Open Knowledge on e-Infrastructures: the BELIEF Project Digital Library

Donatella CASTELL\textsuperscript{1}, Simon J E TAYLOR\textsuperscript{2}, Franco ZOPPI\textsuperscript{1}
\textsuperscript{1}Consiglio Nazionale delle Ricerche
Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”
Via Moruzzi 1, 56124 Pisa, Italy
Tel: +39 050 3153470, Fax: +39 050 3152810,
Email: donatella.castelli@isti.cnr.it, franco.zoppi@isti.cnr.it
\textsuperscript{2}School of Information Systems, Computing and Mathematics,
Brunel University
Uxbridge, Middx UB8 3PH UK
Tel: +44 1895 265994, Fax: +44 1895 251686, Email: Simon.Taylor@brunel.ac.uk

Abstract: The BELIEF Project is a Coordination Action funded by the European Commission in the context of the FP6 and FP7 Programmes. It aims to create a platform where e-Infrastructures providers and users can collaborate and exchange knowledge, ensuring that e-Infrastructures are developed and effectively used worldwide, filling the gap separating the e-Infrastructures providers from the users, and thus contribute to the emergence of a competitive knowledge-based economy. To create this synergy among multi-disciplinary communities, BELIEF created a one-stop-shop providing a Portal and a Digital Library with a huge number of e-Infrastructures open access publications. The Digital Library offers uniform access to multimedia documentation providing continuously updated information on e-Infrastructures-related projects, initiatives and events. The contents are harvested from different sources, such as projects web sites, repositories and databases. The DL - implemented on top of the OpenDLib Digital Library Management System - provides services to support the submission, description, searching, browsing, retrieval, access, preservation and visualization of multimedia documents. Although designed to meet the needs of the e-Infrastructures community, the technology adopted by BELIEF can be easily adapted to meet the information and collaborative needs of other scientific communities.

Keywords: Digital Library (DL), Digital Library Management System (DLMS), e-Infrastructures, Open Access, Sustainability.

1. Introduction

The BELIEF (Bringing Europe’s eLectronic Infrastructures to Expanding Frontiers) Project aims to create an effective open workspace where e-Infrastructures providers and users can collaborate and exchange knowledge, ensuring the development and adoption of e-Infrastructures on a worldwide scale. The BELIEF DL plays a key part in the project, bringing a range of benefits to e-Infrastructures stakeholders across the globe by facilitating the exchange of knowledge and experiences through a single and easily accessible tool. The BELIEF Project arose from the awareness that a gap existed between Research Infrastructure providers and users. In order to bridge this gap, a complete and common source of information on e-Infrastructures was needed, both for users demanding provision and resources and for providers intending to extend their user base and develop their systems. The BELIEF DL responds to this demand by providing users with documentation.
matching their search criteria accurately and according to their interests and professional profile. This paper focuses on the implementation of this key component.

Section “2 - Objectives” introduces the objectives of the project which specifically apply to the DL, giving an outline of the expected outcomes from both a qualitative and quantitative point of view.

Section “3 - Methodology” introduces the main characteristics of the methodology applied to the design of the DL.

Section “4 - Technology Description” points out the rationale behind the adoption of the OpenDLib DLMS and introduces its main characteristics.

Section “5 - Developments” focuses on the main developments carried out on the basic OpenDLib system to fully cope with the e-Infrastructures community’s requirements. This is necessarily an overview as work in this area is substantial. For further details, please refer to [4].

Section “6 - Results” focuses on the analysis of the impact of the DL on its target audience both from a qualitative and quantitative point of view, and summarizes the most relevant usage data as per the statistics gathered during the whole projects lifetime, from 2006 onward.

Section “7 - Business Benefits” gives a brief introduction to the main objective achieved by the project on this side - i.e. sustainability - and highlights the role of the OpenDLib DLMS in that.

The last section summarises the main conclusions, results achieved and future work.

2. Objectives

The DL was designed to serve the needs of e-Infrastructures research and industrial users that want to keep up to date with existing projects and the latest developments in e-Infrastructures. The DL had to offer its user community advanced services to uniformly access multimedia documents such as technical reports, presentations, videos, manuals, online tutorials, etc. These documents contain the very latest details on e-Infrastructures related projects, initiatives and events.

The material maintained in the DL had to be regularly harvested from different sources, such as web sites, repositories and databases of e-Infrastructures Projects, Initiatives and Organisations. The DL organises the harvested information according to the information needs of the user communities rather than according to its physical format, structuring and distribution on diverse sources. This means that it is capable of providing users with multiple virtual views of its content. To this end, an extremely accurate collection and analysis of the requirements of potential users was made before the DL was created.

The DL had to provide services to support the submission, description, searching, browsing, retrieval, access, preservation and visualization of multimedia documents. Users can define the information space which they want to search/browse upon in terms of collections (i.e. sets of documents) selected from those managed by the DL. Collections can be created interactively, based on the archives the documents are to be selected from. Different search/browse options are offered: Google-like or fielded (with fields selected from a variety of known metadata formats). Users can search/browse any information associated with digital documents and their parts.

Two others main objectives were requested to be addressed by the DL: the implementation of open access policies, which had to be achieved via the compliance to the OAI-PMH protocol, and the clear identification of a sustainability model. This latter implied the selection of a proper platform on which the DL had to be implemented in order to minimize both training and installation and operation costs for both the consortium and the community.
Finally, from the quantitative point of view, the objective was to serve a community of nearly one hundred organizations, with thousands of objects selected from the source archives and setting up an accurate statistical tool for the monitoring of the operations performed on the knowledge base, to serve for the evaluation of the DL impact on the e-Infrastructures community.

3. Methodology – Design of the DL

Whilst the technological dimension of the project benefited from a sound and deep experience ([14], [25], [26], [27]), particular relevance was given to the target audience of the DL - the whole EU e-Infrastructures community – whose magnitude and heterogeneity was one of the recognized risk element of the project.

To such end, a design phase based on a deep analysis of the audience needs was carried out with a selected set of e-Infrastructures Entities (Projects, Organisations and Initiatives). The adopted methodology was basically structured in the following activities:

• Drawing up a Questionnaire to collect requirements on functional needs, documents to be managed and their related metadata (descriptive data).
• Identifying proper interfaces within Entities, contacting them, obtaining and discussing requirements.
• Matching similar requirements from diverse Entities.
• Verifying the quality of the metadata and making the semantics and use of metadata conform.
• Verifying the quality of documents, analysing the types of documents used by diverse Entities with different semantics and making them conform.
• Drawing up a Memorandum of Understanding to be signed with each of the Entities collaborating in subsequent releases to ensure clear, comprehensive and effective interaction and collaboration.

4. Technology Description – The OpenDLib DLMS

In the arena of the several existing DLMS, different implementation solution can be adopted according to the flexibility and cost constraints of a specific application scenario. In the following we point out the rationale behind the adoption of the OpenDLib DLMS and give a brief introduction to its main characteristics.

The need of a Digital Library Infrastructure

Traditionally, organizations, such as public research institutions, need to set up Digital Library Systems (DLSs) to support the activities of their research communities. Such systems provide researchers with the technology and the administration needed to collect, manage and expose community-relevant data through customized portals. The design and development of a DLS has a considerable initial technological cost to be then summed up to the cost of maintaining and administering the system during its life-time. Maintenance costs can be high, due to the technical support of software systems and to their potential changes, in turn dependent on the rather dynamic nature of modern research communities, characterized by end-users with evolving functional requirements.

Digital Libraries (DLs) are DLSs sustained at the level of the single organization, typically targeting the local community of a research institution or even a project.

Several well-known DLS technologies (in the following we will rather speak of Digital Library Management Systems – DLMSs) with a variety of functionalities are available on the market: DSpace [18], ePrints [19], Fedora [20], etc. [23]. Some of them adopt a general purpose approach and require further high-costs for ad-hoc software customization, while others tend to provide a static set of features and only entail initial installation efforts.
However, costs are generally equally high in terms of system administration staff and frequently tend to be underestimated.

The diffusion of DLs naturally called for the aggregation of their content, to serve larger research communities, geographically or subject driven. Aggregation Systems are DLMS whose functionality is that of gathering data from heterogeneous data sources – e.g. Repository Systems of different institutions – form a uniform Information Space of data, and then offer researchers customized services on top of such space; e.g. search, inference of references between publications, citation calculation, etc.

Examples of these DLMS are well-known and include: BASE [15], DAREnet [16], DART Europe [17], OAIster [22], Zentity [24], etc. Typically, the relative technology is financed by one or more institutions, built ad-hoc by teams of developers, and maintained by a team of administrators and developers as far as gathering the data and machine installations are concerned. Of course, any further change to the kind of functionality offered to the end-users will require a new design and development phase, whose cost has to be carefully evaluated in its trade-off with benefits. Those systems have high maintenance cost, in hardware, software development, and administration staff. Most of the time, end-users tend not to be satisfied in their requests for new functionality, because funds are hardly enough to maintain the current system alive. Thus, only wealthy organizations can afford advanced Aggregation Systems, while others renounce in principle to their construction; e.g. the majority of European Countries do not have a national Aggregation System through which all national research products can be reached.

In the last three Framework Programme calls, the European Union initiated the so called knowledge infrastructure vision, inspired by the same goal of unifying data resources of all kinds available in Europe. The idea was that of devising e-infrastructures, that are environments through which existing DLs’ functional and data resources can be integrated and combined to then build similar, sustainable DLs. The approach is different from that of Aggregation Systems, which focused on ad-hoc solutions targeting special organization/community scenarios; e-infrastructures should enable the construction of sustainable DLs for any joining organization, by delegating technical costs to the infrastructure maintenance and minimizing the administration costs. In principle, the inspiring model is that of known real-life infrastructures, such as electric power or water systems, centrally maintained by a government through (hopefully) low citizen taxation.

The EU funded a number of projects experimenting in that direction, all of them devising general-purpose software and data infrastructures: Bricks, ScholNet [30],

---

**Figure 1 – DL Implementation Cost vs. DLMS Flexibility**

[Diagram showing cost and flexibility relationship for different DLMS, including Fedora, OpenDLib, Zentity, Eprints, DSpace, Greenstone, with Fedora having the highest cost and flexibility, and Eprints having the lowest cost and flexibility.]
DILIGENT [26], D4Science [27], DRIVER [14], CLARIN [28], and EGEE [29] are known examples.

Of particular interest to Digital Libraries is the OpenDLib DLMS (deriving from the experience of the ScholNet and DILIGENT projects), whose main features are:

- The integration of data sources into a uniform global Information Space.
- The integration of existing functionalities as infrastructure components.
- The capability of BUILDING sustainable DLs infrastructure with a set of interacting components operating over the resulting Information Space.

Figure 1 summarizes the relationship between OpenDLib and some of the most known DLMS with respect to the key parameters of “flexibility” and “implementation cost”.

The OpenDLib System

Digital libraries are instruments for supporting communication and collaboration among worldwide distributed user communities. OpenDLib is a DLMS that makes it possible to satisfy this demand by supporting a cost-effective digital library creation and operational model.

OpenDLib consists of a federation of Services (see the Figure 2) that can be customised to meet the requirements of a target user community (an Application in the following figure). This federation can be expanded at any time by adding other community specific services. The entire set of services can be managed and hosted either by a single or by a multitude of organisations collaborating on the maintenance of the shared digital library, each according to their own computational and human resources. An orthogonal system facility (the Control and Management Tools in the following figure) enables different user groups to define their own virtual view of the shared digital library (the Repositories in the following figure), tailored to the specific needs and policies of the group.

The basic release of OpenDLib provides services to support the submission, description, indexing, search, browsing, retrieval, access, preservation and visualisation of documents. In addition, a number of administration functions are also given to support the preservation of documents, document reviewing process, introduction of new collections, and handling of users and user group profiles.
From an architectural point of view, a generic service of OpenDLib can be distributed over different servers, replicated, or if necessary centralized. The OpenDLib Conceptual Model defines an entity for each service type. Although the presence of multiple instances of a service increases fault tolerance, reduces the overload of each instance, and makes it possible to dynamically reorganize the environment when a server hosting a service instance is not reachable, the replication and distribution of the services is not mandatory and therefore each of the services outlined in Figure 2 can be instantiated as a single instance. This means that the level of distribution and replication, and the physical location of the service instances may be freely chosen to better satisfy the needs of the specific digital library context.

Supporting Portals on the OpenDLib Toolkit

Organizations can build DLs concentrating on user functionalities and exploiting the harvesting and aggregation efforts performed by stronger or richer organizations. The User Interface Services of OpenDLib can be configured to access a subpart, i.e. a virtual collection, of one Information Space and to activate one or more functionality offered by the Service. Organizations are supplied with two typologies of User Interface Service, advanced and lite, the first of those is exemplified in the following BELIEF real case scenario.

5. Developments – The BELIEF Real Case

The BELIEF Consortium was one of the several organizations to build an Aggregation System - the BELIEF DL - on OpenDLib. The resulting Information Space contains records in BELIEF Metadata Format (BMF) describing Open Access publications available from a variety of Repositories in European Countries. BMF records are produced by a number of Aggregation Service instances, virtually running at different sites and managed by OpenDLib Administrators delegated by the BELIEF Consortium and by Content Providers’ Correspondents to be responsible of the aggregation process: harvesting records from an assigned sets of Repositories and define the mappings to convert them into BMF and therefore populate the BELIEF DL.

The BELIEF DL currently counts nearly 14,000 documents from all European Countries harvested from more than 90 selected repositories. These are all accessible from User Interface Services providing advanced functionalities [3]. Each document is described by a BMF record in a uniform way (i.e. using the same vocabularies for the same fields) in terms of its provenance (name of the original repository, etc.) and bibliographical description.

General Characteristics of the Implemented Solution

The BELIEF DL is capable of offering the following services to its users:

- Creation, submission, search, browse, access, and preservation of multimedia documents.
- Definition of their personal Information Space which they want to search/browse upon in terms of collections selected from those managed by the DL. Collections can be interactively created defining:
  - the condition that is satisfied by the members of the collection;
  - which archives the documents are to be selected from.
- Different search/browse options: Google-like or fielded (with metadata elements selected from a variety of known metadata formats). Users can search/browse any information associated with digital objects and their parts. As a result of their
search/browse operations, users obtain a set of result pages with the list of digital objects that satisfy their request. By clicking on an object, users can access any of its multiple manifestations. In particular, they can select the one that is compatible with the software installed on the computer they are working with.

- Full compliant Open Access via the OAI-PMH protocol.

In the implementation of the DL particular effort has also been devoted to the following issues:

- Harmonization of concepts and practices (e.g. use of metadata and terms and of different types of document) to be benefited by the whole Community in facilitating knowledge communication and document exchange.
- Definition and/or homogenisation of vocabularies used both for metadata and document content description, categorization and search.
- Implementation of “document models” and of a web-based interface allowing users to easily submit metadata and documents to the DL using those models.
- Implementation of a set of APIs to be easily used by programmers to interface basic functions of the DL.

Albeit referring to [4] for a complete documentation of the technical characteristics of the OpenDLib Software System, this section briefly summarises the most relevant aspects of the BELIEF implementation, namely:

- User Interface
- Definition of the metadata structure implemented by the DL.
- Definition and implementation of protocols and tools for the submission of metadata and documents to the DL.
- OAI-PMH compliance.

The User Interface and its usage

The DL’s User Interface (UI) - built on top of the OpenDLib User Interface Advanced Service - has been designed to reflect the most recent advances in UI usability:

- The look and feel let users have an extremely comfortable access to functionality and content.
- The overall navigation structure has been designed to minimise the number of clicks the users need to access any content.
- All most common functions can be easily accessed via one-click commands.
- Most of the relevant information related to a document are shown in the same window.

The overall work area of the DL (the DL Desktop) is clearly organised into three sections [2]:

- On the left side is the "Community" section. According to each specific account permissions – after having executed a Login – a number of functions are accessible here: "News", "Personal Profile", "Users & Groups Management", "Documents Management", and "Information Space Management".
- On the right side is the "Content Access" section, where the content access functions (Browse, Search) are accessible. A "News" section is also shown, presenting relevant information about the DL from different points of view (events, user services, technical, etc.).
- In the middle is the "Information Space" section of the Desktop where available Collections are listed and the content resulting from Browse and Search operations is shown.
The following two figures show respectively the overall look of the UI and the way for accessing the actual content of a document – once it has been retrieved via a Browse/Search operation. Please refer to [3] for details, or access directly [2].

**Figure 3 – The DL Desktop organization**

**Figure 4 – Accessing content**

**DL Metadata**

With OpenDLib, resources can be catalogued with multiple metadata formats. The BELIEF DL uses Dublin Core Qualified (DCQ) encoding for the purpose of interoperability since DCQ enables the enhanced sharing of information between Information Sources adopting different coding with no loss of semantics [8].
A total of 17 DCQ metadata are currently supported by the BELIEF DL. (The metadata in bold have to be considered mandatory for an effective classification in the DL): **Title**, **Creator**, **Subject**, **Description**, **Publisher**, **Contributor**, **Date**, **Type**, **Status**, **Format**, **Identifier**, **Source**, **Language**, **Relation**, **Coverage**, **Rights**, and **Provenance**.

The detailed description of the semantics and usage of such metadata elements is given in [5]. Implemented or suggested controlled vocabularies are introduced for Type, Status, Format, and Language, whilst the definition of controlled vocabularies has been undertaken for Creator and Subject and will be carried out along with the implementation of a “Metadata Curation” and an “Authority File Control” SERVICES to support librarians and administrators in such arduous task.

Submission of Metadata and Documents

Different methods for submitting documents and metadata to the DL are supported:

- **On-line submission via Document Models (i.e. Web forms).**
  This can be done either using the native DL facility (accessible for authorized users through the “Community” section of the DL Desktop) or via integration with external portals (e.g. BELIEF itself and D4Science [6]).
  In the former, the document models and the underlying metadata structure are designed and implemented by the DL’s Librarian. Different pre-defined models are supplied to cope with the “standard user” needs. In addition, a “Free Model” is supplied, which can be modified by skilled users according to their specific needs.
  In the latter, the models and the metadata are designed in cooperation with the administrators of the external portals, to exactly fit with their requirements.

- **Harvesting from existing Information Sources, implemented by specific modules and interfaces.**
  To this end, a number of different harvester has been implemented for repositories supporting any kind of programmatic access. Currently, the following standard protocols and coding formats are supported by the harvesting tools of the DL:
  - Metadata encoding formats: DC, DCQ (recommended), MARC, UNIMARC, MARC21, MARCXML.
  - Metadata harvesting protocols: OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting (recommended) or any API call returning an XML file containing metadata encoded in one of the above mentioned formats.
  - File formats: XML (recommended), RSS.
  In some relevant case, even HTML parsers have been implemented to access repositories not supporting any of the above.

- **Batch submission via an XML schema based file.**
  Finally, bulk load from repositories not supporting any on-line access facility has been implemented via batch processing of XML files. These are based on an XML schema supplied by the DL’s Librarian and fine-tuned with the administrators of those repositories.

**OAI-PMH Compliance**


These tools will also promote widespread adoption of an Open Access Policy that will lead towards global and seamless dissemination of publicly-funded research results.

To apply all of these, the integration of the DL within the DRIVER Infrastructure [14] was undertaken, implementing the full compliance with the DRIVER Guidelines 2.0 released by the DRIVER Consortium. DRIVER is the well known European data infrastructure connecting hundreds of digital repositories of institutions and research organisations..

6. Results – The BELIEF DL Impact

Outline

BELIEF successfully organised a series of events which broke new ground in the arena of Grid-empowered infrastructures and e-Infrastructures. These events brought together Grid & e-Infrastructures experts and technology developers, IT innovators from both enterprise and research, decision-makers, and scientific policy-makers. Participants came together to share ideas and knowledge, discuss how technological challenges could be tackled, strengthen alliances between business and research, and help unleash the potential of e-Infrastructures.

All of these international events have helped to setup a synergetic network of relationships and to valuably increase the content of the Digital Library with material submitted by all Community’s Entities. Thanks to the outcomes of these events and to the established synergies and relationships, a significant growth of the Community has been achieved, largely exceeding the original planning of the project. This is summarized by the Figure 4 and Table 1.

![Projects, Initiatives & Organisations](image)

**Figure 4 – The BELIEF Community growth**

<table>
<thead>
<tr>
<th></th>
<th>Planned</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y1</td>
<td>Y2</td>
</tr>
<tr>
<td>Information Sources</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Documents</td>
<td>2,000</td>
<td>6,000</td>
</tr>
<tr>
<td>Collections</td>
<td>15</td>
<td>40</td>
</tr>
</tbody>
</table>

**Table 1 – DL’s actual vs. planned growth (sources, documents and collections)**

The rich and differentiated content offered by the DL is reported by the following chart, with a clear prevalence of material produced by Conference and Technical Meetings, then Presentations, Articles, Training Material, Deliverables, Technical Documents, etc.
It is meaningful to point out the trend of accesses as logged by the DL’s statistical tools (see the Figure 5). The chart clearly shows a number of peaks related – on one side – to the aforementioned events and – on the other side – to key dates corresponding to milestones in projects’ lifetime (reviews, call for proposals, etc.). The decrease in the last two months of 2009 is related to the particular conjunction of the year’s period and of most of projects’ lifecycle period.

It is worthwhile also mentioning that both the average and the maximum number of hits registered a growth of one order of magnitude from the first phase of the project (marked as “BELIEF” in the figure) to the second one (marked as “BELIEF-II”) (see Figure 6).

Statistics analysis

Among the amount of statistical data gathered starting from 2006, some are particularly relevant to evaluate the impact of the DL on the e-Infrastructures community. Among them the following was chosen:
• **Users’ provenance**: to show the real spread of the community on the different countries and to evaluate the interest on the DL from a quantitative point of view.
• **Top sites**: to show the relevant sites from which the DL is accessed.
• **Top operations**: to evaluate the type of use of the DL’s resources.
• **Yearly trend**: to analyse the change overtime from a quantitative perspective, to be cross-interpreted with the previous ones.

The analysis of **users’ provenance** shows that - prescinding from the high number of registered accesses - a huge number of those come from outside the EU and the group of countries directly involved in the projects being part of the community. We should consider this an evidence of the impact of the DL outside of its “original environment”, and of its diffusion in the scientific community “in the large”.

Among the **top sites** (excluding generic name-servers, and accesses coming from known EU projects and projects’ partner sites) from which the DL is accessed the following should be mentioned:
• Google, with the GoogleBot web spider
• Yahoo, with its Web Crawler
• A number of telecoms’ portals via their search services, e.g. Fastweb, Telecom Italia, Vodafone, On Telecoms, etc.
• A number of search engines, e.g. MSN Search, Cuil, Dotbot, etc.

It is worth notice that these data, joined to the info related to the provenance of accesses, highlight the wide spreading and variety of consumers achieved by the DL, also outside of the simple e-Infrastructures community.

Then, the **top operations** as registered by the statistical tools are the following:
• Query (simple)
• Browse
• OAI access
• Query (complex)
• Submission of new content

These data manifestly show that the most frequent operations on the DL are queries performed on both all Metadata Elements and Full Text, followed by generic browse of the content and by accesses from sites/repositories making use of the OAI/PMH protocol to query the DL. Query focused on specific Metadata Elements and the insertion of new content in the DL follow, as expected to be.

Although less meaningful than the former for a proper impact evaluation, these latter statistics are extremely useful to monitor the kind of usage of the DL’s resources as made by users, and to consequently tune the underlying system, in order to implement the most appropriate distribution and configuration of software services on the different hosts (see also [5]).

Finally, specific attention should be given to the **yearly trend** (see [6] for summary data), in order to properly evaluate the meaning of the different mix and balance of collected statistics data with respect to both the achievements of the BELIEF project advertising and dissemination activity, and the actual use of the DL as made by the community.
7. Business Benefits – Sustainability

In this section we summarize the cost benefits derived by organizations building DLs on top of OpenDLib. In particular, we shall describe how different communities can build their specific systems and what kinds of costs they have to face in this process. We shall show that these are affordable also by organizations and institutions whose communities cannot count on adequate resources for building traditional DLs. As a term of reference we should mention that the time scale for the basic system availability is a matter of days, whilst the implementation cost of the BELIEF DL was limited to a few man-months (nearly 4) as to the Aggregation System, and even less for the customization of the Portal/User Interface.

Sustainability Flavours for DLs

Sustainability has become a keyword also for digital library projects and implementations. It has a number of aspects and can be used to refer to a wide variety of concepts. In the digital library context, sustainability is a broad term, referring to everything from technical issues about the choice of an adequate DLMS and the implementation of a DL fully compliant to the users’ needs, to the digital preservation of materials, to the social questions related to the organization and maintenance of a network of producer and consumer of the DL’s content, to the long-term accessibility of resources to the public at large, etc.

Focusing on our project’s DL, it is worthwhile noticing that sustainability has been an integral part of its development. The starting point for our approach towards sustainability was to exploit a product [4] with characteristics of openness, maintainability, reliability and scalability, for which simple customization were needed to cope with the diverse requirements from the Community. Required enhancements to the underlying technology were carefully evaluated to avoid the attractions and risks of a technology driven project. This was already discussed in Section “4 -Technology Description”.

On the non-technological side of the problem, one of the basic goals was actually to let the DL became an integral part of the Community created around the project, and an uncuttable part of its organisation. The key to sustainability, in this context, was to achieve a situation where the digital library is not considered as a simple add-on to the Community’s cooperation tools, but as a core part of the Community itself. Albeit good results were achieved on this side, the goal of a self-sustaining organisation still awaits for a concrete solution, also encompassing the economical aspects. We refer to these issues as the “social” ones (see [7] for details).

Building Sustainable DLs over OpenDLib

In the following, we discuss the costs for organizations willing to build their DLs on OpenDLib and the overall costs of maintenance of the Infrastructure. We shall see how organizations can reduce or eliminate both technological and administrative costs for organizations and how the underlying infrastructure can be maintained by a limited group of experts. From the point of view of DLs construction, two main categories of users and related costs can be identified in the OpenDLib Infrastructure:

• Content Providers: developers of organizations willing to build their DLs in this environment.
• OpenDLib Administrators: OpenDLib staff, in charge of the OpenDLib Infrastructure administration, supporting content providers in building DLs.

Given the current set of Services available in OpenDLib, Content Providers have to deal with two main cost categories:

• Building Aggregation Systems
The OpenDLib DLMS enables multiple organizations to define different Information Spaces building their own Aggregation System on top of the OpenDLib Services. The costs for organizations are generally lower than in traditional Aggregation Systems and depend on the kind of technical involvement they are willing to achieve. Organizations can decide to depend on existing installations of the Services. This scenario, to be evaluated and approved by the OpenDLib Administrators, entails only administrative costs for the organization, that is staff capable of configuring Aggregator Services through their user interfaces. If the organization is willing to offer new hardware and install OpenDLib Software, then local technicians are needed, with support from OpenDLib Administrators. Software installation is rather standard and involves common technology, but some cost is required. Finally, organizations may be willing to use the OpenDLib Software but also to include new typologies of Services into OpenDLib or provide different implementations of existing typologies. This level of engagement is similar in some aspects to that of traditional Aggregation System construction. Skilled organization developers need support from OpenDLib Administrators in order to understand the OpenDLib Infrastructure framework and correctly contribute with compatible software.

- Building Portals on the OpenDLib Toolkit

The Advanced User Interface of OpenDLib (see Section 4) is designed to be fully integrated with a number of user functionalities, based on user registration and preferences. The UI offers simple free-keyword search and advanced metadata fields search and it can be configured to automatically adapt to any metadata format. Most importantly, users can run searches narrowing down to collections, i.e. limiting the range to a subset of the Information Space determined by a predicate on the metadata at hand. Collections are, together with users, managed by DL administrators with simple user interfaces.

The administrative costs are that of staff in charge of managing users, their communities and collections, while the technological costs are those of customizing the aspect of the user interface according to the organization community preferences. Again, the effort is not comparable with the technical costs required to design and develop the same interface with part or all such functionalities.

Such costs are independent from those of OpenDLib Administrators, who work under the supervision of the CNR organization. These operate through the OpenDLib Services, which: (i) offer admin-user interfaces to monitor and configure the behaviour of all running DLs and (ii) perform automatic management of critical aspects of the DLs. Their effort is typically higher when initiating and configuring a new DL for an organization, and become then ordinary administration related to system’s quality-of-service. Of course, due to the distributed nature of the architecture, Services’ automatic controls might not be enough and in that case administrators must have the adequate skills to solve problems. Accordingly, technical and administrative costs of the infrastructure framework can be compared to those typical of a distributed application, thus sustainable for one or more organizations working in synergy, such as those of the BELIEF Consortium.

Summing up, and generalizing the BELIEF experience to any community, the implementation and operation costs of a DL is therefore that of (i) training Content Providers’ Correspondents, (ii) keep alive the network of liaisons necessary to promote the community leveraged by the DL and (iii) perform harvesting operations, which are not delegated to Correspondents, but to OpenDLib Administrators. Technological costs are null to the community, which can exploit hardware and Services provided by the OpenDLib Infrastructure deployment. Overall, such costs are not even comparable to those that would
be required for building an ad-hoc Aggregation System of the same broad impact. Both technological and administrative costs are diluted by the support of the infrastructure framework, which is capable of automatically ensure storage allocation, indexing, and robustness according to the organization needs, and by the power of Aggregation Services, which are capable to deal with the generic pattern of harvesting, cleaning, transforming metadata records through admin user interfaces.

8. Conclusions

The effective implementation and use of a Digital Library within the scope of the BELIEF Coordination Action was shown. Specific implementation issues were highlighted, as well the conceptual aspects characterising this application context.

Focus was put on harmonisation of metadata coming from different Information Sources and on harvesting rules, protocols and formats as well as on specific harvesting tools implemented to cope with the diverse characteristics of those Sources. Compliance to open standards as well openness of the DL’s architecture was introduced.

Results achieved in terms of community growth and richness of the collected content were presented to highlight the impact of the DL usage on the e-Infrastructures community. It’s worth remind that the BELIEF DL currently counts more than 14.000 publications on e-Infrastructures from all European Countries, harvested from more than 90 selected repositories.

Special emphasis was given to the sustainability characteristics of the project, and to the cost elements, in the light of a reuse of this solution in the application context of different scientific communities.

Finally, importance of a proper and controlled usage of metadata elements and values was discussed. Implemented or suggested controlled vocabularies were introduced, whilst the implementation of “Metadata Curation” and “Authority File Control” Services was envisaged. Their detailed design and implementation will be one of the main objective in the furtherance of the project.

References

[14] DRIVER - Digital Repository Infrastructure Vision for European Research
http://www.driver-repository.eu/
http://www.driver-community.eu/
[18] Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G.,
124 – 138
[22] OAster Official Site http://www.oaister.org
[24] Zentity - MSR’s Research Output Repository Platform
[27] D4Science Project http://www.d4science.eu/
[29] EGEE Project http://www.eu-egee.org/