

The BELIEF Digital Library

Michele Artini, Federico Biagini, Franco Zoppi

Consiglio Nazionale delle Ricerche,
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo",
via Moruzzi 1, 56124 Pisa, Italy

Abstract – BELIEF is an international project to facilitate knowledge-exchange on eInfrastructures a one-stop home for public eInfrastructure documentation. This information will be readily accessible to BELIEF Community Members through the BELIEF Digital Library especially developed to provide a central repository for eInfrastructure Information. BELIEF brings together not just individual experts and potential users but also other eInfrastructure 'communities' and initiatives, from both research and industry worldwide. This document explains the implementation of the BELIEF DL using the Digital Library Management System OpenDLib.

General Concepts

The BELIEF DL is an instance of OpenDLib, a Digital Library Management System (DLMS) comprising a federation of services that can be customised to meet the requirements of a target user community. 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.

The main services offered by a OpenDLib based system are:

- Submission, description, searching, browsing, retrieval, access, preservation and visualisation of multimedia documents;
- Definition of user-specific Information Space for searching/browsing in terms of collections selected from those managed by the DL. Collections can be created interactively;
- Different search/browse options: Google-like or fielded (with fields selected from a variety of known metadata formats), with or without relevance feedback.

Different methods for feeding the DL with new metadata and documents are supported:

- Harvesting from existing Information Sources implemented by specific modules and interfaces.
- On-line submission via Document Models.
- Batch submission via an XML Schema.

- Native repository integration.

BELIEF classifies the DL Collections in terms of:

- Projects
- Initiatives and Organisations
- Networks of Excellence
- National Research and Education Networks
- Interest Areas

While Projects, Organisations and Initiatives are continuously increasing, the following Document Types are currently provided by BELIEF DL:

- Article
- Audio
- Authored Book
- Brochure
- Case Study
- Conference & Meeting
- Deliverable
- Dissemination
- eCourse
- Edited Volume
- Manual & User Guide
- Handbook
- Module
- News
- Poster
- Presentation
- Proceeding
- Project Management Document
- Project Management Meeting
- Publication
- Report
- Software
- Summer School
- Technical Document
- Thesis
- Training
- Tutorial
- Video
- Working Notes

Mapping between these DL Document Types and Document Types used by Information Sources is given in 0 “Collections and document types cross-reference”.

The following two paragraphs define respectively:

- Collections and Sub-collections currently implemented within the DL,
- Suggestions for areas of interest to be proposed to BELIEF Community members as “Area of Application Interest” and “Area of Topic Discussion Interest” for which corresponding collections could be defined.

At present, these areas of interest are shown in the on-line Community registration form of the Portal. Each area hereafter is described in terms of its most relevant subjects. Both areas and related descriptions can be extended and/or modified based on the suggestions made by Community member through the Portal.

Implemented Collections

Collections & sub-collections	Details								
• Document Types	<table> <tr> <td>Name</td><td>Document Types</td></tr> <tr> <td>Image/Logo</td><td><BELIEF Project Logo></td></tr> <tr> <td>Description</td><td>This collection contains all public documents produced by collaborating Entities (Projects, Organisations and Initiatives) which will be accessible through the BELIEF DL. Documents in this collection are further organized in sub-collections, according to the pertinent document type.</td></tr> <tr> <td>Sub-collections</td><td> - Articles - Case Studies - Conferences and Meetings - Deliverables - Dissemination & Training - eCourses - Manuals and User Guides - Outreach - Presentations - Proceedings - Project Management - Publications - Reports - Software - Technical Documents - Working Notes </td></tr> </table>	Name	Document Types	Image/Logo	<BELIEF Project Logo>	Description	This collection contains all public documents produced by collaborating Entities (Projects, Organisations and Initiatives) which will be accessible through the BELIEF DL. Documents in this collection are further organized in sub-collections, according to the pertinent document type.	Sub-collections	- Articles - Case Studies - Conferences and Meetings - Deliverables - Dissemination & Training - eCourses - Manuals and User Guides - Outreach - Presentations - Proceedings - Project Management - Publications - Reports - Software - Technical Documents - Working Notes
Name	Document Types								
Image/Logo	<BELIEF Project Logo>								
Description	This collection contains all public documents produced by collaborating Entities (Projects, Organisations and Initiatives) which will be accessible through the BELIEF DL. Documents in this collection are further organized in sub-collections, according to the pertinent document type.								
Sub-collections	- Articles - Case Studies - Conferences and Meetings - Deliverables - Dissemination & Training - eCourses - Manuals and User Guides - Outreach - Presentations - Proceedings - Project Management - Publications - Reports - Software - Technical Documents - Working Notes								
○ Articles	<table> <tr> <td>Name</td><td>Articles</td></tr> <tr> <td>Image/Logo</td><td><BELIEF Project Logo></td></tr> </table>	Name	Articles	Image/Logo	<BELIEF Project Logo>				
Name	Articles								
Image/Logo	<BELIEF Project Logo>								

Collections & sub-collections	Details	
	Description	This is a collection of official articles produced by researcher involved in a variety of topics related to the goals of the communities. Article(s)
○ Case Studies	Values from Information Sources:	
○ Conferences and Meetings	Name Image/Logo Description	Case Studies <BELIEF Project Logo> This is a collection of e-Infrastructures case studies submitted freely through the BELIEF portal by registered members of BELIEF Community. This represent a 'best practice' collection, a window onto the pioneering e-Infrastructures activities of researchers and businesses worldwide. Case Study
○ Deliverables	Values from Information Sources:	
○ Dissemination and Training	Name Image/Logo Description	Conferences and Meetings <BELIEF Project Logo> This collection contains information related to conferences and meetings of interest hosted by or signalled by collaborating Entities. Conferences Events Meetings Workgroups Workshops Deliverables <BELIEF Project Logo> This collection contains public deliverables produced by collaborating Entities. Deliverables Dissemination and Training <BELIEF Project Logo> This collection contains available public dissemination and training documents produced by collaborating Entities.

Collections & sub-collections	Details	
	Values from Information Sources	eCourses Modules Presentations Training Tutorials Videos
○ eCourses	Name Image/Logo Description Values from Information Sources	eCourses <BELIEF Project Logo> This collection contains available public eCourses produced by collaborating Entities. eCourses
○ Manuals and User Guides	Name Image/Logo Description Values from Information Sources	Manuals & User Guides <BELIEF Project Logo> This collection contains official use documentation produced by collaborating Entities. Documentation Guidelines Handbook(s) Manuals User Guides
○ Outreach	Name Image/Logo Description Values from Information Sources	Outreach <BELIEF Project Logo> This collection contains official information and dissemination materials produced by collaborating Entities. Brochures Flyers Information Sheets Leaflets Maps News Newsletters Posters Press
○ Presentations	Name Image/Logo Description Values from Information Sources	Presentations <BELIEF Project Logo> This collection contains public presentations produced by collaborating Entities. Contributions Presentations Talks

Collections & sub-collections	Details	
○ Proceedings	Name Image/Logo Description Values from Information Sources	Proceedings <BELIEF Project Logo> This collection contains public proceedings produced by collaborating Entities. Proceedings
○ Project Management	Name Image/Logo Description Values from Information Sources	Project Management <BELIEF Project Logo> This collection contains documents related to the managements of the contributing Entities. Consortium Agreements Contracts Meetings Minutes Plans Quality Plans Technical Annexes
○ Publications	Name Image/Logo Description Values from Information Sources	Publications <BELIEF Project Logo> This collection contains official documents produced by researcher involved in a variety of topics related to the goals of the communities. Book(s) Publications Thesis Volume(s) White Paper(s)
○ Reports	Name Image/Logo Description Values from Information Sources	Publications <BELIEF Project Logo> This collection contains public reports produced by collaborating Entities. Report(s)
○ Software	Name Image/Logo Description	Software <BELIEF Project Logo> This collection contains software produced by researcher involved in a variety of topics related to the goals of the communities.

Collections & sub-collections	Details	
	Description	<short description>
	Web Site	<Entity's web site>
• Interest Areas	Name	Interest Areas
	Image/Logo	<BELIEF Project Logo>
	Description	This collection contains all public documents accessible through the BELIEF DL organized according to the most relevant Areas of Interest. This collection is further organized in sub-collections, one for each Area of Interest.
	Sub-collections	• AAI (Authentication and Authorization Infrastructure) • Applications • Business • Bioinformatics • Digital Libraries • Interoperability • IPV6 • Network Monitoring • Quality of Service • Security • Testbeds • Virtual Organizations
○ (for each Area)	Name	<Area's Name>
	Description	<short description>

Suggested Collections

Collections & sub-collections	Details	
• Application Areas	Name	Application Areas
	Description	This collection contains the application areas to which the DL contents belong. Each application area is in turn qualified by a set of specific subjects.
	Sub-collections	• Astrophysics • Business • Bioinformatics • Biophysics • Chemistry

Collections & sub-collections	Details	
		• Digital Libraries • Earth Science • Financial • Geophysics • Mathematical and computer methods in mechanics • Material processing & Sciences • Medicine & Biology • Physics • Etc.
○ Application Areas (for each)	Name Description Subject	<Application Area Name> <short description> <main topics>
• Discussion Topic Areas	Name Description Sub-collections	Discussion Topic Areas This collection contains the discussion topic areas to which DL the contents belong. Each discussion topic area is in turn qualified by a set of specific subjects. GRID/Middleware: • AAI (Authentication and Authorization Infrastructure) • Applications of parallel/distributed/GRID computing • Data Management • Dependable networks & middleware • GRID & eInfrastructure in Industry • Grid Access • GRID On Demand • Information & Monitoring • Interoperability & Interfaces • Methods and tools for dependable distributed systems • Metrics for trustworthiness • Middleware Security • Performance Monitoring, Analysis & Prediction • Resources Management & Allocation • Scheduling, mapping, load balancing • Virtual Organisation(s) • Workload Management • Etc.

Collections & sub-collections	Details
	Networking: • Advanced Network Monitoring • Bandwidth on Demand • End to End Service Provision • IPV6 • Layer 2 VPN • Multicast • Network Security • Performance Measurement and Monitoring • Quality of Service • Testbed, technology testing • Etc.
○ Discussion Topic Areas (for each)	Name <Discussion Topic Area Name> Description <short description> Subject <main topics>

Subject	Description
Area of Application Interest in eInfrastructure/GRID	
Astrophysics	Addresses the following issues: ○ Active Galactic Nuclei ○ Air Showers ○ Cosmic Background Radiation Field ○ Gamma Ray Bursts ○ High Energetic Primary Cosmic Rays ○ High Energy Gamma Rays ○ High Energy Neutrinos ○ Imaging Atmospheric Cherenkov Telescope (IACT) ○ Large Area Water Cherenkov Detector ○ Microwave Sky ○ Etc.
Business	<List of issues>
Bioinformatics	<List of issues>
Biophysics	<List of issues>
Chemistry	Addresses the following issues: ○ Chemical Systems Dynamics ○ Molecular Modelling ○ Simulation of Complex Chemical Systems ○ Etc.
Digital Libraries	Addresses the following issues: ○ Application Specific Management ○ Content and Metadata Management ○ Digital Library Creation and Management

Subject	Description
	○ Grid Protocol Architecture ○ Index and Search Management ○ Open Grid Service Architecture ○ Open Grid Service Infrastructure ○ Process Management ○ Scalable Virtual Organisations ○ Web Services Architecture ○ Web Service Definition Language (WSDL) ○ Etc.
Earth Science	Addresses the following issues: ○ Climatology ○ Earth Observation ○ Earthquake Simulation ○ Environmental Modelling ○ Flood Forecasting ○ Hydrology ○ Meteorology ○ Ozone Profiles ○ Solid Earth Physics ○ 3D Geological Models ○ 3D Density-dependant Groundwater Flow and Salt Transport Models ○ Etc.
Financial	Addresses the following issues: ○ Authorization and Accounting ○ Best Quote Data ○ Data Management ○ Econophysics ○ High Frequency Data ○ Transaction Data ○ Virtual Organisation ○ Etc.
Geophysics	Addresses the following issues: ○ Seismic Inversion ○ Seismic Processing ○ Seismic Tools ○ 3D Surface-related Multiple Modelling ○ 3D Tomography Modelling ○ Etc.
Mathematical and computer methods in mechanics	<List of issues>
Material processing & Sciences	<List of issues>
Medicine & Biology	Addresses the following issues: ○ Clinical Decisions Support ○ Contrast Transfer Function (CTF) ○ DNA Sequences Similarities ○ Drug Discovery ○ Electron Microscopic Images Manipulation ○ Genome Sequencing ○ Human Genetic Development ○ Mammography

Subject	Description
	○ Molecular Docking Analysis ○ Molecular Interactions Screening ○ MRI Simulation ○ Protein Analysis ○ Radio Therapy Planning ○ Spatial and Temporal Coalescences ○ Tomographic Emission ○ Etc.
Physics	Addresses the following issues: ○ Energy Particle Accelerator ○ Hydrogen Plasma Torus ○ High Energy Particle Collisions ○ High Energy Physics (HEP) ○ Large Hadron Collider (LHC) ○ Particle Accelerator ○ Physics of Strongly Interacting Matter ○ Plasma Physics ○ Etc.
Area of Topic Discussion Interest in eInfrastructure (GRID/Middleware)	
AAI (Authentication and Authorization Infrastructure)	Addresses the following issues: ○ Interworking AAIs ○ Scalability ○ Management of federations and confederations ○ Legal issues: privacy of user attributes, liability, etc. ○ Attribute-based authorisation in Grids ○ Full Single-Sign-On ○ Roaming and AAI ○ Etc.
Applications of parallel/distributed/GRID computing	<List of issues>
Data Management	Addresses the following issues: ○ Data Catalogs ○ Data granularity ○ Data movement ○ Data naming ○ Data scheduling ○ File placement ○ File transfer ○ Indexing ○ Metadata ○ Replication ○ Storage resource management ○ Storage element ○ Transfer service ○ Etc.
Dependable networks & middleware	<List of issues>
GRID & eInfrastructure in Industry	<List of issues>
Grid Access	<List of issues>

Subject	Description
GRID On Demand	<List of issues>
Information & Monitoring	Addresses the following issues: ○ Grid applications status ○ Grid Monitoring Architecture ○ Resources availability ○ Service discovery ○ Etc.
Interoperability & Interfaces	This subject concerns the essential interfaces between the various stakeholders in the Network-Grid-Application community. It addresses the following issues: ○ Grid interface to network resources ○ Coordination of network resources with Grid resources (e.g. CPU, storage) ○ Interoperability among different Grid infrastructures ○ Standardisation of application interfaces to different grids ○ Etc.
Methods and tools for dependable distributed systems	<List of issues>
Metrics for trustworthiness	<List of issues>
Middleware Security	<List of issues>
Performance Monitoring, Analysis & Prediction	This subject covers the role of performance monitoring in and between different eInfrastructure environments. It addresses the following issues: ○ Multi-domain network monitoring ○ Grid monitoring ○ Tools for measurement collection and visualisation ○ Summarisation of the monitoring information ○ Sharing monitoring information: policies, anonymisation, privacy ○ End-to-end performance ○ Definition of operational procedures in case of a fault ○ Etc.
Resources Management & Allocation	Addresses the following issues: ○ Job prioritisation (between grid and non-grid jobs and between grid-jobs) ○ Advance reservation tools ○ Routing and scheduling for dynamic bandwidth control ○ Bandwidth on Demand ○ Developments in lower layer technology ○ Developments in integrated control planes (optical/SDH/SONET/Ethernet/IP) ○ Time frame possible applicability ○ Vision of place of pure optical networking ○ Etc.
Scheduling, mapping, load balancing	<List of issues>

Subject	Description
Security	Addresses the following issues: ○ Certification Authority ○ Global security architecture ○ Incident handling ○ Incident response ○ Intrusion detection ○ Operational security ○ Public Key Infrastructure ○ Grid security event ○ Grid security incident ○ Security infrastructure ○ Security policies ○ Site access control ○ Site usage control ○ VO systems ○ Web services security ○ Etc.
Virtual Organisation(s)	<List of issues>
Workload Management	Addresses the following issues: ○ Computing elements ○ Grid accounting ○ Job provenance ○ Logging and bookkeeping ○ Scheduling policies ○ Etc.
Area of Topic Discussion Interest in eInfrastructure (Networking)	
Advanced Network Monitoring	Addresses the following issues: ○ Bandwidth ○ Delay measurement ○ Delay variations (jitter) ○ Multi-domain Monitoring ○ Packet loss ○ Trend analysis ○ Etc.
Bandwidth on Demand	Addresses the following issues: ○ Dedicated “light paths” ○ Layer 1 wavelengths on fibre (native) ○ Layer 2 channels (native and emulated) ○ MPLS label-switched paths (LSPs) ○ Packet-based QoS ○ Time-division multiplexing (TDM) channels ○ Etc.
End to End Service Provision	Addresses the following issues: ○ Automatic network elements configuration ○ Multi-domain provisioning ○ Etc.
IPV6	Addresses the following issues: ○ Addressing plan ○ Dual-stack network

Subject	Description
	○ IPV4 and IPV6 coexistence ○ Migration policy ○ Network interoperability ○ Network reliability ○ Network stability ○ Routing policy ○ Service continuity ○ Etc.
Layer 2 VPN	Addresses the following issues: ○ Point-to-point connections ○ Tunnelling ○ Multi-Protocol Label Switching (MPLS) ○ Label Switched Paths (LSPs) ○ Resource Reservation Protocol (RSVP) ○ Etc.
Multicast	Addresses the following issues: ○ Access circuits ○ Broadcast ○ Distribution trees ○ Multicast topology ○ Unicast ○ Protocol Independent Multicast-Sparse Mode (PIM-SM) ○ Multi-Protocol Border Gateway Protocol (MBGP) ○ Etc.
Network Security	Addresses the following issues: ○ Anomaly detection ○ Detection and mitigation of denial of service (DoS) attacks ○ Proactive security services ○ Security policies ○ Service security ○ Etc.
Performance Measurement and Monitoring	Addresses the following issues: ○ End-to-end management ○ End-to-end service ○ Monitoring metrics ○ Etc.
Quality of Service	Addresses the following issues: ○ Differentiated Services Model ○ IP Performance Metrics ○ Network congestion ○ Network performance level ○ Service differentiation ○ Etc.
Testbed, technology testing	<List of issues>

Supported Standard Protocols and Formats

The BELIEF DL supports harvesting information from any Information Source in which metadata are represented using a number of different protocols and formats. The following metadata are currently supported by the BELIEF DL harvesting tools:

Metadata encoding protocols:

- DC
- DCQ (recommended)
- MARC, UNIMARC, MARC21, MARCXML

Metadata harvesting protocols:

- OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting (recommended)
- Any API call returning an XML file containing metadata encoded in one of the above mentioned encoding

File formats:

- XML (recommended)
- RSS

The following character encoding is strongly recommended (virtually mandatory):

- UTF-8 (Unicode Transformation Format-8)

BELIEF DL Metadata Definition

In 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 informations between Information Sources adopting different coding with no loss of semantics [DCMI].

However, BELIEF metadata can be modified (see Paragraph 0 “Managing DL Maintenance”).

The following (17, 9 of which qualified) DCQ metadata are currently supported by the BELIEF DL:

DL Metadata list

- | | | |
|-----|-----|---|
| 1) | (*) | Title (<i>qualified</i>) |
| 2) | (*) | Creator |
| 3) | (*) | Subject |
| 4) | (*) | Description (<i>qualified</i>) |
| 5) | (*) | Publisher |
| 6) | - | Contributor (<i>qualified</i>) |
| 7) | (*) | Date (<i>qualified</i>) |
| 8) | (*) | Type (<i>qualified</i>) |
| 9) | - | Status |
| 10) | - | Format |

- | | | |
|-----|-----|--------------------------------------|
| 11) | (*) | Identifier <i>(qualified)</i> |
| 12) | - | Source <i>(qualified)</i> |
| 13) | - | Language |
| 14) | - | Relation <i>(qualified)</i> |
| 15) | - | Coverage <i>(qualified)</i> |
| 16) | - | Rights |
| 17) | - | Provenance |

The 8 (eight) metadata marked with (*) should be considered mandatory for effective classification in the DL.

DL Metadata details

The following definitions are given in [DCMI].

1) Title

A name given to the resource.

Typically, Title will be a name by which the resource is formally known.

- **Qualifiers**

- **Alternative**

Any form of the title used as a substitute or alternative to the formal title of the resource.

This qualifier can include Title abbreviations as well as translations.

2) Creator

An entity primarily responsible for making the content of the resource.

Examples of Creator include a person, an organisation, or a service. Typically, the name of a Creator should be used to indicate the entity.

It is strongly recommended to follow the <Lastname, Firstname> convention (this is required when submitting to the DL via submission form).

The following are examples of BELIEF DL typical values:

- Von Braun, Werner
- BELIEF, Project
- ACME, Partner

3) Subject

A topic of the content of the resource.

Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

4) Description

An account of the content of the resource.

Examples of Description include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

- **Qualifiers**

- **abstract**

- A summary of the content of the resource.

- **fulltext**

- A summary of the content of the resource.

- **tableOfContents**

- A list of subunits of the content of the resource.

5) **Publisher**

An entity responsible for making the resource available.

Examples of Publisher include a person, an organisation, or a service. This should be used to indicate the Information Source (i.e. the name of the Project, Organisation or Initiative which owns the resource).

6) **Contributor**

An entity responsible for making contributions to the content of the resource.

Examples of Contributor include a person, an organisation, or a service. Typically, the name of a Contributor should be used to indicate the entity. In BELIEF DL this element may be used with the following qualifiers:

- **Qualifiers**

- **Coordinator**

- **Organizer**

- **Participants**

- **Organisationtype**

- At present, the BELIEF specific defined possible values are as follows:

- Academia
 - EU Funded Project
 - Industry
 - Research
 - RTD (Research and Technological Development)
 - SME (Small and Medium Enterprise)
 - NGO (Non Governative Organisation)
 - Government

- **Projecttype**

- At present, any text string is allowed.

7) **Date**

A date of an event in the lifecycle of the resource.

Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.

- **Qualifiers**

- **Available**

- Date (often a range) that the resource will become or did become available.

- **Created**
Date of creation of the resource.
- **Issued**
Date of formal issuance (e.g., publication) of the resource.
- **Modified**
Date on which the resource was changed.
- **Valid**
Date (often a range) of validity of a resource.

8) **Type**

The nature or genre of the content of the resource.

Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCM1 Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.

In BELIEF DL this element is used with the set of values provided in [DCT1] plus a new set of new types that may be given as values for this element.

The DCT1 possible values are currently as follows:

- Collection
- Dataset
- Event
- Image
- Interactive resource
- Moving image
- Physical object
- Service
- Software
- Sound
- Still image
- Text

The BELIEF specific defined possible values are currently as follows:

- Article
- Audio
- Authored Book
- Brochure
- Case Study
- Conference & Meeting
- Deliverable
- Dissemination
- eCourse
- Edited Volume
- Handbook
- Manual & User Guide
- Module

- News
- Poster
- Presentation
- Proceedings
- Project Management Document
- Project Management Meeting
- Publication
- Report
- Software
- Summer School
- Technical Document
- Thesis
- Training
- Tutorial
- Video
- Working Note

- **Qualifiers**

- **Informationsource**

The types of the content used in the information source repositories.

The following values are currently accepted in the BELIEF DL:

- Article(s)
- Audio
- Authored Book(s)
- Book(s)
- Brochure(s)
- Case Study
- Code Exemplar
- Conference(s)
- Consortium Agreement(s)
- Contract(s)
- Contribution(s)
- Course(s)
- Deliverable(s)
- Discussion Document
- Dissemination
- Documentation
- Edited Volume(s)
- eCourse(s)
- EGEE Document
- EU Deliverable
- Event(s)
- Events &/and Meetings
- Exercise
- Flyer(s)

- Guideline(s)
- Information Sheet(s)
- Leaflet(s)
- Manual(s)
- Manuals &/and User Guides
- Map(s)
- Meeting(s)
- (Meeting) Minute(s)
- Module(s)
- News
- News Releases
- Newsletter(s)
- Paper
- Phone Conference(s)
- Plan(s)
- Poster(s)
- Presentation-Publication
- Presentation(s)
- Press
- Proceedings
- Project Management Document
- Project Management Meeting
- Publication(s)
- Quality Plan(s)
- Report
- Research Publication(s)
- Series
- Slide Presentation(s)
- Software
- Summer School
- Talk(s)
- Technical Annex(es)
- Technical Document(s)
- Technical Note(s)
- Technical Report(s)
- Thesis
- Training
- Training material
- Tutorial(s)
- User Guide(s)
- Video(s)
- Website(s)
- White Paper(s)
- Working Notes

- Workgroup(s)
- Workshop(s)

9) **Status**

The working status of the content of the resource.

This element is currently used with the following values in the BELIEF DL:

- Approved
- In Work
- Obsolete
- Released
- Under Approval

10) **Format**

The physical or digital manifestation of the resource.

Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).

11) **Identifier**

An unambiguous reference to the resource within a given context.

Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

- **Qualifiers**

- **URL**

The Uniform Resource Locator.

12) **Source**

A Reference to a resource from which the current resource is derived.

The current resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

- **Qualifiers**

- **Activity**

A value from a controlled vocabulary (RTD Innovation, Training, Demonstration, Management, Functional Design, Architectural Design, Implementation, Test, Dissemination, Exploitation, Project Management, etc.).

- **Event**

A value from a controlled vocabulary (Meeting, Review, Conference, etc.).

13) Language

A language of the intellectual content of the resource.

Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian", and "en-GB" for English used in the United Kingdom.

14) Relation

A reference to a related resource.

Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

- **Qualifiers**

- **isPartOf**

- The described resource is a physical or logical part of the referenced resource.

15) Coverage

The extent or scope of the content of the resource.

Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.

- **Qualifiers**

- **Spatial**

- Spatial characteristics of the intellectual content of the resource.

- **Temporal**

- Temporal characteristics of the intellectual content of the resource.

16) Rights

Information about rights held in and over the resource.

Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.

17) Provenance

A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.

The statement may include a description of any changes successive custodians made to the resource.

Collecting Documents and Metadata from Information Sources into Collections

This section summarises document types correspondences between BELIEF DL and the Information Sources and metadata correspondences, similarities and differences with respect to the BELIEF DL Metadata Definition.

Please note that:

- The table in paragraph 0 “Collections and document types cross-reference” summarizes both the mapping between BELIEF DL collections and document types and the mapping between document types, considering all Information Sources as a whole.
- The Information Sources in paragraph 0 “Metadata cross-reference” are the only ones for which a metadata cataloguing format or convention has been formally defined. However, for all Information Sources included in the DL a screening of the used metadata has been performed to ensure coherence with the DL used ones.

Collections and document types cross-reference

BELIEF Collections	BELIEF Document Types	Information Sources Document Types
Articles	Article	Article(s)
Presentations	Audio	Audio
Publications	Authored Book	Authored Book(s), Book(s)
Outreach	Brochure	Brochure(s), Flyers, Information Sheets, Leaflets
Case Studies	Case Study	Case Study
Conferences and Meetings	Conference & Meeting	Conference(s), Event(s), Events &/and Meetings, Meeting(s), Workshops, Workgroups
Deliverables	Deliverable	EU Deliverable, Deliverable(s),
Dissemination and Training	Dissemination	Dissemination
eCourses	eCourse	eCourse(s)
Publications	Edited Volume	Edited Volume(s)
Manuals and User Guides	Handbook	-
Manuals and User Guides	Manual & User Guide	Manuals, Manuals &/and User Guides, User Guides
Presentations	Presentation	Audio, Contribution(s), Presentation(s), Presentation-Publication,

BELIEF Collections	BELIEF Document Types	Information Sources Document Types
		Slide Presentations, Talk(s)
Outreach	Poster	Maps, Poster(s)
News	News	Bulletin(s) News, News Releases, Newsletters, Press
Proceedings	Proceedings	Proceedings
Project Management Documents	Project Management Document	(Meeting) Minute(s), Project Management Document, Quality Plan(s)
Project Management Meetings	Project Management Meeting	Project Management Meeting
Publications	Publication	eInfrastructures Roadmap, Book(s), Opportunities List, Publication(s), Research Publication(s), White Paper(s)
Reports	Report	Report(s)
Software	Software	Code Exemplar, Software
Dissemination and Training	Summer School	Summer School
Technical Documents	Technical Document	EGEE Document, Technical Document(s), Technical Note(s), Technical Report(s)
Publications	Thesis	Thesis
Dissemination and Training	Training	Course(s), EGEE Training Resource, Exercise, Module(s), Series, Training, Website
Dissemination and Training	Tutorial	Tutorial(s)
Presentations	Video	Video(s)
Working Notes	Working Note	Working Note(s)

Table 1 - Collections and document types cross-reference (1/1)

Metadata cross-reference

BELIEF	EGEE	DANTE	DILIGENT	SEE-GRID
1. Title	Title	Title	Title	Name
2. Creator	Creator	Creator	Creator	
3. Subject		Subject	Subject	
4. Description .abstract .fulltext .tableOfContents	Description.abstract	Description	Description.abstract	Description Programme
5. Publisher		Publisher	Publisher	
6. Contributor .coordinator .organizer .participants .organisationtype .projecttype		Contributor	Contributor	Organizer, Contact Participants
7. Date	Date	Date	Date	Date
8. Type .informationsource	Type	Type	Type	
9. Status	status			
10. Format		Format		
11. Identifier .URL	Identifier Identifier.URL	Identifier Identifier.URL		Link
12. Source .Activity .Event				
13. Language		Language		
14. Relation				
15. Coverage .Spatial .Temporal				
16. Rights				
17. Provenance				

Table 2 - Metadata cross-reference (1/2)

BELIEF	CONDOR	eConcertation	Research Infrastructure
1. Title	Title	Title	Title
2. Creator	Creator	Author	Creator
3. Subject	Subject	Subject/Keywords	Subject
4. Description .abstract .fulltext .tableOfContents	Description.abstract	Description Agenda	Description.abstract
5. Publisher	Publisher		Publisher
6. Contributor .Coordinator .Organizer .Participants	Contributor	Coordinator Organizer Attendees/Members	
7. Date .Created .Issued .Modified	Date	Date Creation Date Modification	Date Creation Date Publication
8. Type .informationsource	Type	Type	Type
9. Status			Status
10. Format			
11. Identifier			
12. Source .Activity .Event		Related Workgroup Related Workshop	
13. Language			Language
14. Relation			Event
15. Coverage .Spatial .Temporal		Venue Data	Event Place Event Period
16. Rights			Availability
17. Provenance			

Table 3 - Metadata cross-reference (2/2)

Developed Tools for Document and Metadata Harvesting

At the present time, the automatic harvesting process has been implemented for the following Information Sources:

- EELA
- EGEE
- AIM@SHAPE (via batch submission capability)
- DANTE and related projects:
 - ALICE
 - EUMEDCONNECT

GEANT2

GEANT

TEIN2

- DILIGENT
- ETICS
- EuQoS (via batch submission capability)
- ICEAGE
- iSGTW
- MUPBED (via batch submission capability)
- SEE-GRID
- SEE-GRID-2
- SEEREN2

Among them, a daily automatic harvesting has been scheduled for the following:

- DANTE (and related projects)
- iSGTW
- SEE-GRID
- SEE-GRID-2
- SEEREN2

Documents are collected also from the following repositories, although automatic harvesting has not yet been implemented on these:

- 6DISS
- CONDOR
- DRIVER
- e-IRGSP
- EUChinaGRID
- EUIndiaGRID
- EUMEDGRID
- Grid@Asia
- GridCC

The following Entities perform direct submission of their documents to the DL:

- BioinfoGRID
- eConcertation
- int.eu.grid
- ISSeG
- LOBSTER
- Research Infrastructure Unit, European Commission

Finally, the following repository are natively integrated within the DL, relying on a OpenDLib based repository:

- DELOS
- DILIGENT

The table below shows how the documents and metadata are collected from each Entity, highlighting specific actions to be performed either by individual Entities or BELIEF as appropriate.

Provider	Harvesting protocols and coding formats	Actions to be performed		Tools developed by BELIEF
		On Provider side	On BELIEF side	
6DISS	Proprietary	None	Temporarily manually performed by DL Administrator	Models for submission
AIM@SHAPE	Proprietary, XML	XML file for batch submission	XML file processing	Specific Harvester
BIOINFOGRID	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
CONDOR	OAI, DCQ	None	Temporarily manually performed by DL Administrator	Models for submission
EELA	Proprietary, XML	None	API invocation	Specific Harvester
EGEE Training	OAI, DC	None	WS invocation	Specific Harvester
EGEE Conferences	Proprietary, MARCXML (alternatively OAI, DC)	None	API invocation	Specific Harvester
EGEE Public	Proprietary, XML	None	API invocation	Specific Harvester
DANTE <i>ALICE</i> <i>EUMEDCONNECT</i> <i>GEANT2</i> <i>GEANT</i> <i>TEIN2</i>	Proprietary, XML	None	API invocation	Specific Harvester
DRIVER	Proprietary	None	Temporarily manually performed by DL Administrator	Models for submission
e-IRGSP	None	None	Temporarily manually performed by DL Administrator	Models for submission
ETICS	Proprietary, XML	None	API invocation	Specific Harvester
DELOS	OAI, DCQ	None	Direct access	None
DILIGENT	OAI, DCQ	None	Direct access	None
eConcertation	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
EUChinaGRID	OAI, DCQ	None	Temporarily manually performed by DL Administrator	Models for submission
EUIndiaGRID	OAI, DCQ	None	Temporarily manually performed by DL Administrator	Models for submission
EUMEDGRID	OAI, DCQ	None	Temporarily manually performed by DL Administrator	Models for submission
EuQoS	Proprietary, XML	XML file for batch submission	XML file processing	Specific Harvester

Provider	Harvesting protocols and coding formats	Actions to be performed		Tools developed by BELIEF
		On Provider side	On BELIEF side	
Grid@Asia	Proprietary	None	Temporarily manually performed by DL Administrator	Models for submission
GridCC	Proprietary	None	Temporarily manually performed by DL Administrator	Models for submission
ICEAGE	OAI, DC	None	WS invocation	Specific Harvester
iSGTW	RSS, XML	None	RSS access	Specific Harvester
int.eu.grid	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
ISSEG	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
LOBSTER	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
MUPBED	Proprietary, XML	XML file for batch submission	XML file processing	Specific Harvester
Research Infrastructure Unit	None – Hosting of the DL on BELIEF	None – Direct submission	None – Native support	Models for submission
SEE-GRID	RSS, XML	None	RSS access	Specific Harvester
SEE-GRID-2	OAI, XML, DCQ	Mapping from proprietary to standard protocols	API invocation	Specific Harvester
SEEREN2	RSS, XML	None	RSS access	Specific Harvester

Table 4 - Harvesting modes & tools (1/1)

Document Submission and Search

Models for on-line Document Submission

OpenDLib supports a very flexible document model. This makes it possible to manage a considerable number of different document structures and types, as well as metadata formats in the same DL.

According to this model, a digital object is identified by one or more **editions**. Each edition is split into a set of **views** that are logical partitions of the object. As the name suggests, a view is a “point of view” of the digital object; for example, an object representing the presentation of a paper at a conference could be partitioned into the following views:

- the PowerPoint presentation,
- the video of the talk,
- the audio of the talk,
- a reference to the paper (already submitted in the library or stored elsewhere).

Descriptive metadata can be associated with both editions and views. These metadata are used by the Search Services to perform discover operations on the DL content. The metadata contain information about the edition or view which enables them to be identified and retrieved.

Each view contains its own **manifestations**. A manifestation is a physical representation of the object or one of its parts. The manifestations contained in a view are semantically equivalent (i.e. a PDF or a PostScript version of the same text), and the system can be configured to migrate automatically from one format to another.

The administrator that creates the DL can define a number of predefined domain specific document structures. The DL provides a user-friendly interface enabling authorized authors to submit their documents by choosing one of the predefined structures and then uploading their data or by creating their own documents structure(s) that fulfil individual needs.

The following basic predefined document structures are available for submission in the current release of the BELIEF DL:

- Document
- Event
- Free Model

Each of them requires a set of mandatory metadata and supports also user driven [customisation](#) encompassing the fields described in Paragraph 0 “BELIEF DL Metadata Definition”.

Each document structure (e.g. index, chapters and appendix for a book, or agenda, opening, sessions, closing and report for events, etc.) acts as a guideline for the creation of the document to be submitted, suggesting basic parts of the document to be submitted and allowing customisation as needed by users.

Tool for batch Document Submission

In order to help collaborating projects to submit metadata and documents to the DL in a batch mode, a submission capability has been added.

This new capability has been designed for all projects that cannot supply a programmatic access to their repositories and/or cannot guarantee the support of a librarian or repository manager for on-line submission.

The capability is based on specific harvesting modules - customised for each Information Source - which perform the processing of an XML file containing all documents metadata.

An XML-schema has been produced, to be used as a sample by each project willing to use this functionality.

Search Types

Two basic search types (the Simple and Multiple Search shown below) are available with the first DL release.

- Simple Search (single condition)
 - Field (List of fields available for search)
 - Operator (Contain, Equal, Phrase)
 - String
- Advanced Search (multiple conditions)

A combination of the above with the following operators:

 - Boolean-and
 - Or
 - Andnot
 - Filter-accept
 - Filter-reject

Portal and DL Integration

The BELIEF DL has been integrated with the BELIEF Portal in order to provide a user-friendly interaction and integration between the User, the Portal and the DL, allowing the seamless transition between the outer and the inner level of the system.

Portal and DL integration via APIs

An OpenDLib based system enables interaction with the DL in two different modes: a human-to-application interaction and an application-to-application interaction. The first mode is obtained through the OpenDLib UI Service that comes with the OpenDLib Toolkit. The second mode is implemented through a set of APIs that enable an application to interact with an instance of OpenDLib. The APIs are also provided as part of the UI Service since usually this is the only service exposed to public access. Typically other services are protected by a firewall and they are not directly accessible.

Since the BELIEF DL was created as an instance of OpenDLib, the implementation of the integration between the BELIEF Portal and the DL is based on the use of the APIs. Part of these APIs were available when the project was implemented as part of the OpenDLib Toolkit [OpenDLib]. Nevertheless, in order to cope with new requirements and/or suggestions coming from the project, the following new APIs have been implemented and included in the OpenDLib Toolkit:

- Browse
- BrowseInfo
- BrowseStandard
- GetDocumentsPerTerm
- ListCollections
- Submit

The following APIs are thus now available (see [OpenDLib] for a complete description):

- AdvancedQuery
- Browse
- BrowseInfo
- BrowseStandard
- CreateSession
- CreateUser
- DeleteUser
- DescribeVerb
- DisplayUserInfo
- Disseminate
- GetDocument
- GetDocumentsPerTerm
- Identify
- ListCollections
- ListVerbs

- Load
- Main
- ModifyUser
- Navigate
- SimpleQuery
- Submit

Accessing the DL via the BELIEF Portal

The main functionalities of the DL have been “wrapped” by the Portal to supply user-friendly interaction and seamless integration between the User, the Portal and the DL, enabling the smooth transition between the outer and the inner level of the system.

Basically, administration functionalities have been hidden and a library of functions have been designed and implemented as part of the Portal environment which can be plugged-in with the corresponding DL APIs.

Direct Access to the DL

Since the BELIEF DL is an instance of OpenDLlib, it inherits all of its functionalities and services. Even if these are usually filtered by the Portal, a direct link to the DL is available whereby access is controlled by the access rights defined for each Community’s member.

Functionalities and services of OpenDLlib are briefly summarised hereafter (for more detailed information see Appendix C).

1.1.1. Information space management

Any user can define the information space which s/he wants to search/browse. The information space is composed of all collections the user has selected from those managed by the DL.

Information space selection

Users can choose the collections to be included in their information space by means of the system menu options. Users can view their private collections and/or DL public collections and insert one or more collections into their information space; they can also view their current information space and update it by removing collections or inserting new ones.

Creation of new collections

A collection consists in a set of documents selected from the archives of the DL publishing institutions or from pre-existing collections. The user who wants to create a collection interactively defines i) which conditions the document metadata should meet in order to be selected; ii) which archives documents are to be selected from. Collections are defined as publicly or privately accessible. Public collections can only be created by the library administrator while any registered user can create their private collections.

1.1.2. Search services

The DL provides a set of dedicated advanced services (referred to as “Search Services”) for discovering operations on DL content. These services enable users to access the digital objects maintained in the DL adopted. The user can search any information associated with digital objects and their parts, namely:

- Content representation (metadata)
- References to other objects
- Textual content of digital objects

Users can refine search results through additional search and feed-back operations. The search function enables the user to signal which retrieved documents are relevant to their interests or targets. Accordingly, the DL uses these signalled documents to repeat search operations that produce results that are as close as possible to user relevance feedback.

Search results are presented as a list of references to documents meeting the user search query. References comprise information extracted from one or more metadata elements of the related documents and customisable by the user. These references enable users to obtain document previews so that they can request access specifically to documents of interest, avoiding the need to peruse less relevant documents.

Simple search

The simple search operation enables users to specify one or more words to be searched in all the indexed pieces of information. No operator is allowed among words. The system finds all the documents containing those words and orders the search results according to the word frequency in these documents.

In the current release, the simple search operation enables users to search by a “phrase”, i.e. to specify that the set of words they have entered is a “phrase”. This search type, which usually produces more precise results, is currently made available on the DL as an option of the advanced search operation; however this option is recommended for users with at least a minimum knowledge of indexed metadata formats.

Advanced search

The advanced search operation enables users can specify a complex query in which the search terms, matching operators between search fields and search terms (such as “contains”, “equal to”, etc), and relation operators between simple queries (such as “boolean-and”, “and-not” etc.) may have to be entered.

This search type enables users to set up structured queries but is recommended for users with at least a minimum knowledge of document metadata formats. These queries are very similar to those possible with the “where” clause of a SQL query in a relational database. However, the BELIEF DL offers the additional option to use Boolean operators, or probabilistic operators, or a combination of both.

Browse

Users can browse any information associated with digital objects and their parts, namely:

- Content representation (metadata)

- References to other objects
- Textual content of digital objects

Browsing can also be used as a pre-search operation. It visualises a portion of any indexed information while arranging its items in alphabetical order. Users can select the pieces of information they are interested in (whether personal names, or subjects, titles, publication dates, etc.) and define the portion of the alphabetical order they want to browse. Users can also customise the visualisation format. The ordered list of references is shown by the system accordingly.

1.1.3. Visualisation - User Interface

As result of their search or browse operations, users obtain a set of results pages listing those documents that meet their request. The DL manages digital objects that are compliant with the OpenDLib DoMDL model and visualises them as a graphical rendering of this model. The visualisation of these objects is more complex than the visualisation of simple file. Two visualisation modes are available to users: *tab page* and *top level*. With the tab page visualisation, all object manifestations are displayed in the same window; with the top level mode, each manifestation is displayed in a new window. Users can choose either mode, depending on their preferences and the type of digital objects selected for visualisation

In both modes, a tree representing the structure of the retrieved digital object appears on the left of the first visualisation page. Although object data can be maintained in very different ways (stored locally or remotely, or as a reference to other data), this is transparent to users accessing them in the same way. By navigating the tree and clicking on the object manifestations, users can access the object data.

The DL is capable of managing every type of manifestation, regardless of its type or dimension. Multiple physical manifestations of the same view make it possible to select the manifestation compatible with the software installed on the local machine.

Basic DL Configuration

Supporting Hardware

The following hardware is used to support the BELIEF DL:

- Sun Fire V20z 1U Rack Mount x86 Opteron 250 Server with:
 - 2x 2.4GHz CPUs
 - 4GB Memory (4x1GB DDR1/400 DIMMs)
 - 2x 73GB 10K RPM Ultra320 SCSI disk
 - 2x 10/100/1000 Ethernet ports
 - 1x Full-Height/Full-Length 64 bit/133 MHz PCI-X slot
 - 1x Half-Height/Half-Length 64 bit/66MHz PCI-X slot
 - CD/floppy incl.
 - 1x Internal AC Power Supply

Software Configuration

The following software is used to support the BELIEF DL:

- Basic:
 - Red Hat Enterprise Linux Ver. 3, WS Edition, (for x86 32bit) Workstation/Client Partner with 2 CPU maximum
- Environment:
 - PERL (Ver. 5.8.6)
 - Apache (Ver. 1.3.34)
 - Perl interpreter (Ver. 5.8.0)
 - GCC compiler (Ver. 3.0)
 - Mod_Perl (Ver. 1.29)
 - EmbPerl Framework (Ver. 2.1.0)
 - BerkeleyDB (Ver. 4.2.52)
 - ImageMagick (Ver. 6.2.6)
 - Expat (Ver. 2.0.0)
 - libxml2 (Ver. 2.5.10)
 - libxslt (Ver. 1.0.33)
 - pdftohtml (Ver. 0.36)
 - Xpdf (Ver. 3.01)
- DL specific:
 - OpenDLib (Ver. 1.0)
- User specific:
 - Microsoft Internet Explorer (Ver. 6.0 or higher)
 - Netscape (Ver. 6.0 or higher)
 - Any other Mozilla-based browser

Managing DL Maintenance

Maintenance operations to be performed on - and supported by - the DL are listed hereinafter. Only the main activities to be performed by users and/or administrators are described.

Managing Collections

Creating new Collections

A personal or public collection can be created by any authorised user at any time according to one of these four modes:

- Create from scratch
- Derive from an existing collection
- Grouping existing collections
- Create collections from Repository Sets

Removing existing Collections

A collection can be removed by any authorised user at any time. Documents belonging to that collection remain unaffected.

Managing Metadata Changes

Adding new Metadata

New metadata can be added by any authorised user at any time changing a document model for submission. Please note that such changes on the document model are not permanent. Existing documents submitted via the original model remain unaffected by the temporary changes.

Removing existing Metadata

Non mandatory metadata can be removed by any authorised user at any time changing a document model for submission. Please note that such changes on the document model are not permanent. Existing documents submitted via the original model remain unaffected by the temporary changes.

Managing Document Models

Changing existing Models

As mentioned above, non-permanent changes to existing document models can be performed by any authorised user at any time.

Adding new Models

We suggest either to:

- Define a specific model which copes with your particular needs by using the “Free structure” model. Once you have defined your new model, ask the BELIEF support staff for having such model inserted among the commonly available models.

or to:

- Directly ask the BELIEF support staff for the new model, having even informally drafted a description of the desired structure and contents.

Removing existing Models

Document models can be removed by the BELIEF support staff only.

Adding a Sub Community

BELIEF provides all the functions needed to support a “Sub Community” facility to:

- Create a Sub Community within the BELIEF Community for a particular project.
- Use the BELIEF DL as a document repository accessible only by this Sub Community.

From a User perspective, the following steps must be performed:

1. Users must sign up with the BELIEF Community, in order to become Sub Community members;
2. A responsible for the Sub Community must be identified. This person must contact the BELIEF Project requesting that 'Project X' have its own community repository, communicating the names and document types that Sub Community members intend to submit;
3. Project members submit the documents with the correct metadata according to the submission form(s) available from BELIEF UI;
4. The Sub Community responsible approves the submitted documents using the facility provided by the DL;
5. The Sub Community documents are placed in a reserved area where only Sub Community members can share it;
6. If the Sub Community wants to make documents available to the entire community, they must re-submit them (provided properly authorised). The BELIEF DL Administrator is responsible for authorising this procedure.

From a BELIEF DL Administrator perspective, the following steps must be performed:

1. Proper DL structures creation: a specific Authority, Private Set and Collection for each Sub Community are needed.
2. Proper DL functions authorisation assignment (Submission for all members, Review and Approval for the user(s) responsible for the Sub Community.
3. Definition of document models which must be used by Sub Community members for submission.

References

- [DCMI] The Dublin Core Metadata Initiative Open Forum
<http://www.dublincore.org/>
- [DCT1] DCMI Type Vocabulary. DCMI Recommendation
<http://dublincore.org/documents/dcmi-type-vocabulary/>
- [ISO639] ISO 639-2 - Codes for the representation of names of languages
<http://www.loc.gov/standards/iso639-2/langhome.html>
- [RFC3066] Tags for the Identification of Languages, Internet RFC 3066.
<http://www.ietf.org/rfc/rfc3066.txt>
- [OpenDLib] <http://www.opendlib.com>
- [SAV] <https://gna.org/projects/savane>
- [TGN] Getty Thesaurus of Geographic Names
<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>
- [W3CDTF] Date and Time Formats, WRC Note
<http://www.w3.org/TR/NOTE-datetime>

Appendix A – OpenDLib

Nowadays digital libraries are instruments for supporting communication and collaboration among worldwide distributed user communities. OpenDLib is a digital library management system 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 that can be customised to meet the requirements of a target user community. 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 enables different user groups to define their own virtual view of the shared digital library, tailored to the specific needs and policies of the group.

Architecture

From the architectural point of view, OpenDLib consists of an open federation of services that can be distributed and replicated. This architecture provides a great flexibility in the management of the digital library. For example, an institution can decide to maintain an instance of the repository service in order to have local control over its own documents (), but to share all the other services with other institutions. The architectural configuration is chosen when the digital library is set up, but it can also be changed later to satisfy new emerging needs. For example, a replication of an Index instance can be created to reduce workloads when the number of search requests exceeds an established threshold, whereas an Index instance, able to serve queries in a language not previously supported, can be added to satisfy the needs a new community of users. All these expansions can be done on the fly, i.e. without switching off the digital library. The OpenDLib architecture has been designed to be highly interoperable with other libraries. In particular, an OpenDLib library can act as both an OAI-PHM data and service provider. This implies that the metadata maintained by an OpenDLib digital library can be open to other libraries and, vice-versa, the OpenDLib services can access the metadata published by any other OAI-PHM compliant library.

The document model

OpenDLib can handle a wide variety of document types (the term “document” is used to mean “multimedia digital object”) with different format, media, language and structure. The same OpenDLib library can maintain, for example, a collection of journals and conference proceedings consisting of articles; a collection of theses in different languages, organised in chapters and sections; a collection of videos, structured into sequences and shots, and a collection of other documents represented only by the set of their bibliographic records. OpenDLib can also manage new types of documents that have no physical counterpart, such composite documents consisting of the slides, video and audio recordings of a lecture, seminar, or courses. It can also maintain multiple editions, versions, and manifestations of the same document, each described by one or more metadata records in different formats.

The OpenDLib document model, DoMDL plays the role of the logical document model that is shared by all the services. Documents submitted by the authors or harvested from different sources are logically represented and known to all the OpenDLib services as DoMDL documents.

DoMDL has been designed to represent structured, multieditions and multimedia documents that can be disseminated in multiple manifestation formats. It can be customised according to the DL content to be handled. For example, it can be used to describe a lecture as the composition of the teacher presentation together with the slides and the summary of the talk transcript; this lecture can be disseminated as the MPEG3 format of a video or the SMIL document synchronizing its parts.

In order to be able to represent documents with completely different structures, DoMDL distinguishes four main aspects of document modelling and, using terms and definitions very similar to those coined in the IFLA FRBR model, represents these aspects through the following entities: Digital Object, Edition, View, and Manifestation.

The Digital Object entity is an abstract entity that represents the document as a distinct intellectual creation, capturing the more general aspect of it in very abstract terms. The Edition entity represents a specific expression of the distinct intellectual creation, thus being able to model an instance of the document along the time dimension. A Document can be recognised only through its individual editions or expressions. The boundaries of Edition are defined, however, so as to exclude aspects of physical form, such as text or image formats, that are not integral to the intellectual or artistic realization of the document as such.

The View entity is functional to model the different ways in which a digital information content can be organised, viewed and disseminated. For example, the View entity related to the original edition of the proceedings of a workshop might be modelled in such a way that this edition be perceivable under three different views: a) a “structured textual view” containing a “Preface” created by the conference chairs, and the list of thematic sessions containing the papers presented to the conference, b) a “presentation view”, containing the list of the ppt slides used in the presentations, and c) a “metadata view”, containing a structured description of the proceedings. This is due to how the View entity is specialised, as described in detail below.

Both Edition and View can be associated with their metadata, namely descriptive, structural and administrative metadata.

The different medium in which any View may be physically embodied constitutes a manifestation of that View, thus the Manifestation entity models a physical format by which a document is disseminated. Examples of manifestations are: the MPEG file containing the video recording of the lecture made at a conference, the AVI file of the same video, the postscript file of a lecture given by another teacher at the same school, etc. Physical formats are accessible via URIs, used to associate local or networked file locations.

These entities are semantically connected by means of a set of relationships. The relationships Has edition, Has view, and Has manifestation link the different aspects of a document. Note that these

relationships are multiple, i.e. there can be several objects in the range associated with the same object in the domain. This means that there can be multiple editions of the same document, multiple views of the same edition and multiple manifestations of the same view.

The information space model

A digital library information space consists in the set of information resources that are accessible by the library users. This space usually concerns digital objects that usually are heterogeneous as for structure, formats, semantic content and access rights, and their potential audience also. For example, some documents might only be made accessible by the users of a given institution; other documents only by the users of a given library service; others might be made accessible only through those machines that have been authorised by the digital library. Because of this, the information space is not to be statically defined by the software, nor to be accorded to any pre-established organisational principle; instead it must be customisable by the DL administrator as for its structure and organisation, as well as for how it can be exploited by the users. For example, the users could want to personalise their own working space by selecting only the information resources they are interested in, thus creating the organisational model that meets their needs.

OpenDLib allows its information space to be structured, organised and personalised by means of the collection mechanism. A collection only consists in the set of the documents selected by the user-specific selection criterion. Collections can be structured and their documents copyrighted. Moreover collections, being virtual containers, need no extra-work by the system; therefore they can be created, deleted or reorganised according to the actual digital library's or users' needs even if they are identifying hundreds of thousands of documents.

The collection mechanism makes the information space a virtual one and models it according to which information or functions a user or a community requires. This allows users to become acquainted with the digital library content in a simple manner and to make searching and browsing operations on information tailored to their interests.

Services

The basic release of OpenDLib provides services to support the submission, description, indexing, search, browsing, retrieval, access, preservation and visualisation of documents.

The search service offers different search options: Google-like or fielded (with fields selected from a variety of known metadata formats); single or cross-language; with or without relevance feedback. Documents retrieved can be navigated across all their editions, versions, structure, metadata and documents formats. All the above services can be customised according to several dimensions such as, for example, metadata formats, controlled vocabularies, and browsable fields. OpenDLib also provides other digital library specific functionality, such as the control of access policies on documents, and the management of "user-shelves" able to maintain document versions, result-sets, session results, and other information, etc..

In addition, a number of administration functions are also given to support preservation of documents, document reviewing process, introduction of new collections, and handling of users and user group profiles.

Functionalities offered to the users to exploit the digital library services are introduced in the following.

Information space management

Any user can define the information space which s/he want to search/browse upon. This information space is composed of all collections the user has selected from those managed by OpenDLib. If no information space is defined, user operations are carried out on all “Public collections”.

Information space selection

The user can choose the collections to be included in her/his information space by means of system menu options. The user can view her/his private collections and/or OpenDLib public collections and insert one or more collections into her/his information space; s/he can also view her/his current information space and update it by removing collections or insert new ones.

Creation of new collections

An OpenDLib collection consists in a set of documents selected from the archives of the OpenDLib publishing institutions or from pre-existing collections. The user who wants to create a collection interactively defines i) which conditions the document metadata have to satisfy in order to be selected; ii) which archives documents are to be selected from. Collections are defined as publicly or privately accessible. Public collections can only be created by the library administrator while any registered user can create her/his private collections.

Search services

OpenDLib provides a set of dedicated advanced services (referred to as “Search Services”) for discover operations on DL content. These services include:

- Query Mediator
- Index
- Browse
- and, of course, User Interface

Thanks to these services, a user can access the digital objects maintained in the DL that they are using. The user can search any information associated with digital objects and their parts, namely:

- Content representation (metadata)
- References to other objects
- Textual content of digital objects

The user can refine search results by means of both additional search operations and feed-back operations. The latter function allows the user to signal which of the retrieved documents s/he judges as being relevant to her/his interests. Accordingly, OpenDLib uses the signalled documents to repeat its search operations which will produce results very much influenced by the user's relevance feedback.

Search results are presented as a list of references to the documents satisfying the user search query. References consist of information extracted from one or more metadata elements of the related documents and are customisable by the user. References allow the users to get document previews so that they can request to access a document only if it is of interest. In this way they don't lose any time in looking at less relevant documents.

Search operations and searchable elements will become customisable according to user categories starting from OpenDLib version 2.1.

Simple search

With the simple search operation the user can specify one or more words and let them be searched in all indexed pieces of information. No operator is allowed among words. The system finds all the documents containing those words and orders the search results according to the word frequencies in such documents. In the OpenDLib version 1.3 the simple search operation will allow users to search by a "phrase", i.e. to specify that the set of words they have entered is a "phrase". This search type, that usually produces more precise results, is currently made available by OpenDLib as an option of the advanced search operation; however this option is exploitable only by users with, at least, a minimum knowledge of the indexed metadata formats.

Advanced search

With the advanced search operation the user can specify a complex query in which search terms, matching operators between search fields and search terms (such as "contains", "equal to", etc), and relation operators between simple queries (such as "boolean-and", "and-not" etc.) may have to be entered. This search type requires users have at least a minimum knowledge of documents metadata formats, but allows them to set up structured queries. Such queries are very similar to those permitted by the "where" clause of a SQL query in a relational database, but OpenDLib offers the additional option to use Boolean operators, or probabilistic operators, or a combination of them.

Query Mediator

Dispatches queries to Index service instances, according to availability.

Index

Accepts queries and returns documents matching those queries. The Index function is parametric with respect to the metadata formats, to the set of indexed fields, to the set of result sets formats and the language of the terms.

Browse

The user can browse any information associated with digital objects and their parts, namely:

- Content representation (metadata)
- References to other objects
- Textual content of digital objects

Browsing can also be seen as a pre-search operation. It visualises a portion of any indexed information while arranging its items in alphabetical order. Users can select the pieces of information they are interested in (whether personal names, or subjects, titles, publication dates, etc.) and define the portion of the alphabetical order s/he want to browse; s/he can also customise the visualisation format. The ordered list of references is shown by the system accordingly.

Visualization - User Interface

As result of their search or browse operations, users obtain a set of results pages with the list of documents that satisfy their request. OpenDLib manages digital objects that are compliant with its DoMDL model and visualises them as a graphical rendering of this model. The visualisation of these objects is more complex than the visualisation of simple file. Two visualisation modes are available to the users: tab page and top level. With the tab page visualisation, all object manifestations are displayed in the same window; with the top level mode, each manifestation is displayed in a new window. Users can choose either mode, according to their preferences and to the type of digital objects to be visualised. In both modes, a tree representing the structure of the retrieved digital object appears on the left of the first visualisation page. Although object data can be maintained in very different ways (stored locally or remotely, or as a reference to other data), this is transparent to the users who access them in the same manner. According to the DoMDL specifications, a digital object is composed by a number of views and manifestations. By navigating the tree and clicking on the object manifestations, users can access the object data. OpenDLib is able to manage every type of manifestation, regardless of the nature or dimension. Multiple physical manifestations of the same view make it possible to select the one that is compatible with the software installed on the local machine. This is useful when no specific viewer for a given format (e.g., Microsoft Word) is available, but a suitable one (e.g., Acrobat Reader) is installed.