Dig/talia

Rivista del digitale nei beni culturali

The European Library Beyond Access

Jill Cousins
The European Library Office

TEL, The European Library, is an EU project turned operational service. Its main aim is to access the collections and digitised items of the National Libraries of Europe and to make them searchable in one place. This aim is being achieved with 15 national libraries making their collections accessible as of the first quarter of 2006 and at least another 7 national libraries coming on stream during 2006. It is expected there will be at least 30 National Libraries contributing their collections by 2007.

So the vision of accessibility is achievable and in its execution issues of interoperability and standards are by practice enforced. However to be a truly useful resource for the user and to move onto become the European Digital Library some of the reasons why it has achieved access to so many collections, in a relatively short time, that would otherwise not be available, have to be reinvestigated. There are very low barriers to participation for the national libraries, but these impose restrictions on the usability of the site. This paper looks at what the libraries need to do now to meet user needs and what The European Library itself can do, using other technologies, to facilitate the next steps.

The History

The European Library (www.TheEuropeanLibrary.org) went live as a Beta site in March 2005. This was the culmination of nearly 4 years work, 3 years as a EU funded project led by the British Library and 9 months of creating an operational site, using a staff of 3. The result is a portal that accesses the collections of Europe's National Libraries, searching simultaneously across more than 150 collections. The founder libraries were Britain, Finland, Germany, Italy, Netherlands, Portugal, Slovenia, Switzerland. Since becoming operational the digitised and born digital collections of the national libraries of Austria, Croatia, Denmark, Estonia, France, Latvia and Serbia have been added. By the end of 2006, the libraries of Cyprus, the Czech Republic, Lithuania, Malta, Poland and the Slovak Republic will add their collections.

In the project there were several work packages, most of which have had lasting effects on the current structure, look and feel and *raison d'être* of The European Library.
The eight national libraries involved in the European Union funded project between them created a metadata registry, a metadata standard that is extended Dublin Core, a test bed "portal in the browser" architecture and unusually for a European project a business plan to create an operational entity.

Project Results

The reports for these work packages can be found in the archive of www.TheEuropeanLibrary.org. The work packages laid a solid foundation for the development of an operational service. The proposed Portal in the Browser solution created under the Interoperability work package was proven and has formed the basis of the development work since. The Metadata Development package gave birth to the TEL library application profile as well as a standard for metadata formation and submission across the partner libraries. The work package Business Plans and models was also instrumental in giving the operational service a kick-start, setting up a management structure and business model for the formation of the European Library Office. All of the relevant results are detailed further below in the sections covering the operational service.

The European Library service

In June 2004 the Office of The European Library was established, with a Head of Office, Technical Assistant and a Webmaster. A project brief and plan were written and a formal structure for the organisation ratified.
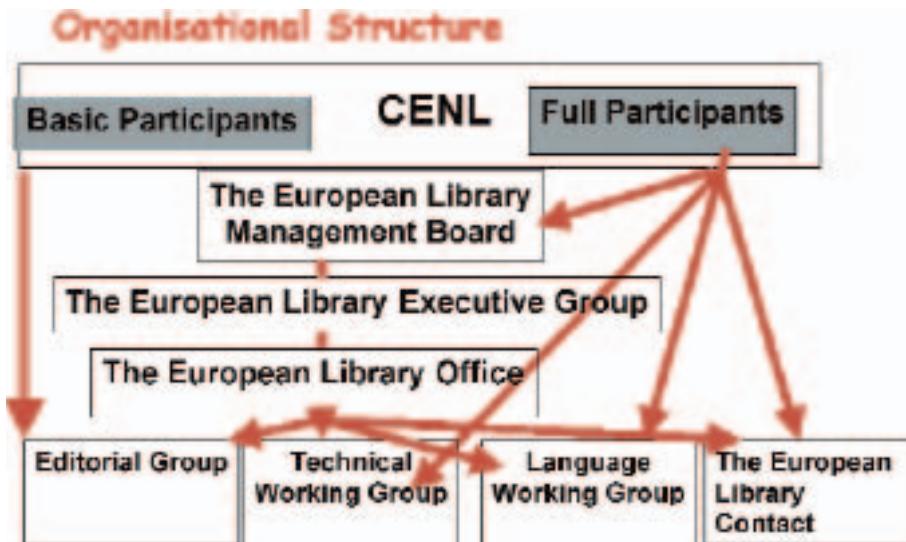

Figure 1. Organisational Structure of The European Library.

Management & Governance

The agreed structure for the management and governance of The European Library is indicated in the diagram above. Overall ownership of the service sits with the Conference of European National Librarians (CENL), all members therefore have a stake

in the service. A Management Board, consisting of each full member, to oversee policy on behalf of the CENL was set up at the end of January 2004. The Management Board nominated a smaller group, the Executive Group, to take decisions on their behalf and speed up the decision-making process. Each full participant of the service, contributes practical and technical expertise as well ensuring access to their collections. This structure continues to work even with expanding participation. It is extremely important to keep the strong sense of ownership that the libraries of CENL have for The European Library and the use of the network helps promote this feeling at the top and bottom of the national libraries. Everyone has been very co-operative and committed in creating and maintaining The European Library.

Business Model

The business model of The European Library is simple. It relies on payment from its full participants as well as all the members of CENL. Subscriptions range from 500 euros per year to 50,000 depending on the library's ability to pay. In addition, under TEL-ME-MOR (www.telmemor.net), an EC funded initiative, the metadata and infrastructures of the National Libraries of the New Member States have been prepared for entry into The European Library. This initiative has considerably speeded up the process of some national libraries being able to join as full participants. The European Library is available worldwide as a free resource for all.

The vision of The European Library is:

«Provision of equal access to promote world-wide understanding of the richness and diversity of European learning and culture.»

And its derived mission is:

«The European Library exists to open up the universe of knowledge, information and culture of all Europe's national libraries.»

The role of The European Library Office is to create a stable portal that allows the user to search within and across the deep Web collections of the National Libraries. It is aimed at the academic researcher and the informed citizen but is available to all. Covering subjects of historical interest from famous cities in Italy, shipping in the Baltic to gardens in Lombardy as well as the cultural, intellectual heritage of a country in its poets, artists and musicians, The European Library is a rich source of material for scholars worldwide.

Available Content

The content of The European Library reflects that of the National Libraries, largely arts and humanities, and is representative of the cultural heritage of Europe. As well as giving everyone in the world access to some of its richest material, The European Library also highlights how much more of our heritage needs to be digitised and made accessible.

Figure 2. Access to ICCU Edit 16 Catalogue on The European Library, La declaratione de le città famose de l'Italia.
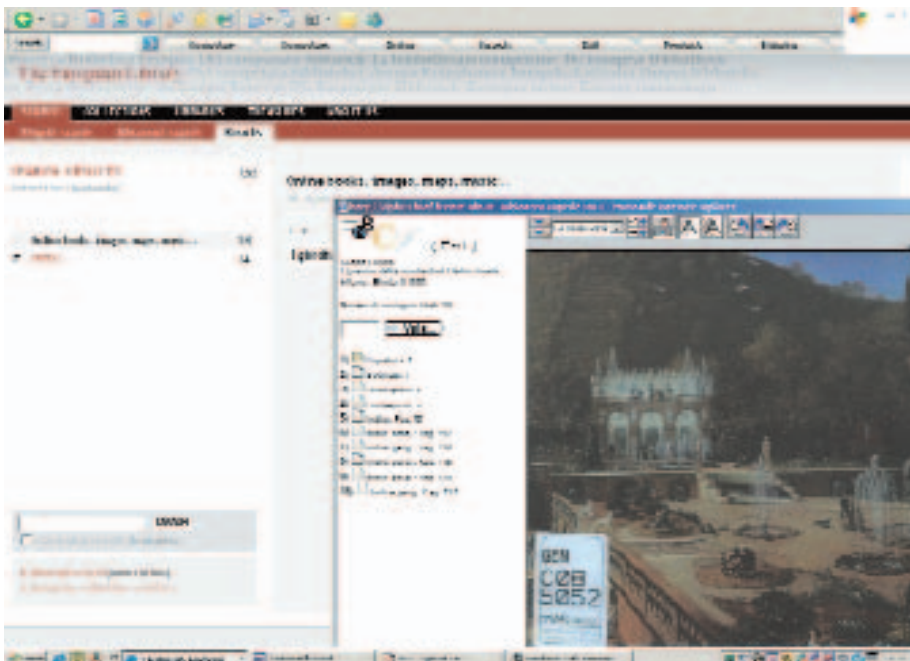

Figure 3. Biblioteca Nazionale Centrale di Firenze, I giardini della Lombardia.

Some of the stronger areas are maps and atlases, music scores, bibles and historical documentation. There remains a scarcity of digitised content. The predominant digital material is the catalogues and national bibliographies, most of which take the user only to the digitised reference not to the object which is still in its physical form in the library. Many of the National Libraries are involved in national digitisation schemes aimed at the digitisation of newspapers or books in the main. Some are in public/private partnerships such as the British Library with Microsoft and some have announced large scale public funding to digitise cultural heritage. The European Library, at the moment, accesses about 80% of the available digitised material from its full partner National Libraries. This is probably less than 5% of the total holdings of the libraries. User expectations are high and they are surprised when a catalogue record does not turn into the display of the digitised item itself. The creation of a virtuous circle where the knowledge of what users are searching for and not finding, might inform digitisation programmes and lead to greater use of the site. So it is important from The European Library's viewpoint that more content is digitised more quickly so the service can continue to meet user expectations. There is therefore a huge need to digitise more of the content held in the National Libraries. Currently not even the tip of the iceberg is in digital form.

CENL has launched a Content Working Group across its partner libraries, aiming to create more coordinated European collections in items such as newspapers, maps, music, bibles and in other areas beneficial to research. This working group will also help to ensure standards and to communicate about achieving funding for digitisation.

Figure 4. Virtuous Circle – showing how more digitised content will benefit users.

So far everything available on The European Library is free of charge, because the data are pre copyright and publicly available. It is possible that some of the libraries may look to charge for downloading their treasures and manuscripts and if access to the academic journal collections is included at some point then the normal payment rules and processes within each national library will apply.

Some users are likely to become frustrated using a resource such as The European Library as most of its content currently takes the user as far as the library doors, but to use it they have to visit the library itself. Most of the collections are references to items rather than digitised content itself. However others will see the enormous benefits this portal brings. For the first time researchers can search across the libraries of Europe discovering texts of Verlaine set to music by Diepenbrock, maps of Singapore in the Koninklijke Bibiotheek or maybe that a famous German philosopher wrote letters that are kept in a Nordic library. The research potential that is opened up by the portal is huge. The portal will be the most comprehensive supra catalogue of the content of these libraries that can be searched jointly.

Technological Background

A cheap and elegant solution using an innovative architecture and the concept of a «portal in the browser» has been developped by Theo van Veen of the Koninklijke Bibliotheek, in The Netherlands, and Bill Oldroyd of the British Library (Figure 5).
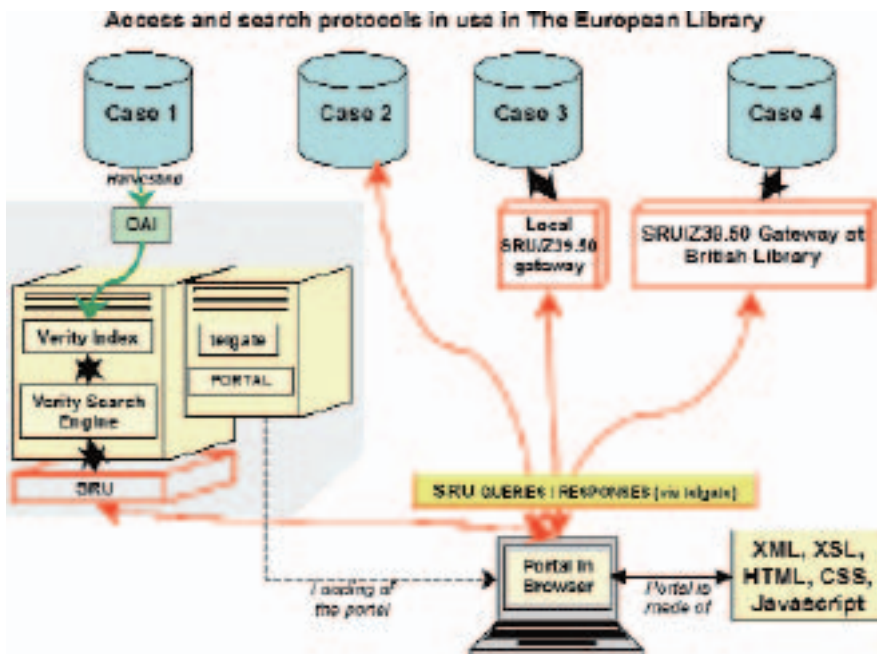

Figure 5. The European Library Portal Architecture.

There is a deliberately very low barrier to entry for the national libraries. They are required to meet 3 criteria. An access protocol such as z39:50 or OAI (Open Archives Initiative), Metadata that complies to Dublin Core, and NISO standard Collection Descriptions.

The European Library enables cross collection and cross-library searching. The portal uses the search protocols already in place in the respective libraries. Therefore libraries are not even asked to alter their current access mechanisms. So a library using z39:50 to search its catalogue will be accessed by The European Library using z39:50 which is translated into the common protocol, Search and Retrieval via URL's (SRU) via the central gateway or local versions of it.

The European Library requires that all data submitted to it is standardised on the TEL Application Profile. This is largely Dublin Core with some extensions to allow for the addition of services such as Open URL linking.

The European Library portal architecture

A complete technical explanation for the architecture is given by van Veen and Oldroyd in their paper published in «D-Lib» at the beginning of 2004 (Oldroyd, van Veen, D-Lib Feb, 2004) and is précised in «Serials Librarian» (Cousins, May 2005).

Technically The European Library is clever. Holding very little data it gives access to panoply of resource. Its architecture allows for the use of legacy protocols to access digital material across all the libraries and does not insist on the use of OAI harvesting over closed protocols such as z39:50 (the most used search protocol in library systems). So a user may search simultaneously for Dante the collections of France, Finland, the Netherlands and Switzerland. They may make connections that were not possible previously and may discover new objects relevant to their research. The portal allows exposure of information to the wider Web that otherwise lay deep within the libraries, possibly accessible via their own portals but rarely via the big search engines. However it is precisely this flexibility that causes usability problems in both search and display.

In a portal a user enters a search term and the query is launched across a number of targets simultaneously. The number of targets and the speed of the target server determine the speed of return of the results. Therefore the greater the number of targets the slower the response is likely to be. Ideally the user first picks their collections, by search or browse, and then enters their search query against these chosen collections. The European Library felt that few users would be prepared to do this without any knowledge of what the site might give them. The advantage the search engines and other centralised databases have in this area is that they are only searching one database, or one target. They are not trying to search tens of remote databases at the same time. This allows faster retrieval and better sorting of results, but requires that the data is harvested into a central database. Much of the data held in the national libraries cannot yet be harvested.

In The European Library, to be able to present the user with "some" results against a two-word search some choices were made. Simple search currently defaults to all the fully digitised items that have been harvested into the central index and the on line catalogues of the National Libraries. From the results produced, it was hoped that the user could be made to realise that there are more collections to search and/or that more intelligent searching can be executed in the background. There is a danger however that the user does not go beyond these results thinking that The European Library is purely a super catalogue for the national libraries of Europe.

So the upside of low barriers to entry (i.e.: accepting all protocols for search) is that the national libraries now provide access to their digital information from a central point, anytime, anywhere, and this is a vast improvement on the complete lack of Web access that existed previously. The downside of not imposing more restrictions, particularly on how the data is retrieved, is for the user. With asynchronous multi-target retrieval it is difficult to sort or rank the data as it comes in at different rates from different targets. It is also problematic for any portal to search more than 15-20 remote targets at once. Therefore the 150 collections (which represent the targets) held on the site cannot all be searched at the same time for results to be retrieved in a reasonable timescale. The user needs to choose their collections before launching their search. As mentioned this goes against the search patterns of most users in this world of large search engines.

The European Library has provided access to information that would otherwise be very difficult to collate in one place, but what should it do next to improve its usability?

The problems stem from the form of access to the digitised data of the libraries, the fact that the data is largely metadata and not full text and from the lack of fully digitised material (discussed above).

The data is largely metadata, and the user wants to see the full text and much of this metadata is held in proprietary databases that do not lend themselves to sharing or harvesting. This means that The European Library cannot provide the access to the full text that the user is looking for (because it does not exist), nor can it hold the records in a way that allows for the search and results to conform to the "norm" of one ranked list.

The responsibility for change needs to come from both sides. Libraries must start to put more resources into the creation of full text, based on XML and make use of the meta data extraction tools, and the indexing capabilities of search software. This does not mean an abandonment of manual creation of metadata, but the lessening of its perceived importance. Metadata very definitely makes information more retrievable, and for non text images and objects provides the best way of finding such material. It adds structure to data and it is this structure that can be used by The European Library to improve on what the big search engines are delivering.

This is obviously not going to happen tomorrow, so meanwhile The European Library will continue to work on making it easier for users to use the site. This is being looked at both in the presentation layer (see below under the next version) and in the architecture, also below.

Figure 6. Version 1.2 of The European Library.

The next version

According to the first user survey, held May/June 2005, there was a fairly even split in users between those who were happy to dig deeper and understood the concept of collections and the ad hoc users who assumed everything could be accessed from simple search and did not bother to look further. However analysis of the log files showed that more than half of users do not go beyond the "Simple, Default" search and are probably therefore disappointed. In an attempt to readdress this problem a new version of the site that pushes the idea of collections on the home page is being user tested and will be launched in April 2006 (see Figure 6). A change from the current version in Figure 7. The European Library is therefore looking at ways to get the user to decide what they are going to search in prior to executing the search. But we are also trying to put more intelligence into the system. For instance some background searching on behalf of the user at the same time as the default collections are being searched, so that 2 sets of results are shown, those of the default collections and those of the other collections it is worth searching in. Presentation of the collections has also changed in the new version with more emphasis being placed on how to find the collections, i.e. by subject, by collection description or by browsing through them all as can be seen in Figure 8. The new version tries to use the knowledge contained in the portal collection descrip-

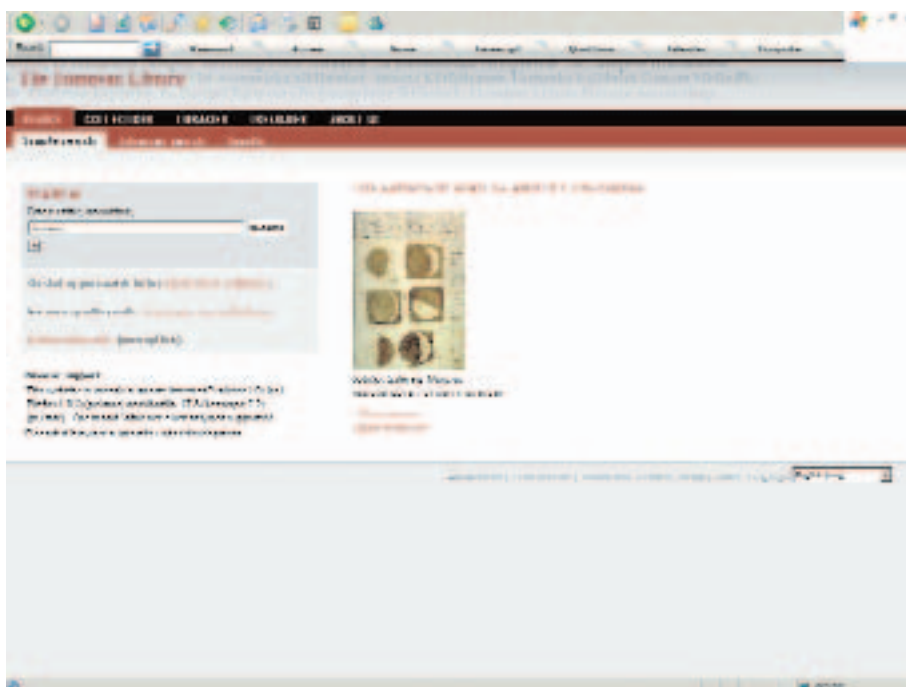Figure 7. Version 1.1 of The European Library.

Figure 8. Version 1.2 trying to deliver more information on collections.

tions to show the user what the possibilities are. Numbers of collections searchable by subject will be possible in version 1.2.

The "portal problem" also raises its head in the presentation of search results. Each target returns a separate set of results, so a user must pick a collection to review the results – this does seem to be understood, but requires more patience than the search engine *habitué* is geared for. As mentioned the application of relevance ranking is impossible without adding another stage into the process, where all the results are pulled into a temporary database and an algorithm runs against them. Not only does this adds time to an already time-pressed delivery, but it would be difficult to know what the algorithm should calculate. Most of the data presented by the libraries is metadata, therefore the number of times a word occurs in an "article" cannot be applied. Date currency is less relevant on the kind of data and information contained in the National Libraries. The humanities' scholar is not after the latest articles on pain relief, but quite often original or precursory material such as maps showing navigation prior to those of the Venetians and Portuguese. The item of the 16th century could well therefore be of more interest than that of the 18th.

As well as the known problems of portals such as the integration of results and searching across large numbers of targets within Internet timescales, The European Library deals with a couple of other issues.

Improving the architecture

As well as trying to optimise the front end so users can find the information and encouraging libraries to move towards XML full text and OAI compatible data, investigations are proceeding into how technology might help the process. The two original architects of The European Library are now looking to extend its usefulness in ways that will benefit the user.

Bill Oldroyd of the British Library has been looking at the use of persistent identifiers and the creation of summary records. Based on the concepts of FRBR, the aim is to use the processing power of the big search engines with a field search to bring together associated items in a meaningful order. An example would be placing all works of a creator in an ordered list by date. This would enable The European Library to structure the holdings of partners' collections into summary records. The A user searching in search engine for a particular title will find a single entry from The European Library which leads to a full display of all the holdings of the that work within the European Libraries.

In addition, a simple, multilingual search language that could be used by the search engines to provide field search, built with the expertise of libraries and organisations such as NISO, could enable the federated search function of The European Library to become a standard piece of browser functionality (Oldroyd, 2006).

From Theo van Veen of the Koninklijke Bibliotheek, the addition of services, to enable the user to go beyond the library record and find out where an item can be seen, whether digitally or phyically, is the basis of his current research and soon to be published article in «Ariadne» (van Veen, 2006).

Next Steps

The portal in the browser solution has allowed CENL and its The European Library to test very cheaply (the cost of a couple of servers and a search engine license) whether such a portal will really work. But having achieved access we need to move into satisfying the user requirements more. To this end we are encouraging more of the partner libraries to implement OAI and to digitise the full text where possible rather than purely creating metadata directing the user to the physical article.

We need more content and this will come from new full members and newly digitised collections. We need non-text content to be described more thoroughly and in a way that will work for the big search engines. And we need to innovate to make what we have more usable and add value in the digital world.

TEL, The European Library, è un progetto dell'Unione Europea divenuto un servizio operativo. Obiettivo principale di TEL è rendere le collezioni di oggetti digitalizzati delle biblioteche nazionali d'Europa fruibili ed esplorabili da un unico punto di accesso; obiettivo verso cui si procede con successo, con 15 biblioteche nazionali che hanno messo a disposizione le proprie collezioni nel corso del primo trimestre del 2006 e un minimo di altre 7 biblioteche nazionali pronte a fare altrettanto nel corso dell'anno. Entro la fine del 2007, si prevede di poter contare sulle collezioni di trenta biblioteche nazionali.
Il traguardo dell'accessibilità si dimostra così a tutti gli effetti raggiungibile, attraverso un percorso in cui le sfide dell'interoperabilità e degli standard trovano risoluzione nella pratica. Ma perché questa risorsa si dimostri effettivamente utile agli utenti e porti alla nascita della Biblioteca digitale europea, occorre tornare a ragionare sui fattori che hanno permesso a TEL di accedere, in un arco relativamente limitato di tempo, a un numero così elevato di collezioni, altrimenti non disponibili. La partecipazione delle biblioteche nazionali al progetto è soggetta a pochi vincoli, ma questi comportano una serie di restrizioni quanto all'utilizzo del sito. In quest'articolo si tenta di capire quali misure le biblioteche dovrebbero intraprendere oggi per rispondere alle esigenze degli utenti, e cosa andrebbe fatto nell'ambito di TEL stesso per facilitare attraverso nuove tecnologie le prossime fasi di questo percorso.

TEL, The European Library, constitue un projet de l'Union Européenne qui est aujourd'hui un système opérationnel. L'objectif principal du TEL est de permettre d'utiliser et d'explorer les collections des objets numérisés des bibliothèques nationales européennes à partir d'un seul point d'accès. Le projet avance avec succès, grâce aussi aux 15 bibliothèques nationales qui ont mis à disposition leurs collections pendant le premier trimestre 2006 et aux 7 autres bibliothèques nationales qui ont été prêtes a en faire autan au cours de la dernière année. Il a été prévu que d'ici 2007 l'on pourra compter sur la présence des collections de trente bibliothèques nationales.

L'accessibilité souhaitée se révèle donc parfaitement réalisable car il est possible de résoudre par un travail pratique les défis posés par l'interopérabilité et les standards. Cependant, pour que cette ressource soit réellement utile aux usagers et puisse donner naissance à la Bibliothèque numérique européenne, il faut mieux réfléchir sur les facteurs qui ont permis à TEL d'accéder assez rapidement à un nombre si élevé de collections qui ne sauraient être disponibles autrement. La participation au projet des bibliothèques nationales n'est sujette qu'à peu d'obligations. Ces dernières comportent néanmoins des restrictions concernant l'utilisation du site. Cet article tente de comprendre quelles mesures devraient adopter les bibliothèques aujourd'hui pour satisfaire les exigences des utilisateurs et le travail nécessaire au sein du TEL même pour faciliter les étapes de ce parcours par de nouvelles technologies.

REFERENCES

Jill Cousins, *Will the 'Portal in the Brower' Solution Work?*, «The Serials Librarian», 49 (2005), n. 3.

Bill Oldroyd – Teo van Veen, *Search and Retrieval in The European Library: a New Approach*, «D-Lib Magazine», 10 (2004), n. 2.

Bill Oldroyd, *TEL: the Big Picture*, work in progress (2006).

Teo van Veen, *Serving Services for Web 2.0*, «Ariadne», 2006, n. 47.