



## LEAF - Linking and Exploring Authority Files

IST-2000-26323

---

# Model Requirements Analysis Document - 1: Analysis and Evaluation of Input Data

DATE	02/05/2002
AUTHOR	José Borbinha, Nuno Freire
QUALITY ASSURANCE	R. Bull
WORK PACKAGE	5.1
KEYWORDS	LEAF requirements, LEAF system design, user scenarios
DELIVERABLE NUMBER	D 5.1, part 1

### DOCUMENT HISTORY

Version	Date	Comments	Status	Distribution
0.1	02.11.2001	First version	Draft	all project participants
0.2	04.01.2002	Second version	Draft	all project participants
0.3	05.03.2002	Third version including comments to version 0.2 from ONB, SBB and SNL	Draft	all project participants
0.4	02.05.2002	Revised version, consolidated with the other MRAD parts.	Draft	all project participants

1.0

08.05.2002

Final version.

Final

all project participants

## Table of Contents

<b><u>Table of Contents</u></b>	<b>3</b>
<b><u>Table of Figures</u></b>	<b>4</b>
<b><u>1 Introduction</u></b>	<b>5</b>
1.1 <u>Scope</u> .....	5
1.2 <u>Document Overview</u> .....	5
<b><u>2 Generic scenario</u></b>	<b>6</b>
<b><u>3 LEAF Data Providers</u></b>	<b>8</b>
<b><u>4 Generic requirements and model</u></b>	<b>13</b>
4.1 <u>Central concepts</u> .....	13
4.1.1 <u>LAF, LAR and LOID</u> .....	13
4.1.2 <u>Grouping LARs in SLARs with links</u> .....	14
4.1.3 <u>The CNAF - Central Name Authority File</u> .....	15
4.1.4 <u>LEAF Annotations</u> .....	15
4.1.5 <u>LEAF complete static structure</u> .....	17
4.2 <u>Actors and requirements for interoperability</u> .....	17
4.2.1 <u>Internal interoperability</u> .....	19
4.2.2 <u>External interoperability</u> .....	19
4.3 <u>Use cases</u> .....	19
4.4 <u>Requirements for the design</u> .....	21
<b><u>5 LEAF key concepts</u></b>	<b>23</b>
<b><u>6 Acronyms</u></b>	<b>25</b>
<b><u>7 References</u></b>	<b>27</b>

## Table of Figures

<a href="#"><u>Figure 1: Main components in the LEAF general scenario</u></a> .....	6
<a href="#"><u>Figure 2: LAF, the main authority file in LEAF</u></a> .....	14
<a href="#"><u>Figure 3: The SLAF, Shared LEAF Authority File</u></a> .....	14
<a href="#"><u>Figure 4: The CNAF - Central Name Authority File</u></a> .....	15
<a href="#"><u>Figure 5: Annotations in LEAF</u></a> .....	16
<a href="#"><u>Figure 6: LEAF static structure</u></a> .....	17
<a href="#"><u>Figure 7: Users modelling in LEAF</u></a> .....	18
<a href="#"><u>Figure 8: Detailed use cases</u></a> .....	21
<a href="#"><u>Figure 9: The LEAF model</u></a> .....	22

# 1 Introduction

## 1.1 Scope

This document is the first part of the Model Requirements Analysis Document (MRAD-1) prepared for the Project LEAF (Linking and Exploring Authority Files; project number IST-2000-26323).

This document was prepared under Work Package 5.1 (Model Requirements Analysis), Task 5.1.1 (Analysis of input data).

The purpose of this document is to identify requirements to contribute to the identification and definition of model user scenarios and, following the analysis of these requirements, to identify their implications for the LEAF model. The work here presented was based on the analysis of input data provided by the partners (authority file information, data management systems, export formats, interfaces and protocols, usage models), and in a long period of discussion in the consortium. The content of this document is the result of that process, with contributions to the overall system design.

The intended readerships of this document are all project participants and the European Commission.

## 1.2 Document Overview

This document presents part of the results of the first phase of the LEAF project.

The contents here presented are the final result of the identification and analysis of the main requirements for the system, and are based in the discussions held during the first phase of the project. This task was supported by several discussion documents, namely the references presented in the end of this document).

The relevant parts of this document are:

- **Generic scenario:** This section formulates the main assumptions for the generic scenario for LEAF, presented as a main requirement resulted from the earlier discussions.
- **LEAF Data Providers:** This section presents the local systems of the LEAF data providers, and the characteristics of their requirements.
- **Generic requirements:** LEAF intends to promote a vision for a specific kind of service and a model of collaboration. Associated with this vision is a set of key concepts that must be taken into account as requirements for the design and development of the system. These concepts, as also the associated requirements and the use cases are presented in this section.
- **Requirements for the design:** This section presents the requirements for the general system's design resulting from the previous analysis.

## 2 Generic scenario

LEAF aims the development of a service illustrating two main purposes:

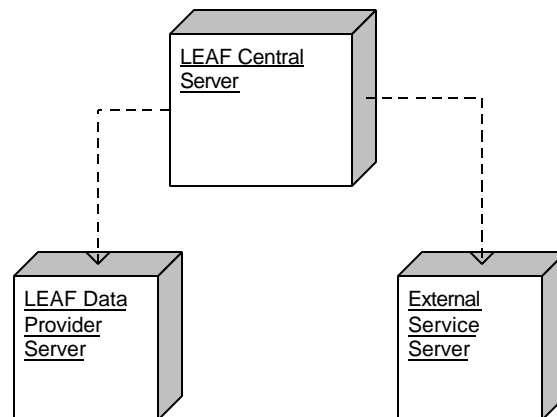
- A model of collaboration for different organizations providing information about persons and corporate bodies to offer in the Internet a common service for the discovery, annotation and exchange of their data;
- Technological approaches for the development of services of that kind.

This service is intended to become useful to a wide range of users, namely (see MRAD-2 for more detailed description of the user groups [6]):

- Professional Users from libraries and archives, using LEAF to create or enhance records
- Professional Users from libraries and archives, using LEAF to provide reference services
- Commercial providers of Services / Products (e.g. manuscript dealers)
- End users looking for information about authorities and links to other services related to them, namely for bibliographic information (such as MALVINE, etc.).

The LEAF model is based on the existence of a LEAF Central Server and a number of local servers (i.e. Data Provider Servers). The local servers contain the authority records of the data providers in their original format. That data will be updated regularly at the central server, where it is processed in order to make it possible to the end-users to search, link and annotate it.

An important purpose of LEAF is to see its model and final service recognized as a reference for the European and, in general, for the international communities dealing with this kind of resources (researchers in history, literature, etc.; libraries; archives; museums; documentation centres; etc.). This is intended to happen due to the quality of the resources, the reputation of the data providers, the simplicity of the functionality and the malleability of the technical solutions.



**Figure 1: Main components in the LEAF general scenario.**

The main components of this service will be (as illustrated by the deployment diagram in Figure 1):

- **LEAF Central Server:** The generic scenario for the LEAF Service assumes the existence of a centralized server with active interfaces for each LEAF Data Provider, represented by remote systems (i.e. LEAF Data Provider Servers). Functionally, the

users will understand the central service as a gateway for distributed data sources, but with centralized functions for optimal interaction and specific user's tasks.

- **LEAF Data Provider Servers:** These are the local servers of the LEAF partners that provide authority data to the service. Those partners are, for this purpose, recognized as LEAF Data Providers.
- **External Service Server:** These are third-party services complementary to LEAF, such as OPAC services or others. A specific example of this class of services is MALVINE.
- **LEAF Data Providers:** These are organizations (libraries, archives, etc.) with collections of biographic records about individuals or corporations that are used in LEAF. The structures of the data collections are heterogeneous, not only from different LEAF Data Providers but also from the same data provider (which can give access, in this sense, to different collections of data). Therefore, each LEAF Data Provider has different requirements, models and techniques for the definition, representation and maintenance of those collections.

The main actors in LEAF will be the Users, the **LMA - LEAF Maintenance Agency** (in charge of the management of the central LEAF Service), and the Local System Administrators. This document will focus mainly on the Users (their modelling and associated requirements).

### 3 LEAF Data Providers

The local scenarios of the LEAF partners that are potential LEAF Data Providers for the final system were analysed in the context of WP4. That data was also compared and reported in detail in [4], in the context of the analysis work for WP4.

Those potential LEAF Data Providers and their systems are resumed in Table 1.

BL	British Library - MOLCAT - Manuscripts Online Catalogue (MLAR) <a href="http://molcat.bl.uk">http://molcat.bl.uk</a>
BN	Biblioteca Nacional – PORBASE (UNIMARC) Central system: WEBPAC (HORIZON), <a href="http://webpac.bn.pt">http://webpac.bn.pt</a> Alternative systems: PACWEB (PORBASE 5): <a href="http://pacweb.bn.pt">http://pacweb.bn.pt</a> SIRIUS: <a href="http://sirius.bn.pt">http://sirius.bn.pt</a>
DLA	Deutsches Literaturarchiv – Kallias <i>Not available online</i>
FDÖP	Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie <i>Not available online</i>
GSA	Goethe- und Schillerarchiv – Briefdatenbank; RetroSig <i>Not available online</i>
IMEC	Institut Mémoire de l'Édition Contemporaine <i>Not available online</i>
NUK	National and University Library, Ljubljana <i>Not available online. Metadata of the electronic catalogue in preparation</i>
ÖNB	Austrian National Library – NAK: Nachlaß- und Autographenkatalog, - Allegro-HANS <a href="http://nak.onb.ac.at/cgi-bin/allegro/nak/nak.pl">http://nak.onb.ac.at/cgi-bin/allegro/nak/nak.pl</a>
RA	Riksarkivet – Nationell Arkivdatabas, National Archival Database of Sweden <a href="http://www.nad.ra.se/NAD.html">http://www.nad.ra.se/NAD.html</a>
SBB	Staatsbibliothek zu Berlin – Kalliope <i>Soon available online</i>
SNL-H	Swiss National Library – Helveticat (USMARC/MARC21) <a href="http://www.vtls.sn.ch/gateway_01/german/vtls.html">http://www.vtls.sn.ch/gateway_01/german/vtls.html</a>
SNL-RHN	Swiss National Library – Repertorium der handschriftlichen Nachlässe <a href="http://www.sn.ch/repertoi/alsrep.htm">http://www.sn.ch/repertoi/alsrep.htm</a>
UCM	Biblioteca de Universidad Complutense de Madrid – CISNE (USMARC/MARC21) <a href="http://cisne.sim.ucm.es/">http://cisne.sim.ucm.es/</a>

**Table 1: LEAF Data Providers and their systems.**

	Personal Names		Corporate Bodies		Total		Future Total
	Number of Records	File Size	Number of Records	File Size	Number of Records	File Size	Number of Records
<b>BL</b>					<b>431000</b>	<b>23</b>	
<b>BN</b>	528566		107480		<b>636046</b>	<b>600</b>	
<b>DLA</b>	109210		30266		<b>139476</b>	<b>75</b>	
<b>FDOEP</b>					<b>7300</b>	<b>4</b>	
<b>NUK</b>							10000
<b>ONB</b>	4274	0.6	763	0.1	<b>5037</b>	<b>0.7</b>	
<b>RA</b>					<b>170000</b>	<b>400</b>	
<b>SBB</b>	236263		21272		<b>257535</b>	<b>300</b>	
<b>SNL-H</b>							1000000
<b>SNL-RHN</b>					<b>7200</b>	<b>500</b>	
<b>UCM</b>	3003	1.4	1323	0.6	<b>4326</b>	<b>2</b>	
<b>TOTAL</b>	<b>881316</b>		<b>161104</b>		<b>1657920</b>		<b>1010000</b>
	$\pm 85\%$		$\pm 15\%$				

**Table 2: Data availability in LEAF in short-term (Total) and long-term (Future Total).**

The data sizes of these systems are reported in Table 2.

These LEAF Data Providers represent a wide range of different cases, with a high level of heterogeneity. They represent also different realities for the maintenance of those data sets, as also different requirements for export procedures, as represented in Table 3 and Table 4.

In general, all partners with their systems actually in production report the possibility to export their data sets in formatted ASCII files. Those partners report average ASCII pure sizes for their registers between 100 and 300 bytes, which represent 500 to 1000 bytes in formatted ASCII files (XML, ISO2709, etc.).

Almost all partners referred also that they can provide all or almost all of their records in the conditions to be agreed in LEAF. The exceptions or other notes were:

- NUK will be able to provide only 100 records in January 2002.
- In order to be able to provide records to LEAF, SNL needs to check previously the LEAF conditions with other departments in the library.
- For SNL-H it is impossible to export records relevant only for MALVINE.
- DLA claims that only  $\pm 50\%$  of their records will make sense for LEAF.
- DLA presented also the requirement of making it technically possible to withdraw any provided record(s) at any time.

	Size of the local authority files, in number of records and number of bytes	Yearly rates of creation of records	Yearly rates of modification of records
BL	431.000 (±23MB )	±8000	±500 <sup>1</sup>
BN	528566 Personal Names 107480 Corporate Bodies 636046 Total (±600MB)	24000 Personal Names 4000 Corporate Bodies 28000 Total	70000 Personal Names 6000 Corporate Bodies 76000 Total <sup>2</sup>
DLA	109210 Personal names 30266 Corporate names 139476 Total (±75MB)	±16000 Personal name ±2000 Corporate name	±4000 Personal name ±600 Corporate name
FDOEP	7300 (±4MB)	Not estimated...	
NUK	Computing system under installation. Actually, 10000 in paper, 50% related with manuscripts		
ONB	4274 Personal names (±600KB) 763 Corporate names (±100KB) 5037 Total (±700KB)	±900 Personal Names ±100 Corporate Bodies ±1000 Total	±500 Personal Names ±30 Corporate Bodies ±530 Total
RA	170000 (±400MB)	Not estimated...	
SBB	236263 Personal names 21272 Corporate names 257535 Total <sup>3</sup> (±300MB)	>1200 person names >1200 corporate names	>2500 person names >1200 corporate names
SNL-H	±1000000	±20000	
SNL-RHN	7200 (±500KB)	±2000 (expected for 2002)	
UCM	3003 Personal names (±1.3MB) 1323 Corporate names (±600KB) 4326 Total (±2MBKB)	±1400 Personal Names ±500 Corporate Bodies ±1900 Total	±2000 Personal Names ±1000 Corporate Bodies ±3000 Total

Table 3: Data description of the Data Providers.

<sup>1</sup> BL is correcting and editing the Name Authority file in a Retrospective Conversion project

<sup>2</sup> BN is undertaking an important extra effort to clean and improve the quality of its authority file. This task will proceed until the middle of 2003.

<sup>3</sup> Both files will considerably increase as the union catalogue is starting to co-operate with more institutions and will integrate some big data files from different providers.

	Preferred exporting data format	Preferred exporting protocol or process to transfer the local authority records
BL	Local MS-Access database. Can be supplied in MDB (common delimited ASCII)	File transfer...
BN	ISO2709, or other required formats, including any documented XML based model.	FTP, OAI (possible), HTTP/HTML-form Z39.50 (depending of LEAF attribute set)
DLA	Kallias format <sup>4</sup> in tokenised line oriented ASCII files. Character encoding is based on ISO-8859-1 plus proprietary enhancements for special characters (8 bit).	FTP upload, maybe Z39.50 in the future...
FDOEP	Line oriented ASCII files in local tokenized export format possible to adapt local format (such as to EAC)	FTP upload
NUK	COMARC/UNIMARC (ISO2709?)	FTP download, maybe Z39.50 in the future
ONB	Offline, line oriented ASCII files in local tokenized format. Character encoding based on DOS Codepage 437 plus proprietary enhancements for special characters (8 bit). Planned export in MAB2 (ISO 5426-1982).	FTP, OAI (possible), HTTP/HTML-form, Z39.50 (depending of LEAF attribute set)
RA	XML or other text formats. Preferred format will be EAC.	SOAP, OAI, FTP
SBB	(Same as DLA). Preferred format will be EAC.	Z39.50
SNL-H	ISO2709. Any other required formats might be possible by the end of 2002	File transfer...
SNL-RHN	Local, non-standard, format in XML/Unicode. Preferred format can be EAC if conversion takes place locally instead of centrally	FTP download, Z39.50 in the future
UCM	Line oriented ASCII files in local tokenized export format, with enhancements for special characters (8 bit)	FTP, Z39.50 (depending of LEAF attribute set)

Table 4: Preferred exporting formats and processes of the Data Providers.

<sup>4</sup> See reference [5] for technical details about Kallias. Kallias records can be exported offline, in line oriented in tokenized ASCII files.

From this data the following conclusions and requirements can be extracted:

- The information maintained by the potential LEAF Data Providers is managed in relatively heterogeneous systems and solutions. In face of the wide range of organisations, this was expected! However, all data providers refer to the possibility to export that data in acceptable solutions, suitable for posterior processing (ASCII tokenized files).
- The compatibility between the different formats has to be confirmed by the analysis to be done in WP4, but a preliminary analysis shows it to be potentially possible (see [4] for more details).
- The possible solutions concerning the transport of the local data sets to external services are resumed in Table 5. These options clearly show a requirement for a flexible and heterogeneous model, thus confirming the discussion registered in [2] and [3].

Solutions	Preferred	Possible	Maybe possible, or conditional
<b>FTP upload</b>	DLA FDOEP	BL BN ONB RA SNL-H UCM	
<b>FTP download</b>	NUK SNL-RHN	BL BN ONB RA SNL-H UCM	
<b>Z30.50 export</b>	SBB		Maybe in the future: DLA, NUK, SNL-RHN Depending of the LEAF profile: BN, ONB, UCM
<b>OAI export</b>		BN RA ONB	
<b>HTTP</b>		BN ONB	
<b>SOAP download</b>		RA	BN

**Table 5: Possible solutions to transport the data in LEAF**

## 4 Generic requirements and model

In this chapter, the LEAF model is represented in the Unified Modelling Language (UML). Please refer to <http://www.omg.org/uml/> for reference.

The main actors in LEAF will be Expert Users, Public Users, Service Providers, the LEAF Maintenance Agency (in charge of the management of the central LEAF Service) and the local System Administrators.

Since LEAF intends to explore a specific perspective of a new kind of service and a model for collaboration, a group of key concepts was developed to fit those purposes. Those concepts are central to the envisaged model, and thus their definition must be perceived as part of the generic requirements of the system.

### 4.1 Central concepts

The definition of the functional model and of the technical solutions to be developed in LEAF must take into account not only the requirements of the present LEAF Data Providers, but also the requirements of further future LEAF Data Providers that the consortium intends to attract to the service.

The actual partners of LEAF represent mainly archives (of different profiles) and libraries. However, the number of LEAF observing partners has increased in an impressive way<sup>5</sup>. At this point in time it can be concluded that the requirement for the supporting of heterogeneity in the data model is mandatory. In other words, the LEAF model needs to be sufficiently generic to answer the needs of institutions and users currently not part of the LEAF consortium.

During the project's lifespan the results will be implemented in the MALVINE Services<sup>6</sup>, extending these into the "Integrated MALVINE / LEAF Services". Thus this said requirement will be tested in a specific context and it therefore needs to be taken into account not only in the final design and implementation, but also in the early definition of the central concepts.

#### 4.1.1 LAF, LAR and LOID

The central concepts that need to be stressed at this moment are fully illustrated in the diagram of classes presented in Figure 6. To describe those concepts we will start with the set of classes in Figure 2.

The first concept defined in this model is the **LAR - LEAF Authority Record**.

A LAR is any single authority record in the LEAF Central Server that was sent or retrieved from a specific local LEAF Data Provider Server. Each LAR is marked with a **LOID - LEAF Organisation Identifier**, the identifier of the organisation that provided the LAR. In this sense, a LAR belongs only to one LEAF Data Provider. A LEAF Data Provider might be able to contribute to LEAF with more than one data set. The LARs will be centrally stored in a data model defined accordingly the EAC format.

The collection of all LARs forms the **LAF - LEAF Authority File**.

---

<sup>5</sup> MRAD-3 [8] presents in detail some characteristics of the LEAF observing partners and their expressed needs. Apart from libraries and archives, museums answered the survey in considerable numbers.

<sup>6</sup> MALVINE: <<http://www.malvine.org>>

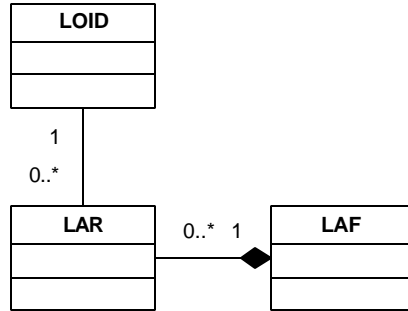


Figure 2: LAF, the main authority file in LEAF

#### 4.1.2 Grouping LARs in SLARs with links

A second important concept is the **SLAR - Shared LEAF Authority Record** (Figure 3).

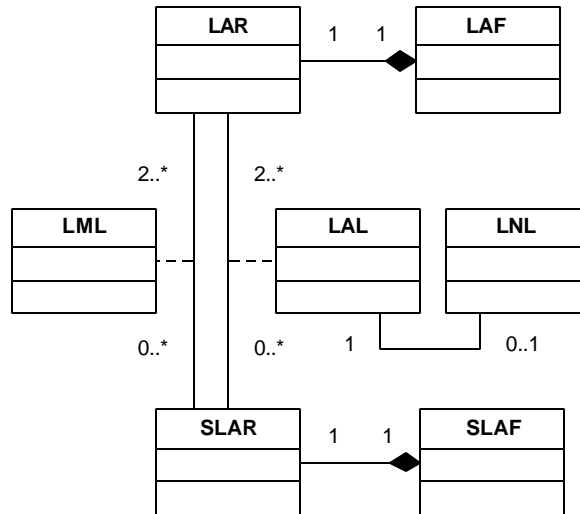


Figure 3: The SLAF, Shared LEAF Authority File

A SLAR is a mechanism by which the LEAF Service intends to aggregate all LARs provided by different LEAF Data Providers but referring potentially to the same entity (person or corporate body).

One SLAR is composed by any group of two or more LARs that were linked as a result of a linkage action. That action can be performed manually by a registered User, who creates a **LML - LEAF Manual Link**. Alternatively, that linkage can be performed automatically by the system, creating a **LAL - LEAF Automatic Link**. A LAL is established automatically by the linking system according to a specific set of linkage rules. Those rules will be based on parameters to be defined by the consortium within WP 5.2, to endeavour to ensure with a high probability that the records in question do cover the same specific entity. The relevance and feasibility of these heuristic approaches will be tested within the project. If a LAL proves to be wrong, it can be overruled manually by a **LNL - LEAF Negative Link**. A LNL can be created by a registered User.

Finally, all SLARs in the LEAF Central Server make the **SLAF - Shared LEAF Authority File**.

### 4.1.3 The CNAF - Central Name Authority File

A purpose of LEAF is not only to bring together the authority files of LEAF Data Providers, but also to promote their excellence.

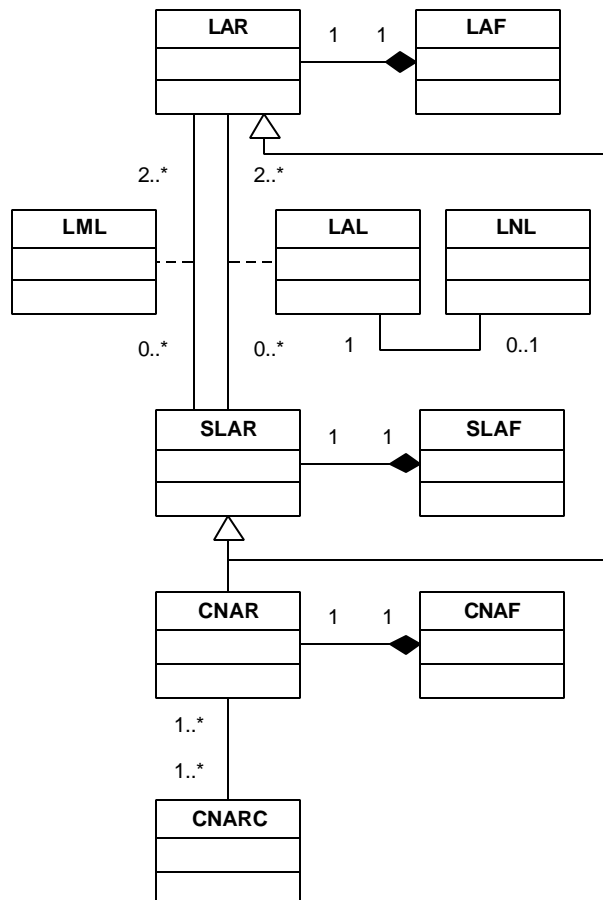


Figure 4: The CNAF - Central Name Authority File.

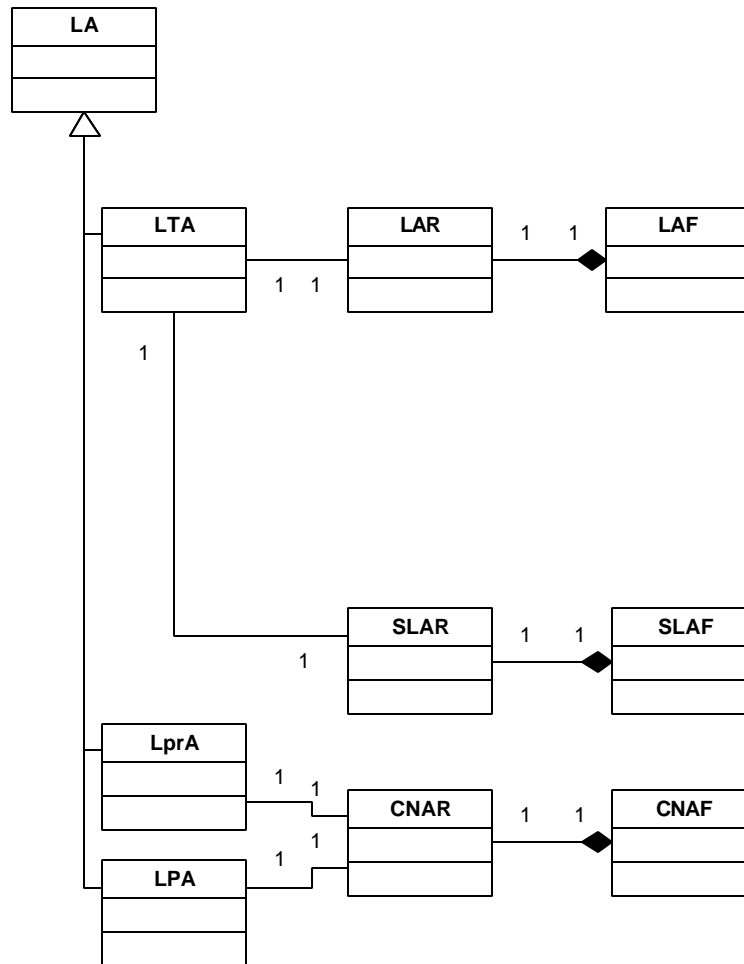
An important concept for that purpose is the **CNAF - Central Name Authority File**, which is composed by all records identified as CNARs. A **CNAR - Central Name Authority Record** is a LAR or a SLAR that was marked with a **CNARC - CNAR Condition** (Figure 4).

A CNARC will be typically a search query submitted by a User. If that query returns one or more records (LARs or SLARs), and if in that list there are results recognised as correct by the User, then those records can be marked by that query and become CNARs.

In this sense the CNAF will represent the records that have been looked for by the users. Therefore, these will be understood as the most relevant records in LEAF. This can be relevant for the LEAF Data Providers to incite them to pay a special attention to those records, to invest in the improvement of their quality, or just to confirm their contents.

### 4.1.4 LEAF Annotations

The last important concept in the generic model is the generic concept of **LA - LEAF Annotation**. In LEAF we have three classes of annotations (Figure 5).



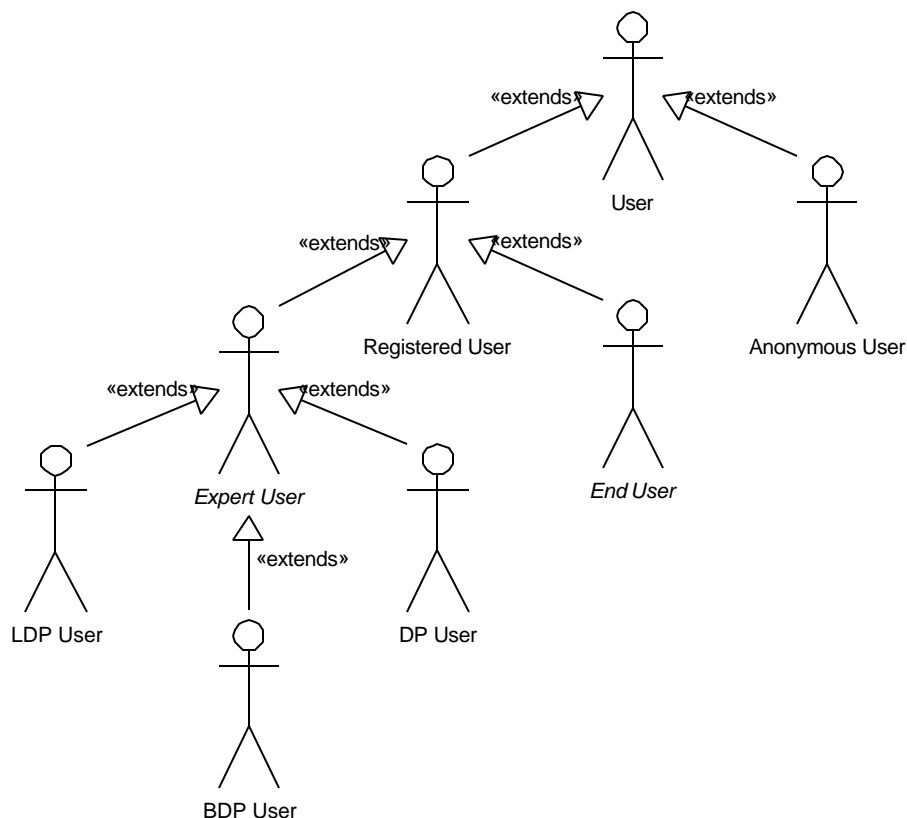
**Figure 5: Annotations in LEAF.**

An **LTA - LEAF Temporary Annotation** is an annotation to a LAR or SLAR created by a registered User with the intention to communicate with the LEAF Data Providers (it can mean a note about a possible error, an intention to provide complementary information, etc.). The data providers can be alerted and verify those annotations, and as a result they can confirm or deny their relevance.

Other class of annotation is the **LPA - LEAF Persistent Annotation**. Those are intended to provide complementary information about a CNAR, and are the responsibility of a specific User.

Finally, there is the **LprA - LEAF Private Annotation**, which is a private annotation that a User can make to a CNAR.





**Figure 7: Users modelling in LEAF.**

Users can submit queries to search and retrieve records from the biographic collections, and have also access to other complementary services. In that sense, LEAF provides services for two main classes of Users: anonymous users and experts. Different user groups may be distinguished by the different rights allocated to them.

**Anonymous Users** who have not registered with LEAF, are normal Users who can interact with the system with a minimal set of functions (typically, to search the system).

**Registered Users** will have more rights than Anonymous Users, namely they will be able to post annotations.

An **Expert User** is a Registered User who can have even a higher level of interaction with the system, namely to submit annotations and links. Expert Users may be:

- Staff members of LEAF Data Providers;
- Staff members of organisations which do not provide their data to LEAF but which use LEAF much in the same way as LDPs for the creation or enhancement of records. Those organisations are identified by LEAF as potential future LDPs, therefore they are defined as **DP - Data Providers**. Clearly members of LDPs will have more rights than members of DPs - see MRAD-2 for more detailed description;
- Staff members of organisations that provide a bibliographic system/database connected to the LEAF system, therefore they are defined as **BDP – Bibliographic Data Providers**. These users use LEAF much in the same way as DPs and LDPs.

Each Registered User (End User or Expert User) will be identified in the LEAF System through a specific **LUID - LEAF User Identifier**.

These scenarios raise requirements for interoperability. In LEAF this problem of interoperability has two dimensions: internal interoperability and external interoperability.

### 4.2.1 Internal interoperability

Internal interoperability is related to the requirement of bringing together the different data models, supported by the different LEAF Data Providers, in a coherent model suitable to support the upper layer of services. Those requirements are:

- The possibility to process the highest possible number of different representation formats (coming from the LEAF Data Providers) in the highest possible number of different presentation formats.
- To enable any Data Provider (LDP or DP) to retrieve, convert and copy records from other LEAF Data Providers.
- This retrieval must be possible in syntactical and coding formats specified by the Data Provider (LDP or DP) copying the records, independent of the formats of the originating LEAF Data Provider. The coding and syntactical transformations must be performed by the LEAF Central System, which will try to offer the highest possible number of semantic cross-walking transformations between the different syntactic (UNIMARC, MARC21, MAB2, EAC, etc.) and coding (ISO 2709, XML, etc.) formats proposed by all LEAF Data Providers. To bring together the different data models, supported by the different LEAF Data Providers, the LEAF model uses EAC as the interoperable communication format for authority records. The usage of EAC is derived from the requirement for a common format for authority records to support the LEAF Services and especially these requirements. A LEAF Data Provider must be able to specify the conversion of its own format to and from all other formats by specifying the conversion of its format to and from EAC. This way LEAF Data Providers will not be required to have extensive knowledge of all the formats used by other data providers, instead they only need knowledge of their own format and EAC. However, and because conversions between formats may result in a loss of information or in semantic misinterpretations, the specific usage of EAC in LEAF will require that it will always be possible to present a record in its original format.

### 4.2.2 External interoperability

One scenario of usage is to use LEAF to find authoritative forms for reference to be used in other complementary external systems. External interoperability is related to this purpose of making it possible to jump to those systems from the central LEAF System.

One relevant external system for LEAF is MALVINE. A desired scenario will be to enable a user to start a query for bibliographic information in MALVINE using as arguments the results of a previous query in LEAF.

Another example of the external interoperability is a user going directly to a local system to look for possibly updated information about a specific record found in LEAF. If that action is possible with that Data Provider Service, this means that such Data Provider Service can also be an External Service.

Other possible similar scenarios include the search in other bibliographic services, namely OPAC systems.

## 4.3 Use cases

Figure 8 details the generic LEAF use case (an example of an External Service can be MALVINE). This case stresses the following requirements:

- **Authentication:** All types of Users will have to be able to register into the service (which can be a formal process submitted for approval, for example, according to a specific exploitation model).
  - **Registration:** Registered users receive a unique LUID - LEAF User Identifier.
  - **Login:** Registered Users can provide their identification, becoming known by the LEAF System.
- **Interact:** Users can interact with the system to perform a set of actions, namely (note that not all user groups will have equal access to all search spaces -for details see chapter 4 of MRAD-2 [6]):
  - **Search Records:** Users can search for records using names of authorities and restrict their searches to specific spaces such as the LAF, CNAF, SLAF, or a specific collection of a specific LEAF Data Provider.
  - **Retrieve Records:** Users can retrieve records found as a result of a query. The results of that retrieval will be presented in a set of syntactical and coding options, which will be as complete as possible (such as, for example, free text, UNIMARC coded in ISO 2709, UNIMARC coded in XML, EAC coded in XML, etc.).
  - **Search Annotations:** Various types of Users can search the annotations made by other Registered Users in records in the LAF, SLAF or CNAF.
- **Annotate:** Registered Users can **Create**, **Modify** or **Delete** annotations to selected records in their working space. The annotations created in one working space will be available to all the other users.
- **Check Annotations:** LEAF Data Providers can learn about annotations made by other users to their records. This is important for collaboration with the other users, and to support their tasks of quality assurance for their records.

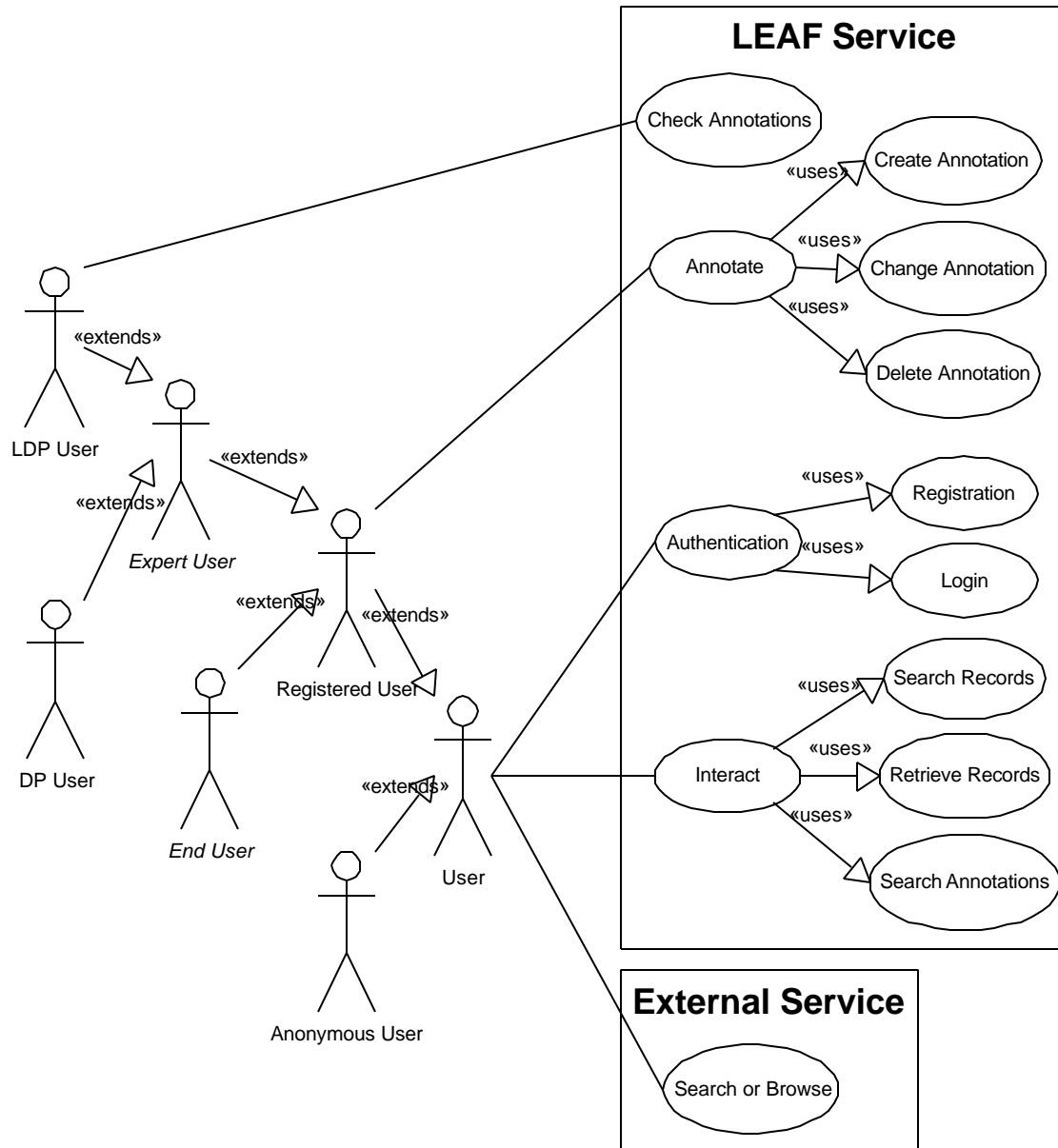


Figure 8: Detailed use cases

#### 4.4 Requirements for the design

Figure 9 presents the design requirements to be taken into consideration for the design of the final system.

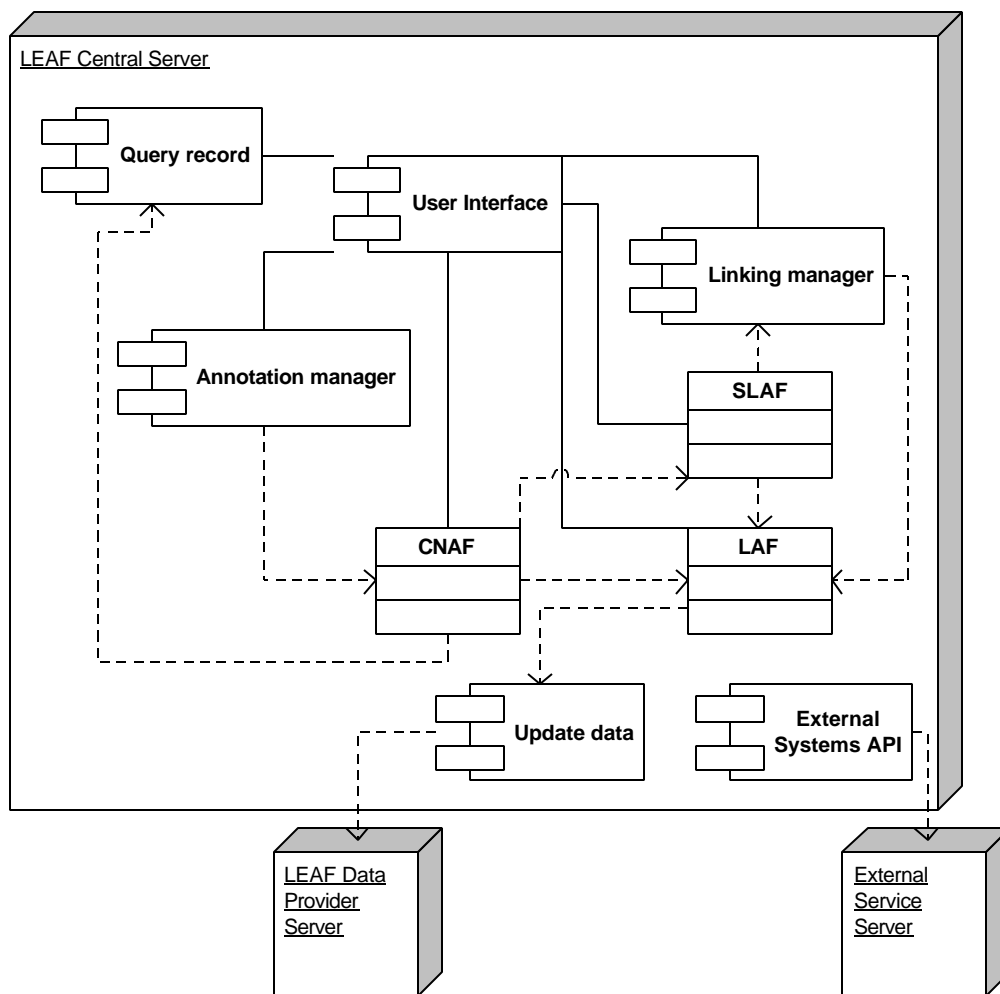


Figure 9: The LEAF model.

According to this model, the Central Server must comprise the following components:

- **User Interface:** This component handles the interaction between the Users and the LEAF Service. Possible indexing sub-components are supposed to be included here.
- **Linking Manager:** This component provides the mechanisms to maintain links between authority records. Links may be established by Registered Users, or may be created automatically by special functions of the Linking Manager.
- **Annotation Manager:** This component provides a way for Registered Users to annotate records.
- **Update data:** This component is responsible for keeping the LAF updated, by harvesting or receiving the authority records from the servers of the LEAF Data Provider. Records are received in local formats, and converted to the internal format used in LEAF, the EAC – Encoding Archival Context

## 5 LEAF key concepts

The formal definitions of the key concepts of the LEAF complete static structure presented in the Figure 6, as well as the definitions of the other important complementary concepts, are:

- **Anonymous User** is a user that can interact with the system with a minimal set of functions (typically, searching).
- **BDP – Bibliographic Data Provider** is an organisation that provides external bibliographic services connected to LEAF. These organizations also use LEAF much in the same way as DP and LDP for the creation or enhancement of records.
- **BDP User** is a registered user of a BDP.
- **DP - Data Provider** are organisations which do not provide their data to LEAF but which use LEAF much in the same way as and LDP for the creation or enhancement of records.
- **DP User** is a registered user of a DP.
- **External Service** is a service complementary to LEAF, namely bibliographic services such as MALVINE.
- **Expert User** is a Registered User who can have a higher level of interaction with the system, namely to submit annotations and links. Expert Users may either be staff members of LDPs or DPs..
- **LA - LEAF Annotation**, is an annotation that Registered Users can add to LEAF records. A LA can be of three classes: LPA - LEAF Persistent Annotations; LTA - LEAF Temporary Annotations; LprA - LEAF Private Annotations.
- **LAF - LEAF Authority File** is a generic authority file in the LEAF Central Server with the collection of all the LARs - LEAF Authority Records.
- **LAL - LEAF Automatic Link** is a link between two or more records created automatically by the LEAF system and according to the Linkage Rules. These records become thus a SLAR - Shared LEAF Authority Record.
- **LAR - LEAF Authority Record** is any single authority record in the LEAF Central Server that was sent or retrieved from a specific LDP - LEAF Data Provider.
- **LDP - LEAF Data Provider** is the entity providing data for the LAF – LEAF Authority File. Staff members of LDPs form a special class of Expert Users who have access to high level functions related to the authority records.
- **LDP User** is a registered user of a LDP.
- **LMA - LEAF Maintenance Agency** is the entity in charge of the management of the central LEAF Service.
- **LML - LEAF Manual Link** is a link between two or more records created by a LDP or the LMA.. These records become by this mean a SLAR - Shared LEAF Authority Record.
- **LNL - LEAF Negative Link** is a link created by LDP to overrule a LAL - LEAF Automatic Link. This mechanism makes it possible to ignore a LAL that is automatically created by the system because of the Linkage Rules.
- **LOID - LEAF Organisation Identifier** is the identifier of a LDP, DP, SP or BDP.
- **LPA - LEAF Persistent Annotations** are annotations intended to provide complementary information about a CNAR. An LPA is created by a specific LDP or the LMA.

- **LprA - LEAF Private Annotations** are private annotations that any Registered User can make to a CNAR - Central Name Authority Record.
- **LTA - LEAF Temporary Annotations** are annotations created by any Registered User to a LAR - LEAF Authority Record or a SLAR - Shared LEAF Authority Record, with the intention to communicate with the LDP - LEAF Data Provider.
- **LUID - LEAF User Identifier** is the unique identifier of each any Registered User in the system.
- **Registered User** is a user authenticated by the system that can have a higher level of interaction with the system, namely to submit annotations or links. Every Registered User is identified in the system through a LUID - LEAF User Identifier.
- **SLAF - Shared LEAF Authority File** is the collection of all SLARs - Shared LEAF Authority Records in the LEAF Central Server.
- **SLAR - Shared LEAF Authority Record** is a mechanism by which the LEAF Service intends to aggregate all LARs - LEAF Authority Records provided by different LEAF Data Providers but referring potentially to the same entity (person or corporate body).

## 6 Acronyms

<b>BDP</b>	Bibliographic Data Provider
<b>CNAF</b>	Central Name Authority File
<b>CNAR</b>	Central Name Authority Record
<b>CNARC</b>	CNAR Condition
<b>DP</b>	Data Provider
<b>EAC</b>	Encoded Archival Context
<b>EDU</b>	Expert Data Users
<b>LA</b>	LEAF Annotation
<b>LAF</b>	LEAF Authority File
<b>LAL</b>	LEAF Automatic Link
<b>LAR</b>	LEAF Authority Record
<b>LDP</b>	LEAF Data Provider
<b>LEAF</b>	Linking and Exploring Authority Files
<b>LMA</b>	LEAF Maintenance Agency
<b>LML</b>	LEAF Manual Link
<b>LNL</b>	LEAF Negative Link
<b>LOID</b>	LEAF Organisation Identifier
<b>LPA</b>	LEAF Persistent Annotation
<b>LprA</b>	LEAF Private Annotation
<b>LTA</b>	LEAF Temporary Annotation
<b>LUID</b>	LEAF User Identifier
<b>MALVINE</b>	Manuscripts And Letters Via Integrated Networks In Europe
<b>OAI</b>	Open Archives Initiative
<b>SLAF</b>	Shared LEAF Authority File
<b>SLAR</b>	Shared LEAF Authority Record
<b>UNIMARC</b>	Universal Machine Readable Cataloguing

### LEAF Partner's Acronyms:

<b>BL</b>	British Library
<b>BN</b>	Biblioteca Nacional
<b>CNS</b>	Crossnet Systems Ltd.
<b>DLA</b>	Deutsches Literaturarchiv
<b>FDÖP</b>	Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie
<b>GSA</b>	Goethe- und Schiller-Archiv
<b>IMEC</b>	Institut Mémoire de l'Édition Contemporaine
<b>JRS</b>	Joanneum Research
<b>NUK</b>	National and University Library, Ljubljana, Slovenia

**ÖNB** Österreichische Nationalbibliothek  
**RA** Rijksarkivet  
**SBB** Staatsbibliothek zu Berlin  
**SNL** Swiss National Library  
**UCM** Biblioteca de Universidad Complutense de Madrid  
**UoB** University of Bergen

## 7 References

- [1] **Definitions, scenarios, use cases and analysis of requirements for LEAF: An analytical discussion.** José Borbinha (BN), Nuno Freire (BN); August 8, 2001.
- [2] **Contribution for the Model Requirements Analysis (Part 1).** Borbinha (BN), Nuno Freire (BN); September 25, 2001
- [3] **Minutes of the LEAF Working Group Meeting in Caen, France, 27.-28.09.2001.** Ulrike Beermann (SBB); October 16, 2001.
- [4] **LEAF Metadata Tables: Person Names and Corporate Bodies.** Ulrike Beermann (SBB), Max Kaiser (OeNB); November 15, 2001.
- [5] **KALLIAS Token Export Format: Technical Description.** Heinz Werner Kramski. December 7, 2001.
- [6] **Model Requirements Analysis Document - 2: User Groups Statements and Special User Scenarios.** Hans-Jörg Lieder (SBB), Max Kaiser (OeNB); January 7, 2002.
- [7] **Synopsis of WG1.** Hans-Jörg Lieder (SBB); November 15, 2001.
- [8] **Model Requirements Analysis Document - 3: User survey.** Hans-Jörg Lieder (SBB), Max Kaiser (OeNB); December 21, 2001.