

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

Franco Zoppi Federico Biagini
franco.zoppi@isti.cnr.it federico.biagini@isti.cnr.it

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Via G. Moruzzi 1, 56124 Pisa, Italy

Abstract

The BELIEF (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers) 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. This will help ensuring that e-Infrastructures are developed and effectively used worldwide, filling the gap separating the Research Infrastructure providers from the users, and thus contribute to the emergence of a competitive knowledge-based economy.

To create this synergy among multi-disciplinary Research Infrastructure communities, BELIEF created a one-stop-shop for e-Infrastructures communities providing a Community Portal and a Digital Library (DL) with a huge number of e-Infrastructures open access publications. Moreover, it has organised events including brainstorming, networking workshops and international conferences and publications, since BELIEF's values are firmly rooted in international cooperation with the emerging economies, particularly in Latin America, India and South Africa.

The BELIEF DL – implemented on top of the OpenDLib Software System – offers uniform access to multimedia documentation (e.g. presentations, videos, technical reports, manuals, on-line tutorials, etc.) providing continuously updated information on eInfrastructure-related projects, initiatives and events. The contents are harvested from different sources, such as projects web sites, repositories and databases. The DL provides services to support the submission, description, searching, browsing, retrieval, access, preservation and visualization of multimedia documents. Although designed to meet the needs of a specific community, the technology adopted by BELIEF can be easily adapted to meet the information and collaborative needs of other scientific communities.

1 Introduction

The BELIEF 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 play 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.

General outcomes of the design phase are reported in [7]. In the following (Section 2) the main characteristics of the DL are described. General concepts regarding the implementation of the DL are outlined (Paragraph 2.1). These general concepts represent the logical link between basic requirements and the components of the implemented solution (Paragraph 2.2). Then (Paragraph 2.3) the organization of the User Interface is introduced.

Section 3 summarizes the most relevant usage data as per the statistics gathered during the whole projects lifetime, from 2006 onward.

This paper won't cover all of the relevant topics of the implementation work, for those please refer to [4]. Hereafter are summarized the ones covered in this paper (Section 4):

- 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.
- Integration with external portals.

Conclusions and results achieved are finally outlined in Section 5.

2 The Digital Library

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 offers its user community advanced services to uniformly access multimedia documents such as technical reports, presentations, videos, manuals, on-line tutorials, etc. These documents contain the very latest details on e-Infrastructures related projects, initiatives and events.

The material maintained in the DL is 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 the existing documentation. To this end, an extremely accurate collection and analysis of the requirements of potential users was made before the DL was created.

The DL provides 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), with or without relevance feedback. Users can search/browse any information associated with digital documents and their parts.

2.1 Preliminary Design

A design phase based on a deep analysis of the Project's target audience needs was carried out with a selected set of Research Infrastructure Entities (Projects, Organisations and Initiatives). The methodology adopted for requirements collection was basically structured in the following activities:

- Drawing up a Questionnaire to collect requirements on functional requirements, 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.

2.2 DL's General Concepts

This section introduces the main concepts characterising the DL from the perspective of its end-users. These concepts represent the logical link between the basic users' requirements and the components of the solution implemented, as follows:

Users of the BELIEF **DL** access the DL's **Resources** according to their **Role** and to the **Access Rights** they hold on **Collections**. These are grouped in user-specific **Information Spaces**. A **User Profile** is then associated with each User, recording his/her characteristics in terms of Role and Access Rights.

Where:

Users Are the actors entitled to interact with the DL. Different types of Users can be provided.

The following types of Users are currently supported:

- **Policy Maker**
- **Researcher**
- **Manager/Executive**
- **General Public**

DL From a high-level perspective, a DL is an organisation capable of collecting, creating, accessing and managing Resources.

Resources The resources that can be accessed by Users, depending on their Role and Rights. Diverse types of Resources can be provided.

The BELIEF DL currently supports the following types of Resources:

- **Content** Metadata and Documents (i.e. Information Sheets, News & Events, Training, Deliverables, Presentations, Meetings, User Guides, Technical Reports, Technical Notes, etc.).

- **Services** DL Administration, Document Submission, Document Approval, Collection Creation, etc..

Resources of type Content are gathered from **Information Sources**.

Information Source This term is used to indicate a repository belonging to an Entity (Project, Organisation or Initiative) from which metadata and documents are collected to form the DL. Information Sources are characterised by the type, structure and semantics of the metadata and documents they contain.

Role A function within the context of an organisation (in our case the BELIEF DL) with some associated semantics regarding the authority and responsibility conferred to the user. Different types of Roles can be provided.

The DL currently supports the following roles:

- **Consumers** Users that access the DL to utilise its Content.
- **Providers** Users that provide new Content to be stored on the DL or update Content that already exist.
- **Librarians** The DL Administrators.

Access Rights The actions allowed on Resources (Create, Modify, Delete, Search/Browse, Use/Perform, etc.).

Collection A set of Resources of type Content defined according to characterisation criteria. Collections represent the mechanism to organise Content in order to provide focused views of the Information Space, and allow DL end-users to access to thematic parts of the whole Information Space.

Collections can be created by the Librarians. Additionally, they can be created by authorised Users to implement their own personal views of the DL Information Space. A set of characterisation criteria defines and identifies the Content forming a Collection. These criteria can range from the enumeration of the Content belonging to the Collection to membership conditions that specify the properties that Content must meet in order to become part of the Collection.

The following criteria are suggested:

- Different **abstraction levels**:
 - By Project (EGEE, DANTE, etc.)
 - By Main Topic (Digital Libraries, High Parallel Computing, Security, etc.)
 - By Specific Keywords (examples related to the topic Security follows: Certification Authority, Public Key Infrastructure, Security policies, Network Security, Web services security, etc.)
 - Etc.
- The most common **information needs** of defined Users:
 - “At a glance” (e.g. Brochure + Presentation, etc.)
 - “General Executive” (e.g. Project Management Document + Project Management Meeting, etc.)
 - “Developer” (e.g. eCourse + Software + Technical Document, etc.)
 - “General Admin” (copying the actual organisation of resources)
 - Etc.

Information Space Corresponds to the **set of Collections of the DL** that each user can access. It defines the scope of Content against which the Browse/Search operation are executed. The Information Space can be dynamically modified by the user to restrict/enlarge such scope.

User Profile Contains information about the user. It may be viewed as a set of characteristics with associated values. A profile may store different types of information, such as personal data (e.g., name, address), demographic data (such as marital and economic status), education and training level, group membership, preferences (e.g., look and feel of the user interface, preferred Content and Services of the DL), experience, knowledge, or skills, disabilities (e.g., vision problems), permissions, search history data, navigation history and so forth. Each piece of user information alone and in conjunction with others determines system behaviour towards the user.

The profile is strictly coupled to the information, preferences and options managed by the BELIEF Project's Portal and expressed by a user on registering with the BELIEF Community through the Portal. The initial User Profile is created based on such information and can be further modified by the user.

2.3 General Characteristics of the Implemented Solution

The BELIEF DL is created as an instance of the OpenDLib [4] Software System, thus being capable of offering the following services to its users:

- Creation, submission, search, browse, access, and preservation of multimedia documents.
- Definition of the 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 been devoted to the following issues:

- Implementation of harvesters supporting standard formats and specific conventions on Information Sources side. Presently, 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.

- 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.
- 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.

The DL will thus cover a double role:

- it will facilitate cross-fertilization and synergies between e-Infrastructures performers by supporting the integration of the produced documentation, and
- it will promote pervasive penetration of the state-of-the-art technology in the research infrastructures by providing a more friendly access via the availability of advanced services for the definition and navigation of personalized information spaces.

The benefits coming from the use of the DL are twofold:

- On the Information Provider’s side:
 - e-Infrastructures projects’ documents can be accessed by a wider users’ community.
 - This wider audience is also a more targeted audience. The DL provides users with the most accurate selection for their information needs, in a way that bridges them directly with the projects’ material.
 - It gives projects’ audience a one-stop-shop to find the precise documentation that they require about the projects, thus it is a valuable target to send inquirers to, should they be searching quickly for projects’ documentation.
 - Documents are accessible via different formats.
 - The DL supports document preservation.
 - Projects can submit their documents directly to the DL. BELIEF offers a submission interface and a document hosting service. To ensure that any new documents can be easily collected by the DL a standard method is provided to describe them.
- On the Information Consumer’s side:
 - While accessing a website and browsing a project’s material, users get material from that project/website only,. using the DL they get the relevant information from many projects’ sites in a single shot. This give them a wider range of information on e-Infrastructures, how they work, how to use them, etc..
 - User can search documents in a uniform way, accessing via a unique interface documents produced by different providers.
 - The material accessible is extremely well focused on the users’ needs. This is not like Google which may often come up with nothing or thousands of results or it yields material that is not relevant or up to date.

2.4 The User Interface and its usage

The DL's User Interface (UI) 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 simplified 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.
- A new On-line User Guide has been implemented by means of a friendly Wiki capability.

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] and [6] for details, or access directly [2].

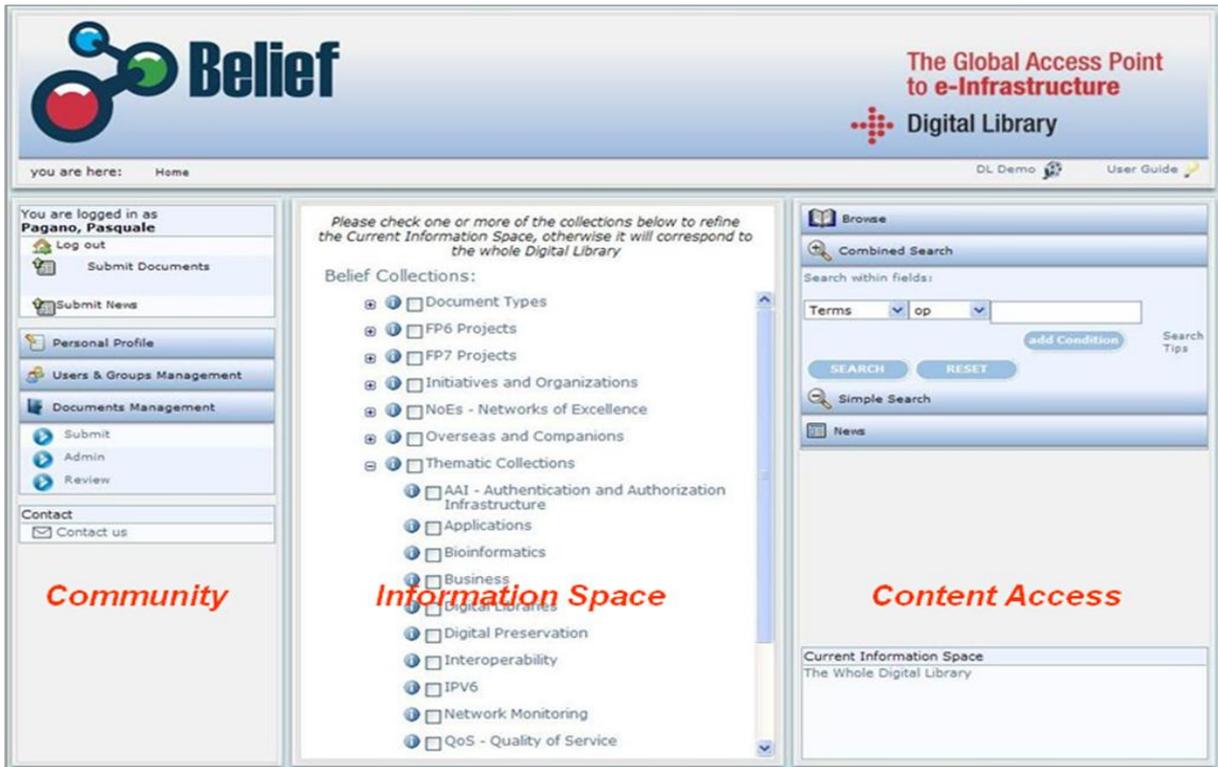
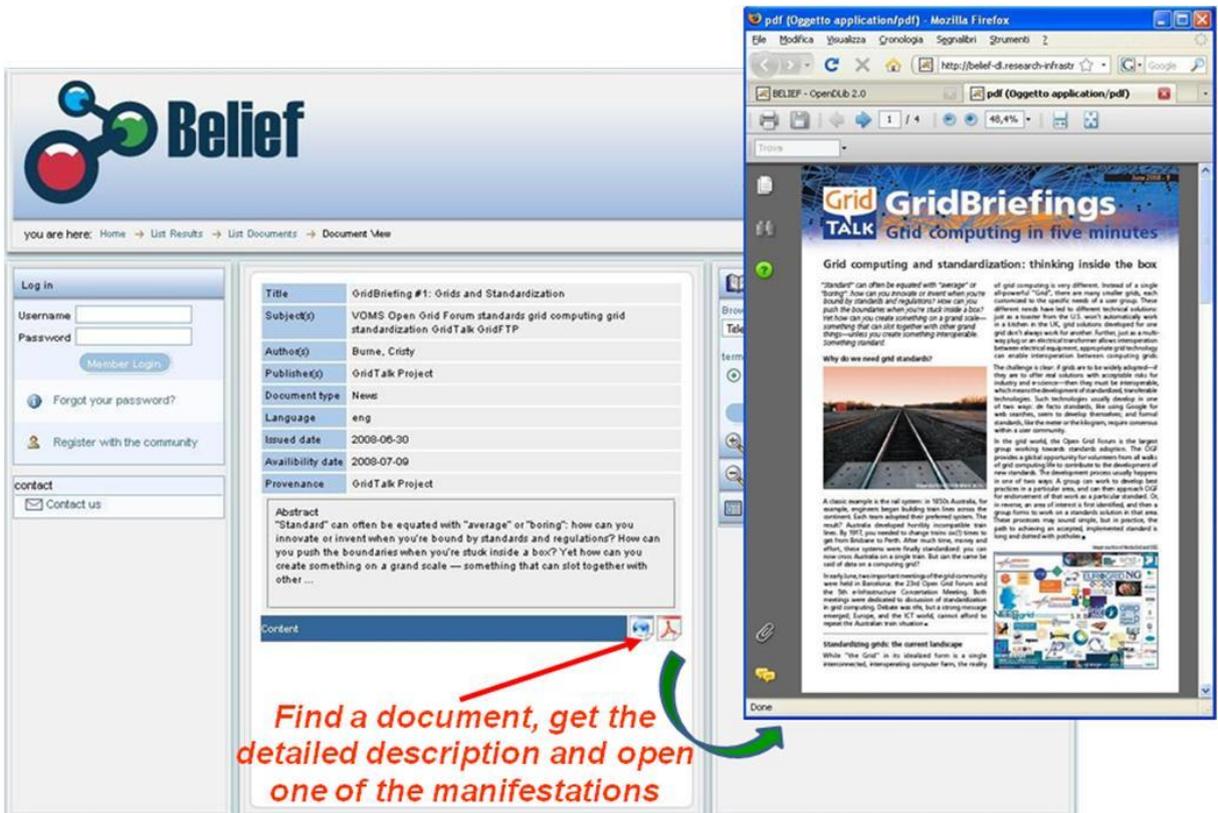


Figure X – The DL Desktop organization



Find a document, get the detailed description and open one of the manifestations

Figure X – Accessing content

Within each of the three parts of the DL Desktop, a number of capability can be accessed by users. Hereafter the main capability of the UI are summarized:

Community Section

- News
 - **Submit** to publish news on the DL.
 - **View & Edit** to modify news already published.
- Personal Profile
 - **View & Edit** to modify the personal profile data.
 - **Unregister** to resign from the Community.
- Documents Management
 - **Submit** to submit requests for the publication of new documents, new editions of documents already published, correction of documents already published, and withdrawal requests.
 - **Review** to check submission requests before approval (or rejection).
 - **Admin** to manage all submission requests.
- Users & Groups Management
 - **View & Modify Users** to create/delete users and manage their access rights.
 - **View & Modify Groups** to create/modify/delete groups of users and manage their access rights.
- Information Space Management
 - **Create & Modify Sets** to create/edit/delete sets of data (see [4] for details) on which collections are then created.
 - **Create & Modify Collections** to create/edit/delete public and/or personal collections.

Content Access Section

- **Browse** to browse the content of the current Information Space according to a selected Metadata Element.
- **Simple & Combined Search** to search the content of the current Information Space as a whole or according to a structured condition on multiple Metadata Elements.

Information Space Section

- **Collection Selection** (via a check-box) to modify the actual scope of the Information Space.

3 BELIEF Events and Community's DL growth and usage

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.

The events comprised:

- 5 Brainstorming Workshops on multidisciplinary topics. The output from these events was published in four Technological Reports available to the public. A Concept Paper was developed for each event and made available on line beforehand.
- 3 Research and Industry Networking Sessions offering opportunities for interaction between these two communities. The output from these events contributed to dedicated Research and Industry Handbooks, tailored for different user communities, highlighting how e-Infrastructures can be deployed and sustained. A Concept Paper was developed for each event and made available on line beforehand.
- 2 International Conferences dedicated to e-Infrastructures across key domains. The 1st BELIEF International Conference "European & Indian Research Infrastructures Conference" took place 14-15 December 2006 in New Delhi, India and spotlighted themes that are high on business and research agendas and that have recently been the focus of Indian presidential addresses. The 2nd BELIEF International Conference "Connecting the knowledge of today for the value of tomorrow" was jointly organised with the EELA project and was held in Rio de Janeiro, Brazil 25-28 June 2007. This event helped to pave the way to a better understanding and development of e-Infrastructures in Latin America.
- 3 Concertation Meetings held in Sophia Antipolis, Barcelona and Lyon on the themes of standardisation and sustainability for e-Infrastructures.
- 1 International Symposium held in New Delhi, India, 28-29 January 2009 on the theme "e-Infrastructures for Distance Learning: A challenge for the Indian Society".

More events are planned within the current year, namely:

- 2 International Symposia, one to be held in São Paulo, Brazil on 16-17 July 2009 on the theme "Future trends and e-Infrastructures application: towards sustainable development", the other to be held in Johannesburg, South Africa, in December 2009.
- 1 Concertation Meeting and 1 Brainstorming Workshop whose location and themes will be shortly agreed with the European Commission.

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. The following table summarizes the DL actual growth vs. the initially planned trend during the first 3 years of the project.

	Planned		Actual		
	Y1	Y2	Y1	Y2	Y3
Projects, Initiatives & Organizations	8	30	14	35	78
Collections	15	40	27	60	112
Documents	2.000	6.000	3.000	13.000	14.000

Table 1 – The BELIEF Community’s DL growth

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..

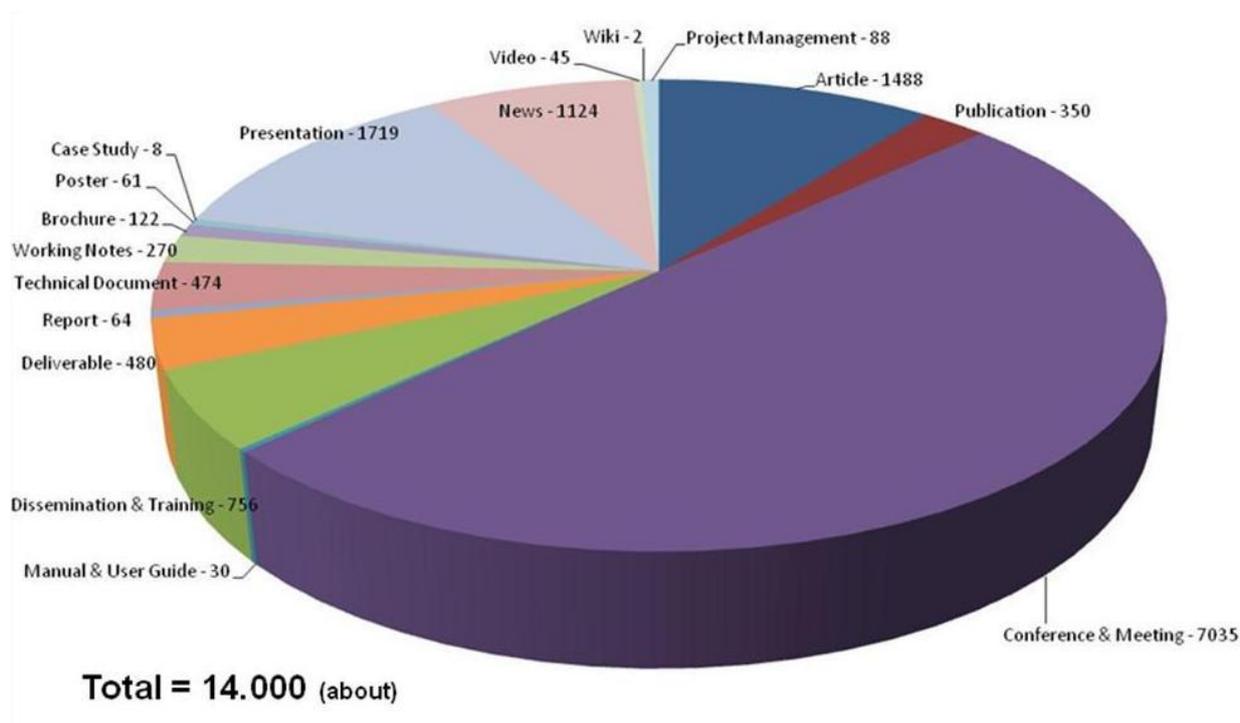


Figure X – DL’s content type

It’s meaningful to point out the trend of accesses as logged by the DL’s statistical tools (see following figure). 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.).

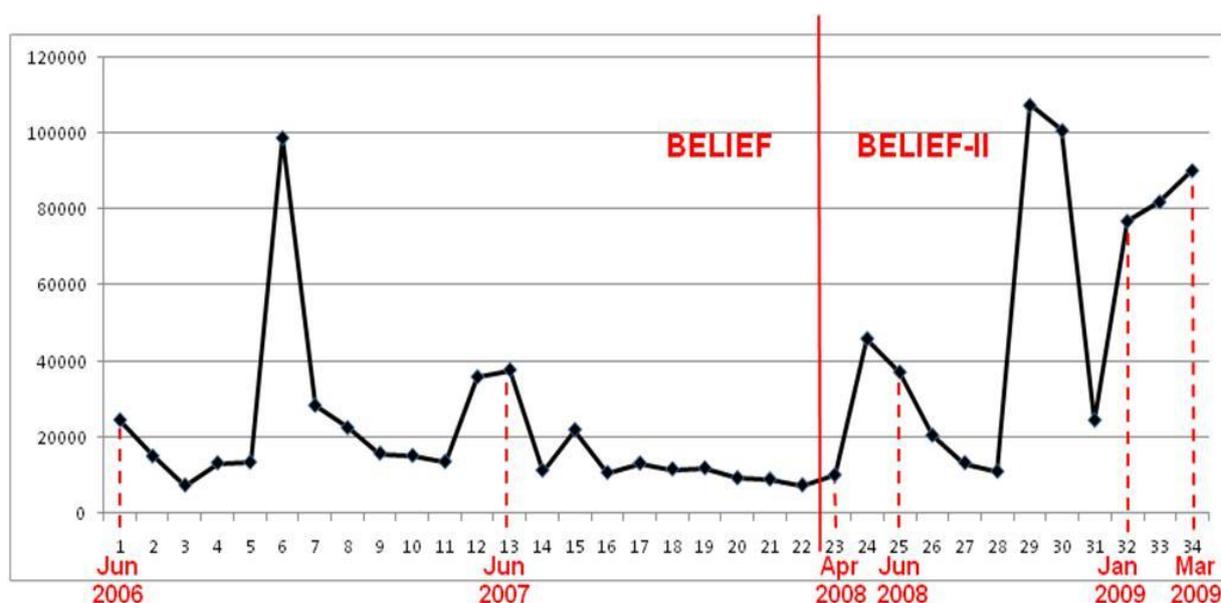


Figure X – DL hits per month from June 2006 to March 2009

It's 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 previous figure) to the second one (marked as BELIEF-II). This is exemplified by data related to a single month in the table below.

March 2009	Avg	Max
Hits per Hour	121	3225
Hits per Day	2911	8043
Files per Day	1547	5862
Pages per Day	514	1127
Visits per Day	29	73

March 2008	Avg	Max
Hits per Hour	10	665
Hits per Day	241	896
Files per Day	195	421
Pages per Day	6	21
Visits per Day	4	10

March 2007	Avg	Max
Hits per Hour	20	1442
Hits per Day	491	3138
Files per Day	295	786
Pages per Day	13	53
Visits per Day	5	14

Table 2 – Growth of accesses to the DL (Avg and Max values per month)

The detailed analysis of users' provenance is shown in the table below (figures are referred to one month – May 2009):

Top 30 of 34 Total Countries							
#	Hits		Files		KBytes		Country
1	38264	46.43%	8287	17.01%	189470	35.36%	BELIEF Project Portal
2	36486	44.27%	33297	68.33%	84740	15.81%	US Commercial
3	1905	2.31%	1660	3.41%	25003	4.67%	Unresolved/Unknown
4	1662	2.02%	1656	3.40%	113684	21.22%	Germany
5	1124	1.36%	1015	2.08%	32000	5.97%	Network
6	605	0.73%	565	1.16%	17032	3.18%	Brazil
7	230	0.28%	221	0.45%	14773	2.76%	International (int)
8	189	0.23%	178	0.37%	4470	0.83%	United Kingdom
9	188	0.23%	183	0.38%	1416	0.26%	Netherlands
10	178	0.22%	174	0.36%	1047	0.20%	Poland
11	154	0.19%	149	0.31%	3578	0.67%	Greece
12	153	0.19%	149	0.31%	1387	0.26%	Spain
13	153	0.19%	142	0.29%	13954	2.60%	India
14	142	0.17%	137	0.28%	911	0.17%	Sweden
15	125	0.15%	122	0.25%	882	0.16%	Luxembourg
16	109	0.13%	106	0.22%	13696	2.56%	Czech Republic
17	96	0.12%	91	0.19%	685	0.13%	France
18	91	0.11%	90	0.18%	579	0.11%	Austria
19	81	0.10%	78	0.16%	13507	2.52%	Hungary
20	79	0.10%	76	0.16%	564	0.11%	US Educational
21	79	0.10%	78	0.16%	472	0.09%	Indonesia
22	68	0.08%	60	0.12%	438	0.08%	Jordan
23	64	0.08%	62	0.13%	470	0.09%	Belgium
24	63	0.08%	63	0.13%	470	0.09%	Ireland
25	59	0.07%	57	0.12%	435	0.08%	Norway
26	28	0.03%	0	0.00%	19	0.00%	Chile
27	19	0.02%	18	0.04%	38	0.01%	Portugal
28	12	0.01%	6	0.01%	17	0.00%	Russian Federation
29	4	0.00%	2	0.00%	6	0.00%	Old style Arpanet (arpa)
30	4	0.00%	3	0.01%	55	0.01%	Japan

Table 3 – Provenance of accesses to the DL (May 2009)

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's 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.

Finally, 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.

4 Technical insights

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:

- 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.
- Integration with external portals.

4.1 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].

The following (17, 9 of which qualified) DCQ metadata are currently supported by the BELIEF DL. The metadata marked with (*) have to be considered mandatory for an effective classification in the DL:

- 1) (*) **Title** (*qualified*)
- 2) (*) **Creator**
- 3) (*) **Subject**
- 4) (*) **Description** (*qualified*)
- 5) (*) **Publisher**
- 6) - **Contributor** (*qualified*)
- 7) (*) **Date** (*qualified*)
- 8) (*) **Type** (*qualified*)
- 9) - **Status**
- 10) - **Format**
- 11) - **Identifier** (*qualified*)
- 12) - **Source** (*qualified*)
- 13) (*) **Language**
- 14) - **Relation** (*qualified*)

- 15) - Coverage (*qualified*)
- 16) - Rights
- 17) (*) 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 an “Authority File Control Service” to support librarians and administrators in such arduous task.

4.2 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, as shown in the following section).
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. As to the formats, both DC/DCQ and different “flavors” of MARC have been used, whilst both OAI-PMH and a number of proprietary interfaces are supported to implement the communication protocol. 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.

4.3 OAI-PMH Compliance

The DL features an OAI-PMH [9] fully compliant implementation which let the DL provide effective tools to support the Principles of the European Commission’s Communication of Scientific Information [10] and the subsequent Competitiveness Council Conclusions on Scientific Information in the Digital Age [11].

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 (publications and data), as set out by the ERC Scientific Council Guidelines for Open Access [12] and the Open Access Pilot [13] launched by the European Commission.

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. Such integration produced a first implementation which is now going to be refined with a close interaction with the DRIVER Consortium.

4.4 Integration with External Portals

An advanced and powerful characteristic of the BELIEF DL is its capability of being integrated in an external portal, that is the portal implemented by another project, willing to strictly interact with the DL to let this behave like a component/service of the portal itself.

Such integration can be achieved via either the use of the APIs provided by the underlying OpenDLib System, or by incapsulating specific code modules – especially developed for the BELIEF DL – within the external portal.

The former approach was implemented in the BELIEF Portal, which interact with the DL by means of APIs like “List Authorities” (i.e. the Entities authorized to submit content), “Submit Document”, etc..

The latter was used in the D4Science project’s portal, since this required the DL to run in a sort of “sand box” within the Portal itself.

The following figures show the implementation of the D4Science project, where the authentication of the user and the submission of a document to the DL are embedded in the portal. In a similar way a “plug-in-search” capability was integrated in the portal to implement a direct access to the BELIEF DL content.

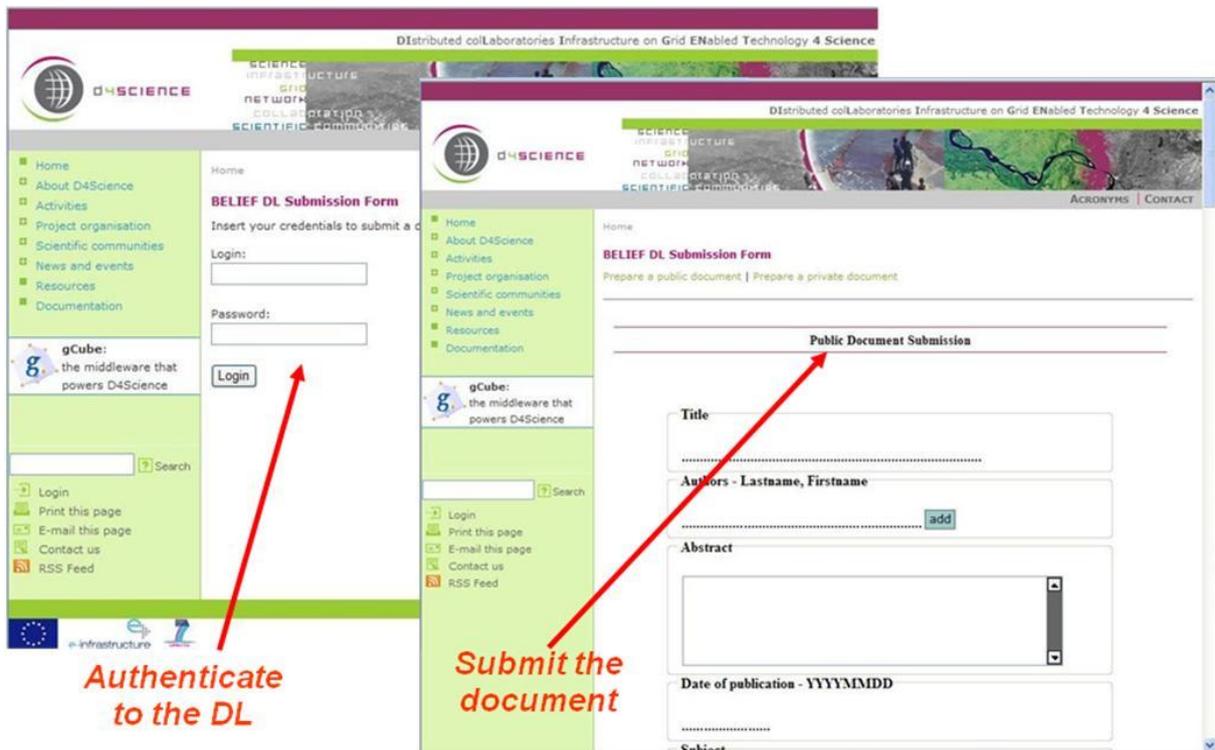


Figure X – DL integration with D4Science (Authentication and Submission)

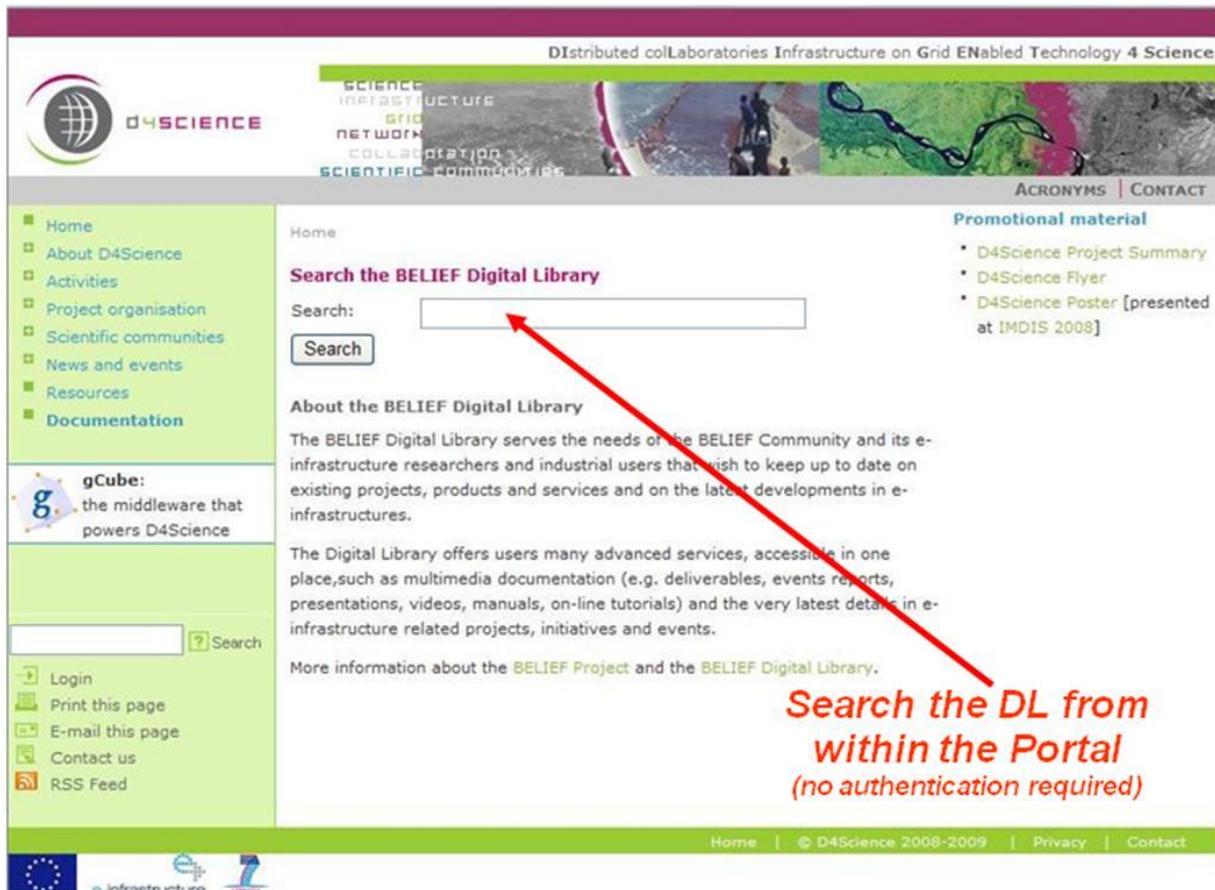


Figure X – DL integration with D4Science (Search)

5 Conclusions and Future Work

The effective 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, with special emphasis on the DRIVER and D4Science examples respectively.

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 an "Authority File Control Service" was envisaged. Its detailed design and implementation will be the main goal of this activity.

References

- [1] BELIEF Project <http://www.beliefproject.org>
- [2] BELIEF Digital Library <http://belief-dl.research-infrastructures.eu/>
- [3] BELIEF Digital Library Wiki User Guide
<https://userguide.wiki.belief.research-infrastructures.eu/>
- [4] OpenDLib website <http://opendlib.research-infrastructures.eu/>
- [5] Zoppi F., Biagini F. *BELIEF-II D3.4.1 DL - Release Notes*. Bringing Europe's eElectronic Infrastructures to Expanding Frontiers - Phase II (BELIEF-II) - FP7-Infrastructures-2007-2 - 223759. Deliverable BELIEF-II D3.4.1, 2008
- [6] Zoppi F., Baldacci M. B., Biagini F. *BELIEF-II D3.3.1 DL - User Guide*. Bringing Europe's eElectronic Infrastructures to Expanding Frontiers - Phase II (BELIEF-II) - FP7-Infrastructures-2007-2 - 223759. Deliverable BELIEF-II D3.3.1, 2008.
- [7] Zoppi F., Artini M., Baldacci M. B., Biagini F. - *BELIEF - WP1D5.4 - Bridging Information Consumers & Providers*. BELIEF (Bringing Europe's eElectronic Infrastructures to Expanding Frontiers) - FP6-2004-Infrastructures-6 - 026500. Deliverable BELIEF WP1D5.4, 2008.
- [8] The Dublin Core Metadata Initiative Open Forum
<http://www.dublincore.org/>
- [9] The Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [10] Principles of the European Commission's Communication of Scientific Information
http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf
- [11] Competitiveness Council Conclusions on Scientific Information in the Digital Age
http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97236.pdf
- [12] ERC Scientific Council Guidelines for Open Access
http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf
- [13] Open Access Pilot
ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf
- [14] DRIVER - Digital Repository Infrastructure Vision for European Research
<http://www.driver-repository.eu/>
<http://www.driver-community.eu/>