LEAF. Linking and Exploring Authority Files (www.leaf-eu.org)

Jutta Weber
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz

The project LEAF is funded by the EU under the 5th framework (IST program). It started in March 2001 and will run for three years.

The scope of LEAF reflects the new role which authority data get in libraries, archives and museums: Their internal use in the cataloguing process, the authority control which is needed when authority data originating from different sources are used in a distributed retrieval process and the access control which in fact is dependent from the quality of authority data: All this can be summed up under the question: How can millions of existing authority data originating from very different sources be used together by everybody (librarian, archivist, museologist, public user) in a way that there will be no loss of information, no data accumulation without quality check but an automated linking between high quality information driven by actual user needs?

LEAF will try to enhance search and retrieval facilities by providing high quality access to authority information for everybody. For this purpose LEAF is developing a model architecture for collecting, harvesting, linking and providing access to existing name authority information, independent from their creation in libraries, archives or museums and independent from national differences. The scenario will be build using authority files about person names and is integrating the user directly into the establishing of a Central European Name Authority File.

A LEAF demonstrator will be built, it will be integrated into the MALVINE search engine.

The LEAF consortium consists of :
- The Coordinator: Staatsbibliothek zu Berlin, Berlin,
- The Manager: Crossnet Systems Ltd., Newbury ,
- The System developer: JOANNEUM RESEARCH, Graz,

and Biblioteca Nacional de Portugal, Lisbon; Biblioteca de Universidad Complutense, Madrid; British Library, London; Deutsches Literaturarchiv, Marbach; Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie, Graz; Goethe- und Schiller-Archiv, Weimar; Institut Mémoires de l'édition contemporaine, Paris; Österreichische Nationalbibliothek, Vienna; University of Bergen, Bergen; Swiss National Library, Bern; National and University Library, Ljubljana and Riksarkivet, Stockholm.

There is a big number of observing partners which do not get any funding but which want to accompany the project as institution working in the same field and being interested in sharing also preliminary results and giving advice when needed. There are: Archives de France pour les technologies de l'information et de la communication, Paris; Det Arnamagnæanske Institut, Copenhagen; Biblioteca Nacional de España, Madrid; Biblioteca Nazionale Centrale di Roma (National Central Library of Rome), Rome; Jagiellonian Library, Krakow; Warsaw Public Library, Warsaw; Torun University Library, Torun; Warsaw University Library, Warsaw; Bibliothèque Nationale de France, Paris; CIMI consortium, Chicago; Constantijn Huygens Instituut foor Tekstedities en Intellectuele Geschiedenis, 's-Gravenhage; Dansk Biblioteks Center, Copenhagen; Die Deutsche Bibliothek, Frankfurt am Main; Dokimas Group Holdings Ltd., Nottingham; EKT, (National Documentation Centre), Athens; Franz-Michael-Felder-Archiv der Vorarlberger Landesbibliothek, Bregenz; GTAA, Grupo de Trabajo de Autoridades de Aragón, Zaragoza; Interparty, Boston Spa, Wetherby; Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche, (ICCU), Rome; Jewish National & University Library, Jerusalem; Det Kongelige Bibliotek, Copenhagen; Koninklijke Bibliotheek, The Hague; Library of Congress, Washington; Magyar Tudományos Akadémia, Budapest; Onderzoekssteunpunt en Databank Intermedaire

Structuren in Vlaanderen; Online Computer Library center (OCLC), Dublin/Ohio; Research Libraries Group (RLG), Mountain View/California; Tartu University, Tartu; University of Virginia, Charlottesville/Virginia; Wiener Stadt- und Landesbibliothek, Vienna and .....

There are also sponsoring partners which agreed to provide test data when needed:
K.G.Saur Verlag, Munich; J.A. Stargardt, manuscript dealer, Berlin; Library of Congress, Washington.

MALVINE (www.malvine.org) is a search engine harvesting data bases providing information about letters written by famous persons and which are kept in very different institutions in Europe. Due to the lack of better information provided by the participating institutions only names, not individuals can be searched for. Scholars working in the sector of hand written resources of our cultural heritage and searching in MALVINE are happy with the results provided as far as the names they are searching for are used only by one single person. When ever two or more persons are using the same name scholars want to distinguish them. This term of distinction may consist of a different data of birth, profession etc., with other words all that information that national authority files try to provide. The problem is that small institutions normally don't have access to these national authority files. So it will not be surprising that only very few MALVINE participants are able to provide this sort of authority information:

The LEAF model will be based on a very simple idea: Data from different providers are stored in a central server, this server will be uploaded regularly or a harvesting mechanism will add new data every time a user is doing a search in LEAF. All data will be linked automatically, the quality of these links will only be verified when a search is submitted. This verification may be done by the user her/himself or by the data provider who shall automatically be informed about an annotation. The following features are planned:[1]

"1.Upload distributed authorities to a central system.
Local authority records (LAR) 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 as Local Authority Files (**LAF**). 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" (SLAR) which themselves will form the "Shared Name Authority File" (**SLAF**). 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.

---

[1] The following text is cited and in some points updated from the LEAF newsletter, "LEAF in a nutshell" by Hans-Jörg Lieder

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" (**CNAF**). 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 diagram of the simplified LEAF System Architecture illustrates the main components of the LEAF system:
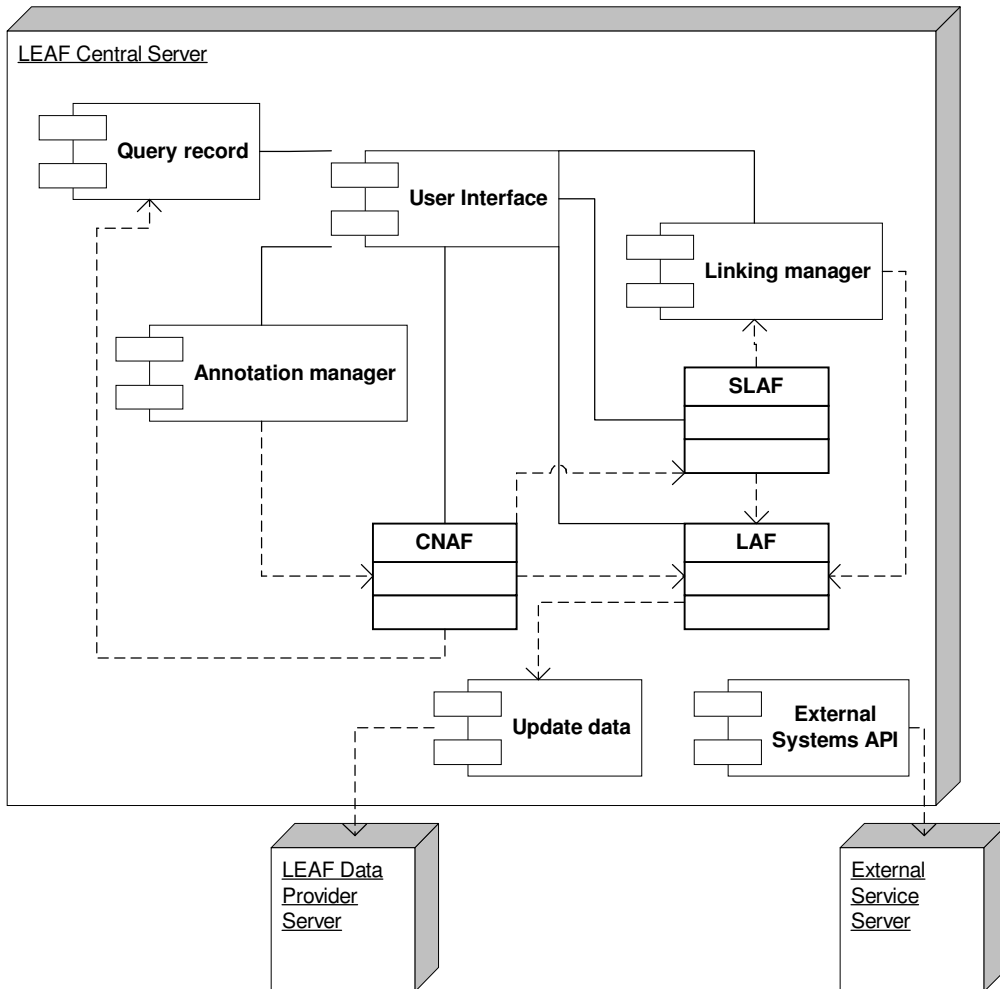

The Update Manager will transfer local data from the LEAF Data Provider Servers to the central LEAF EAC Database where the records will be stored in the EAC format. The Linking Manager will process the automatic linking of records referring to the same entity, the Annotation Manager will deal with the 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:
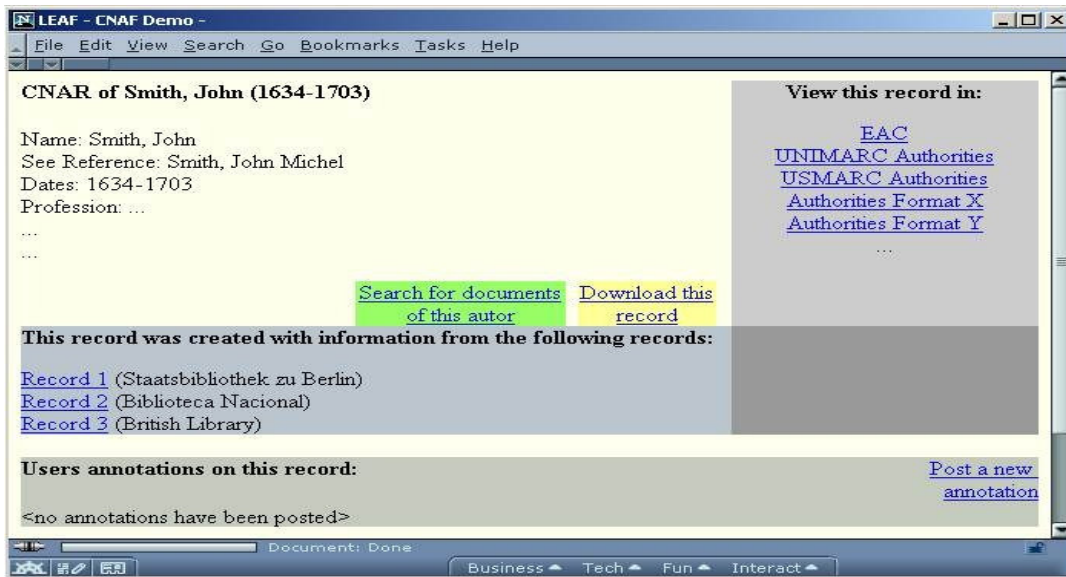
The technical implications will be multiple, the LEAF demonstrator will be developed upon this draft within the next months. The following diagram shows the main features:
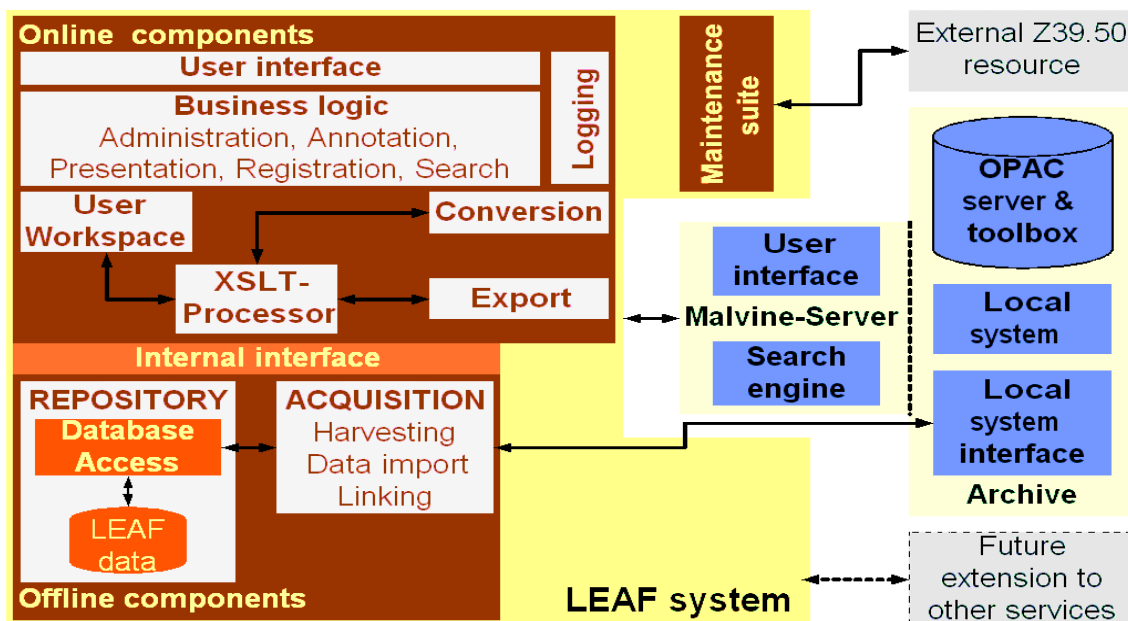
Please visit www.leaf-eu.org, where within short more detailed information about the demonstrator and the LEAF testbed will be available.