

LEAF Newsletter

In this issue:

- [Editorial](#)
- [General Project Information: Factsheet](#)
- [Observing Partners in LEAF](#)
- [LEAF in a Nutshell](#)
- [LEAF User Survey](#)
- [Encoded Archival Context \(EAC\) - Recent Developments](#)



Linking and Exploring Authority Files

Editorial

Welcome to the first issue of the LEAF Newsletter. It will not be published at fixed dates but whenever the project consortium wants to inform the public about project proceedings and results.

Besides giving some general information about the project, this first issue illustrates the activities and achievements that were made during the first year. In that context we will also present to you the main results of the LEAF User Survey that was running from late July until mid September 2001 in which many of you may have participated.

In the context of the work recently done on the system architecture Per-Gunnar Ottosson, senior archivist of the LEAF project partner Riksarkivet in Stockholm and member of the EAC working group elaborates on the emerging format EAC (Encoded Archival Context) which will play an important role within the future model of LEAF.

In this Newsletter we will also introduce to you our observing partners network.

Thank you for your interest in LEAF. This first Newsletter hopefully will deliver a closer insight into the project. Comments or questions are very welcome, please send them to:
co-ordinator@sbb.spk-berlin.de.

In order to keep up-to-date in between Newsletter issues, please make sure to visit the regularly updated LEAF website at www.leaf-eu.org.

Ulrike Beermann, Berlin State Library

General project information: Factsheet

- Objectives: LEAF will develop a model architecture for a distributed search system harvesting existing name authority files (person names and corporate bodies). When a user searches for a name string, LEAF will search the records of all LEAF Data Providers and combine these records to one single LEAF authority record. This record will automatically be stored in a "Central Name Authority File" which will thus contain name information of high quality and high user relevance, as it will only contain records that were actually searched for.
- Benefits: Registered users will be able to annotate records and thus enhance the name information. Commercial agents like manuscript dealers may add customised offers. Institutions without electronic data may indicate that they have relevant information about a specific person. Having the LEAF record as a starting point, the user will also have the possibility to view and download this record in a variety of formats, to view the local records the LEAF record was built of, and to search for documents related to the person or corporation. This latter facility will be demonstrated by integrating LEAF into the MALVINE service (Manuscripts and Letters via Integrated Networks in Europe www.malvine.org).
- Expected results: The main result to be expected is that access to authority files will no longer be reserved to professional users from large libraries and archives, but be available to any interested person. Furthermore, the professional use of authority files will also be enhanced, as the workflow between the participating institutions will be established, authority records can easily be exchanged and improved and small institutions will be able to provide their information, too.
- Project partners: The LEAF consortium consists of 15 partners - libraries, archives, documentation and research centers - in 10 European countries. Besides, a network of observing and sponsoring partners was set up which presently consists of 32 partners. The coordinator of the project is the Staatsbibliothek zu Berlin (Berlin State Library).
- Duration: LEAF is a three year project that started in March 2001 and will end in February 2004.
- Funding: LEAF is co-funded through the [Information Society Technologies](#) (IST) Programme within the European Union's [Fifth Framework](#).

- Contact:

Jutta Weber
Department of Manuscripts
Head of the German Union Catalogue of Modern Manuscripts and Letters
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
Potsdamer Str. 33, D - 10785 Berlin
Email: jutta.weber@sbb.spk-berlin.de
Tel.: +49-30-266-2416
Fax: +49-30-266-3007

Hans-Jörg Lieder
Department of Manuscripts
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
Potsdamer Str. 33, D - 10785 Berlin
Email: hans-joerg.lieder@sbb.spk-berlin.de
Tel.: +49-30-266-2249
Fax: +49-30-266-2842

Observing Partners in LEAF

Ulrike Beermann, Berlin State Library

A lot of work has already been and is presently being done in the sector of authority files. In order to avoid any overlap of efforts the LEAF consortium has set up a network of observing partners presently consisting of 32 institutions from various European countries, including Eastern Europe, and from the USA and Israel. This network also includes three sponsoring partners who will support the project by providing their data as test data.

Observing partners do not get any funding money and therefore do not have any responsibilities or obligations. They receive all project documents and information and are invited to communicate via a mailing list particularly set up for them, serving as their central communication channel to communicate with the LEAF consortium and among each other. Observing partners do also have access to the LEAF observers mailing list archive. On the project document store a special section for the observers has recently been established, containing all public project deliverables and further project publications in various languages. Furthermore, sponsoring and observing partners can be invited to all LEAF task meetings.

Thus, LEAF will not only consider authority files from its full partners but will also benefit from current projects in the field of authority files via the observing partners and their activities. This cooperation aims at enhancing the standards of quality in LEAF. Through the network of observing partners it is to be hoped that especially during the test phase (December 2002 to May 2003) LEAF will also benefit from experts of those countries that are not represented in LEAF by full or associated partners.

An up-to-date list of our sponsoring and observing partners can be viewed at www.leaf-eu.org (Partners page).

LEAF in a Nutshell

Hans-Jörg Lieder, Berlin State Library

Much has been said and written in the last few years about the use of authority records in cultural heritage organisations (libraries, archives and museums). The shortcomings of the current situation in Europe marked the starting point of the LEAF project.

Most European countries have authority files, the obvious benefits of which, for cataloguing purposes, have been widely acknowledged. The fact that authorities – in this present context we are exclusively dealing with authorities for person names and corporate bodies – can be linked to bibliographic data records points at a crucial benefit for librarians/archivists. Once the identity of any person/corporate body has been established by means of an authority record, cataloguers can link bibliographic records to it with consistency and accuracy. It is obvious that this certainty is also beneficial for search and retrieval operations conducted by users.

There are, however, some clear limitations in the present use of authority information. One major problem is the limited access to such data. Generally only large institutions respectively networks of institutions have unlimited online access to permanently updated authorities. Smaller institutions either have to fall back on less efficient ways of obtaining authority information or have no access at all. Additionally, access to authorities is mainly restricted to professional users, typically staff working in institutions which use these authorities. Public users are, by and large, excluded from access. Another crucial problem is the fact that authorities, if at all, are usually shared on a national level only. There is presently no pan-European or international framework available which would ensure shared access. Last but not least: existing authorities are frequently of rather poor quality and in need of improvement.

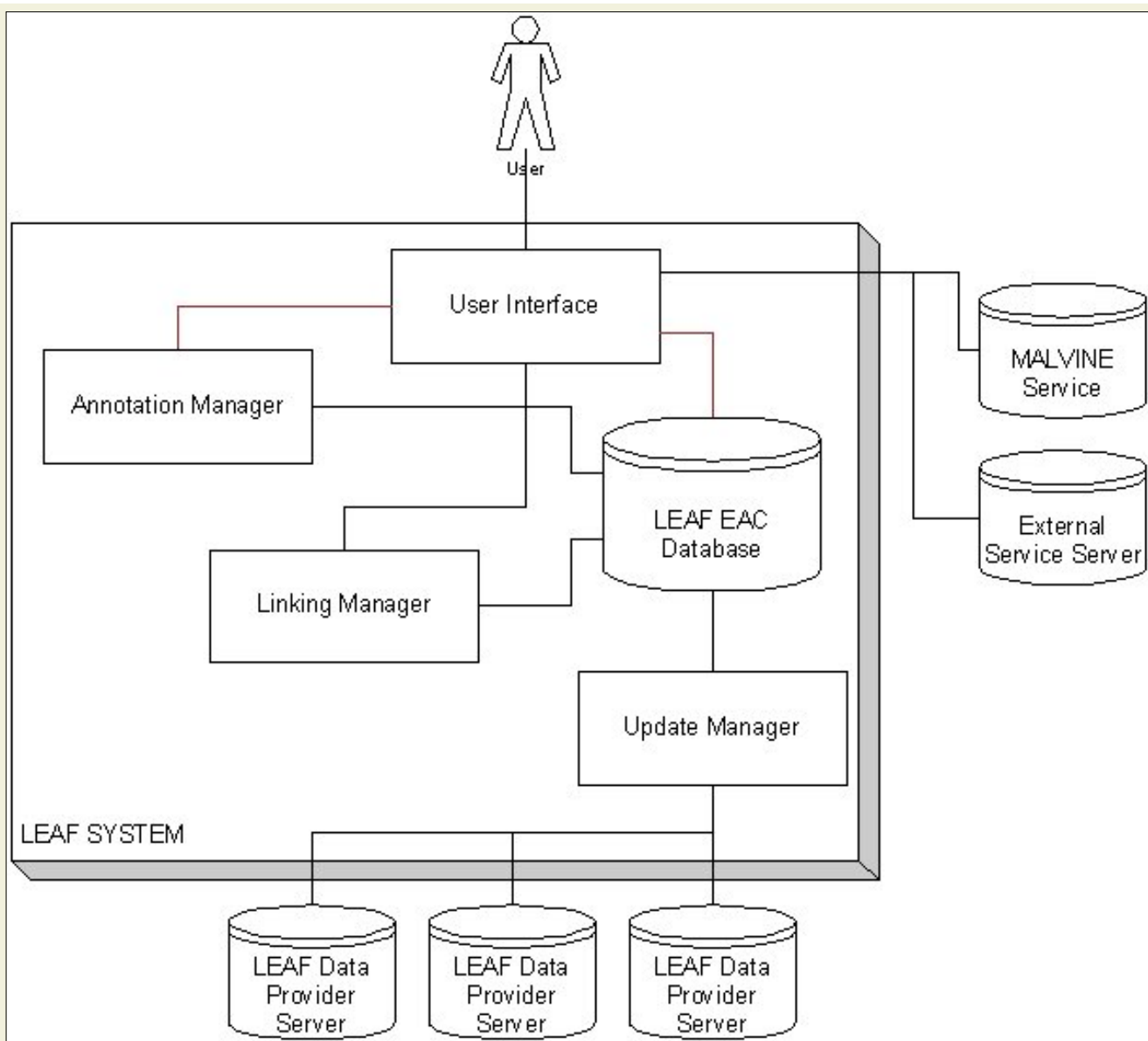
The LEAF project tries to overcome these present shortcomings by striving for these three major objectives:

1. provide shared access to authority information for all involved (cataloguers, reference librarians, end users etc.),
2. improve the quality of existing authorities,
3. improve search and retrieval functionalities of a variety of applications.

The methods/steps chosen to reach these objectives are:

1. Upload distributed authorities to a central system.
Local authority data will be uploaded from the local servers of the participating organisations to the central LEAF system where it is stored in the currently emerging EAC (Encoded Archival Context) format. Regular updates of the uploaded data will ensure that data in the central LEAF system is as up-to-date as possible.
2. Link authorities which refer to the same entity.
With the help of automated linking rules defined within the project those authority records which refer to the same entity are linked together. Of course, it will be possible to check these automatically created links and overrule them manually if necessary. Whenever a user queries the LEAF system using a name string as search argument, this name string will thus represent an entity - or, in LEAF's jargon, various local authority records representing the same entity will be aggregated to form a "Shared Name Authority Record". It is crucial to note that local name authority records will not be merged into one definitive "corporate" record, but grouped or linked in recognition that, despite whatever local differences might exist, they refer to the same entity. In this way, maintaining local authority traditions (which has many practical advantages) may be seen to be compatible with a desire for greater accuracy and consistency for the end user.
3. Annotate authorities with a view on improving content and providing additional information.
All registered users of the LEAF system will be able to post annotations to particular data records in the LEAF system. This functionality is mainly geared towards the improvement of local authority records and is expected to require some negotiation between the annotating user and the owner of the data record in question. LEAF will provide a framework in which such negotiation processes can be easily carried out. Further to this it will be possible to attach additional information to a specific data record, e.g. small institutions without an electronic data offer of their own can thus inform users of LEAF that manuscripts related to a specific entity can be found in that particular institution. Furthermore manuscript dealers can indicate that manuscripts of a particular person are on sale etc.
4. Support external services.
Existing Internet applications could, in many cases, clearly benefit from the integration of authority information. Since names represent the most common access point to bibliographic databases and networks, online retrieval will be greatly improved by the linking of authority name records to bibliographic records. To demonstrate this, LEAF will be integrated into the existing MALVINE Service (www.malvine.org).
5. Save search results in a pan-European "Central Name Authority File".
Information which is retrieved as a result of a query submitted to LEAF will be stored in a pan-European "Central Name Authority File". Since every new query will generate a new record to be saved, this "Central Name Authority File" will grow with each query and at the same time will reflect precisely which data records the users of LEAF were interested in. Libraries and archives wanting to improve authority information will thus be able to prioritise their editing work.

The following graphic of the simplified LEAF System Architecture illustrates the main components of the LEAF system:



The Update Manager will transfer local data from the LEAF Data Provider Servers to the central LEAF EAC Database where the records will be stored in the EAC format. The Linking Manager will process the automatic linking of records referring to the same entity, the Annotation Manager will deal with the processing and administration of annotations to records that were posted by users of LEAF. Through a User Interface all types of users, including providers of external services, will be able to interact with the central LEAF database.

An example of a fictitious search operation may illustrate the main functionalities:

A user searches for "Smith, John". The local LEAF Data Provider Servers contain a number of authority records referring to "Smith, John". Via the Linking Manager these records are grouped in a way that may look like this:

Smith, John (1542-1598)
 Smith, John (1634-1703)
 Smith, John (1712-1788)
 Smith, John (fl. 1912)

Each entry in this short list will be expandable to display an aggregated "Central Name Authority Record" which will be centrally stored and will look similar to this example:

CNAR of Smith, John (1634-1703)

Name: Smith, John

See Reference: Smith, John Michel

Dates: 1634-1703

Profession: ...

...

...

[Search for documents
of this autor](#) [Download this
record](#)

View this record in:

[EAC](#)
[UNIMARC Authorities](#)
[MARC21 Authorities](#)
[Authorities Format X](#)
[Authorities Format Y](#)
...

This record was created with information from the following records:

[Record 1](#) (Staatsbibliothek zu Berlin)

[Record 2](#) (Biblioteca Nacional)

[Record 3](#) (British Library)

Users annotations on this record:

[Post a new
annotation](#)

<no annotations have been posted>

LEAF User Survey

Max Kaiser, Austrian National Library

Since "user orientation" is one of the key phrases of the LEAF project much effort was spent in designing and carrying out a user survey. Two main objectives were tied to this survey: on the one hand the project needed to understand better the actual status quo regarding the use of authorities (persons and corporate bodies only) in libraries, archives and museums. On the other hand we wanted to know from potential future users what kind of functionalities they might expect from a future LEAF system.

A detailed report on the survey results will soon be available at the LEAF Website (www.leaf-eu.org Publications page).

For practical reasons the survey was carried out online, with the questionnaires being available at the LEAF Website. The survey was launched on July 31st and was online until September 14th 2001. It was completed by 568 users worldwide, a number which indicates the interest with which the LEAF project meets.

A majority of survey participants come from German-speaking countries (Germany 26%, Austria 13%), a slightly smaller group from Anglo-Saxon countries (UK 12%, USA 11%). Many other countries – including a number of non-European countries – were represented with much smaller numbers of participants. Since four different questionnaires were conceived for different LEAF user groups, survey participants were firstly asked to describe themselves either as "Librarian",

"Archivist", "Museum Specialist" (all these used one questionnaire for "professional users"), "End User", "Commercial Provider of Services", or, indeed, as "None of the Above, but still interested". Since the mentioned types of institutions – Library, Archive, Museum – engulf a variety of activities that may differ considerably in scope it was decided to additionally ask for specific professional functions. The majority of all participants – 54% - labelled themselves as 'librarian' when given a limited choice only; another 17% labelled themselves as archivists. Notably, 15% of all participants were "End Users".

Status Quo Analysis Questions regarding the present status quo yielded results which can be summarised thus:

- A vast majority of institutions actually do use authorities. Percentage figures range according to user types from 79% (libraries using person name authorities) to 58% (archives using corporate body authorities).
- The most commonly used authority files are:
 - The widespread use of the Library of Congress-Name Authorities (LoC-NA) came as no surprise (41% of all organisations use it for person names, 37% for corporate bodies). It is most frequently used in the USA (more than 80 %), but also consulted in a number of other countries.
 - The Personennamendatei (PND) is used by 42% of the participants, mostly in German-speaking countries but also in a number of other countries.
 - The Gemeinsame Körperschaftsdatei (GKD) is almost exclusively used in German-speaking countries (42%).
 - The Union List of Artists Names (ULAN) is only used by 7% of participating institutions, these mainly being museums.
 - Additionally a large number of local or smaller national authority files are being used (see the detailed report for exact figures).
- Most frequently used cataloguing standards are:
 - Anglo-American Cataloguing Rules (AACR2), used by 23% of all survey participants.
 - Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB), used by 23% of all participants.
 - General International Standard Archival Description (ISAD(G)), used by 42% of all archivists.
 - International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF)), used by 27% of all archivists.
 - Encoded Archival Description (EAD), used by 25% of all archivists.
 - In the museum sector several standards are almost equally represented (RAK-WB, ISAD(G), EAD, Dublin Core, SPECTRUM, CDWA, CIDOC are each used by 18% of the museum specialists whereas the CIMI Standard and the AMICO Data Specification are only used by 9% of all museum specialists.
 - But: more than one third of all librarians and archivists - and even a larger number of museum specialists are using "other" standards.
- Access to national authority files is only in about half of the cases provided through the Internet, less practical means of access have to suffice for the other half: regularly updated CD-ROMs or, indeed, "other" means of access. A considerable part of professionals is only connected to local, not to national authority files.
- Roughly a quarter of all survey participants do not import or manually copy authority records into their local databases, but link their local bibliographic resources to an authority information pool, which can be considered to be "best practice". Note though, that this figure gives no indication as to what kind of authority pool the records are linked to. In fact, more than half of those who link resources do so with a local authority pool and almost 10% stated they use "another link" to authority records, suggesting that a variety of 'non-standard' procedures are in

use. Another quarter of all participants currently transfer (parts of) authority records manually from an external source into the local database. Again, a small part of these participants use local authority files only. A little less than a quarter of all participants declared that they import the required authority files from an external system. Such external systems are mainly pools of national name authority records but also local ones. Lastly, slightly less than 20% of the participants stated that their name authority records are both locally created and used, and 5% are provided with different, unspecified kinds of access.

- In the librarian and archival sector the most commonly used data formats for the export or exchange of data are those of the MARC family (around 45%) and of MAB2 (around 20%). The most frequent single format in use by archivists is the EAD standard (around 30%). Only within the museum community are "other" standards used on a significant scale (around 50%).
- Only a third of all local systems make use of the Z39.50 data exchange protocol.

Summarising the above it is obvious that the present status quo leaves ample room for improvement: the use of large, i.e. national authority files needs to be encouraged, a framework for shared, international access to a variety of authority records is clearly a desideratum, the use of standards still needs to be promoted – to name but a few lacking possibilities.

Desired Features of the LEAF System

Desired features that were most frequently named by survey participants include:

- Facilities to download search results (93%).
- Conversion tool for the transformation of search results into other formats (83%).
- Facilities to attach annotations to data records (78%).
- Online work space (requested by a clear majority of end users, not so much by professional users).

When asked for the desired general results of the LEAF project, it is interesting that most participants have a broad picture in mind and are not exclusively interested in improving their own authority records. By far the highest percentage – more than 90% – considers a "model for a new way to collaborate with different institutions using an always available, on-line Name Authority File" to be important or even very important. These figures drop sharply when asking whether the model should be applicable for a single institution (more than 58%) or a single department (more than 46%). A majority of over 60% also wishes for "a solution for a specific application", e.g. MALVINE. General benefits like optimisation of workflow and improvement of records, again, were wished for by a significant majority.

Following the results of the LEAF user survey the LEAF project consortium has formulated a number of additional system requirements all of which were identified in the relevant project documentation.

Encoded Archival Context (EAC) - Recent Developments

Per-Gunnar Ottosson, Riksarkivet, Stockholm

EAD ([Encoded Archival Description](#)) is an SGML/XML format that is about to become a de facto standard for communication of archival data. EAD has elements for names of corporate bodies and persons with attributes allowing for links to authority files. There are also elements for the narrative administrative histories and biographies, as well as elements for controlled access in terms of functions and geographic names. However, EAD does not provide support for separate files of authority and context information.

In response to this need, an international group of archivists and information scientists met in Toronto in March 2001 to lay down the principles for governing such an encoding standard. The group prepared for the meeting by drafting and reviewing a set of principles and criteria to direct its work, and agreed that the standard needs to address more than traditional authority control of headings and that accompanying documentation is needed for contextual information.

The name of the format became the Encoded Archival Context, thereby stressing its wider scope:

- Archival context information consists of information describing the circumstances under which records (defined broadly here to include personal papers and records of organisations) have been created and used. This context includes the identification and characteristics of the persons, organisations, and families who have been the creators, users, or subjects of records, as well as the relationships amongst them.

For the development of the DTD, a special working group was assigned consisting of Daniel Pitti (University of Virginia), Joanne Evens (University of Melbourne), Stephan Yearl (Yale University), and, from LEAF, Gunnar Karlsen (University of Bergen) and P-G Ottosson (National Archives of Sweden). During a meeting in Charlottesville in June, the group came up with a draft DTD, which was ready for circulation to the full group in the middle of July. The DTD has been successfully tested on LEAF data by Gunnar Karlsen.

Relations to other authority standards

The EAC DTD is adopted to librarian standards for authority records, such as UNIMARC/Authorities. Especially when it came to the elements of the header and the entry elements it was regarded as crucial to keep a compatibility with MARC records. A special attribute (ea= encoding analog) documents the relation between an EAC element and the MARC field of the source. The Committee for Description Standards of the International Council of Archives is now reviewing the ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families . Some of the members of the committee took part in the development of EAC, and it is proposed that the new version of ISAAR(CPF) shall accommodate the structure of EAC.

Overview of structure

- The header of the EAC record, containing elements for maintenance history, and declarations of languages, rules, and source.
- The identity area, which contains elements necessary for identifying the person, corporate body or family, such as names and additions to names.
- EAC relations: elements for linking and explaining the relations between EAC records.
- Resources relations: links to resources, such as the archival descriptions, catalogue records, or web pages.
- Links to controlled vocabulary and description of the functions or activities of the person or corporate body.
- A systematic description of the entity and its environment.
- A biography or administrative history in the form of an essay or a chronological list.
- The rescue for all legacy data not fitting into the EAC structure: other context description.

The functionality of EAC will be thoroughly tested within the LEAF project. LEAF will thereby be able to contribute to a development of the format that is applicable in practice and responds to user needs.