

**ECP-2005-CULT-038074**

**EDLproject**

**D1.2. Final Report on Usability Developments  
in The European Library**

<b>Deliverable number</b>	<i>D1.2</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>19 March 2008</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>Name(s): Georgia Angelaki – The European Library Office – The Netherlands - <a href="mailto:georgia.angelaki@kb.nl">georgia.angelaki@kb.nl</a></i>

The EDLproject is funded under the European Commission *eContentplus* Programme.



**eContentplus**

## Contents

1. Executive summary.....	3
2. Structure of the current report .....	4
3. Registration and authentication .....	4
4. Log files Analysis and Personalisation of Services .....	5
4.1 Findings and Summary of Interim Report .....	5
4.2 Usability developments in The European Library .....	8
4.3 Considerations about the log files analysis .....	9
4.4 Further work on the log files: towards true personalisation .....	10
5. Promotion of OAI-PMH.....	11
5.1 Achievements .....	11
5.2 Future work .....	13

## 1. Executive summary

The present report is the final deliverable D1.2 of [EDLproject](#) WP1 T3 about usage and usability developments in The European Library.

The task lasted from September 2006 to February 2008, that is, for the whole duration of EDLproject. The goal was to investigate and promote three specific areas of usage and usability of The European Library:

- the development of a registration and authentication procedure for the users of the portal,
- the preliminary analysis of The European Library log files in view of identifying usage habits and patterns in order to enhance the usability of the portal and to be able to provide personalised services to the users and
- the promotion of [OAI-PMH](#) protocol to EDLproject partners

The registration and authentication function serves a dual role: it helps offer additional services to authenticated users as well as gather information about the users which we aim at using in the future in order to provide personalised services.

The initial goal as regards the log files analysis task was to investigate the intuitive actions performed by the users in TEL in order to identify user groups and usage patterns which would feed into usability developments in TEL and eventually into providing personalised services for the users. In this framework a small project was held with two external expert teams on the analysis of The European Library log files.

The project was very informative about the use of the log files in monitoring users' actions and the performance of TEL and provided general recommendations for improvements of particular aspects of the portal. It also proved that richer sources of information than the existing TEL ones -both rich log data and explicit feedback from the users- are necessary in order to identify individual users and therefore to provide true personalised services. It finally brought out the importance of the registration of users which helps identify them on an individual basis and track down their actions and click stream history.

The promotion of the Open Archives Protocol for Metadata Harvesting is in general deemed successful. All the EDLproject partners have realised its importance for speeding up and enhancing metasearch in TEL and for providing more usable and relevant results to their users and have either implemented it or are willing to adopt it soon. Additionally, concerted efforts from the TEL office and the Austrian partners have put pressure on digital library systems vendors to include OAI-PMH modules in their products contributing in the wider adoption of the standard.

In continuation to the [interim report](#)<sup>1</sup> delivered in September 2007 (M1.4), the current report summarizes the activities and the outcomes of the task and its impact for TEL. The activities on the

---

<sup>1</sup> The report is uploaded in the password-protected area of the EDLproject website

log files analysis and the promotion of OAI-PMH are continued in a more systematic way in [TELplus](#), a new project that kick-started in October 2007 and which is anticipated to develop significantly the usability of The European Library.

## 2. Structure of the current report

The three subtasks are loosely connected and are presented separately in the report. Some information included in the interim report is considered given. In each section there is a summary of activities of the subtask, results and follow-up actions.

## 3. Registration and authentication

The [registration](#) and login functionality in The European Library was set up in August 2006. It serves a dual role: to gather personal information about our users (sex, education, profession, preferences) in order for us to be able to develop and give them in return personalised information and services.

The outcome of the reports of [UNIPD](#)<sup>2</sup> and [MPII](#) was that the little extra functionality offered to the logged users at present does not make it attractive for them to log in. Both of them concluded that special care should be taken in promoting the registration functionality to the users, possibly by offering them additional services.

The table below shows how the number of registered users has increased the last months:

Month	Number of registrants
January	519
February	581
March	661
April	767
May	826
June	880
July	1.134
August	1.259
September	1.344
October	1.446
November	1.660
December	1.830

---

<sup>2</sup> The two teams analysed The European Library log files and data from the registered users in the framework of the subtask presented in the next section and delivered reports can be found online at password-protected area of the edlproject website. More information about this project is included in the next chapter.

The registered users have more than doubled in the past months. This is an encouraging development taking into account that it has not been possible<sup>3</sup> in the meantime to schedule, develop and offer additional services to the users to make it more attractive for them to login. The following services are though under investigation and are going to be developed very soon:

Provide RSS feeds and email alerts to the users based on their declared interests. The functional requirements are currently under investigation and the service will be implemented as soon as these are clear.

Offer an annotation service for images. This is a tool developed by [ARC](#) research centers in Austria and helps users select, annotate, store fragments or whole pictures with the annotations as well as search the annotated text. This is a very interesting tool for the users. The tool is ready to be added and user requirements for its integration in TEL are being investigated. Some user testing might be needed before it is made available.

Additionally, the following development will be added:

Insert in the authentication process the possibility to automatically recognise the user without the request of user name and password. This will be done with the use of the cookies.

## 4. Log files Analysis and Personalisation of Services

### ***4.1 Findings and Summary of Interim Report***

In spring 2007 The European Library launched a collaboration with the University of Padua- [Department of Information Engineering](#) and [Max Planck Institute for informatics](#) on the analysis of the portal log files. The aim was to analyse the two types of log files of The European Library, the http log files showing traffic to the portal and a custom made recording system that logged the interaction of the users with The European Library- called the TEL Action Logs (previously the TEL server logs). The data from the registered users database were also used in order to combine users' clickstream history with their declared preferences and personal characteristics. The goal was to analyse the portal from a usability point of view, to identify where users stumble and to identify user groups and usage patterns. Ultimately, the aim was to be able to trace individual users and to provide to them personalised services.

The main outcomes of this project were the following:

---

<sup>3</sup> The priority set was to replan and rebuild the back-end technical infrastructure of The European Library in a way that it offer a lighter, more stable, modular and scalable architecture which would also be able to accommodate the incremental content stored and indexed centrally and to be able to offer more services and more personalised services to the users. This is the version 2.0 of the portal which is due to be released in April 2008. Any developments concerning the front-end such as the registration functionality and the services offered to the registered users will be planned for release 2.1 of the portal which is due in June 2008.

- The University of Padua analysed mainly the http logs from October 1<sup>st</sup> 2006 to 30 April 2007. It provided an initial analysis of the http log files and the user findings that result from this analysis. General recommendations about some usability improvements in the portal were made. Some of these and their implementation in the TEL portal are presented later in this report.
- Max Planck Institute analysed the Action (server) logs from 19 December 2006 to 31<sup>st</sup> May 2007. They analysed mainly the action logs posing questions related to search habits. Some interesting observations arose and some recommendations were given especially as regards better presentation and ordering of the collections descriptions in the portal. Additionally, MPII provided the methodology towards personalised search and retrieval of results as regards automatic query expansion and query-driven selection of collections. Unfortunately, the limited availability of data from the registered users and from the logs made it difficult to provide statistically valid and applicable recommendations.
- UNIPD brought out the limits of building profiles based on the analysis of the log files only. Such efforts should be based on a combined analysis of the logs with user studies in order to build group profiles or explicit feedback from the user in order to build individual profiles that can be used to provide personalised services.
- Granularity of possible personalisation on the portal depends on the richness of information from (direct explicit feedback) and about the users (implicit feedback), ie rich log files especially from returning users, ample information from the registered users, as well as expressed interests and feedback from these users.

#### **Findings about the users:**

##### Demographic information:

- users are coming mainly from Europe and the States (though there are a lot of crawlers from the States). France, Germany, Poland, Italy and the Netherlands were the 5 main countries of origin.

##### Social characteristics:

- they are higher education students and education professionals as well as information professionals
- there was also a lot of information about the “non-real” users, that is software agents such as crawlers but this was not very interesting for us to present in the context of the current analysis.
- Technical characteristics: most of them use Windows (74%) and Explorer (60%).

##### Registered users:

- They form a particular group. The fact that they register might show that they have some higher expectations than the non-registered users.

- They register but a very small percentage returns as a logged in user. 349 out of 614 for this period returned at least once while 365 registered but never logged in which means that their expectations are probably not met. Of the logged in users almost half did not perform any query.

Usage habits:

- Most of the users (82,35%) are one-time users. Among the rest, only a 0,47% accessed the site more than 15 times.
- Most of the users perform one query and spend less than a minute on the portal
- There is though a sizable number of users that spend 2-10 minutes on the portal
- Most of the users find us via websites where the Mini Search Box is placed, via google or blogs that make a mention to The European Library.
- There is a peak in usage from the countries whose national libraries recently joined TEL
- 84% does not change the language of the interface regardless of the country of origin
- The vast majority performs a query on the default list of collections. Which means that they do not understand the existence and use of the collections

Search-related observations (which are more obvious from the action logs):

- Most of the users that visit TEL do not perform any query.
- Most queries effectuated are one keyword search. The total average is around 2,55 queries per session.
- Advanced search is little used.
- There is occasionally use of advanced query syntax in the simple search box.
- In 38% of the (1<sup>st</sup>) queries, the users rephrase their query.
- 25% views another resultlist
- Only 15% requests the next 10 results
- 17% views a detailed record information
- As regards the last state of a session:
  - o Around 18% of the sessions end in result or follow-up action states
  - o 22% end in a search state
  - o 46% end in result list
  - o 4% end in collection selection

- Keywords relating to a nameplace or to an author appear often: ie Poland, France, polska, Zagreb, Shakespeare, Kochanowski, Dante, Goethe, etc.
- When the users select collections the vast majority prefers the thematic over the selection by country, subject or description.
- Users most often view the results in the default order but are looking more thoroughly at collections that are not always ranked higher in the default list.

Towards true personalisation:

- Furthermore, a model was considered to identify query –specific rankings of collections. The model considers user-related and query-related parameters such as collection preferences, country of provenance, language, the query associated with a collection preference, etc. The model will prove its usefulness as more data accumulate. The above constitute mainly areas of future work.

## ***4.2 Usability developments in The European Library***

The findings of the two institutions shed some light into some usage habits but also raised more questions: for example, the users do not change the default language because they are competent enough with the English language or because they do not see the “select language” option?

Furthermore, the analysis of the log files proved that the data are still too scarce to provide a rich granularity of information that permits also identifying individual users, their actions and behaviors. However, it proved a valuable source of information for identifying “weak” points of the portal which fed into overall developments.

The following developments in The European Library portal are directly emanating from the log files analysis or are in line with them:

- As regards the registration procedure the recommendations and relevant improvements are mentioned earlier
- As regards the search and the collections selection the following improvements have been made on the portal or are scheduled to happen:

In The European Library release 1.6 (December 2007)

- o In the portal release v1.6 (December 2007) collections were regrouped under certain themes: “catalogues from national libraries”, “digital collections”, “maps”, “photographs, posters and images”, “portraits”, “newspapers and periodicals”, “manuscripts”, “theses and dissertations”. These categories were added under the search box and can be selected using a radio button. This is a major development to help the selection and use of collections “Personal selection” is not yet activated but will be in the portal release v2.0 in April 2007.

- Collections containing digital items were grouped and form a separate thematic group offering search possibility only among digitised objects- catalogues of national libraries.
- A more appropriate and self-explanatory naming was adopted for the collections: ie: “catalogues from national libraries” has replaced “default list of collections”
- After the user has performed a query he is given the possibility to further narrow down his search within the results of the last query.

#### **In The European Library release 2.0 (April 2008)**

- Another scheduled development is the automatic query expansion functionality that is currently being investigated and implemented with the use of subject headings, name authorities and vocabularies from the partner libraries. This aims to improve query entries with wrong spellings and provide the user with search variants.

Further developments are considered to improve the presentation of the results such as clustering and ordering, etc. Any major such developments are suspended for the moment as a new architecture is being developed for TEL to allow for greater speed in retrieval of results and to provide the backbone for personalised services to be developed. The new technical architecture will be live in the portal release v2.0 in April. Major services and content related developments are scheduled for v2.1 of the portal (July 2008).

### ***4.3 Considerations about the log files analysis***

- Great number of registered users is needed to observe individual users and usage habits that can form patterns and can be used in a way to provide truly personalised services. Log files do not always refer to one individual; many institutions use proxy servers that can collapse several users into one. Therefore, the more information about the registered users (explicit feedback) there is the better log files (implicit feedback) can be used to analyse individual usage habits and preferences.
- Collaborating teams on the log files analysis have to agree upon the analysis approach and harmonise the methodology used. Sometimes the approach adopted does not provide very usable findings. For example UNIPD based its findings on the reconstruction of sessions with the use of heuristics which collapsed human and software agents blurring the results as regards human users and their behavior.
- The European Library has two to three releases per year. This is not so much a limitation but it definitely imposes extra difficulty in the analysis of the logs as files migrate, merge or are renamed and it takes some time to sort out the new logged information.
- It is becoming clear that all sources of information from user feedback are needed to inform us about the usability of the site. These are: the feedback forms, statistics, user studies and evaluation studies. Via these we can identify user groups and this information can be processed and be inserted in usage profiles.

- Priorities need to be set for the investigation of the characteristics and habits of particular user groups; the academic users are the main target group set in The European Library product and service plan for 2008.
- For this reason UNIPD parallel to the log files analysis also performed a small scale user study with students from the University to understand better the motives of users (explicit feedback) and analyse these in combination with their log files (implicit feedback)

#### ***4.4 Further work on the log files: towards true personalisation***

Following up this initial work, there is a new project that started in October 2007 called TELplus and which will provide further insight into the analysis of the users of TEL. [TELplus](#) is an eContentplus funded project coordinated by the national library of Estonia and EREMO srl. The project aims at improving access to digital content and the overall usability of The European Library. WP5 in particular is the continuation of the work that started on the log files analysis and is led by Maristella Agosti at the University of Padua. Other partners to the project are: the National Kapodistrian University of Athens (NKUA), the national libraries of France(BnF), the Netherlands (KB) and of Italy in Florence (BNCF) as well as the Austrian Research Centers GmbH (ARC). The workpackage aims at:

- Analysing the user requirements for building personalised services in The European Library
- Data mining and analysing the log files (http and action logs) for user modeling with different levels of granularity
- Investigating and building of the prototype for a profile service that makes use of the extracted knowledge and which in turn
- Will be used to invoke personalised search-related services: personalised search and context-aware query expansion; personalised alerting and notification; personalised retrieval of results (ranking, presentation-ie personalised summaries)
- Additionally, the user requirements will feed into the development of an annotation sharing prototype tool.

The workpackage kick- started with a meeting in Padua on February 18 and 19. The national libraries presented the work they have done on analyzing user requirements and how these feed into the development of services and of the usability of their respective national portals. The European Library will make available to the research team the past user surveys it has realized as well as any other information that can help into identifying user groups that it has in hand such as evaluation studies of the portal performed by third parties. The project ends in December 2009 and most of the above results will be made available then.

## 5. Promotion of OAI-PMH

It can be considered that the OAI-PMH<sup>4</sup> protocol was successfully promoted to the EDLproject partner libraries. This was the result of the following actions:

- Holding of a workshop about implementing OAI-PMH during the EDLproject and Technical Working Group joint meeting and workshops (January 10, 11 and 12 2007).
- Development of a platform independent installer package (server-side). This package installs a stand-alone java server that allows metadata in xml format to be made available via the OAI-PMH protocol. Additionally, an OAI-PMH harvester (client-side) was made available to the partners. It is a java tool that The European Library Office uses to harvest the OAI-PMH targets. It is quite simple and straightforward to use and is publicly available. These two tools are available via [The European Library Handbook](#) and significantly lowered the barrier of implementation for our partners.
- Systematic promotion of the protocol by Olaf Janssen, EDLproject WP1 manager via the teleconferences he held during the project with the EDLproject partner libraries.
- Successful problem solving with the support of the relevant technical developer of TEL.

### 5.1 Achievements

The greatest achievement is that initial fears of partners were overcome and national libraries are more ready to accept to give away their metadata and provide alternative ways to their users to find the content(via TEL and other search engines)as well as integrated and more usable results to their queries.

Additionally, it was realised that it was not so difficult to shift away from Z39:50 to a more web-fit protocol and that implementation was relatively easy with the use of the tools and the support that TEL provided.

The developments in TEL and in digital libraries at large have resulted in market pressure to the library management systems vendors such as EXLIBRIS to integrate an OAI-PMH module in their library management system [ALEPH](#) to facilitate further their library customers in making their records harvestable via the protocol.

Throughout 2007 the national library of Austria has worked on adaptation of the Aleph OAI-module to cater for DCqualified and later on the TEL Application Profile. Having been adapted by Austria, the other Aleph using libraries(\*) should be able to configure the OAI-module relatively easily for TEL needs. On 9th October 2007 TEL Office distributed the adapted Aleph configuration files among the

---

<sup>4</sup> An introduction to OAI-PMH and the reasons why The European Library strongly supports it are included in the interim report.

libraries<sup>5</sup>. Luxembourg and Sweden have been collaborating to tackle common problems/issues/bugs for the correct configuration of the OAI-provider

Spain has integrated many browse-only collections into a Digitool environment to form a digital library. This new environment supports OAI-PMH. In future the records in the digital library will be harvested into the TEL Central Index<sup>6</sup> as well.

All in all, all the EDLproject partners (and not only) have provided their collections for harvesting to TEL or are ready to do this<sup>7</sup>.

Concurrently, TEL has been preparing to host all the harvested records addressing the implementation and interoperability issues which became obvious gradually as well as the huge storage needs for all the records of the (soon to be 45) national library partners of The European Library. TEL servers and central repository is currently hosted in the national library of the Netherlands (KB). Addressing the hosting issue of the records of all the national libraries had to be worked out thoroughly<sup>8</sup>. A solution of hosting external repository(s) was considered to accommodate for future needs as well. Investigations with the KB and with hosting companies are under way and a solution will be put in place in spring 2008.

Additionally, the records of the partners had to be checked manually for their appropriateness for harvesting and problem solving had to be done on an individual basis.

As a result of this, harvesting of records is queued for the moment and is addressed on an individual country basis. This process will be automated once the storage requirements are met and the partners make use of the OAI harvesters and servers.

---

<sup>5</sup> Aleph using libraries : Austria, Luxembourg, Liechtenstein, Sweden, Iceland, Latvia, Czech Republic, Denmark, Russia State Library, British Library.

<sup>6</sup> In TEL we call Central Index the central repository of the harvested records that is being built and which are also indexed with using the Verity search engine

<sup>7</sup> For an analytic presentation on a per country basis of the EDLproject partners collections and the access protocols used for each one of them see final EDLproject WP1 D1.3 deliverable: "Implementing Full Membership of The European Library for 9 European national libraries" , March 2008.

<sup>8</sup> To give an idea about the hosting needs; at the end of November 2007 there were 1,8m records in the TEL central index while more than 40m records is expected to be hosted before the end of 2008.

## **5.2 Future work**

The ongoing promotion of OAI-PMH to the TEL partners has been endorsed by The European Library Management Board as a key objective for 2008 in The European Library Product and Services Plan 2008. In the follow up of this task a number of actions are foreseen:

- Address the issue of the huge storage needs that are the result of harvesting the distributed metadata in a central repository in order to index them in a successful and sustainable way. Contract with an external hosting company. This issue is expected to be addressed in March 2008 and will help make automatic and standardise the harvesting process.
- Hold a workshop for "OAI-PMH new implementers" for the 13 new member libraries of The European Library that are joining TEL via various projects (FUMAGABA, TELplus and the libraries of Turkey, Russia-Moscow and Russia- St. Petersburg) during the Welcome New Members! Meeting and Workshops in March 26-28 in the national library of The Hague.
- Major work is undertaken under TELplus WP2 which is led by the Instituto Superior Tecnico in Lisbon, Portugal. The aim of the workpackage is to provide the participating partner libraries with tools, guidelines and resources for the implementation of OAI-PMH on their catalogues/collections. The following libraries are partners in the work package and are expected to implement the protocol before the end of the project: National Library of Latvia, Slovak National Library, Hungarian National Library, National Library of Lithuania, Slovenian National and university library, National Library of Portugal, French National Library, National Library of Poland, National and University Library of Iceland. Additionally, at the kick-off meeting of the work package in the national library of the Netherlands in The Hague (March 26) all the members of The European Library technical working group are invited to participate. The workshop will focus on best practice guidelines and tools for OAI-PMH implementation as well as will help problem solve any major issues the partners are facing with the implementation of the protocol. Other tasks in the project involve: the implementation of a generic OAI-PMH infrastructure (Task leader: Nuno Freire, BNP). This work package will result in the implementation of an OAI-PMH solution made of two information systems packages: a local service that will make it possible to install local metadata repositories comprising a local OAI-PMH server and components necessary to manage and interface it with the local databases; an OAI-PMH central service that will manage all the synchronization in a network of OAI-PMH servers, and provide a central metadata repository service to be used by other functionalities of the TEL portal. A gateway will be developed by TEL (T2.6).
- A subgroup of The European Library Technical Working Group is foreseen to be established after the end of TELplus to take over the results and continue the work of the promotion of the adoption of the standard, the exchange of know-how and problem-solving as regards implementation.