered as independent variables and the regression model was performed on a dependent variable that measured the overall quality of each category. Table 6 presents the results of stepwise multiple regressions for each category. The multicollinearity values of VIF and Tolerance were calculated for each variable of the regression model. This check showed that all values were below 10 (min.=1.496, max.=3.369) and Tolerance were found more than .2, hence data multicollinearity was evaded.

Relevance (t(125)=4.697, p<.001), Level (t(125)=3.956, p<.001) and Coverage (t(125)=2.887, p<.01) had a significant effect on the prediction for DL usefulness. The three attributes are responsible for the 65% of the observed variance. Four out of five usability attributes of the usability category produce the 63% of variance and have a significant effect at the p<.01 level, while the fifth construct, Navigation, was excluded from the analysis (t(126)=.534, p>.05). The data in the Performance category reveal that Precision (t(127)=8.050, p<.001) and Recall (t(127)=4.467, p<.001) cause 66% of variance, when the remaining construct, Response Time (t(127)= 1.713, p>.05), has a non-significant effect.

5.4. E-LIS functionalities evaluation

Table 7 presents the participants' opinion on the questions of the third part of the questionnaire regarding E-LIS functionalities and characteristics and how they affect the system's overall usefulness and usability. The selected functionalities reflect adequately the E-LIS nature and its information management processes, which are the information discovery methods (e.g. search and browse), personalized delivery of services, peripheral services provision (e.g. e-mail alerts, cross-reference), its open access (OA) nature and finally the procedural phases (e.g submission, review, edit, delete of contributions, etc.). It is notable that more than fifty per cent of the participants strongly believe that E-LIS is both useful and usable to them due to its open access (OA) nature. In contrast, participants do not think that the personalized functionalities of E-LIS are useful and usable.

Stepwise multiple regression analysis was repeated including as variables the system functionalities in investigation, to predict the overall usefulness and usability. Despite the fact that five of six functionalities (browsing, search, personal account, services and OA) are considered to cause the 55% of variance (R^2=0.556, F=31.312, p<.001), only two of the functionalities possess significant strength to predict satisfaction on usefulness. In particular participants believe that the services provided by E-LIS (t(125)=3.227, p<.05) and the fact that it is an OA system (t(125)= 4.152, p<.01) are enough to predict usefulness.

The OA nature of E-LIS was strengthened even more as it was conceived by the participants as a feature that could predict the overall usability (t(125)=2.878, p<.01), together with the personalized functionalities that the e-print archive provides (t(125)= 2.317, p<.05). Once again the above mentioned five functionalities are considered to capture 57% of usability variance (R^2=0.579, F=34.331, p<.001).

6. Discussion

The present study reported differences in users' opinions as their experience changes. Users' experience, reflected in the role they have in the system and the reported usage, affects their perceptions about E-LIS usefulness and usability. This conclusion validates prior studies' findings (Koohang & Ondracek, 2005) and makes them applicable in the field of e-print systems. Differences between experienced and novice users were also reported for performance of e-prints usage (Silva, Laender & Gonçalves, 2005). Whilst one can support that there are studies claiming no actual differences in the user performance between novice and experienced users of DLs (Kengeri et al., 1999), this doesn't infer a clear conflict among the various experimental findings, but indicates the need to exploit a wide range of methods and metrics, especially in the usability field. Additionally the differences shown between the different roles suggest that the more active a user is the more favourable is towards E-LIS' usefulness and usability. This fact introduces a new parameter, totally depended to DL's nature, which are the engaging roles of the users that represent different needs, requirements and behaviours.

Concerning the other users' characteristics, E-LIS significance to the work tasks is not considered very important. Users appoint mediocre importance to E-LIS when they execute their work tasks and this low level of significance shows respectable differences only for the factor of usage (F=40.399, p=0.001). This leads to the conclusion that system usage affects preference. Participants who are making frequent use of a DL consider it as important for their work task. Furthermore, users are committed to spend as much time and effort is needed in using E-LIS to find the information they want. Once again, users' behaviour was significantly different between these two parameters (for Time F=96.456, p=0.001, while for Effort F=95.400, p=0.001). Finally the introduction channels suggest an elaboration of the promotional and marketing means that the administration team employs. These results, which are in agreement with the results from other studies (Hitchcock et al., 2002), request much more attention in the area of support and documentation, especially in such cases of subject-oriented e-print repositories, where spatial distribution restricts physical interaction with the served community.

Most of the usefulness attributes were highly appreciated by the users. However participants are not satisfied by E-LIS Coverage. Apparently they prefer more uploaded content and their preference is in accordance to the remarks of the Citebase subjects (Hitchcock et al., 2002). As the authors of the Citebase research note, the responsibility for content growth is transferred to the hands of the users themselves. Raw statistic evidence, provided by the E-LIS administrative team, showed a steady increase in the collection size. Therefore, despite the users' arguments and the fact that these systems are still in their infancy, the collection growth rate can be characterized high.

Users seem to prefer systems that provide them a structured and levelled presentation of information, which is relevant to their information tasks. However they do not appoint significance to the role of Format and Reliability. In the former attribute, this interesting exclusion can be credited to the deposit policy of the DL, which allows only the PDF file format. The result for the latter attribute can be explained by the fact that the only prerequisite for the documents deposit in E-LIS is relevance and document integrity. Therefore it can be supported that users of such systems do not assign significance to reliability, due to their awareness that the content review and approval procedure does not employ reliability criteria. It is remarkable that both attributes are highly correlated (r=0.74).

The five usability features scored high above the medium. Users reported their preference to Easiness of Use and Learnability, while understandable Terminology and Aesthetics do not attain high levels of users' preference. This allocation of preference is in accordance with conclusions by other studies, such as Kani, Ghinea and Chen (2006). According to their findings, users give higher priority to the easy discovery of information and the high degree of familiarization with DL functionalities. Moreover, according to ANOVA, Learnability is the only widely preferred attribute, regardless the participants' roles. Despite Navigation gained high scores by participants, it was not perceived as an important feature that affects overall opinion on DL usability. However literature reveals that navigation is considered an important factor that influences many of the traditional work tasks of librarians and information scientists (Moyo, 2002). One possible explanation might be that the information scientists with the assistance of appropriate navigation tools and aids (such as indices, menu bars, etc.) have gained the experience needed and may overtake any obstacles in their navigation routes. This experience is underlined by some DL development studies that highlight the role of experienced users, in the successful information discovery through navigation tools (Xu, 2004).

Although not a primary aim of this study, following the theoretic foundations of ITF, we investigated the performance axis. Precision and Recall are considered the main factors that affect users' opinion on system performance. Despite the dependence on factors that influence Response Time (e.g. traffic, webpage size etc.) and the critical position in the evaluation of web-based systems, users expressed higher levels of preference to the "traditional" measurements of Precision and Recall.

It is very intriguing the fact that participants in this study believe that the open access nature of e-print archives makes them both useful and usable. This may be associated with the current trends of the scientific community to advocate open access.

Participants in this study didn't appreciate the provision of personalized services, but they were satisfied with the usefulness and usability of the browsing and searching features. These features allow users to retrieve documents in a wide number of methods, such as author, country or year indices and advanced search interfaces. Therefore, participants' appreciation of this wealth of tools outweighs the inefficiency of E-Prints software, upon which E-LIS is build, to search on a full-text level, as reported by Goh et al. (2006).

Furthermore participants in this study correlated very highly the Usefulness and Performance categories, something that can be imputed to their profile characteristics (Tsakonas & Papatheodorou, 2006). Librarians and information scientists are paying much attention to issues concerning usefulness and performance, since these are closer to tasks performed in everyday activities.

7. Conclusions

While commercial DLs have proved their degree of self-sustainability, e-print archives operation is dependent on many issues, either political, or economic. One of the major challenges that e-prints face is to become self-sustainable systems, closely linked with users' work tasks, instead of being gradually transformed into graveyards of invaluable documents.

Although the issue of open access systems has gained significant attention and the consensus for their evaluation strengthens among the scientific community, there are not many studies concentrating on usefulness and usability issues. In this study we applied a theoretical model for DL evaluation to assess the usefulness and usability of an open access digital library and we proved experimentally that in general the proposed ITF fits to needs for the open access repositories evaluation. In particular this study attempted to analyze the significant content and system attributes and their interactions, and how they affect user interaction and satisfaction. These attributes request research attention and in depth analysis on each one will reveal more details and contribute to the overall evaluation of DLs.

Acknowledgements

The authors wish to thank Antonella DeRobbio, Imma Subirats, Zeno Tajoli and the national editors of E-LIS for their invaluable help and support during the realization of this study, Dimitris Gavrilis, Department of Electrical Engineering, University of Patras, for his technical assistance, and Tasos Tsagkos for his constructive comments on the statistic analysis of this study.

References:

Allen, M. (2002). A case study of the usability testing of the University of South Florida's virtual library interface design. *Online Information Review, 26*(1), 40-53.

Bertot, J.C., Snead, J.T., Jaeger, P.T. & McClure, C.R. (2006). Functionality, usability, and accessibility: iterative user-centered evaluation strategies for digital libraries. *Performance Measurement and Metrics, 7*(1), 17-28.

Center for Research Libraries (2006). General factors to consider in evaluating digital repositories. *Focus, 25*(2), 3-4.

Cockrell, B.J. & Jayne, E.A. (2002). How do I find an article? Insights from a web usability study. *Journal of Academic Librarianship, 28*(3), 122-132.

Correia, A.M.R. and Castro Neto, M. (2002). The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA-a case study by the National Library of Portugal. *Journal of Information Science, 6*(28), 231-241.

Delone W.H. and McLean E.R. (1992). Information systems success: the quest for the depending variable. *Information Systems Research, 3*(1), 60-95.

Ebenezer, C. (2003). Usability evaluation of an NHS library website. *Health Information and Libraries Journal*, *20*(3), 134-142.

Ferreira, S.M. & Pithan, D.N. (2005). Usability of digital libraries: a study based on the areas of information science and human-computer interaction. *OCLC Systems and Services, 21*(4), 311-323.

Fidel R. & Green, M. (2004). The many faces of accessibility: engineers perceptions of information sources. *Information Processing & Management, 40*(3), 563-581.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A. & Sølvberg, I. (2002). Digital libraries: a generic classification and evaluation scheme. In *Proceedings of the 5th European Conference on Digital Libraries 2001: LNCS Vol. 2163*, (pp. 187-199). Berlin: Springer.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.P., Kovács, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C. & Sølvberg, I. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries* (in press).

Fuller, D.M. & Hinegardner, P.G. (2001). Ensuring quality website redesign: the University of Maryland's experience. *Bulletin of the Medical Library Association, 89*(4), 339-345.

Goh, D.H., Chua, A., Khoo, D.A., Khoo, E.B., Mak, E.B. & Ng, M.P. (2006). A checklist for evaluating open source digital library software. *Online Information Review*, *30*(4), 360-379.

Hartson, R.H., Shivakumar P. & Pérez-Quiñones, M.A. (2004). Usability inspection of digital libraries: a case study. *International Journal on Digital Libraries, 4*(2), 108-123.

Heery, R. & Anderson, S. (2005). Digital repositories review. Report by UKOLN and AHDS. Retrieved January 24, 2007 from http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf.

Hitchcock, S., Woukeu, A., Brody, T., Carr, L., Hall, W. & Harnad, S. (2002). Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service. Retrieved January 24, 2007 from http://opcit.eprints.org/evaluation/Citebase-evaluation/evaluation-report.html.

Hong, W., Thong, J.Y.L., Hong, W. & Tam, K. (2002). Determinants of user acceptance of digital libraries: an empirical examination of individual differences and system characteristics. *Journal of Management Information Systems, 18*(3), 97-124.

Jackson, M. (2001). A user centered approach to the evaluation of a hybrid library. *Performance Measurement & Metrics, 2*(2), 97-107.

Järvelin, K. & Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information Research, 10*(1). Retrieved January 24, 2007 from http://InformationR.net/ir/10-1/paper212.html.

Jeng, J. (2005). What is usability in the context of digital library and how it can be measured? *Information Technology and Libraries, 24*(2), 47-56.

Kani-Zabihi, E., Ghinea, G. & Chen, S.Y. (2006). Digital libraries: what do users want? *Online Information Review, 30*(4), 395-412.

Kengeri, R, Seals, C.D., Harley, H.D., Reddy, H.P. & Fox, E.A. (1999). Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal on Digital Libraries, 2*(2-3), 157-169.

Kobayashi, M. & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys, 32*(2), 144-173.

Koohang, A. & Ondracek, J. (2005). Users' views about the usability of digital libraries. *British Journal of Educational Technology, 36*(3), 407-423.

Krottmaier, H. (2002). Current and future features of digital journals. In *M.H. Shafazand and A.M. Tjoa (Eds), EurAsia-ICT 2002: Information and Communication Technology: First EurAsian Conference: LNCS Vol. 2510*, (pp. 495-502). Berlin: Springer.

Kurata, K., Matsubayashi, M., Mine, S., Muranushi, T. & Ueda, S. (2006). Electronic journals and their unbundled functions in scholarly communication: views and utilization by scientific, technological and medical researchers in Japan. *Information Processing & Management* (in press).

Kyrillidou, M. & Giersch, S. (2005). Developing the DigiQUAL protocol for digital library evaluation. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries,* (pp. 172-173). New York, NY: ACM Press.

Liaw, S. & Huang, H. (2003). An investigation of user attitudes towards search engines as an information retrieval tool. *Computers in Human Behavior, 19*(6), 751-762.

Liu, Z. (2004). Perceptions of credibility of scholarly information on the web. *Information Processing & Management, 40*(6), 1027-1038.

Majid S. & Tee Tan, A. (2002). Usage of information resources by computer engineering students: a case study of Nanyang Technological University, Singapore. *Online Information Review, 26*(5) 318-325.

Mat-Hassan, M., Levene, M, (2001). Can navigational assistance improve search experience? A user study. *First Monday, 6*(9). Retrieved January 24, 2007 from http://firstmonday.org/issues/issue6_9/mat/index.html.

McMullen, S. (2001). Usability testing in a library Web site redesign project. *Reference Services Review, 29*(1), 7-22.

Moyo, L.M. (2002). Collections on the Web: some access and navigation issues. *Library Collections, Acquisitions, and Technical Services, 26*(1), 47-59.

Nicholas, D., Huntington, P. & Rowlands, I. (2005). Open access journal publishing: the views of some of the world's senior authors. *Journal of Documentation, 61*(4), 497-519.

Nunnaly, J.C. & Bernstein, I.H. (1994). *Psychometric theory*, 3rd ed. New York: McGraw-Hill.

Park, S. (2000). Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries. *Journal of the American Society for Information Science, 51*(5), 456-468.

Silva, L.V., Laender, A.H.F. & Gonçalves, M.A. (2005). A usability evaluation study of a digital library self-archiving service. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, Denver, CO, USA*, (pp. 176-177). New York, NY: ACM Press.

Sutcliffe, G., Ennis, M & Hu, J. (2000). Evaluating the effectiveness of visual user interfaces for information retrieval. *International Journal of Human-Computer Studies, 53*(5), 741-763.

Thong, J.Y.L., Hong, W. & Tam, K. (2002). Understanding user acceptance of digital libraries: what are the roles of interface characteristics, organizational context, and individual differences? *International Journal of Human-Computer Studies, 57*(3), 215-242.

Toms, E.G. (2002). Information interaction: providing a framework for information architecture. *Journal of the American Society for Information Science and Technology, 53*(10), 855-862

Tsakonas, G. & Papatheordorou, C (2006). Analysing and evaluating usefulness and usability in electronic information services. *Journal of Information Science, 32*(5), 400-419.

Tsakonas, G., Kapidakis, S. & Papatheodorou, C. (2004). Evaluation of user interaction in digital libraries. In *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries, Padua, Italy*. Retrieved January 24, 2007 from http://dlib.ionio.gr/wp7/workshop2004_program.html.

Vakkari, P. & Hakala, P. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation, 56*(5), 540-562.

Van House, N.A., Butler, M.H., Ogle, V. & Schiff, L. (1996). User-centered iterative design for digital libraries: the Cypress experience. *D-Lib Magazine, 2*(2). Retrieved January 24, 2007 from http://www.dlib.org/dlib/february96/02vanhouse.html.

Wolfram, D. & Xie, H. (2002). Traditional IR for web users: a context for general audience digital libraries. *Information Processing and Management, 38*(5), 627-648.

Wyles R., Maxwell, M. & Yamog, J. (2006). Technical Evaluation of selected Open Source Repository Solutions. Retrieved January 24, 2007 from https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf .

Xie, H. (2006). Evaluation of digital libraries: criteria and problems from users' perspectives. *Library and Information Science Research, 28*(3), 433-452.

Xu, Q. (2004). Content management and resources integration: a practice in Shanghai Digital Library. In *Proceedings the 7th International Conference on Asian Digital Libraries, ICADL 2004, Digital Libraries: International Collaboration and Cross-Fertilization: LNCS Vol. 3334*, (pp. 25-34). Berlin: Springer.

Yang, Z., Cai, S., Zhou, Z. & Zhou, N. (2004). Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Information & Management, 42*(4), 575-589.

CVs

Giannis Tsakonas. Giannis Tsakonas is PhD candidate student in Department of Archives and Library Sciences, Ionian University, Corfu, Greece. He also is librarian in the Library and Information Service, University of Patras, Greece. His research interests include user centred digital library evaluation, information behaviour and visual communication.

Christos Papatheodorou. Christos Papatheodorou holds a BSc. and a Ph.D in Computer Science from the Department of Informatics, Athens University of Economics and Business. He is Assistant Professor at the Department of Archive and Library Sciences, Ionian University, Corfu, Greece, where he teaches Information Systems, Metadata and Information Retrieval. His research interests include, user-centred digital library evaluation, user modelling, web mining and metadata interoperability.

Tables

Table 1
Cronbach α

Usefulness		Usability		Performance	
Group	.910	Group	.939	Group	.854
1.1	.902	2.1	.922	3.1	.795
1.2	.895	2.2	.932	3.2	.858
1.3	.886	2.3	.924	3.3	.823
1.4	.886	2.4	.929	3.4	.775
1.5	.906	2.5	.927		
1.6	.888	2.6	.931		

Table 2
Demographic data

Role		Area		Introduction Channel	
	n (%)		*n* (%)		*n* (%)
Editors	17 (12.98)	Africa	2 (1.53)	Announcements	29 (22.14)
Registered Users	44 (33.59)	Asia	14 (10.69)	By my own	31 (23.66)
Unregistered Users	66 (50.38)	Europe	76 (58.01)	Papers	20 (15.27)
Other	4 (3.05)	N. America/Caribbean	22 (16.79)	Personal communication	33 (25.19)
		S. America	12 (9.16)	Other	18 (13.74)
		Oceania	5 (3.82)		
Total	131 (100.00)		131 (100.00)		131 (100.00)

Table 3
Descriptive statistics (1= disagree, 5=agree)

	1 *n* (%)	**2** *n* (%)	**3** *n* (%)	**4** *n* (%)	**5** *n* (%)	**M**	**S.D.**
Usage	24 (18.32)	24 (18.32)	38 (29.01)	20 (15.27)	25 (19.08)	2.98	1.359
E-LIS Significance	21(16.03)	32 (24.43)	45 (34.35)	20 (15.27)	13 (9.92)	2.79	1.183
Time	7 (5.34)	18 (13.74)	43 (32.82)	31 (23.66)	32 (24.43)	3.48	1.159
Effort	7 (5.34)	18 (13.74)	32 (24.43)	39 (29.77)	35 (26.72)	3.59	1.176

Table 4
Descriptive statistics for questions in features categories (1= disagree, 5=agree)

	1 *n* (%)	**2** *n* (%)	**3** *n* (%)	**4** *n* (%)	**5** *n* (%)	**M**	**SD**
Usefulness							
1.1. Relevance	8 (6.11)	9 (6.87)	30 (22.90)	47 (35.88)	37 (28.24)	3.73	1.129
1.2. Format	5 (3.82)	8 (6.11)	32 (24.43)	42 (32.06)	44 (33.59)	3.85	1.075
1.3. Reliability	8 (6.11)	7 (5.34)	34 (25.95)	55 (41.98)	27 (20.61)	3.66	1.058
1.4. Level	6 (4.58)	14 (10.69)	25 (19.08)	43 (32.82)	43 (32.82)	3.79	1.150
1.5. Coverage	7 (5.34)	16 (12.21)	47 (35.88)	36 (27.48)	25 (19.08)	3.43	1.096
Usability							
2.1. Ease of use	5 (3.82)	3 (2.29)	25 (19.08)	44 (33.59)	54 (41.22)	4.06	1.021
2.2. Aesthetic	5 (3.82)	8 (6.11)	39 (29.77)	48 (36.64)	31 (23.66)	3.70	1.021
2.3. Navigation	6 (4.58)	6 (4.58)	23 (17.56)	54 (41.22)	42 (32.06)	3.92	1.045
2.4. Terminology	4 (3.05)	7 (5.34)	21 (16.03)	56 (42.75)	43 (32.82)	3.97	0.992
2.5. Learnable	4 (3.05)	3 (2.29)	23 (17.56)	50 (38.17)	51 (38.93)	4.08	0.966
Performance							
3.1. Precision	8 (6.11)	22 (16.79)	48 (36.64)	39 (29.77)	14 (10.69)	3.22	1.047
3.2. Response Time	4 (3.05)	5 (3.82)	31 (23.66)	48 (36.64)	43 (32.82)	3.92	0.997
3.3. Recall	6 (4.58)	20 (15.27)	54 (41.22)	40 (30.53)	11 (8.40)	3.23	0.965

Table 5
Inter-item correlation matrix

	1.1.	1.2.	1.3.	1.4.	1.5.	2.1.	2.2.	2.3.	2.4.	2.5.	3.1.	3.2.	3.3.
1.1.	**1**												
1.2.	**0.62**	**1**											
1.3.	**0.63**	**0.74**	**1**										
1.4.	**0.57**	**0.65**	**0.72**	**1**									
1.5.	**0.44**	**0.50**	**0.61**	**0.65**	**1**								
2.1.	0.58	0.63	0.61	0.57	0.49	**1**							
2.2.	0.51	0.61	0.52	0.48	0.44	**0.74**	**1**						
2.3.	0.54	0.61	0.59	0.52	0.37	**0.79**	**0.76**	**1**					
2.4.	0.58	0.67	0.59	0.64	0.47	**0.70**	**0.65**	**0.71**	**1**				
2.5.	0.54	0.62	0.57	0.50	0.47	**0.76**	**0.63**	**0.73**	**0.78**	**1**			
3.1.	0.56	0.53	0.67	0.63	0.59	0.54	0.51	0.54	0.54	0.50	**1**		
3.2.	0.52	0.67	0.58	0.58	0.48	0.61	0.54	0.55	0.66	0.66	**0.52**	**1**	
3.3.	0.57	0.40	0.53	0.50	0.56	0.45	0.35	0.38	0.40	0.45	**0.58**	**0.49**	**1**

$p<0.01$

Table 6
Multiple regression analysis

	B	***S.E. B***	***β***	***p***
Usefulness				
R^2= 0.653, *Adjusted* R^2= 0.639				
Relevance	0.346	0.074	.338	<.001
Format	0.053	0.090	.049	>.05
Reliability	0.034	0.102	.031	>.05
Level	0.342	0.086	.340	<.001
Coverage	0.220	0.076	.208	<.01
Usability (+)				
R^2= 0.634, *Adjusted* R^2= 0.622				
Easy to use	0.245	0.091	.267	<.01
Aesthetic	0.159	0.077	.173	<.01
Terminology	0.187	0.087	.198	<.01
Learnability	0.248	0.096	.256	<.01
Performance				
R^2= 0.665, *Adjusted* R^2= 0.657				
Precision	0.494	0.061	.539	<.001
Response Time	0.104	0.061	.108	>.05
Recall	0.292	0.065	.294	<.001

(+) : Navigation variable removed

Table 7
Descriptive statistics for functionalities questions (1= disagree. 5=agree)

	1 *n* (%)	**2** *n* (%)	**3** *n* (%)	**4** *n* (%)	**5** *n* (%)	**M**	**SD**
Browsing usefulness	5 (3.8)	9 (6.9)	24 (18.3)	52 (39.7)	41 (31.3)	3.88	1.053
Browsing usability	5 (3.8)	8 (6.1)	35 (26.7)	43 (32.8)	40 (30.5)	3.80	1.063
Search usefulness	5 (3.8)	7 (5.3)	33 (25.2)	46 (35.1)	40 (30.5)	3.83	1.046
Search usability	5 (3.8)	8 (6.1)	37 (28.2)	41 (31.3)	40 (30.5)	3.79	1.067
Personal Account usefulness	19 (14.5)	16 (12.2)	33 (25.2)	31 (23.7)	32 (24.4)	3.31	1.354
Personal Account usability	18 (13.7)	18 (13.7)	31 (23.7)	34 (26.0)	30 (22.9)	3.31	1.335
Services usefulness	15 (11.5)	5 (3.8)	31 (23.7)	42 (32.1)	38 (29.0)	3.63	1.260
Services usability	15 (11.5)	6 (4.6)	32 (24.4)	45 (34.4)	33 (25.2)	3.57	1.241
OA usefulness	9 (6.9)	3 (2.3)	23 (17.6)	29 (22.1)	67 (51.1)	4.08	1.183
OA usability	9 (6.9)	5 (3.8)	26 (19.8)	30 (22.9)	61 (46.6)	3.98	1.202
Procedures usefulness	10 (7.6)	7 (5.3)	36 (27.5)	44 (33.6)	34 (26.0)	3.65	1.150
Procedures usability	9 (6.9)	7 (5.3)	41 (31.3)	38 (29.0)	36 (27.5)	3.65	1.143

Figures and captions

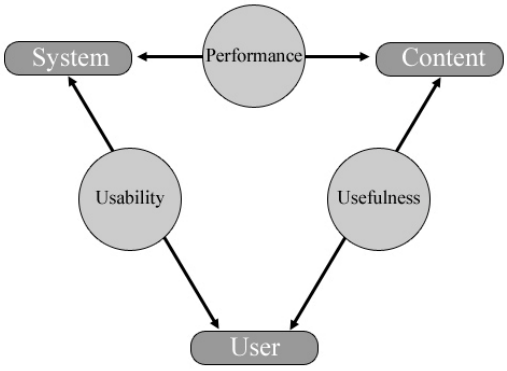

Fig. 1. Interaction Triptych Framework
(See attached file Figure1.tiff)

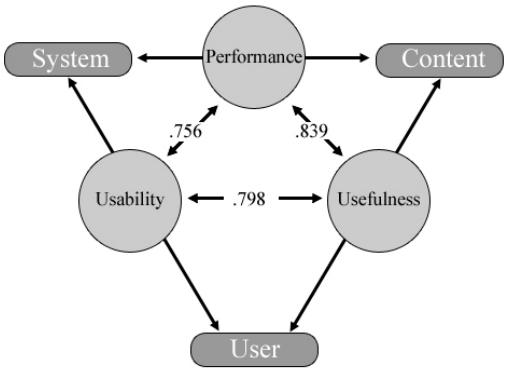

Fig. 2. Correlations between the evaluation axes.
(See attached file Figure2.tiff)