



LEAF - Linking and Exploring Authority Files

IST-2000-26323

Model Application Requirements Document

DATE	22.03.2003
AUTHORS	José Borbinha (BN), Max Kaiser (ÖNB), Per-Gunnar Ottosson (RA)
QUALITY ASSURANCE	R. Bull
WORK PACKAGE	5.2
KEYWORDS	Status Quo Analysis, Partnerships, Analysis of Practice
DELIVERABLE NUMBER	D 5.2

DOCUMENT HISTORY

Version	Date	Comments	Status	Distribution
0,1	22.01.2003		Draft version	LEAF consortium EC Reviewers
1,0	21.03.2003	QA read	Version 1.0	LEAF consortium

Table of Contents

Table of Contents	2
1 Introduction	4
1.1 Scope	4
1.2 Document Overview	4
2 MARD-1: Status Quo Analysis	4
2.1 Analysis projects	4
2.1.1 AUTHOR: Transnational Application of National Name Authority Files	4
2.1.1.1 Management details	4
2.1.1.2 Summary of Objectives	4
2.1.1.3 Outcome / Results	4
2.1.2 DELOS/NSF Working Group on "Actors in Digital Libraries"	4
2.1.2.1 Management details	4
2.1.2.2 Summary of Objectives	4
2.2 Activities, technology, standards and frameworks	4
2.2.1 Hong Kong Chinese Authority (Name) Workgroup (HKCAN).....	4
2.2.1.1 Management details	4
2.2.1.2 Summary of Objectives	4
2.2.1.3 Proceedings	4
2.2.1.4 Outcome / Results	4
2.2.2 HKUST XML Name Access Control Metadata Repository	4
2.2.2.1 Management details	4
2.2.2.2 Summary of Objectives	4
2.2.2.3 Outcome /Results	4
2.2.3 Authority Control among Chinese, Korean and Japanese Languages Project	4
2.2.3.1 Management details	4
2.2.3.2 Summary of Objectives	4
2.2.3.3 Outcome / Results	4
2.2.4 MACS - Multilingual Access to Subjects	4
2.2.4.1 Management details	4
2.2.4.2 Summary of Objectives	4
2.2.4.3 Outcome / Results	4
2.2.5 vCard.....	4
2.2.5.1 Management details	4
2.2.5.2 Summary of Objectives	4
2.2.6 eduPerson Object Class	4
2.2.6.1 Management details	4
2.2.6.2 Summary of Objectives	4
2.2.6.3 Relevant references	4
2.2.7 OAI - Open Archives Initiative.....	4
2.2.7.1 Management details	4
2.2.7.2 Summary of Objectives	4
2.2.7.3 Relevant references	4
2.2.8 Z39.50 / Bath Profile Maintenance Agency	4
2.2.8.1 Management details	4
2.2.8.2 Summary of Objectives	4
2.2.9 DC-Agents	4
2.2.9.1 Management det ails	4
2.2.9.2 Summary of Objectives	4
2.2.9.3 Outcome / Results	4

2.2.10	OASIS Extensible Resource Identifier (XRI) Technical Committee	4
2.2.10.1	Summary of Objectives	4
2.2.10.2	Relevant references	4
2.3	Other resources	4
2.3.1	ULAN - Union List of Artist Names	4
2.3.1.1	Management details	4
2.3.1.2	Summary of Objectives	4
2.3.1.3	Outcome / Results	4
2.3.2	EUAN - European Union Archive Network	4
2.3.2.1	Management details	4
2.3.2.2	Summary of Objectives	4
2.3.2.3	Relevant references	4
2.3.3	Authority Control in the 21st Century: An Invitational Conference.....	4
2.3.3.1	Management details	4
2.3.3.2	Summary of Objectives	4

3 MARD 2: Identification of Partnerships – Context and Scope of Relevant Projects **4**

3.1	MALVINE - Manuscripts and Letters via Integrated Networks in Europe	4
3.1.1	Management details	4
3.1.2	Summary of Objectives	4
3.1.3	Outcome / Results	4
3.2	IFLA Working Group on Functional Requirements And Numbering of Authority Records (FRANAR)	4
3.2.1	Management details	4
3.2.2	Summary of Objectives	4
3.2.3	Proceedings	4
3.2.4	Outcome / Results	4
3.3	MARBI Multilingual Record Task Force	4
3.3.1	Management details	4
3.3.2	Summary of Objectives	4
3.3.3	Outcome / Results	4
3.4	Authority Control in the Online Environment Interest Group (LITA/ALCTS CCS).....	4
3.4.1	Management details	4
3.4.2	Summary of Objectives	4
3.4.3	Proceedings	4
3.4.4	Outcome / Results	4
3.5	NACO Program for Cooperative Cataloguing (NACO PCC)	4
3.5.1	Management details	4
3.5.2	Summary of Objectives	4
3.5.3	Proceedings	4
3.5.4	Outcome / Results	4
3.6	Bibliographic Control of Web Resources. A Library of Congress Action Plan.....	4
3.6.1	Management details	4
3.6.2	Summary of Objectives	4
3.6.3	Proceedings	4
3.6.4	Outcome / Results	4
3.7	Virtual International Authority File (VIAF).....	4
3.7.1	Management details	4
3.7.2	Summary of Objectives	4
3.7.3	Outcome / Results	4
3.8	AFNOR Experts group „Métadonnées d'autorité”	4
3.8.1	Management details	4
3.8.2	Summary of Objectives	4
3.8.3	Proceedings	4

3.8.4	Outcome / Results	4
3.9	ALCME NDLTD Authority Linking Proposal	4
3.9.1	Management details	4
3.9.2	Summary of Objectives	4
3.9.3	Outcome / Results	4
3.10	International Council on Archives – Committee on Descriptive Standards (ICA –CDS)	4
3.10.1	Management details	4
3.10.2	Summary of Objectives	4
3.10.3	Proceedings	4
3.10.4	Outcome / Results	4
3.11	Encoded Archival Context (EAC) Working Group	4
3.11.1	Management details	4
3.11.2	Summary of Objectives	4
3.11.3	Proceedings	4
3.11.4	Outcome / Results	4
3.12	Interoperability of Data in E-Commerce Systems (<indec>)	4
3.12.1	Management details	4
3.12.2	Summary of Objectives	4
3.12.3	Proceedings	4
3.12.4	Outcome / Results	4
3.13	InterParty – A Framework for the Unique Identification of Parties in E-Commerce	4
3.13.1	Management details	4
3.13.2	Summary of Objectives	4
3.13.3	Proceedings	4
4	MARD 2: Identification of Partnerships – Specific Relevance and Contact Details	4
4.1	MALVINE (Manuscripts And Letters Via Integrated Networks in Europe).....	4
4.1.1	Relevance for LEAF project	4
4.1.2	Contact Persons	4
4.1.3	LEAF Liaison	4
4.2	The National Register of Archives (NRA)	4
4.2.1	Relevance for LEAF project	4
4.2.2	Contact Persons	4
4.2.3	LEAF Liaison	4
4.3	IFLA Working Group on Functional Requirements And Numbering of Authority Records (FRANAR)	4
4.3.1	Relevance for LEAF project	4
4.3.2	Contact Persons	4
4.3.3	LEAF Liaison	4
4.4	MARBI Multilingual Record Task Force	4
4.4.1	Relevance for LEAF project	4
4.4.2	Contact Person	4
4.5	Authority Control in the Online Environment Interest Group (LITA/ALCTS)	4
4.5.1	Relevance for LEAF project	4
4.5.2	Contact Person	4
4.6	NACO Program for Cooperative Cataloguing (NACO PCC)	4
4.6.1	Relevance for LEAF project	4
4.6.2	Contact Person	4
4.6.3	LEAF Liaison	4
4.7	Bibliographic Control of Web Resources. A Library of Congress Action Plan	4
4.7.1	Relevance for LEAF project	4
4.7.2	Contact Person	4

4.7.3	LEAF Liaison	4
4.8	Virtual International Authority File (VIAF).....	4
4.8.1	Relevance for LEAF project	4
4.8.2	Contact Persons	4
4.8.3	LEAF Liaison	4
4.9	AFNOR Experts group „Métadonnées d'autorité”	4
4.9.1	Relevance for LEAF project	4
4.9.2	Contact Person	4
4.9.3	LEAF Liaison	4
4.10	ALCME NDLTD Authority Linking Proposal	4
4.10.1	Relevance for LEAF project	4
4.10.2	Contact Person	4
4.10.3	LEAF Liaison	4
5	MARD-3: Analysis of Practice	4
5.1	Introduction.....	4
5.1.1	Scope.....	4
5.1.2	Document Overview	4
5.2	Data Quality and Model Requirements.....	4
5.2.1	Data Quality and Authority Records	4
5.2.2	Model Requirements	4
5.3	Descriptions of Systems.....	4
5.3.1	The Data Providers: An Overview	4
5.3.2	Description of Local Systems	4
5.3.3	National Registries of Private Archives	4
5.4	Rules and Principles.....	4
5.4.1	Mandatory and Essential Elements.....	4
5.4.2	Local Procedures	4
5.4.3	Names and Entities: Differentiated and Undifferentiated Records	4
5.4.4	Dates and Date Formats	4
5.5	Problems and Samples.....	4
5.5.1	Language Dependent Name Conventions	4
5.5.2	Matching Names and Dates	4
5.5.3	Contents of Essential Elements	4
5.5.4	Complementary Authority Records	4
5.6	Conclusions.....	4
5.6.1	Identification and Automatic Linking.....	4
5.6.2	Selection of Data Providers	4
5.6.3	Completeness	4
5.6.4	The Concept of Authority Record.....	4
5.6.5	Minimal Requirements for Authority Records in LEAF.....	4
6	Bibliography	4
	Definitions, Acronyms and Abbreviations	4

1 Introduction

1.1 Scope

This document is the *Model Application Requirements Document* (MARD). The document is prepared for the Project LEAF (Linking and Exploring Authority Files; project number IST-2000-26323) and was produced under Work Package 5.2 (*Focussing the model with authority file data*) of the LEAF project.

1.2 Document Overview

The original design of this document included 4 parts. However, due to a slight delay in the project course which led to WP 5.1 being started late, it was decided to specify all user scenarios and system requirements in the *Model Requirements Analysis Document* (MRAD), namely in MRAD-2. Thus the timely start of WP 6 (Functional Specifications of the Demonstrator) was ensured. All specifications concerning aspects of uniformity and interoperability of name authority information are detailed in section 5 of this present document. For this reason the plan to write a fourth part of this MARD (part 4 was specified as "System Requirements) was dropped entirely. The present document therefore consists of three parts which are presented as 3 different chapters of one integral document. The single parts of the MARD are:

- MARD-1: Status Quo Analysis (by José Borbinha of BN and Max Kaiser of ÖNB)
- MARD-2: Identification of Partnerships (by Max Kaiser of ÖNB)
- MARD-3: Analysis of Practice (by Per-Gunnar Ottosson of RA)

2 MARD-1: Status Quo Analysis

This section, being part 1 of the MARD, provides the identification and analysis of previous or running initiatives and projects relevant for focussing the general LEAF model with name authority data.

2.1 Analysis projects

This section presents projects aiming at providing a step forward or just a better understanding of a specific problem related with authority control and data.

2.1.1 AUTHOR: Transnational Application of National Name Authority Files

2.1.1.1 Management details

Status: This project was supported by the European Commission (Program "Telematics For Libraries"), in the scope of the CoBRA+ activities (a consortium of national libraries).

Duration: January 1995 – December 1997 (3 years)

Members:

- Bibliothèque nationale de France (co-ordinator)
- Bibliothèque Royale Albert 1er (Belgium),
- Biblioteca Nacional (Spain),
- British Library (UK),
- Biblioteca Nacional (Portugal),
- Bureau Van Dijk (France); subcontractor; provided technical expertise,
- Index Data (Denmark), subcontractor for the implementation of the prototype.

Website: <http://portico.bl.uk/gabriel/projects/pages/cobra/author.html>

2.1.1.2 Summary of Objectives

The objectives of AUTHOR were defined as follows [Zillhardt / Bourdon 1998, 4f]:

- “1. To investigate the feasibility for the transnational exchange of national name authority files for personal and corporate names:
 - to give access to authority files for names of persons and corporate bodies, at the international level by means of a test bed platform and to define a target technical architecture.
 - to convert authority data produced by the national libraries partners of the Project to the international exchange format for authority data prepared by IFLA : UNIMARC/Authorities
 - to re-use authority data so made available in the current practice of cataloguing.
2. To implement and promote results of previous European projects

The Project UseMARCON (User-Controlled Generic MARC Converter):

This Project managed by the Koninklijke Bibliotheek (Netherlands), with The British Library and the Biblioteca Nacional (Portugal) as partners, aimed at the development of a Generic MARC Converter, that is to say a software which allows a librarian to state himself/herself the conversion rules necessary to convert bibliographic records from any source MARC format into another target MARC format. This "toolbox" uses UNIMARC as a pivot format.

The Project EUROPAGATE:

Ended in 1995 this Project developed a gateway between a Z39.50 client and a ISO SR server, and vice versa between a Z39.50 server and a ISO SR client, in order to give distant access to bibliographic databases. It solved also the technical problems raised by the access to multiple servers each of them having its own characteristics. This software, easily portable, offers a standardised interface between servers which give access to bibliographic databases and largely facilitate international connections. Project AUTHOR implements EUROPAGATE to get a suitable technical architecture.

The expected results were the following:

- to elaborate conversion tables for the partners' authority files, from their national format to the UNIMARC/Authorities format ; these tables should be re-usable later on.
- to examine problems raised by the elaboration of conversion tables and to propose recommendations to the IFLA UNIMARC Permanent Committee to have the UNIMARC updated according to the requirements.
- to give access to authority data through the Z39.50 protocol and the WEB.
- to propose to IFLA a definition of the minimal content of an authority record intended for international exchange, in close relation with the IFLA UBCIM Working Group created in May 1996 to work on the same topic and of which The Bibliothèque Nationale de France and The British Library are also members.
- to infer a target technical architecture able to be opened to other libraries from the test bed platform.”

2.1.1.3 Outcome / Results

The project AUTHOR was one of the first international projects addressing these problems. The results were not very relevant at the consequential level, but it was very important by providing one of the earliest analysis of the difficulties of the problem of reusing authority information.

2.1.2 DELOS/NSF Working Group on “Actors in Digital Libraries”

2.1.2.1 Management details

Status: This is an international initiative supported by the NSF – National Science Foundation (USA) and DELOS - Network of Excellence for Digital Libraries (supported by the IST Program of the European Commission).

Duration: September 2002 – June 2003 (ongoing).

Members:

- The working group is composed by individuals representing complementary scopes (e.g., libraries, archives, museums, computer engineering, and publishing industry).

Website: <http://www.delos-nsf.actorswg.cdlib.org/>

2.1.2.2 Summary of Objectives

“The purpose of this working group will be the development of a reference model describing the different classes of actors related to the problem of the Digital Library, including their roles and implications.

This work will try to identify and model as many cases as possible, covering not only the traditional entities, such as authors, editors and publishers, but also emerging classes or classes traditionally not modelled. The latter encompasses both familiar actors that tend to play their roles without formal specification (users, cataloguers, reference librarians, etc.), as well as less familiar entities (software agents, mediators, special rights holders, etc.).

The scope comprises the requirements of main players in libraries, museums, archives, the publishing industry, higher education (and possibly other specific user communities), and technology development.

This work has application in planning research and development efforts in areas such as the linking of metadata to sources of authority (for resource description and discovery), to the management of rights and intellectual property, and to user privacy protection, etc.”¹

¹ <http://www.delos-nsf.actorswg.cdlib.org/work.html>

2.2 Activities, technology, standards and frameworks

This section presents activities related with the definition or development of standards, framework or technology for the recording, transport or usage in general of authority information.

2.2.1 Hong Kong Chinese Authority (Name) Workgroup (HKCAN)

2.2.1.1 Management details

Status: In January 1999, a group of academic libraries in Hong Kong agreed to set up the Hong Kong Chinese Authority (Name) (HKCAN) Workgroup for establishing a union database that would reflect the unique characteristics of the Chinese authors and organizational names. This project was spearheaded by both Lingnan and Chinese University Libraries.

Duration: January 1999– (ongoing)

Members:

- Chinese University of Hong Kong Library (Database host institution)
- Hong Kong Baptist University
- Hong Kong Institute of Education Library
- Hong Kong Polytechnic University Library
- Lingnan University Library (Convener)
- University of Hong Kong Library
- City University of Hong Kong (Observer)

Website: <http://hkcan.ln.edu.hk/index.htm>

2.2.1.2 Summary of Objectives

The terms of reference for HKCAN are²:

- To facilitate a collaborative project among the participating libraries, for the development of "Hong Kong Chinese Authority (Name) database – HKCAN" that would reflect the characteristics of authors/organizations in Hong Kong.
- To create a multilingual, multi-script, centralized, and union database in a web-server.
- To set up standardization for name headings, including personal names, corporate names, meeting/conference names, uniform and series titles, etc.
- To identify software that would facilitate multilingual retrieval via remote access in HKCAN. This would allow the participating libraries to copy the records and manipulate the data according to their individual cataloguing practices.
- To set up principles for authority record selection.
- To monitor the future maintenance and update of HKCAN, and the distribution of work among the participating libraries.
- To identify related reference tools that would benefit the Project.
- To enhance resource sharing and to promote cross-straits co-operation with other libraries for Chinese name authority control.

2.2.1.3 Proceedings

For an overview of the proceedings see [Kwok 2002].

² see <http://hkcan.ln.edu.hk/about.htm>.

2.2.1.4 Outcome / Results

All non-Roman language headings in LCNAF, which has been widely adopted throughout the world, are rendered in romanized forms without the Chinese script. Chinese names are transliterated with the Wade-Giles romanization system. This kind of transliteration treatment presents no morphological meaning or exact identity to the original Chinese script. In order to tackle this problem, many libraries in Asia have begun, recently, building their own individual authority databases that support Chinese script. The problems with the LCNAF in Asia are [see Chan/Hu/Lo 2000]:

- All Chinese-language headings in LCNAF are rendered in romanized form only. But the word-phonetic romanization of the Wade-Giles or Pinyin scheme presents no morphological meaning to Chinese readers. It is difficult even for Chinese-speaking cataloguers to identify the correct records. As in MARC21 practice the #1xx field (main heading) is not repeatable, it is not possible to have parallel fields for an entry in an alternative script.
- In LCNAF many different Chinese personal names are represented by the same romanization.
- It is often very difficult to determine which is the author's given or family name, judging merely by the romanization in LCNAF.
- Occasionally, a heading is made up of 2 different forms of romanization in LCNAF.
- Important cross-references are not provided in LCNAF.

The main challenge faced by the HKCAN Workgroup lay in the design of a HKCAN authority model, which would accommodate Chinese scripts. After studying several models, the Workgroup finally agreed to adopt the #7XX field as the authority format for the HKCAN Database (see recommendations by the MARBI Task Force, chapter 3.3). This model provides the following advantages [see Hu/Lo/Kuo/Lau 2000 and Lau 2001]:

- The #7xx field is a standard MARC21 practice and is clearly defined in the latest edition of the MARC21 Format for Authority Data (Oct. 1999).
- The #7xx field has already been employed by a number of libraries for many years, e.g., the National Library of Canada. It is therefore an internationally recognized standard, and will facilitate cooperative cataloguing, as well as the sharing and exchanging of authority data on a global scale in the future.
- There is no requirement to have a corresponding authority record for the heading in the #7xx. Furthermore, it is repeatable. In other words, more than one equivalent established forms of headings may be included inside the same authority record, regardless of their languages or scripts.
- In addition, #7xx's synthetic structure enables flexible cross linkages of headings from different thesauri or authority files. The use of authority linking #7xx fields clearly indicates the relationships between headings of different authority files (e.g., LCNAF, Chinese Name Authorities or HKCAN local name headings, etc).

In the HKCAN database, field #1xx is used for the Authorized Chinese Name Form and #7xx (repeatable) is used for equivalent authorized headings (e.g. LC established heading in Pinyin form). As of May 2002, the HKCAN database comprised 116,371 records (40% LCNAF enhanced records, 60 % original HKCAN records) [Kwok 2002].

2.2.2 HKUST XML Name Access Control Metadata Repository

2.2.2.1 Management details

Status: Project at the Hong Kong University of Science and Technology (HKUST) Library.

Duration: 2002– (ongoing).

Members: Staff of HKUST Library.

Websites:

- <http://library.ust.hk/info/reports/xmlnac.html>
- <http://library.ust.hk/repository.pl>

2.2.2.2 Summary of Objectives

The project goals are [Lam/Kwok 2002 and Lam 2001]:

- To experiment with creating an XML-based repository of personal name metadata for Chinese authors.
- To collocate the variant name forms of a personal author so that a user can retrieve all of his/her bibliographic works with optimal precision.
- To provide metadata about a person with both non-roman script (Chinese) and corresponding romanized forms, accessible from either form.
- To establish a connection between the Repository and the HKUST library system.

2.2.2.3 Outcome / Results

The purpose of this project at the Hong Kong University of Science and Technology (HKUST) Library is to investigate the advantages of using XML format for personal name metadata. This project explores an alternative to the model adopted by the HKCAN initiative (see chapter 2.2.1). To enhance identification, access, and exchange of author information internationally, the project recommends that a global name repository should be set up in such a way that both vernacular forms and the corresponding roman transliterations for non-roman authors are included. For this purpose XML would be a useful tool for marking up multi-script authority records because it offers greater flexibility in data presentation, data exchange and interoperability than MARC21. The metadata created in XML format will allow more room for manipulation so that the use of name records in the repository can expand from authority control to access facilitation. In the repository only one XML metadata record will be created for a person, and computer programs can do the job of meeting the needs of individual institutions and the requirements of existing cataloguing standards such as AACR2.

Within the project the local personal authority records from the HKUST library database are converted into XML format using the MARCXML Schema.³ Chinese scripts of different name forms of an entity are added if not already present, including "see from" and "see also" references. The different name forms are then classified with XML attributes.

Each record (an XML instance) in the Repository represents a **Person** as defined in the IFLA FRBR model [FRBR 1998]. A Person can have multiple **Names**. A Name is an expression addressing the Person. A Name can be written in many **Name Forms**, depending on the applicable script, language and romanization schemes. This approach aims to solve the following problems, e.g.:

- The identification problem with romanized names - a one-to-many problem that cannot be resolved without richer links among name forms.
- The multiple "bibliographic identities" problem. According to AACR2, an authority record is created based on bibliographic existence. If an author has more than one identity in bibliographic records, more than one authority record must be created, with one for each identity. **By using the Person Model, one XML instance will only be created for one person. A set of authority records for an author can be generated from this instance as needed.**
- Failure of round trip conversion among existing MARC21 models. A richer name access control metadata format, such as the PERSON Model, will guarantee the possibility of presenting the data in various existing standard formats on demand.

The schema used in the Repository is based on the Library of Congress MARCXML Schema for MARC21 format. In order to hold the extra data content that is not defined in MARCXML, the project extended this schema to include additional attributes which are defined separately in a HKUST maintained namespace. In the Repository, all Name Forms of a Name are linked together by an attribute called "nameid". Each Name Form has an attribute called "reftype", which allows to define whether the Name Form is an established heading, or a "see from" or "see also" reference. Another attribute associated with Name Form is "script", which defines the script, language and romanization scheme being used.

³ <http://www.loc.gov/standards/marcxml/>.

The project also implemented a subset of SOAP (Simple Object Access Protocol⁴) on the Repository in order to demonstrate the capability of remote authority control from OPACs in addition to the built-in search interface. This repository is built on Tamino XML Server⁵. Each Person Record is an XML instance of the database. Most of the programs and Web CGI developed for this project were written in Perl. Additional software was used for data transformation and conversion, including LIBXML2⁶ for XSLT transformation and MARC4J⁷ for MARC formats conversion.

The project summary on the website:

"This project has demonstrated the potential of using XML format for name access control. It allows collocation of various name forms for the same person in a metadata record even if they are in different languages or scripts. It is now, therefore, technologically possible to set up a name access control repository that is useful for and accessible by all libraries in the world. What we need now would be the efforts and contributions from the various concerned parties, such as, national libraries, system vendors, bibliographic agencies, cataloguing utilities, etc. And we believe that this repository, once it is set up, will become an immensely useful information resource to libraries and other related institutions in the world." [Lam/Kwok 2002]

2.2.3 Authority Control among Chinese, Korean and Japanese Languages Project

2.2.3.1 Management details

Status: Two workshops: *Authority Control among Chinese, Korean and Japanese Languages* in the framework of the International Research Program "International Sharing of Japanese Information", co-ordinated by the National Institute for Informatics (NII), Japan.

Duration: Two workshops in Tokyo on Jan 11–12, 2001 and March 28–29, 2001.

Members: Six organizations participated at these workshops:

- National Institute of Informatics (NII), Tokyo
- National Diet Library (NDL), Tokyo
- National Library of China (NLC), Beijing
- National Library of Korea (NLK), Seoul
- Korea Education and Research Information Services (KERIS), Seoul.
- IFLA Universal Bibliographic Control and International MARC Office (IFLA UBCIM)

Website: <http://www.nii.ac.jp/CAT-ILL/PUB/CJK-WS/>

2.2.3.2 Summary of Objectives

The aim of the workshops was stated as "to pursue a standardised or harmonised cumulating of the name authority data of Chinese, Korean and Japanese language in other countries." The scope was stated as "Names are, in the first instance, Author Name, then followed in future if the project factors allow, by proper nouns that are used in other countries in local usage other than original usage." Target was set as "CJK Interchange Format of Authority Data" that conforms to the IFLA UNIMARC Authorities Format." [Naito 2001]

2.2.3.3 Outcome / Results

[NDL Newsletter 2001] summarizes the scope of the workshop as follows:

"Librarians have spent much time and effort in discussing appropriate formats of name authority records to handle CJK entries. In addition to ordinary problems, there are special difficulties in cataloguing CJK materials in CJK countries. For example, in the case of Chinese names, Japanese readers may read them in the Japanese way, and Korean readers in the Korean way, often quite differently from the original Chinese

⁴ <http://www.w3c.org/2000/xp/Group/>.

⁵ <http://www.softwareag.com/tamino/>.

⁶ <http://xmlsoft.org/>.

⁷ <http://marc4j.tigris.org/>.

way. Similarly, the reverse would happen in other CJK countries in the case of Japanese and Korean names. When a foreign publication is being entered into a database, it is desirable to record its original spelling (writing), local adaptation for colloquial use, as well as other forms possible for internal use. When searching for these names, which one is to be used? Original form or local adaptation? There could be a common treatment of foreign names in the three countries. Common conversion and data elements will offer better search capability.”

2.2.4 MACS - Multilingual Access to Subjects

2.2.4.1 Management details

Status: The MACS project was established to take forward the recommendations of the CoBRA+ Working Group on Multilingual Subject Access which conducted a feasibility study on matching headings between the three subject heading languages (SHL's) used in the Bibliotheque Nationale, Die Deutsche Bibliothek, the Swiss National Library and the British Library between Autumn 1997 and February 1999. The directors of the four libraries involved in the original study agreed to finance the development of a prototype system to:

(a) research the technical and organisational issues involved in managing a working system for creating and maintaining links between the three SHLs and (b) demonstrate the effectiveness of the linked SHLs for retrieving results for the end-user.

Duration: April 1999 – May 2001

Members:

- Swiss National Library (project leader)
- Bibliothèque nationale de France
- British Library
- Die Deutsche Bibliothek
- Prototype has been developed by Index Data (Denmark) and Tilburg University Library (Netherlands)

Website: <http://infolab.kub.nl/prj/macs>

2.2.4.2 Summary of Objectives

The project MACS attempts to link entries from three authority files in different languages: LCSH (English), RAMEAU (French), and SWD/RSWK (German).

“MACS (Multilingual Access to Subjects) aims to provide multilingual subject access to library catalogues. MACS enables users to simultaneously search the catalogues of the project's partner libraries in the language of their choice (English, French, German). (...)

This multilingual search is made possible thanks to the equivalence links created between the three indexing languages used in these libraries (...). Topics (headings) from the three lists are analysed to determine whether they are exact or partial matches, of a simple or complex nature. The end result is neither a translation nor a new thesaurus but a mapping of existing and widely used indexing languages.

On the basis of this approach, a prototype has been developed (...) which contains a small subset of data from the indexing languages and the libraries' databases so that link creation and management and subsequent searching can be explored and tested.”⁸

2.2.4.3 Outcome / Results

“The MACS prototype contains:

- around 3,000 headings covering theatre, athletics and those topics from all fields of knowledge which have been most often used in indexing documents
- 1,200 links created between the above concepts

⁸ <http://infolab.kub.nl/prj/macs>

- 30,000 bibliographic records extracted from the catalogues of the 4 partner libraries indexed with the appropriate subjects.
- Additional links are being prepared for headings in the fields of mathematics, psychology and information science.

Two interfaces demonstrate how a multilingual system should work:

- The search interface: allows users to browse headings and retrieve bibliographic records by using the links established between the concepts. The search interface uses the Z39.50 protocol.
- The Link Management Interface: enables the creation and management of links between headings from the indexing languages. Read-only access to these links is available to users outside the project.

Please note, the prototype:

- does not yet offer direct access to the partner libraries' catalogues
- does not yet contain all the concepts held in the indexing languages
- only contains authorised headings (i.e. those used to index documents) and not yet any see references (lead-in terms)
- for better searching, please check the list of headings available in the prototype for each subject heading language.”⁹

2.2.5 vCard

2.2.5.1 Management details

Status: The Versit Consortium developed a comprehensive family of PDI (Personal Data Interchange) technologies based on open specifications and interoperability agreements. The two main technologies that came from the Versit Consortium were vCard, an electronic business card, and vCalendar, an electronic calendaring and scheduling exchange format. Beginning in December, 1996, the Internet Mail Consortium took on responsibility for the development and promotion of these two important technologies. The Internet Mail Consortium is an international organization of vendors focused on expanding the role of mail on the Internet into areas such as commerce and entertainment. IMC works closely with standard organizations such as the IETF (Internet Engineering Task Force).¹⁰

Website: <http://www.imc.org/pdi/>

2.2.5.2 Summary of Objectives

vCard is a “de facto” standard for the interchange of personal information, recognized for example by most of the normal email applications.

“vCard automates the exchange of personal information typically found on a traditional business card. vCard is used in applications such as Internet mail, voice mail, Web browsers, telephony applications, call centers, video conferencing, PIMs (Personal Information Managers), PDAs (Personal Data Assistants), pagers, fax, office equipment, and smart cards. vCard information goes way beyond simple text, and includes elements like pictures, company logos, live Web addresses, and so on”¹¹

2.2.6 eduPerson Object Class

2.2.6.1 Management details

Status: Lack of agreement between institutions of higher education on information in LDAP -accessible directories makes it difficult to deploy multi-institutional networked services and resources. The edu-Person working group, with Educause and Internet 2 sponsorship, has the mission of defining an LDAP object class that includes widely used person attributes in higher education. The group will draw on the work of educational standards bodies in selecting definitions of these directory attributes.

⁹ <http://infolab.kub.nl/prj/mac/prototype.html>

¹⁰ <http://www.imc.org/>

¹¹ <http://www.imc.org/pdi/>

Website: <http://www.educause.edu/eduperson/>

2.2.6.2 Summary of Objectives

The EDUCAUSE/Internet2 eduPerson task force has the mission of defining an LDAP object class that includes widely-used person attributes in higher education. The group will draw on the work of educational standards bodies in selecting definitions of these directory attributes”¹²

2.2.6.3 Relevant references

- About LDAP and X.500:
 - An LDAP Roadmap & FAQ - A tutorial aid to navigating various LDAP and X.500 Directory Services resources on the Internet
 - <http://www.kingsmountain.com/ldapRoadmap.shtml>
 - Current Internet-Drafts
 - LDAP (v3) Revision (Idapbis) - 8 Internet-Drafts
 - <http://www.ietf.org/ids.by.wg/ldapbis.html>
 - LDAP Extension (Idapext) - 5 Internet-Drafts
 - <http://www.ietf.org/ids.by.wg/ldapext.html>
 - LDAP Duplication/Replication/Update Protocols (Idup) - 2 Internet-Drafts
 - <http://www.ietf.org/ids.by.wg/ldup.html>
 - Open source implementation of the Lightweight Directory Access Protocol
 - <http://www.openldap.org/>
 - Understanding X.500 - The Directory
 - <http://www.isi.salford.ac.uk/staff/dwc/Version.Web/Contents.htm>

2.2.7 OAI - Open Archives Initiative

2.2.7.1 Management details

Status: The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program. The fundamental technological framework and standards that are developing to support this work are, however, independent of the both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials. Support for Open Archives Initiative activities comes from the Digital Library Federation, the Coalition for Networked Information, and from National Science Foundation.

Duration: 1999– (ongoing).

Members:

- The OAI Steering Committee is responsible for establishing the general policy direction of the Initiative. A roster is available at <http://www.openarchives.org/news/oaiscpres000825.html>
- The OAI Executive Committee manages the operational details of the Initiative under the oversight of the Steering Committee and with the support of the Technical Committee. The Executive Committee is:
 - Carl Lagoze
 - Herbert Van de Sompel
- The OAI Technical Committee evaluates the effectiveness of the OAI interoperability architecture and proposes changes and enhancements based on community experience. The technical committee is currently being constituted. A roster is available at <http://www.openarchives.org/organization/tech.comm.html>

Website: <http://www.openarchives.org/>

¹² <http://www.educause.edu/eduperson/>

2.2.7.2 Summary of Objectives

"The Open Archives Initiative (OAI) develops and promotes interoperability solutions that aim to facilitate the efficient dissemination of content. The roots of the OAI lie in the E-Print community. Over the last year its focus has been extended to include all content providers. This paper describes the recent history of the OAI – its origins in promoting EPrints, the broadening of its focus, the details of its technical standard for metadata harvesting, the applications of this standard, and future plans. (...)

The Open Archives Metadata Harvesting Protocol consists of six requests or verbs. The protocol is carried within HTTP POST or GET methods. The intention is to make it simple for data providers to configure OAI conformant repositories by using readily available Web tools such as libwww-perl9. OAI requests all have the following structure:

- base-url – the Internet host and port of the HTTP server acting as a repository, with an optional path specified by the respective HTTP server as the handler for OAI protocol requests.
- keyword arguments – consisting of a list of key-value pairs. At a minimum, each OAI protocol request has one key=value pair that specifies the name of the OAI protocol request."¹³

"The Open Archives Initiative Protocol for Metadata Harvesting (...) provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework:

- Data Providers administer systems that support the OAI -PMH as a means of exposing metadata; and
- Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services."¹⁴

The actual version of the OAI protocol (OAI-PMH V2.0) supports Dublin Core as the exchange format. However, other formats are also possible:

"Communities should adopt guidelines for sharing of metadata Prefixes, metadata schema and XML namespace URI's of metadata formats. Such guidelines are outside of the scope of the OAI-PMH. The accompanying Implementation Guidelines document provides some sample XML Schema and instance documents for common metadata formats such as MARC and RFC 1807."¹⁵

The OAI initiative has been especially focused on bibliographic metadata, but the potential to use it also for the exchange of authority metadata has been in the agenda.

2.2.7.3 Relevant references

Open Archives Forum: <http://www.oaforum.org/>

"The Open Archives Forum provides a Europe-based focus for dissemination of information about European activity related to open archives and, in particular, to the Open Archives Initiative. The aim of the Forum is to facilitate clustering of IST projects, national initiatives and other parties interested in the open archives approach. In order to do so, the Forum brings interested parties together to build a community of interest, enable exchange of information and establish a web-based European information source for open archives. In addition, the Forum undertakes comparative reviews of technical and organisational issues."

2.2.8 Z39.50 / Bath Profile Maintenance Agency

2.2.8.1 Management details

Status: "Z39.50" refers to the International Standard, ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", and to ANSI/NISO Z39.50. The Library of Congress is the Maintenance Agency and Registration Authority for both standards, which are technically identical (though with minor editorial differences).

¹³ <http://www.openarchives.org/documents/oai.pdf>

¹⁴ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

¹⁵ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Websites:

- **Z39.50 Maintenance Agency:** <http://www.loc.gov/z3950/agency/>
- **Bath Profile Maintenance Agency:** <http://www.nlc-bnc.ca/bath/>
- **ZING:** <http://www.loc.gov/z3950/agency/zng.html>

2.2.8.2 Summary of Objectives**Z39.50:**

"Z39.50" refers to the International Standard, ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", and to ANSI/NISO Z39.50.¹⁶

This standard specifies a client/server based protocol for Information Retrieval. It specifies procedures and structures for a client to search a database provided by a server, retrieve database records identified by a search, scan a term list, and sort a result set. Access control, resource control, extended services, and a "help" facility are also supported. The protocol addresses communication between corresponding information retrieval applications, the client and server (which may reside on different computers); it does not address interaction between the client and the end-user.¹⁷

Bath Profile:

"An international Z39.50 specification for library applications and resource discovery, identifies those features of the Z39.50 standard that are required to support effective use of Z39.50 software for a range of library functions, such as basic searching and retrieval of bibliographic records for cataloguing, interlibrary loan, reference, and acquisitions. The profile defines both a core set of basic author, title and subject search and retrieval specifications across a variety of library databases, and a set of more complex searches. The functionality and specifications identified in the profile are intended to be incorporated into more detailed national, regional, provincial/state, and local agreements."¹⁸

MARC authority records exchange have been specified in the new version 2.0 of the Bath profile. A draft design document was published:

- <http://www.nlc-bnc.ca/bath/bath-draft2-2-e.html> (V2.0)
- <http://www.nlc-bnc.ca/bath/bp-functional.htm> (section "Bath Profile: 5.D. Functional Area D: Authority Record Search and Retrieval in Online Library Catalogues").

and the new version 2 was published on 17 March 2003 at <http://www.nlc-bnc.ca/bath/tp-bath2-e.htm>

ZING: Z39.50-International: Next Generation:

"ZING, 'Z39.50-International: Next Generation', covers a number of initiatives by Z39.50 implementors to make the intellectual/semantic content of Z39.50 more broadly available and to make Z39.50 more attractive to information providers, developers, vendors, and users, by lowering the barriers to implementation while preserving the existing intellectual contributions of Z39.50 that have accumulated over nearly 20 years.

Current ZING initiatives are SRW (including SRU), CQL, ZOOM, ez3950, and ZeeRex. Some (for example, SRW/U) seek to evolve Z39.50 to a more mainstream protocol, while for others (e.g. ZOOM) the purpose is to preserve the existing protocol but hide its complexity."¹⁹

"Z39.50-International: Next Generation" is the new generation of the Z39.50 concept, but now fully adapted to a scenario of Web Services.

¹⁶ <http://www.loc.gov/z3950/agency/>

¹⁷ <http://lcweb.loc.gov/z3950/agency/markup/01.html>

¹⁸ <http://www.nlc-bnc.ca/bath/bathfaq.pdf>

¹⁹ <http://www.loc.gov/z3950/agency/zng.html>

2.2.9 DC-Agents

2.2.9.1 Management details

Status: The DC-Agents is one of the several working groups of the DCMI - Dublin Core Metadata Initiative.

Duration: Nov. 1998- (ongoing).

Members: José Borbinha, John Kunze (Chairs); Hans Becker, Warwick Cathro, Robina Clayphan, Becky Dean, Rebecca Guenther, Lyn Marko, Heike Neuroth, Ed O'Neil, Andy Powell, Barbara Tillett, Stuart Weibel, Andrew Wilson.

Website: <http://dublincore.org/groups/agents/>

Archives of the discussion list: <http://www.jiscmail.ac.uk/lists/dc-agents.html>

2.2.9.2 Summary of Objectives

The charter of this group is to:

- "Develop a commentary which sets out the rationale and purpose of an agent record and an agent element set.
- Develop a Recommendation for an agent element set, possibly to include agent element qualifiers.
- Propose qualifiers for the Creator, Contributor and Publisher (the CCP) elements of the Dublin Core metadata element set. This activity is completely independent of the agent element set.
- Provide input to the DCMI Architecture working group concerning a recommended mechanism for linking the CCP elements to agent record."²⁰

2.2.9.3 Outcome / Results

The working group has advanced very little in the last two years, due mainly to a complex discussion about the key requirements for the CCP qualifiers and the uncertainty for the real practical feasibility of a useful agent element core. These issues were deeply discussed in the Dublin Core 2002 conference, from which came out an action plan to readdress these issues with a stronger commitment.

2.2.10 OASIS Extensible Resource Identifier (XRI) Technical Committee

2.2.10.1 Summary of Objectives

OASIS members have formed a new technical committee to establish a common identification scheme for distributed directory services. The Extensible Resource Identifier (XRI) Technical Committee purposes to create a URI scheme and a corresponding URN namespace for distributed directory services that enable the identification of resources (including people and organizations) and the sharing of data across domains, enterprises, and applications. XNS Public Trust Organization (XNSORG) will contribute the Extensible Name Service (XNS) specifications to the TC to serve as a basis for the OASIS committee work. The committee "will define a Uniform Resource Identifier (URI) scheme and a corresponding Uniform Resource (URN) namespace that meet these requirements, as well as basic mechanisms for resolving XRIs and exchanging data and metadata associated with XRI-identified resources."

<http://xml.coverpages.org/ni2003-01-08-a.html>

This is an activity related with distributed directory services and interoperability. Its scope is mainly the corporate and e-business areas (the TC members represent companies such as AMD, CISCO, VISA, etc.). However, its influence in promoting the future development of generic tools and systems might be very relevant in the medium and long term for the LEAF project in general and for each project partner planning local implementations.

²⁰ <http://dublincore.org/groups/agents/>

2.2.10.2 Relevant references

- **Home-page:** <http://xml.coverpages.org/OASIS-XRI-TC.html>
- **TC Proposal:** <http://xml.coverpages.org/XRI-TC-Proposal.html>
- **Announce :** <http://xml.coverpages.org/xns.html>
- **Related resources:** <http://xml.coverpages.org/ni2003-01-08-a.html>

2.3 Other resources

This section presents extra references, related with the creation, control, and usage in general of authorities, in a complementary perspective of the previous sections.

2.3.1 ULAN - Union List of Artist Names

2.3.1.1 Management details

Status: ULAN is maintained by The Getty Research Institute. The Getty Vocabulary Program, working as a unit with the Getty Standards Program, builds, maintains, and disseminates vocabulary tools for the visual arts and architecture.

Website: <http://www.getty.edu/research/tools/vocabulary/ulan/>

Getty Vocabulary Program: <http://www.getty.edu/research/institute/vocabulary>

2.3.1.2 Summary of Objectives

The ULAN contains more than 200,000 names and other information about generic and worldwide individual or corporate body artists, from the Antiquity to the present.

"The vocabularies produced by the Getty are the Art & Architecture Thesaurus® (AAT), the Union List of Artist Names® (ULAN), and the Getty Thesaurus of Geographic Names™ (TGN). These resources are intended to aid in the documentation and retrieval of automated information about art, architecture, and material culture."²¹

2.3.1.3 Outcome / Results

"The AAT, ULAN, and TGN are structured vocabularies that can be used to improve access to information about art, architecture, and material culture.

They may be used as data value standards at the point of documentation or cataloguing. In this context, they may be used as a "controlled vocabulary" or "authority." They provide "preferred" terms (or "descriptors") for concepts, as well as other synonyms that could be used by the cataloguer or indexer. They also provide structure and classification schemes that can aid in documentation.

They may be used as search assistants in database retrieval systems. They are "knowledge bases" that include semantic networks that show links and paths between concepts and these relationships can make retrieval more successful.

They may be utilized as research tools, valuable because of the rich information and contextual knowledge that they contain.

The focus of each vocabulary is art, architecture and material culture. The vocabularies provide terminology and other information about the objects, concepts, artists, and places important to various disciplines that specialize in these subjects. The primary users of the Getty Vocabularies include museums, art libraries, archives, visual resource collection cataloguers, bibliographic projects concerned with art, researchers in art and art history, and the information specialists who are dealing with the needs of these users. In addition, a significant number of users of the vocabularies are students or members of the general public seeking information.

²¹ <http://www.getty.edu/research/institute/vocabulary/index.html>

The users of the vocabularies typically access them in two ways: By using them as they are implemented in a museum collection management system, or by using the Web "browsers" on this site."²²

2.3.2 EUAN - European Union Archive Network

2.3.2.1 Management details

Status: The project was funded by the INFO2000 program of the European Commission

Duration: January 1999 – December 2000

Members:

- National Archives of Scotland (co-ordinator)
- National Archives of Sweden
- National Archives of Italy,
- International Institute of Social History, Amsterdam
- Scottish Archive Network Ltd.

Website: <http://www.euan.org/>

2.3.2.2 Summary of Objectives

EUAN was a two-year's project to address the problems of standardization and multilingual archival data for access to top-level descriptions of the holdings in several national archives.

"EUAN will examine both archival questions: how to ensure consistent and standardised description independent of language, and information technology aspects: how to navigate between different computer systems. The project will produce a prototype user interface together with reports and guidelines on promoting further European standardisation in these areas."²³

2.3.2.3 Relevant references

Conference Report: http://www.euan.org/euan_ead.html (report with interesting statements about the importance of keeping descriptive and contextual information managed in complementary systems, instead of integrated).

2.3.3 Authority Control in the 21st Century: An Invitational Conference

2.3.3.1 Management details

Duration: March 31 – April 1, 1996 -

Members:

- Organizing Committee: Ed O'Neill, OCLC (conference chairman); Jennifer Younger, The Ohio State University (conference co-chair); Kerre Kammerer, OCLC; Doug Perkins, OCLC; Diane Vizine-Goetz, OCLC; Karen Calhoun, OCLC; Bill Crowe, OCLC; Carol Mandel, Columbia University; Wilma Minty, Oxford University; Barbara Tillelt, Library of Congress; Robin Wendler, Harvard University.

Website: <http://www.oclc.org/oclc/man/authconf/confhome.htm>

²² <http://www.getty.edu/research/tools/vocabulary/ulan/about.html>

²³ http://www.euan.org/euan_about.html

2.3.3.2 Summary of Objectives

“This conference is an outgrowth of the program entitled The Future is Now: Reconciling Change and Continuity in Authority Control that OCLC hosted at the 1995 ALA Annual Conference in Chicago. In discussions leading up to this conference, it became obvious that:

- Authority control is expensive.
- Changes in authority control are needed.
- Authority control has not seen the benefits from major productivity increases.
- No consensus has yet emerged on what form of authority control is needed, or even if any authority control is needed.”²⁴

3 MARD 2: Identification of Partnerships – Context and Scope of Relevant Projects

This chapter, being the first part of part 2 of the MARD, sets out to describe in detail both the context and scope of those projects / initiatives the LEAF project has already, or intends in future to establish close working relations with.

3.1 MALVINE - Manuscripts and Letters via Integrated Networks in Europe

3.1.1 Management details

Status: MALVINE was a project supported by the European Commission, within the program Telematics for Libraries of the Fourth Framework Programme.

Duration: July 1998 – December 2000.

Members: Joanneum Research, Austria; University of Bergen - Norwegian Computer Centre for the Humanities, Norway; Fraunhofer Institut Software- und Systemtechnik, Germany; Institut de Textes et Manuscrits Modernes, France; Österreichische Nationalbibliothek, Austria; Biblioteca de Universidad Complutense de Madrid, Spain; Biblioteca Nacional, Portugal; Institut für Wissenschaftstheorie und Wissenschaftsforschung der Universität Wien, Austria; Forschungsstelle und Dokumentationszentrum für österreichische Philosophie, Austria; National Museum of Denmark, Denmark; Deutsches Literaturarchiv, Germany; Goethe- und Schiller-Archiv, Germany; British Library, United Kingdom; Institut Mémoire de l'Édition Contemporaine, France; Swiss National Library, Switzerland; Crossnet Systems Ltd., United Kingdom; K.G. Saur Verlag, Germany.

Website: <http://www.malvine.org>

3.1.2 Summary of Objectives

“MALVINE opens new and enhanced access to disparate modern manuscript holdings that are kept and catalogued in European libraries, archives, documentation centres and museums. MALVINE provides a common multilingual user interface in a clear terminology with the help of metadata. (...)

The main technical outcome of the project will be delivery of the operational MALVINE "retrieval manager" allowing parallel, multi-site searching and navigation at the collection level, via metadata.

The MALVINE-OPAC system (database and Z39.50-target) will be generic for library, museum and archival applications (MALVINE-profile) supported by a data loading facility (...).”²⁵

²⁴ <http://www.oclc.org/oclc/man/authconf/welcome.htm>

²⁵ <http://www.cordis.lu/libraries/en/projects/malvine.html>

3.1.3 Outcome / Results

The project developed a resource discovery service, based on a Z39.50 gateway, for the discovery of manuscripts in several European archives and libraries .

3.2 IFLA Working Group on Functional Requirements And Numbering of Authority Records (FRANAR)

3.2.1 Management details

Status: Working Group of IFLA

Duration: June 1999 – (ongoing)

Members:

- Francoise Bourdon, Bibliothèque nationale de France
- Tom Delsey, National Library of Canada
- Christina Hengel-Dittrich, Die Deutsche Bibliothek
- Olga Lavrionova, Russian State Library
- Andrew MacEwan, The British Library
- Eeva Murtomaa, National and University Library of Finland
- Glenn Patton, OCLC (Chair of the WG)
- Marie-France Plassard, Die Deutsche Bibliothek, IFLA UBCIM
- Henry Snyder, University of California
- Barbara Tillett, Library of Congress
- Hartmut Walravens, Staatsbibliothek zu Berlin, International ISBN and ISMN Agencies
- Mirna Willer, National and University Library, Croatia

Website: No website established.

3.2.2 Summary of Objectives

The following terms of reference were endorsed [Plassard/Ratthei 1999]:

- to define functional requirements of authority records: continuing the work that the "Functional Requirements of Bibliographic Records" [FRBR 1998] for bibliographic systems initiated;
- to study the feasibility of an International Standard Authority Data Number (ISADN): to define possible use and users, to determine for what types of authority records such an ISADN is necessary, to examine the possible structure of the number and the type of management that would be necessary;
- to serve as the official IFLA liaison to, and work with, other interested groups concerning authority files.

3.2.3 Proceedings

As a follow-up to the *IFLA Working Group on Minimal Level Authority Records and the ISADN*, which had completed its task in 1998²⁶, the FRANAR Working Group : (Functional Requirements And Numbering of Authority Records) was created in June 1999 under the auspices of the IFLA Division of Bibliographic Control and the IFLA UBCIM Programme. FRANAR started its work with analysing the feasibility of an International Standard Authority Number (ISADN) [see Bourdon 2001, 2f and 6–8]. It was agreed that no additional standard number should be created, but rather existing numbers should be re-used. The WG examined the feasibility of using numbers which are automatically assigned by systems to authority records and reviewed the international standardized numbers which are already defined by ISO. The WG came to

²⁶ <http://www.ifla.org/VI/3/p1996-2/mlar.htm>

the conclusion that the ISO identifiers probably will not serve for use in authority databases managed by libraries because they serve different purposes. Many of these numbers, FRANAR concluded, are created to facilitate B2B relations and are not supposed to be used as search criteria or as identification element in authority databases managed by libraries. The WG decided that it is necessary to define the requirements for an ISADN first [Bourdon 2001, 7]:

- What are the goals of an ISADN? (To help search? To facilitate authority management?)
- What entities are expected to be identified by an ISADN?
- What are the data elements which characterize these entities?
- What relations are capable to exist between the different entities on the one hand, and between the data elements pertaining to a given entity on the other hand ?
- Is the international standard numbering of authority data still fit for acceptance today?
- Are other technical means possible?

In order to answer these questions FRANAR decided to first move its work towards the definition of functional requirements for authority records. In order to pursue this task the following questions were identified [Bourdon 2001, 8f]:

- What are the entities concerned?
- What are the elements of data which constitute an authority record?
- What is the use of an authority record? / Who are the users of authority records?
- What modelling is appropriate for authority records?

3.2.4 Outcome / Results

In June 2002 FRANAR issued the draft of a conceptual model for functional requirements and numbering of authority records, prepared by Tom Delsey [Delsey 2002]. The purpose of this model is to

- provide a clearly defined, structured frame of reference for relating the data that are recorded in authority records to the needs of the users of those records;
- clarify the function of identifiers or standard numbers for authority data;
- assist in an assessment of the potential for international exchange and use of authority data both within the library sector and beyond [Delsey 2002, 1]

The conceptual model defines (1) the relationships of entities associated with the names and identifiers that are registered in authority files, (2) the relationships of entities associated with the names and identifiers that are used as access points.

The model makes a distinction between the following types of *entities*

1. **“Bibliographic” entities** (person, family, corporate body, work, expression). They reflect intellectual constructs or concepts that are integral to the rules used to create library catalogues. The model depicts also entities that reflect the logical groupings of data that make up an library authority file (heading, target, tracing, explanatory heading, authority record, reference record, general explanatory record): As the model also addresses issues related to the management of authority files per se, it is essential to reflect the key logical groupings of authority data as entities in their own right.
2. **“Real world” entities** (individual, group, event, content). They reflect individuals and groups of individuals, the texts, music, recordings, images, etc. that those individuals and groups create, produce, perform, etc., and the ideas, locations, and periods of time that are reflected in products of intellectual and artistic endeavour. These entities have been included in the model primarily to assist in clarifying the relationship of the “bibliographic” entities associated with the names and identifiers registered in library authority files to those registered in files that serve a similar though not necessarily identical purpose in other sectors (archives, museums, rights administration organizations, etc.).

The model also defines *attributes* for each of the entities at a “logical” level (i.e., as characteristics of the entities to which they pertain, not as specifically defined data elements).

For the purpose of the conceptual model, the *users of authority data* are broadly defined to include [Delsey 2002, 37]:

1. cataloguers and reference librarians who use authority files directly;
2. library patrons who use authority information either through direct access to authority files or indirectly through the headings and references in library catalogues, national bibliographies, etc.;
3. database management and applications software designed to support the creation, maintenance, search and retrieval of data contained in bibliographic and authority files.

The model defines eight *user tasks* [Delsey 2002, 37f]:

- **Resource discovery:** *Search* for an entity / *Identify* an entity / *Control* the form of heading used for entries in a catalogue, bibliography etc. / *Relate*, i.e. establish or clarify the relationship between one entity and another.
- **Data Management:** *Process* a record or heading / *Sort* headings or records / *Display* an entry, heading, or data field / *Integrate* a record, entry, or heading into an existing authority file.

In a table these user tasks are mapped to the entities and their logical attributes which are defined in the conceptual model [Delsey 2002, 38–43].

The model also addresses the issue of *identifiers*. Identifiers are grouped into three conceptually different categories. From a functional perspective, each of the three categories serves a different purpose. In the context of **international exchange of authority records**, a case could be made for the inclusion of identifiers from all three categories in the authority record [Delsey 2002, 46f]

4. **Identifiers for “real world” entities** (individual, group, event, content). These identifiers “could be incorporated into authority records as a means of efficiently identifying the ‘real world’ entity or entities associated with the authority record. The inclusion of such identifiers would facilitate the identification of records from multiple sources associated with the same ‘real world’ entity. A library accessing multiple authority files or a common file could use the identifier as a mechanism for selecting from the file(s) all records associated with a particular *individual*, *group*, or *event* or a particular instance of *content*.” [Delsey 2002, 47].
5. **Identifiers for “bibliographic” entities** (person, family, corporate body, work, expression, manifestation, item). These identifiers “could be incorporated into authority records as a means of efficiently identifying the ‘bibliographic’ entity associated with the authority record. The inclusion of such identifiers would facilitate the identification of records from multiple sources associated with the same ‘bibliographic’ entity. A library compiling a consolidated file of authority records from sources using different languages of cataloguing and cataloguing rules could use the identifier as a mechanism for detecting potential matches or conflicts. Such identifiers would provide a level of identification that is essential in a library context and that could not be achieved through the use of identifiers for ‘real world’ entities alone.” [ibid.]
6. **Identifiers for records** (authority record, reference record, general explanatory record). The identifiers are “essential for processing such records, both in a local file environment and in an international context. Record identifiers alone, however, will not serve to identify records associated with either the same ‘real world’ entity or the same ‘bibliographic’ entity.” [ibid.]

The conceptual distinction between “real world” entities and “bibliographic” entities is also critical to considerations of interfacing / exchanging authority files emanating from different sources of authority systems:

“The files created by archives, museums, rights management organizations, etc. that are frequently associated with library authority files are in fact significantly different inasmuch as the headings carried in records in those files normally represent “real world” entities (*individuals*, *groups*, *events*, *content*) directly. In the authority files created by libraries, on the other hand, those “real world” entities are only indirectly associated with the *headings* carried in *authority records* and *reference records*. The headings in library authority files represent “bibliographic” entities (*persons*, *families*, *corporate bodies*, *works*, *expressions*), which are in fact abstract entities reflecting library practice. Although in many instances there is a one-to-one relationship between an *individual* or *group* and a corresponding bibliographic entity (*person*, *family*, or *corporate body*), there are instances where the relationship is one-to-many or even many-to-many. For that reason, the record created by an archive, museum, or rights management organization that is associated with an *individual*, *group* or

event may not always function as a direct parallel to the record or records created by a library for a *person*, *family*, or *corporate body*. The differences between the two may stem not simply from differences with respect to the structure of headings or the form of name used as the basis for the heading, but more fundamentally from a functional difference with respect to the entities that are represented by those headings. Those differences cannot be overlooked, and must be taken into account when assessing cross-sector compatibility of data and the potential for interoperability of authority systems.

It should also be noted that because the headings in library authority files represent "bibliographic" entities, and because those entities in turn reflect intellectual constructs or concepts that may be interpreted differently from one set of cataloguing rules to another, a specific instance of *person*, *family*, *corporate body*, *work*, or expression represented in an authority file compiled by one library may not directly parallel an instance of a similar bibliographic entity represented in a file compiled by another library. Again, the differences between the two files may stem not simply from differences with respect to the structure of headings or the form of name or title used as the basis for the heading, but from functional differences with respect to the entities that are represented by those headings. Those functional differences may be reflected not only in the heading itself but also more extensively through the reference structure established for the file (i.e., information notes, references and tracings). The nature of those differences must be taken into account when assessing compatibility of data and the potential for interoperability of authority systems even within the library sector." [Delsey 2002, 48]

3.3 MARBI Multilingual Record Task Force

3.3.1 Management details

Status: MARBI Working Group.

MARBI (Machine-Readable Bibliographic Information Committee) is a body of the following divisions of ALA (American Library Association):

- ALCTS (Association for Library Collections & Technical Services)
- LITA (Library and Information Technology Association)
- RUSA (Reference and User Services Association)

Duration: 1999– (ongoing)

Members: A roster is available at <http://www.ala.org/alcts/organization/div/marbi/tf-nrs-1a.html#roster>

Websites:

- MARBI <http://www.ala.org/alcts/organization/div/marbi/marbi.html>
- MARBI Multilingual Record Task Force: <http://www.ala.org/alcts/organization/div/marbi/tf-nrs-1a.html>

3.3.2 Summary of Objectives

MARBI is responsible for developing official ALA positions on standards for the representation in machine-readable form of bibliographic information. MARBI focuses its attention on the development of the MARC format.

The MARBI Multilingual Record Task Force among other things²⁷ has been charged to investigate models and data structures that would allow MARC21 authority records to support global and multilingual OPAC displays more effectively. What language/script functions should the authority record be expected to perform? Should an expansion of the format be oriented towards management of headings, or display in multi-language environments, or both? Is there a conflict between these goals, or can both be accomplished using the same data structures?

²⁷ See <http://www.ala.org/alcts/organization/div/marbi/tf-nrs-1a.html#charge>.

3.3.3 Outcome / Results

The Multilingual Record Task Force presented a discussion paper on June 8, 2001 [MARBI 2001], which proposed changes to the MARC21 coding structure that would support the identification of equivalent headings in multiple languages and/or scripts, and would provide for the syndetic structure needed to support those equivalent headings. The Task Force began by examining two different families of record configurations, those employing **single authority records** ("Model A") and those employing **linked authority records** ("Model B"). Model B is compatible with recent conclusions of the FRANAR working group (see chapter 3.2). To add flexibility, the Task Force expanded the language centricism of Model B to include the whole **context** of a heading ("Model C").

Model A: One authority record is constructed for each entity. The one established heading in the #1XX field is appropriate for a catalogue using the language designated at the record level by the code in subfield #040\$b (Language of cataloguing). Different forms of the #1XX heading (in other languages, scripts, etc.) are treated as variant reference headings and are traced in the #4XX fields (See From Tracings). This model has a number of limitations. It could lead to a very large number of reference tracing fields since several reference structures might be combined in a single record. This makes difficulties both for record management and for displays. Furthermore cooperative maintenance of such records would be complicated. The model does not provide an easy way to distinguish multiple variants in the same language of catalogue, such as a French form used in the National Library of Canada catalogue and the form used in the Bibliothèque Nationale catalogue.

Model B: A separate authority record is constructed for each separate established heading for an entity. The heading in the #1XX field is appropriate for a catalogue in the language designated at the record level by the language code in subfield #040\$b. The #4XX and #5XX traced headings constitute the reference structure for the #1XX heading in that catalogue. As necessary, a cataloguing agency constructs separate records that contain parallel #1XX established headings that differ in the language of the catalogue into which they fit; for example, the French catalogue version of a heading elsewhere represented for an English catalogue. Each authority record contains #4XX and #5XX reference tracing fields and #6XX note fields appropriate to the record level language of catalogue coding in the #040 field. If there are separate records for alternative headings, the records may be **linked** by recording the associated record numbers (subfield \$0) in #7XX fields. If there is no separate record then the #7XX may be primarily intended for alternative displays in OPACs. This model has the advantage of separating the reference structures for the various headings for an entity, facilitating record maintenance and intelligible displays. A version of the model is already used successfully by the National Library Canada.

Model C (recommended by the Task Force): rather than defining the established headings and equivalent established headings solely in terms of the language of the catalogue, they are defined by a **catalogue context**. The primary component of the catalogue context is the body of **rules** under which the heading was formulated (e.g., AACR2, LCSH, RAK, RAMEAU, etc.). Additional components could be added; these might include an explicit indication of the **language of catalogue** into which the heading fits (e.g., hun, eng, ger), or the **audience** for which the heading is useful (e.g., children, popular). As with Model B, a separate authority record is constructed for each separate established heading for an entity. The #4XX and #5XX reference tracings in the record are appropriate to the #1XX heading in that context. A cataloguing agency may decide to include headings for the entity that are suitable in other contexts. These alternative established headings would be recorded in the #7XX fields in the authority record, along with an indication of the context in which they are appropriate. A cataloguing agency may separately decide to create additional authority records in which the alternative heading(s) are the #1XX heading; such authority records would contain reference structures and #6XX note fields appropriate to the alternative context. If there are separate records for alternative headings, the records may be **linked** by recording the associated record numbers (subfield \$0) in #7XX fields. If there is no separate record then the #7XX may be primarily intended for alternative displays in OPACs. Because the notion of "context" may be adjusted to suit various situations, Model C is extensible and flexible, and has the following additional advantages:

- It supports any reference structure required by the cataloguing or thesaurus rules used.
- It supports simple identification of alternative headings for display, in addition to linking to alternative authority records with their full reference structures.
- It is not disruptive of existing MARC21 files and would not require retrospective conversion.

- Systems can optionally strip away or ignore all context markers, or even ignore altogether the #7XX fields in which they reside.

Context marking technique: The context marker is a compound code that represents the context in which an equivalent heading might be substituted for the #1XX established heading in OPAC displays. The components of the context marker proposed by the Task Force are the **rules**, **audience**, and the **language** of the catalogue. The definition of a new subfield \$7 (**Context marker**) in the authority #7XX fields is recommended to hold context information. The context marker specifies the environment in which a heading is appropriate for display. The context marker subfield would need to be repeatable so that a heading appropriate on multiple contexts will not need to be repeated. For example, many name headings without qualifiers might be valid in several different catalogues built for different languages. This argues also for making the record and #1XX heading level context markers repeatable. Thus #040 \$b, \$e, and \$f would need to be made repeatable.

3.4 Authority Control in the Online Environment Interest Group (LITA/ALCTS CCS)

3.4.1 Management details

Status: Interest Group of LITA (Library and Information Technology Association, a division of the American Library Association) and Discussion Group of ALCTS CSS (Cataloguing and Classification Section of the Association for Library Collections & Technical Services, a division of the American Library Association).

Duration: 1985– (ongoing; established in 1985 and renewed in 1988, 1991, 1993, 1997).

Members: Chair: Ann Della Porta (Library of Congress, Assistant Coordinator, Integrated Library System Program), adel@loc.gov

Website: <http://www.lita.org/igs/Acig/acig.htm>

3.4.2 Summary of Objectives

To provide a forum for discussion of a variety of issues related to authority control for online catalogues and for international sharing of authority of data. The goals are to raise the level of awareness on authority control issues, to encourage ideas for new approaches to authority control and to promote significant research on authority control.

3.4.3 Proceedings

LITA established procedures for the formation of interest groups in 1985, to facilitate a more efficient and effective response to its members' interest in and involvement with new library technologies. Interest Groups are expected to meet at each ALA Annual Conference, and may also meet during the ALA Midwinter Meeting.

3.4.4 Outcome / Results

The Interest Group met at the ALA Annual Conference in Atlanta on June 16, 2002. The presentations of this meeting are available online at: <http://www.lita.org/igs/Acig/acig2002.htm>. They include a presentation of Function Area D of the Bath Profile (see chapter 2.2.8) by Larry E. Dixon [Dixon 2002], a presentation of Z39.50 profiles (Bath Profile, U.S. Profile) by Carrol Lunau [Lunau 2002] and an overview on "International Efforts to Improve Interoperability" by Glenn Patton (covering AUTHOR, MACS, MALVINE, LEAF, AFNOR, InterParty, MLAR, and FRANAR) [Patton 2002].

In preparation for the 2002 meeting the IG published a bibliography: "Bibliography of Current Issues in Authority Control (covering January 2000 to July 2001)" [ACIG 2002].

3.5 NACO Program for Cooperative Cataloguing (NACO PCC)

3.5.1 Management details

Status: Established in 1977 as a result of an agreement between the Library of Congress (LC) and the U.S. Government Printing Office (GPO) to use and maintain a common name authority file (Library of Congress Name Authority File, LCNAF). Now part of the Program for Cooperative Cataloguing (PCC) of the Library of Congress which was initiated in 1995.

Duration: 1977– (ongoing)

Members: A list of NACO institutions or projects is available at:

<http://www.loc.gov/catdir/pcc/pccliaisons.html>.

NACO has expanded outside the USA and has now members in Asia, Africa, South America and Europe.²⁸ Partners outside the USA contributed 19,43% of new NARs in 2001.²⁹ As of January 2001 NACO had more than 330 participants.

Website: <http://www.loc.gov/catdir/pcc/naco.html>

3.5.2 Summary of Objectives

The following objectives of the NACO program are stated in [NACO 1996, p 1]:

- To produce a national-level authority file;
- To share the costs of authority work;
- To reduce duplication of effort;
- To increase the timeliness of cataloguing copy;
- To extend cataloguing coverage; and
- To produce cataloguing of national-level quality.

3.5.3 Proceedings

An individual institution may join the NACO program, or a group of libraries may form a “funnel project” to contribute records via a coordinator who is responsible for the joint input. At this time NACO participants must belong to one of the two bibliographic organizations, OCLC or RLG, in order to be able to contribute authorities online. A basic principle of NACO is that all authority contributions are to be formulated using the following documentation [see NACOPARA]:

- Anglo-American Cataloguing Rules, second revised edition (AACR2), Chapters 22-26 and index,
- USMARC Authority Format (including the Z1 addition issued by the Library of Congress),
- Library of Congress Rule Interpretations (LCRIs), Chapters 22-26,
- Library of Congress Subject Cataloguing Manual (SCM), Memo H405.

3.5.4 Outcome / Results

Per year more than 100.000 name authority records are created by NACO participants. Since 1977, NACO participants have contributed about 1,6 million name and series authority records which comprise about one third of all records in the LCNAF.

In 2002 the PCC offered an interim model for alternate authorized headings which uses the 7XX field in the MARC21 format, which may be of relevance also for the LEAF Project: With the expansion of LCNAF via NACO to users worldwide there is an issue of **alternative authorized headings** (#1XX) within LCNAF:

²⁸ For recent statistics see <http://www.loc.gov/catdir/pcc/stats/stats.html>.

²⁹ <http://www.loc.gov/catdir/pcc/intco/graphs01.html>.

"National libraries with well-developed national bibliographies frequently discover that many names established under the NACO program differ from the forms already in use in their national bibliographies. Even if identical cataloguing documentation is used in the creation of an individual country's national bibliography and the NACO name authority file, the application of different options offered by AACR2 and LCR1's, different titles used as the basis for heading construction, and the use of different databases to serve as background for the formation of names can lead to different choices for an authorized form of name (1XX). This leads to a dilemma regarding the choice of name to be used on bibliographic records, with a potential need for extensive bibliographic file maintenance." [NACO 7XX]

As it is not practical to change the affected bibliographic records, **the PCC in 2002 offered an interim model for alternate authorized headings which uses the 7XX field in the MARC21 format [NACO 7xx]**. This model has similarities to the model A discussed in the MARBI Multilingual Record Task Force (see chapter 3.3).³⁰

3.6 Bibliographic Control of Web Resources. A Library of Congress Action Plan

3.6.1 Management details

Status: Each action item in the plan features a lead individual or organization, an Library of Congress Cataloguing Management Team liaison, and suggestions for potential collaborators that the lead might invite to work on the action item.

Duration: 2001– (ongoing)

Members: Lead Organizations (as of November 20, 2002):

- Library of Congress Cataloguing Directorate
- LC CDS (Library of Congress Cataloguing Distribution Service)
- LC CPSO (Library of Congress Cataloguing Policy and Support Office)
- LC NDMSO (Library of Congress Network Development and MARC Standards Office)
- LC PAIG (Library of Congress Portals Applications Interest Group)
- ALCTS Task Force on the LC Action Plan for Bibliographic Control of Web Resources
- DCMI (Dublin Core Metadata Initiative)
- OCLC Digital and Preservation Resources Division
- OCLC Office of Research
- PCC SCA (Program for Cooperative Cataloguing Standing Committee on Automation)
- RUSA (ALA Reference and User Services Association)

Website: <http://lcweb.loc.gov/catdir/bibcontrol/actionplan.html>

3.6.2 Summary of Objectives

The action plan stems from the Library of Congress "Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenge of Networked Resources and the Web", held on November 15–17, 2000.³¹ The primary goals of the conference were³² (1) to develop an overall strategy to address the challenges of improved access to Web resources through library catalogues and applications of metadata; and (2) to identify attainable actions for achieving the overall strategy. The aim of the conference, therefore,

³⁰ Initially, this #7XX model is limited to records created by National Bibliographic Agencies (NBA) and under certain conditions by designated NACO contributors authorized by the NBA from within the same geographic jurisdiction, for the sole purpose of recording the form of heading used in its national bibliography. For guidelines for including #7XX fields abs \$2 (source authority file) in NACO records, see [NACO 7XX].

³¹ <http://lcweb.loc.gov/catdir/bibcontrol/conference.html>.

³² see <http://lcweb.loc.gov/catdir/bibcontrol/actionplan.html>

was to generate recommendations for the Library of Congress, in collaboration with the larger library community, to use as a blueprint for action to improve bibliographic control of the Web. The deliberations of the conference participants resulted in eleven sets of recommendations that have been distilled into an Action Plan. In developing the Plan, LC took into account both the original Conference goals and *LC21: A Digital Strategy for the Library of Congress*.

The plan, which has been updated on November 20, 2002 [BIBCONTROL 2002], identifies six objectives into which the action items were placed: 1) increased availability of standard records for Web resources; 2) enhanced record display and access across multiple systems; 3) collaboration among metadata standards communities for better bibliographic control of Web resources; 4) development of automated tools for harvesting and maintaining metadata; 5) provision of appropriate training for the Web environment; 6) support of research and development to enhance bibliographic control of Web resources. Every objective of the Plan acknowledges the importance of partnerships between libraries and a broad spectrum of other groups: metadata producers; standards developers; systems and software vendors; computing and technology suppliers; scholarly and academic enclaves; publishers; dot.com creators; bibliographic utilities; registration agencies; other information providers; government agencies; other libraries, including national libraries; and other stewards of cultural and historical knowledge, e.g., museums and archives.

Each action item in the plan features a lead individual or organization, an LC Cataloguing Management Team liaison, and suggestions for potential collaborators that the lead might invite to work on the action item. Some organizations will assume lead roles; others will serve in contributing roles. Additionally, experts may be secured to serve as principal investigators. LC also assigned a priority: "near-term" could be accomplished within eighteen months; "long-term" could be accomplished within five years. The aspect of "High," "Medium," or "Low" was assigned on the basis of the benefit each action would bring to the library community relative to the expense involved in carrying it out.

3.6.3 Proceedings

The lead organizations began to work on the 29 individual action items after a kick off meeting on Jan. 20, 2002.³³ A 2nd Library of Congress 2000 Action Plan Forum³⁴ was held on June 16, 2002 during the ALA Annual Conference 2002 in Atlanta [FORUM 2002]. The action plan is intended to be regularly updated.

3.6.4 Outcome / Results

Action Item 2.2 is of relevance for the LEAF Project:

Action Item 2.2. Promote the international sharing and collaborative use of "authority information"

The action item (priority: "long-term/high") is lead by Barbara Tillett LC, Cataloguing Policy and Support Office (LC CPSO).

The work plan identifies the following task: "Promote the international sharing and collaborative use of 'authority' information (authority records for names of persons, corporate bodies, works / expressions / manifestations / items / subjects, etc. – also called controlled vocabularies) starting with personal name authority records, launching a prototype of a virtual international authority" [Tillet 2002b]. As first component of the task a proof concept project with LC/DDB/OCLC on personal name authority files will be initiated (see chapter 3.7).

The following stakeholders for the task are identified in the work plan: IFLA (FRANAR; Section on Cataloguing); Die Deutsche Bibliothek; OCLC Office of Research; RLG; EU-Project LEAF; EU-Project InterParty; DC-Agents Working Group; DELOS/NSF Actors and Roles Working Group; Group consisting of National Institute of Informatics (Japan) / National Diet Library (Japan) National Library of Korea / National Library of China; Program for Cooperative Cataloguing (PCC); W3C Semantic Web Working Group.

³³ <http://www.loc.gov/catdir/lccn/lccn1002.html>.

³⁴ <http://www.loc.gov/catdir/lccn/lccn1009.html>.

3.7 Virtual International Authority File (VIAF)

3.7.1 Management details

Status: VIAF is a joint project of OCLC, Library of Congress and Die Deutsche Bibliothek.

Duration: Autumn 2002 – Autumn 2004 (2 years, ongoing)

Members:

- OCLC
- Library of Congress
- Die Deutsche Bibliothek

Website: <http://www.oclc.org/research/projects/viaf/index.shtm>

3.7.2 Summary of Objectives

The objectives of the project are [Tillett 2002]:

- Assure consistency of authority control (following a standard set of rules and guidelines). This would:
 - Enable cataloguers to use controlled forms to efficiently collocate works in catalogues through the use of controlled forms of headings;
 - Facilitate sharing of the workload in creation of authority records to reduce cataloguing costs to libraries, museums, archives, rights management agencies, etc.
- Enable users to access information in the language, script, form they prefer.
- Simplify creation and maintenance of authority records internationally.

3.7.3 Outcome / Results

VIAF is a joint project of OCLC, Library of Congress and Die Deutsche Bibliothek. In its first phase, starting at the end of 2002, this proof of concept project will test linking personal name authority records of the Library of Congress and Die Deutsche Bibliothek by using software developed by OCLC for linking the retrospective records. Later VIAF will test to maintain the virtual files using OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) to harvest the metadata for new, updated, and deleted records. The project will also address needs of users, enabling the display of languages and scripts that the user can read.

The objectives of the project correspond to a new view on Universal Bibliographic Control (UBC) emerging from several working groups within IFLA (see FRANAR; chapter 3.2). This new perspective reinforces the importance of authority control, yet puts the focus on the user's perspective (i.e. access control). To still get the benefits of shared authority work and creation of bibliographic records that can be re-used worldwide, VIAF aims to develop a prototype to enable linking authorized forms of names through the authority files of national bibliographic agencies and other regional agencies to create a virtual international authority file.

In order to facilitate authority work, a local system could launch a search against the virtual international authority file, if no match was found locally. The found matches could then be edited by the cataloguer and / or merged or linked with the local authority record.

A possible VIAF user scenario [cf. Tillett 2002]:

A cataloguer at the Library of Congress creates a bibliographic record for a book by an Russian author and wants to link the heading with the local authority record. The local system checks the local authority file and finds no match, so it tells the cataloguer that the heading was not found and launches a search against the Virtual International Authority File. A match is found with an authority record created at the National Library of Russia in St. Petersburg. The cataloguer doesn't want all the information but likes a reference and wants a link, so the local system asks the cataloguer if s/he wants the system to create a basic authority record from the one found and to make a link to it. The local system then automatically builds a local basic authority record, grabbing the linking information

from the Virtual International Authority File (the record from the National Library of Russia). The cataloguer then adds the MARC21 field #100 (authorized form) according to the locally used cataloguing rules, in this case AACR2. The cataloguer also adds other fields if needed. The local system adds the MARC21 linking #700 field (the linking authorized form, the record control number and the source information for future linking). This linking of authority files would primarily be among the national or regional authority files of national bibliographic agencies. Thus a link in the virtual international authority file between the authorized form following AACR2 and the Russian record for the same entity is added. The local system updates the local bibliographic record.

A central aspect of the VIAF concept is that an entity can be expressed in several languages and scripts. Parallel authority records for the same entity should be preserved (the MARBI group explores how to integrate this possibility in the MARC21 format for authority records, see chapter 3.3). This allows to reflect the national and cultural needs of the individual users and at the same time allows to set up the syndetic structure of cross references and authorized forms of headings to be used in catalogues intended for a specific audience. In order to be of most benefit the entity should be presented to the user in the language / script s/he prefers to read.

The first phase of the project is a proof of concept to test the linking of existing authority files with the name authority files of Library of Congress (LCNAF) and Die Deutsche Bibliothek (PND, Personennamendatei). The project will use automatic matching algorithms developed by OCLS, building on bibliographic clues for machine matching at a fairly high level of accuracy.

In the second phase of the project a centralized model for a Virtual International Authority File will be tested: It uses the Open Archives Initiative Protocol (OAI-PMH) to harvest metadata from existing authorized authority files. The harvested data will reside in one or more servers and will be refreshed whenever changes are made in any of the national files. This means the day to day record maintenance activities continue to be managed as they are nor by the National Bibliographic Agency.

This shared Virtual International Authority File could also be an integral part of a future "Semantic Web". It would be linked with other controlled vocabularies in the WWW, like the controlled vocabularies of abstracting and indexing services, biographical dictionaries, telephone directories, and other online reference tools and resources and thus help users to navigate in the internet and to improve the precision of searches. These tools would link to their respective databases for bibliographic and other resources. For example the Library of Congress authority records, being part of the Virtual International Authority File, would link to the bibliographic and holdings data bases of the Library of Congress or even to the digital repositories of the linked objects themselves.

3.8 AFNOR Experts group „Métadonnées d'autorité”

3.8.1 Management details

Status: Working Group within AFNOR (Association Française de Normalisation; French organization for standardization) under the leadership of Françoise Bourdon (Bibliothèque nationale de France). The WG (AFNOR/CG46/CN357/GE4) is attached to the Commission "Modelling, Production and Access to documents" (CN357) chaired by Odile Artur (Université de Marne-la-Vallée). This Commission is part of the general commission 46 (CG46) "Information and Documentation" chaired by Marcelle Beaudiquez (Bibliothèque nationale de France)

Duration: June 2000 – (ongoing).

Members:

- Agence Bibliographique de l'Enseignement Supérieur (ABES)
- Association des Utilisateurs du Réseau OCLC en France (AUROC)
- Bibliothèque nationale de France (BNF)
- Bibliothèque Universitaire de Nice
- École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB)
- Bibliothèque de la Fondation Nationale des Sciences Politiques
- Institut national des Techniques de la Documentation (INTD)

- Médiathèque de la Cité des Sciences et de L'industrie
- Direction des Archives de France
- Centre Historique des Archives Nationales (CHAN)
- Direction des Musées de France
- Département des Estampes de la Bibliothèque Nationale de France
- Bibliographie d'Histoire de l'Art (BHA)
- Institut National de l'Audiovisuel (INA), Direction de la Recherche et Département Inathèque
- Société des Auteurs et Compositeurs Dramatiques (SACD)

Website: www.afnor.fr/ → "Espace Normalisation" → "Comités Electroniques" → "Technologies de l'information et de la communication" → "Documentation"

3.8.2 Summary of Objectives

The group consists of experts from different domains (libraries, archives, museums, documentation centres, rights management agencies, multimedia agencies etc.) and aims to carry out an interdisciplinary inquiry on the processing of authority files in different sectors [see Bourdon 2001]. One of the objectives is to propose a model for authority data that is based on the works in progress in this field at the international level.

3.8.3 Proceedings

In order to identify the points of convergence and divergence between the different sectors, the participants contribute reports on the usage of authority data in their respective fields. These different reports make it possible to discover the objectives, procedures and implementations in the various sectors represented in the WG and to develop a common authority knowledge within participants. A lot of problems were pointed out concerning terminology used in different sectors, or concerning the content of a minimal authority record. The WG decided to elaborate a dictionary of authority data, recording all elements of authority data already used or expected by at least one sector, and to aim at a consensus on a common definition of all these data elements.

3.8.4 Outcome / Results

The concepts "person", "corporate body", "object", "place", "support", "technique", "work" are defined so far. The WG is inspired by the CRM model for museums, FRBR, and various other standards. On the basis of understanding the different practices and terminologies (reference, thesaurus, authority list, authority file, lexicon etc.) the WG aims in organizing these data in a model for authority metadata taking into account the needs of all the concerned sectors (libraries, archives, museums, end users, rights managers, etc.), the state of the art at the international level and the models already defined for similar data. The WG will liaise with the FRANAR Working Group to see if the proposed solutions will properly address the needs of the international community.

3.9 ALCME NDLTD Authority Linking Proposal

3.9.1 Management details

Status: Part of the ALCME (Advanced Library Collection Management Environment) at the OCLC Office of Research: Study for the NDLTD project (Networked Digital Library of Theses and Dissertations³⁵).

Duration: 2001– (ongoing).

Members: OCLC Office of Research.

Websites:

- ALCME project: <http://alcme.oclc.org/>

³⁵ <http://www.ndltd.org/>.

- ALCME NDLTd projects: <http://alcme.oclc.org/ndltd/>
- Authority linking proposal: <http://alcme.oclc.org/ndltd/AuthLink.html>
- LAF example: <http://alcme.oclc.org/laf/>

3.9.2 Summary of Objectives

ALCME is a project of OCLC with the goal to create a set of open-source tools to seamlessly integrate the components needed to build a distributed library collection management system. These tools support services as repository interfaces, metadata harvesting, and authority control.

Part of the project was a study for the NDLTd project, see [Suleman et al. 2001] concerning a name authority linking mechanism for NDLTd's theses and dissertation records.

3.9.3 Outcome / Results

Part of the project was a study for the NDLTd project (Networked Digital Library of Theses and Dissertations, see [Suleman et al. 2001]) concerning a name for NDLTd's theses and dissertation records. The proposal for the authority linking mechanism has been published on 12 April, 2001 [Young 2001]. The key assertion of this proposal is, that the problem of authority control is well suited to a decentralized, peer-to-peer networked solution. Design goals are:

- The model must be simple and easy to understand.
- The system must be easy to deploy and maintain.
- Barriers to participation should be made as low as possible by using open standards and open-source software.
- Authority records should be freely created and shared among participants.
- The URIs used for authority links should be persistent and current.
- The URIs used for authority links should be meaningful and useful to anyone outside NDLTd's domain.
- The network should work properly independent of any individual institution.
- The network should be capable of seamlessly supporting centralized service providers that can streamline inefficiencies in the peer-to-peer model.
- The model should be scalable and applicable beyond NDLTd.
- The model should support access control allowing locale and preference-specific variations in the form of the names.

[Young 2001] proposes using the OAI protocol for sharing name authority records and using a Dublin-core based element set for describing name authority data. As DC elements are bibliographic in nature, they aren't appropriate for name authority information. Therefore a set of extension should be defined specifically for name authority information (see DC Agents Working Group). Name authority records in this DC Name Authority Metadata Standard (called NAMS) would be available from OAI repositories via an *oai_nams* metadataPrefix. The repository itself is called Linked Authority File (LAF). Participants will be able to attach annotations to the LAF records. This feature will be accomplished by using RDF metadata (Resource Description Framework), as it is a major feature of RDF to allow for the creation of collaborative metadata about the resource being described. Automated tools could process these annotations to aid in the identification and resolution of data problems by authorized personnel in the network. The representation of such annotations in user views should also help cataloguers to identify and differentiated individuals and institutions.

In NDLTd's Electronic Thesis and Dissertation Metadata Standard (ETDMS) each reference to an individual or institution in an ETDMS field should contain a string representing the name of the individual or institution as it appears in the work. In addition, these references also may contain a URI that points to an authoritative record for that individual or institution.

- LAF records should be freely created and shared among NDLTd participants. While a *central* authority database is an option, the LAF design expects the database to be *distributed* to share cost. Individual participants or groups should be able to host a copy of the LAF database and share

changes they make to local copies of LAF records with other hosts using the OAI-PMH. A new record should be entered in an institution's home repository and harvested by other institutions in a timely manner. Those institutions, in turn, are free to attach annotations to the record within their own repository and redistribute their amended copies to the network asynchronously. Although the official copy used for linking purposes will reside in the originator's repository, the copies that reside in the other repositories are all fully functional facsimiles. The mechanism for keeping records in sync is described in [Young 2001].

- The URIs should be meaningful and useful to anyone outside NDLTD's domain. A benefit of using the OAI-PMH is that individual LAF records will be accessible via an OAI *GetRecord* request.
- The URIs should be persistent and current. This raises a number of challenges, such as duplicate resolution. By using PURLs³⁶ in ETDMS records, the underlying OAI *GetRecord* URLs can be rearranged without affecting the ETDMS records that rely on them.
- The model should be scalable and applicable beyond NDLTD. The LAF model was designed to work entirely with open standards and open-source software.

[Young 2001] summarizes the advantages of a networked collection of name authority data:

- Only the level of participation among the decentralized participants limits the coverage of such a collection.
- Using XML namespaces, XML schemas, and RDF, the content of LAF records can be as broad or as narrow as the group chooses.
- Access to name authority data is trivial and instantaneous from any context given a useful URI such as the PURL scheme described in this document.
- Existing authority records around the world can be combined into a single LAF record that could support locale-specific access control of the name variations.
- Undifferentiated name records would no longer be black holes, sucking up useful information and making it unusable.

3.10 International Council on Archives– Committee on Descriptive Standards (ICA–CDS)

3.10.1 Management details

Status: Committee of the International Council on Archives (ICA)³⁷.

Duration: 1996– (ongoing; successor of the ICA Ad Hoc Commission on Descriptive Standards [ICA/DDS], 1990–1996)

Members: A roster is available at <http://www.hmc.gov.uk/icacds/eng/membership.htm>.

Website: <http://www.hmc.gov.uk/icacds/icacds.htm>.

3.10.2 Summary of Objectives

The Mandate of ICA-CDS is defined as follows: "The Committee shall promote and make possible the exchange of information on archival material at the international level by specifying requirements for standards related to archival description that are acceptable to national archival administrations and the international archival profession."³⁸

Presently ICA-CDS is reviewing *the International Standard Archival Authority Record for Corporate Bodies, Persons and Families* (ISAAR (CPF) 1995). The Committee will elaborate a substantial revision of the standard. This revision will take into account the EAC (Encoded Archival Context) DTD which is presently being developed (see chapter 3.11).

³⁶ Persistent URLs, see <http://purl.oclc.org>

³⁷ <http://www.ica.org/>.

³⁸ <http://www.hmc.gov.uk/icacds/eng/mandate.htm>.

3.10.3 Proceedings

At a meeting in Brussels on 3–6 October, 2001 the Committee decided that in the light of the EAC initiative a substantial revision of ISAAR (CPF) would be necessary in order to accommodate to this emerging standard. In the Brussels meeting the Committee produced a changed ISAAR (CPF) structure based on EAC ([ICA-CDS 2001], Appendix A).

During the next meeting in Madrid on June 12–15, 2002 this draft was further elaborated. The draft of the revision will be completed at a meeting in Rio de Janeiro in November 18–22, 2002 and mounted on the ICA-CSD Website for an open review period of 6 months closing in May 2003. A final version of the standard will be approved at the Committee's Australia meeting in 2003 and presented in published form to the 15th International Congress on Archives, Vienna in August 2004.

3.10.4 Outcome / Results

The structure of the draft for the ISAAR (CPF) revision is described in [MARD-3], Chapter 4.1.2. A crosswalk ISAAR (CPF) / Madrid Draft to EAC / Alpha Version is provided in [Ottosson 2002a].

3.11 Encoded Archival Context (EAC) Working Group

3.11.1 Management details

Status: Ad Hoc Working Group.

Duration: March 2001– (ongoing).

Members: A roster is available at: <http://www.library.yale.edu/eac/members.htm>.

Websites:

- EAC Website: <http://www.library.yale.edu/eac/>.
- EAC Working Group Website: <http://www.iath.virginia.edu/eac/>.
- LEAF EAC Sample Page: <http://helmer.hit.uib.no/~gunnar/>.

3.11.2 Summary of Objectives

“Encoded Archival Context (EAC) is an ongoing initiative within the international archival community to design and implement a prototype standard based on Extensible Markup Language (XML) for encoding descriptions of record creators. The primary developers of this prototype standard are members of the international archival community. The description of individuals, families, and organizations that create records is an essential component of the documentary evidence of human activity. Identifying record creating entities, recording the names or designations used by and for them, describing their essential functions, activities, and characteristics, and the dates and places they were active is an essential component of the management of archival records. Creator description facilitates both access to and interpretation of records. Description of creators is also essential in bibliographic systems, and in museum documentation, and thus EAC may be of interest to other cultural heritage communities as well. As custodians of the records upon which biographies and organization histories are based, archivists are well-placed to develop a standard that will assist in the fulfilment of their professional responsibilities, and at the same time lay the foundation for building international biographical and organization history reference resources. [...]

EAC is intended to extend and complement EAD [i.e. Encoded Archival Description³⁹]. EAC will support the descriptive needs of the archival community, specifically in the creation, maintenance, and publication of creator description. In addition to the international archival community, the EAC effort will be useful in the broader context of libraries and museums, formal authoring and publishing of biographical and organization histories, and the amassing of large genealogical databases. While the EAC working group anticipates interrelations between EAC descriptions and genealogical information, it should also accommodate existing authority, biographical, and historical data. There are several possible sources of authority and biographical and historical data.” [Pitti 2001]

³⁹ <http://www.loc.gov/ead/>.

3.11.3 Proceedings

With support from the Digital Library Foundation (DLF), an international group of archivists and information scientists met at Yale University in December 1998 and laid out a program for the development of an Archival Context encoding standard. On March 3-6, 2001 an international group met in Toronto 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.

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 Per-Gunnar Ottosson (National Archives of Sweden). During a meeting in Charlottesville in June 21-23, 2001, the group came up with a draft DTD, which was ready for circulation to the full group in the middle of July.

The DTD is presently being tested on LEAF data by Gunnar Karlsen, UoB (see [RND] and <http://helmer.hit.uib.no/~gunnar/>).

3.11.4 Outcome / Results

A detailed discussion of the EAC DTD and its applicability for the LEAF project is given in [RND], chapters 5-7.

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 ISAAR(CPF) (see chapter 3.10). 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. A crosswalk ISAAR (CPF) / Madrid Draft to EAC / Alpha Version is provided in [Ottosson 2002a].

Overview of structure [Ottosson 2002b]:

- 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.

A more detailed description is given in [Pitti 2001] and [RND].

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 [see RND].

Since September 2002, the Alpha Version of the EAC Tag Library is available at <http://www.iath.virginia.edu/eac/>.

3.12 Interoperability of Data in E-Commerce Systems (<indecs>)

3.12.1 Management details

Status: Project, supported under the Info 2000 Programme of the European Commission.

Duration: November 1998 – March 2000 (17 months, finished).

Members:

- Authors' Licensing and Collecting Society Limited, London, UK (ALCS), <http://www.alcs.co.uk/>
- Music Copyright Operational Services Ltd., London, UK (MCOS)
- Verwertungsgesellschaft BILD-KUNST, rV, Bonn, Germany, <http://www.bildkunst.de/>
- Société des Auteurs et Compositeurs Dramatiques (SACD), Paris, France, <http://www.sacd.fr/>
- International Federation of the Phonographic Industry (IFPI), London, UK, <http://www.ifpi.org/>
- European Book Sector Electronic Data Interchange Group (EDITEUR), London, UK, <http://www.editeur.org/>
- Kopiosto, Helsinki, Finland, <http://www.kopiosto.fi>
- Centrum voor Dienstverlening Auteurs - en aanverwante Rechten (CEDAR), Hoofddorp, The Netherlands, <http://www.cedar.nl>
- Copyright Agency Limited (CAL), Sydney, Australia, <http://www.copyright.com.au/>
- Muze UK Ltd, Croydon, UK, <http://www.muze.com/>
- MDR Partners, London, UK, <http://www.mdrpartners.com/IE/home.htm>

Website: <http://www.indecs.org/>

3.12.2 Summary of Objectives

The project focused on the practical interoperability of digital content identification systems and related rights metadata within multi-media e-commerce. It was an international effort intended to generate a formal structure for describing and uniquely identifying intellectual property (IP), the people and businesses involved in trading IP on the Internet, and the agreements people make about those online sales. [Indecs 200a, 1]

3.12.3 Proceedings

The 17-month project began work in November 1998 and formally ended as a European Commission funded project in March 2000. At the close of the project, <indecs> partners established a not-for-profit, membership *Indecs Framework* Limited organisation. The Framework organisation holds the <indecs> IPR rights and has the goal of working with partners to move forward the work of the <indecs> initiative.

3.12.4 Outcome / Results

The <indecs> framework encompasses a metadata model, and a high-level metadata dictionary, principles for mappings to other schemas, and a Directory of Parties Proposal. The <indecs> metadata framework recognises [Indecs 200a, 1 and Rust/Bide 2000, 8]:

- metadata relating to all types of creations,
- the integration of descriptive metadata with commercial transactions and rights,
- a generic attribute structure for all entities,
- the central nature of events in the development of interoperable metadata; and
- that as far as possible metadata should be created once and used for different purposes.

The <indecs> meta data framework is based on four guiding principles for the development of “well-formed” metadata to support effective e-commerce [Rust/Bide 2000, 8–11]:

- **Unique Identification:** Every entity should be uniquely identified within a unique namespace.

- **Functional Granularity:** It should be possible to identify an entity whenever it needs to be distinguished.
- **Designated Authority:** The author of an item of metadata should be securely identified.
- **Appropriate Access:** Everyone requires access to the metadata on which they depend, and privacy and confidentiality for their own metadata from those who are not dependent on it.

The <in decs> model elaborates a logical and semantic framework for describing *entities*, their *attributes* and *values* of each. <in decs> defines *entity* a something which is identified and takes into account three distinct views of entities: a general view, a commerce view and an intellectual property view. These three views allow to describe the main metadata concerns in relation to e-commerce. <in decs> asserts that the attributes of any entity may be usefully understood as being of five types: labels, quantities, qualities, types and roles, each of which has its own particular structure and behaviour.

With the increasing granularity of the Intellectual Property being traded, metadata is never likely to come from a single source or to follow a single standard for identification and description. If metadata from different sources is to interoperate successfully, it must, though, be developed within a coherent and consistent view of the things that are being described so that such views can be successfully mapped to others. Central to the analysis is the assumption that it is possible to produce generic mechanism to handle complex metadata for all different types of IP.

A complete and detailed description of the <in decs> Metadata Framework is given in [Rust/Bide 2000].

A further accomplishment of the <in decs> project is a proposal for a “**Directory of Parties**” [Bide 1999]. In e-commerce it is necessary to identify parties (individual human beings or ‘legal people (companies and the like)’ who make and use creations and have rights in them. Names are inadequate because they are not unique and change over time. So a mechanism for the **unique and persistent identification of parties** is an underpinning of the entire <in decs> schema.

Some international identifiers are already established or in development. The <in decs> “Directory of Parties” proposal offers a “proposal” on how to enable these and other existing and emerging initiatives to **interoperate so that people can be identified unambiguously throughout the network**. The aim of this aspect of <in decs> work was to develop consensus on the structure, function and operation of a commonly accessible directory of parties. This has resulted in a general specification for the establishment of such a directory or network of directories. The directory should allow any person participating in the making or use of a creation, or in a deal relating to intellectual property, to be unambiguously identified while at the same time protecting both the privacy of individuals and commercial confidentiality, as far as this is possible and appropriate.

The proposal sets out the issues relating to the structure, function and operation of a commonly accessible Directory of Parties. It provides a discussion of issues, a review of current initiatives and work, the issues involved in striving for interoperability in this area, and a plan for proceeding on creating a Directory of Parties. The proposal is not considered a replacement for existing schemes for the identification of participants in intellectual property domain where these already exist. Rather, it proposes a way forward for achieving interoperability.

In addition to identifying right owners (legal persons), in order to manage intellectual property rights a directory of parties also has to recognise and identify other entities which do not have legal status but which may act or be perceived to act in a creative or commercial capacity with respect to “creations.” Within the <in decs> model, any entity playing an active role in the making or use of a creation is said to be taking an *agent role*. Typically an agent is an individual or group of human beings, but it may also be an animal, or a computer – or it may be any other entity. Parties include the whole range of agents and perceived agents, including stage personae, animals, characters in plays, films or cartoons, and company trading names like publisher imprints and record labels. Legal status plays no direct part in the definition of parties in the <in decs> commerce model. Parties may be identified as playing a creative or commercial role without necessarily having the capability to enjoy or exercise intellectual property rights.

In order to adequately identify parties (and differentiate them from other parties) descriptions are necessary. These metadata elements may include:

- Events associated with a party (date / place formation of a company; the date / place of birth of an individual human being; etc.).

- Associations with particular genres of creation (publisher, actor, author, etc.).
- Relationships with other parties.

Even if the immediate attention of <indec> is exclusively on the identification of rights owners rather than rights users, there are many questions of privacy to be considered. These include [Bide 1999, 10]:

- What metadata relating to parties needs to be made available for the purposes of disambiguation and to whom should that metadata be made available?
- Who has the right to create and maintain this metadata?

The directory of parties proposal identifies the following functions of a Party Directory [Bide 1999, 11f]:

- The use of a unique identifier makes the processes more certain and less error prone.
- The primary function of a party directory within the <indec> schema is to allow multiple instances of the same identity (in multiple databases) to be linked through a single unique identifier.
- Bearing in mind that the same party will inevitably appear in multiple metadata sets, this implies the development of navigational tools which use the unique identifier to link metadata sets together.
- In order to ensure the proper maintenance of a system of unique identity, the directory will need to allow either a party, or another party delegated to act on behalf of that party, to find out whether or not a unique identifier already exists for that party. This implies that a certain amount of metadata must be available for public inspection.
- The most commonly used differentiator between individual human beings is place and date of. It has to be explored to which extent a name and place-and-date-of-birth combination fails to guarantee uniqueness within the global community.
- There may also be issues relating to both individual privacy and commercial confidentiality in the metadata relating to group parties (whether legally constituted or not), and the relationships of individuals with those groups.

The proposal identifies the following high-level functional requirements of a Directory of Parties [Bide 1999, 13f]:

- Registration of identities with the Directory of Parties should be mediated by a range of agencies.
- The Directory of Parties should provide to participating agencies facilities for online checking of identities of parties and the registration of "new" parties.
- Registration will require the two-level approach to identification (base number, name number). Every party registered must have a base number and a minimum of one name number; there should be no constraint on the number of names by which a party is known.
- Some "kernel" metadata will need to be registered alongside every party registered to allow for differentiation of parties who use the same name.
- The process of checking identities and the registration of new identities will need to be designed to allow it to be fully integrated into the workflows of the agencies which use the Directory. This implies not only 24 x 7 access but also the highest possible degree of machine-to-machine interaction.
- The Directory of Parties should provide a link between all participant identifier schemes and a central point of access to participant databases.
- The Directory of Parties should respect the integrity of participant databases and identifiers. This should include the establishment of mechanisms to ensure that amendments made to registrations can be accurately and promptly reflected in the relevant participant databases.
- The Directory of Parties should be operated in such a way that the identity of the agency or individual creating any specific item of personal metadata can be preserved in the record and that any changes to metadata are fully auditable. This is particularly important bearing in mind the different significance that agencies will place on the importance of accuracy in identification (libraries may have fewer concerns about inaccuracy in a name authority than commercial organisations will have in identifying people to whom they owe money).
- It will be necessary for the operation of the Directory of Parties that it be possible for multiple – and potentially conflicting – metadata sets concerning the same party to be declared by different agencies. The Directory of Parties would not be responsible for independent resolution of such

conflicts, but would need to recognise conflicts and bring them to the attention of the agencies concerned so that any dispute can be resolved in a timely and effective manner.

- Operation of the Directory of Parties must be sensitive and responsive to the potential hazards which the Directory creates for personal privacy and commercial confidentiality.

3.13 InterParty – A Framework for the Unique Identification of Parties in E-Commerce

3.13.1 Management details

Status: Project, supported under the Information Society Technologies Programme (IST) of the European Commission.

Duration: April 2002 – March 2003 (12 Months; ongoing)

Members:

- European Book Sector Electronic Data Interchange Group (EDItEUR), London, UK, <http://www.editeur.org/>
- Kopiaisto, Helsinki, Finland, <http://www.kopioisto.fi>
- Book Data Ltd., London, UK, <http://www.bookdata.co.uk/>
- Royal Library, Stockholm, SE, <http://www.kb.se/>
- British Library, London, UK, <http://www.bl.uk/>
- International Federation of Library Associations and Institutions (IFLA), The Hague, NL, <http://www.ifla.org/>
- Library of Congress, Washington/DC, USA, <http://www.loc.gov/>
- OCLC, Dublin, Ohio, USA, <http://www.oclc.org/>
- Corporation for National Research Initiatives (CNRI), Reston, Virginia, USA, <http://www.cnri.reston.va.us/>
- International DOI Foundation, <http://www.doi.org/>

Website: <http://www.interparty.org/>

3.13.2 Summary of Objectives

Building on the work of the **<in decs> Project** (see chapter 3.12) InterParty will specify the blueprint for a network that will provide participating agencies with a means of online, on-demand, checking of identities – a virtual Directory of Parties – and will enable party identifiers to be mapped automatically between domains.

InterParty will develop a framework for the unique identification of parties in e-commerce, which is essential to the effective trading of intellectual property (IP) relating to cultural and information goods. Collecting societies, performer databases, national libraries, bibliographic agencies all have the need to disambiguate between different parties who share the same name and to be aware of the same party using different names. The specification of a Directory of Parties will involve three main activities:

- the establishment and maintenance of the technical infrastructure;
- the quasi-bibliographical function of data management and registration;
- policy and governance of the system.

These might be eventually be separated from each other or might all be undertaken by the same organisation. The first two activities could be carried out on a commercial service basis assuming that an appropriate economic model can be developed. Policy and governance should be managed outside the purely commercial arena (although these activities too will require an appropriate level of funding).

Work on the issue of identification of parties for e-commerce in cultural content has been mainly conducted in two sectors: **collecting societies** — bodies concerned with the collection of royalties on behalf of authors,

creators and performers; and **libraries**, who have sought through name authority control to derive a range of benefits in their acquisitions and cataloguing workflow, procurement activities and in the provision of services to users.

InterParty is concerned to establish a model and working system which can be used by bodies involved at all stages of the e-content lifecycle. The Directory of Parties as proposed by InterParty can be used to underpin rights transactions across a wide range of creative media sectors. <indecs> has already developed a specification for the establishment of a commonly accessible Directory of Parties, but this is just an outline, and does not deal in detail with the technical components of a working system. It will, however, serve as a starting point for InterParty's work.

Requirements for the Directory of Parties:

- The Directory of Parties should not only be neutral in respect of its relationships with existing interests, but must be seen to be neutral. The establishment of an appropriate policy and governance model for sustainable operation, once the InterParty project is completed, will therefore be a central requirement.
- The operation and governance of a centralised Directory of Parties of the type envisaged will need to be distributed (and thus delegated) which increases the need for appropriate governance mechanisms. Ultimately, the most significant question relates to who should be enabled to establish unique identities. This raises the question of authentication of identity.
- Registration of identities with the Directory of Parties should be mediated by a range of agencies. InterParty will help to establish who can be allowed to declare and maintain metadata about a party and in what circumstances. The Directory of Parties will require mechanisms for the identification and resolution of disputes.
- The demonstration of an appropriately secure environment and set of procedures will be an important element. The operation of the Directory of Parties must be sensitive and responsive to the potential hazards which the Directory creates for personal privacy and commercial confidentiality.
- The Directory of Parties is not proposed as a replacement for existing schemes for the identification of participants in the intellectual property domain where these already exist (e.g. national library name authority files or systems oriented towards the needs of rights licensing) but as a means of effecting their interoperation. This will also require the development of navigational tools that use the unique identifier to link metadata sets together.

3.13.3 Proceedings

Within the 12 months of project duration InterParty aims to deliver [Bide 2002]:

- Analysis of existing data models
- Identifiers and party metadata model
- Report on privacy and security mechanisms
- Specification for working demonstrator
- Demonstrator (Alpha System)
- Business model, exploitation plan and governance proposals

4 MARD 2: Identification of Partnerships – Specific Relevance and Contact Details

This chapter, being the second part of part 2 of the MARD, describes the specific relevance for the LEAF project of those projects / initiatives that were listed in the previous chapter.

4.1 MALVINE (Manuscripts And Letters Via Integrated Networks in Europe)

4.1.1 Relevance for LEAF project

- MALVINE will be the test bed for the LEAF system.

4.1.2 Contact Persons

- Jutta Weber (Staatbibliothek zu Berlin, Zentralkartei der Autographen, MALVINE Co-ordinator), jutta.weber@sbb.spk-berlin.de
- Hans-Jörg Lieder (Staatbibliothek zu Berlin, Handschriftenabteilung, MALVINE Co-ordinator), hans-jorg.lieder@sbb.spk-berlin.de

4.1.3 LEAF Liaison

- Most of the MALVINE consortium member institutions are members of the LEAF consortium as well.

4.2 The National Register of Archives (NRA)

4.2.1 Relevance for LEAF project

- The entire NRA data will be used in the test phase of the LEAF system.

4.2.2 Contact Persons

- Dick Sargent (Director of the National Register of Archives), dicks@hmc.gov.uk

4.2.3 LEAF Liaison

- The National Register of Archives is a LEAF observing partner.

4.3 IFLA Working Group on Functional Requirements And Numbering of Authority Records (FRANAR)

4.3.1 Relevance for LEAF project

- Proposal of a conceptual model for the distinction between “bibliographic” and “real world” entities of authority files, which is important for achieving interoperability between library authority files and those of other agencies (archives, museums, rights management organizations, etc).
- Mapping of user tasks to conceptual entities of authority files.
- Analysis of the feasibility of an International Standard Authority Number (ISADN).
- Liaison with other relevant initiatives: dissemination of LEAF findings.

4.3.2 Contact Persons

- Glenn Patton (Chair), OCLC (Director, Metadata Standards and Quality); glenn_patton@oclc.org
- Françoise Bourdon (Bibliothèque nationale de France), francoise.bourdon@bnf.fr
- Andrew MacEwan (British Library), Andrew.MacEwan@bl.uk

- Barbara Tillett (Library of Congress), btil@loc.gov
- Christel Hengel (Die Deutsche Bibliothek), hengel@dbf.ddb.de

4.3.3 LEAF Liaison

- The British Library is member of the LEAF consortium; OCLC, Bibliothèque nationale de France, Die Deutsche Bibliothek and the Library of Congress are LEAF observing partners.
- LEAF has been presented at the FRANAR meeting on 10 June, 2002 by Chris Fletcher (British Library).

4.4 MARBI Multilingual Record Task Force

4.4.1 Relevance for LEAF project

- Proposal for changes to the MARC21 authorities format that would support the identification of equivalent headings in multiple languages and / or scripts ("catalogue context"). This proposal could be considered in LEAF for the transformation of LEAF EAC records into MARC21 format.

4.4.2 Contact Person

- Josephine Crawford (Chair), University of Minnesota, Bio-Medical Library; jocraw@tc.umn.edu

4.5 Authority Control in the Online Environment Interest Group (LITA/ALCTS)

4.5.1 Relevance for LEAF project

- Dissemination of LEAF results.

4.5.2 Contact Person

- Ann Della Porta (Chair); Library of Congress, Assistant Coordinator, Integrated Library System Program adel@loc.gov

4.6 NACO Program for Cooperative Cataloguing (NACO PCC)

4.6.1 Relevance for LEAF project

- Proposed NACO interim model for alternate authorized headings which uses the #7XX field in the MARC21 format. This model could be considered in LEAF for the transformation of LEAF EAC records into MARC21 format.

4.6.2 Contact Person

- Barbara Tillett, Library of Congress, btil@loc.gov

4.6.3 LEAF Liaison

- Library of Congress is a LEAF observing partner.

4.7 Bibliographic Control of Web Resources. A Library of Congress Action Plan

4.7.1 Relevance for LEAF project

- Dissemination of LEAF results.

4.7.2 Contact Person

- Barbara Tillett, Library of Congress, btil@loc.gov

4.7.3 LEAF Liaison

- Library of Congress is a LEAF observing partner.

4.8 Virtual International Authority File (VIAF)

4.8.1 Relevance for LEAF project

- Model for automatic linking of name authority records.
- Using OAI-PMH for sharing of name authority records.
- Model for international maintenance of a virtual authority file
- User-oriented approach.

4.8.2 Contact Persons

- Ed O'Neill, OCLC; ed_oneill@oclc.org
- Barbara Tillett, Library of Congress; btill@loc.gov
- Christel Hengel, Die Deutsche Bibliothek; hengel@dbf.ddb.de

4.8.3 LEAF Liaison

- OCLC, Library of Congress and Die Deutsche Bibliothek are LEAF observing partners.

4.9 AFNOR Experts group „Métadonnées d'autorité”

4.9.1 Relevance for LEAF project

- Cross-domain collaboration in the field of name authority data.
- Cross-domain data model for name authority data.

4.9.2 Contact Person

- Françoise Bourdon, Bibliothèque nationale de France; francoise.bourdon@bnf.fr

4.9.3 LEAF Liaison

- Bibliothèque Nationale de France is a LEAF observing partner.

4.10 ALCME NDLTD Authority Linking Proposal

4.10.1 Relevance for LEAF project

- Using OAI-PMH for sharing of name authority records.
- Model for annotation of shared records.
- Model for de-duplication.
- Model for persistent identifiers.

4.10.2 Contact Person

- Jeff Young, Consulting Software Engineer, OCLC, Office of Research; jyoung@oclc.org;
<http://www.oclc.org/research/staff/young/index.shtml>

4.10.3 LEAF Liaison

OCLC is a LEAF observing partner.

5 MARD-3: Analysis of Practice

5.1 Introduction

5.1.1 Scope

The purpose of this document, being part 3 of the MARD, is to provide an analysis of actual practices regarding name authority information and data structures among LEAF partners and external agents. The analysis is focussed on authority records used as access points to archival material, modern manuscripts, and letters.

It is natural that a development project such as LEAF focuses on the requirements of the data model and technology. However, the outcome of the project will in the end be dependent of the quality and consistency of data provided. The overall question is whether the data that will be provided reach the standard the proposed data model presumes.

This preliminary evaluation of the quality of data provided should be followed up in the forthcoming Work Packages 8 "Test of the Demonstrator" and 9 "Assessment and Evaluation".

The accuracy of links in the document has been tested 5th November 2002.

5.1.2 Document Overview

Ch. 5.2 gives the background and the general requirements on data quality the project presumes. In Ch. 5.3 there is a general survey of the data providers with short descriptions of the local systems. Ch. 5.4 is focussed on general issues concerning rules and principles, while the actual practice is described and discussed in Ch. 5.5 from a selection of samples. The conclusions in Ch. 5.6 include a preliminary evaluation of the quality of data provided for the objectives of LEAF, and suggestions for priorities and quality assurance of local systems.

5.2 Data Quality and Model Requirements

5.2.1 Data Quality and Authority Records

The Data Model Requirements, as defined in the Model Requirements Analysis Documents: *1: Analysis and Evaluation of Input Data* (MRAD1) and *2: User Groups Statements and Special User Scenarios* (MRAD2), implies that the data provided must meet a certain standard of data quality.

ISO 9000 defines quality as **the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs**". When it comes to measure data quality one usually discusses attributes such as accuracy, completeness, consistency, reliability, timeliness, uniqueness, and validity (for a discussion definitions and dimensions of data quality see e.g. ABATE & al. 1998, <http://www.dataquality.com/998abate.htm>).

Also the concept of authority records as such implies that the data should meet very high standards of these characteristics.

A record in an authority file for which the organising element is the authorized heading for an entity (person, corporate body, or work/expression) as established by the cataloguing agency responsible. In addition to the authorized heading, the record contains, as applicable: information notes; a record of all variant and related headings from which references have been made (tracings); notes recording sources consulted, etc.; an identification of the cataloguing agency responsible for the entry; and (when implemented) the International Standard Authority Data Number (ISADN).

Quoted from *Guidelines for Authority Records and references* (München, 2001), 0.2; cf. *Mandatory Data Elements for Internationally Shared Resource Authority Records* (<http://www.ifla.org/VI/3/p1996-2/mlar.htm>)

The user will expect that the cataloguing agency has used the best sources available, and follows professional standards in creating the records. On the other hand, the creation of authority records might develop into an art in its own right without consideration of what the optimal standard is from a user's perspective. Also in the case of authority files we may say that an acceptable level of quality has been achieved when the data conforms to defined requirements reflecting the actual user needs.

5.2.2 Model Requirements

The LEAF project aims to facilitate search, retrieval, and shared access by reducing the redundancy that is likely to appear when records from different files are brought together. This function is described in the report **MRAD1** (ver. 1.0), Ch. 4.1.2 "Grouping LARs in SLARs with links":

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**.

Both the creation of manual and automatic links presumes a very high level of data quality, but the requirements for structure and consistency differ. When users are creating manual or negative links, the structure of data is of less importance.

For the creation of LEAF Automatic Link the requirements of compatible data structures and consistency are very high. The variables must be such that they can be processed in a multilingual environment by an automated procedure.

In **MRAD2**, 3.1, it is presumed that LARs will be linked according to a defined hierarchical set of linking rules, but the actual rules for the linking are not determined. The first requirement is that the number of used linking rules must be flexible. Depending on the actual use of LEAF (either as a stand-alone service or in combination with other applications) the linking rules need to be open to change. The present analysis aims at proposing a number of elements that should be tested. Further quality checks of the linking rules should then be undertaken during the project phase.

The different user groups defined in **MRAD2**, 3.3, must all at least be able to search on a name string for a record concerning a particular person or corporate body, and decide intellectually whether the retrieved record correspond to the entity searched for. A user has the right to expect that a search on a name form generally represented in sources or literature will give result if there is a record of the entity in the system. The added information in the record shall allow for a decisive conclusion.

The criteria for identifying an individual are depending on the context. It is quite a difference between creating data bases covering total populations, and this relatively limited population of people appearing as creators of archives, and authors of literature and letters kept in public repositories.

5.3 Descriptions of Systems

5.3.1 The Data Providers: An Overview

- The current LEAF data providers might be divided into 4 groups
- National registers or union catalogues representing different repositories (SBB, SNL, RA)
- Local systems for archives and manuscripts based on or linking to established authority files (ONB, DLA)
- Authority files for special areas or subjects (FDÖP)
- Authority files for general bibliographic use (BN, UCM)

5.3.1.1 PND and GKD

With the present selection of data providers, German authority files play a crucial role. PND (*Personnamendatei*) is an authority file for personal names consisting of around 1 900 000 records. It started as a project funded by the German Research Council (*Deutsche Forschungsgemeinschaft*) in 1995. Most of these records are gathered from cataloguing projects of *Die Deutsche Bibliothek*, *Die Bayerische Staatsbibliothek*, and, since 1998, from the *Zentralkartei der Autographen* of SBB.

The authority file for corporate bodies GKD (*Gemeinsame Körperschaftsdatei*) started already in the 1970s with data from the database of journals (*Zeitschriftendatenbank ZDB*), *Die Deutsche Bibliothek* Frankfurt am Main, and *Die Bayerische Staatsbibliothek*.

The authority files are available by a Z39.50 gateway for users of the information system PICA/ILTIS of [Die Deutsche Bibliothek](#), and on [CD ROM](#).

5.3.2 Description of Local Systems

The descriptions are based on information available on line or reported by the partners for this or other Work Packages. **Number** of records as reported for [MRAD-1](#)

5.3.2.1 British Library, London, United Kingdom (BL)

Database:

Manuscripts' Catalogue on <http://molcat.bl.uk/>

Rules:

Automated Cataloguing: A Manual and Manuscripts Indexing.
In-house manuals not available on line

Number of records:

Catalogue records: 431 000

The On-line Catalogue is designed to offer a single means of access to the mainstream catalogues of the Department of Manuscripts covering accessions from 1753 to the present day. The greater part of the On-line Catalogue was created by scanning the published volumes. Poor physical condition and print quality have sometimes introduced errors, whilst the quality of the information reflects the varying cataloguing standards of nearly two centuries. The most recent entries represent work in progress which may be edited and expanded (quoted from [About the Manuscripts' Catalogue](#)).

The catalogue has no separate authority file for persons and corporate bodies. Nor has a timescale for the implementation of such a system been reported.

However the manuscript records have entries for personal names that allow for index search on names or additions to names. Information on life dates, place, and nationality are not distinguished but included in a general descriptive adjunct.

5.3.2.2 Biblioteca Nacional (National Library of Portugal), Lisbon, Portugal ([BN](#))

Database:

PORBASE (UNIMARC)
 Central system: WEBPAC (HORIZON) <http://webpac.bn.pt> (not available 5/11/02)
 Alternative systems: PACWEB (PORBASE 5): <http://pacweb.bn.pt>
 SIRIUS: <http://sirius.bn.pt>

Number of records:

Persons: 528 566
 Corporate bodies: 107 480

These refer to the central authority file of the National Library.

The records follow the format of UNIMARC/Authorities, most of them only containing names and dates.

5.3.2.3 Deutsches Literaturarchiv, Marbach, Germany ([DLA](#))

Database:

Kallias (*Katalog des Deutschen Literaturarchivs*)
 URL: <http://www.dla-marbach.de/kallias/aDISWeb/main/index.html>.
 (The data base was not available on line during the preparation of Ch. 5 and 6.)

Number of records:

Persons: 109 210
 Corporate bodies: 30 266

Rules:

Gebrauchsanweisung für die Namensnormdatei (Version 1.5) [i.e. the manual for personal names]
Gebrauchsanweisung für die Körperschaftsnormdatei (Version 1.4) [i.e. the manual for corporate names]
 Not available on line, but provided in electronic form for this WP.

There is a Register of names in the printed guide to the collections in Ingrid Kussmaul: *Die Nachlässe und Sammlungen des Deutschen Literaturarchivs Marbach* (3rd ed. 1998). This guide is also available under the name of *Hyperkuss* on <http://www.dla-marbach.de/kallias/hyperkuss/index.html>. It contains references to 13 800 names in almost 1000 collections.

The complete Hyperkuss has been imported into the new system Kallias, and updates are now only made in Kallias. While Hyperkuss was limited to descriptions of collections in a narrative way, Kallias also has description on item level.

Kallias is based on the library management system aDIS/BMS with special development for the manuscripts department and the department of visual material (museum objects). Kallias has separate authority files and object files in a relational system that also supports multilevel descriptions of the collections.

Kallias provides an integrated online connection to the *Südwestdeutscher Bibliotheksverbund (SWB)* for library material and authority files for personal names and corporate bodies. However, SWB runs its own name authority file only roughly related to the German PND. The GKD, however, is integrated in the SWB, so corporate bodies can be expected to be GKD compatible.

SBB (see [below](#)) recently completed its migration of the ZKA to Kallias-based Software (called Kalliope there), so data formats and other descriptions should be identical. Above all, a Z39.50 gateway will be developed there which is expected to work for Kallias as well.

There are plans to integrate Kallias and Kalliope into a common search engine and to establish an automated replication system of relevant data, but this will not take place in the near future.

5.3.2.4 Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie, Graz, Austria ([FDÖP](#))

Database:

MS-Access database not available on line

Number of records:

Persons and corporate bodies: 7 300

Rules:

Regeln zur Erschließung von Nachlässen und Autographen (RNA)

<http://zka.staatsbibliothek-berlin.de/rna/>

Christoph König, *Verwaltung und wissenschaftliche Erschliessung von nachlässen in Literaturarchiven. Österreichische Richtlinien als Modell* (München: K. G. Saur, 1988)

Biographies with pictures are available on the web page for 13 creators on <http://www.austrian-philosophy.at/bestaende.html>

This database is designed for managing both archival materials and bio-bibliographical data. One of the main objectives of FDÖP is to extend the existing data base file on person names to a comprehensive authority file for Austrian philosophers including related persons and corporate bodies.

5.3.2.5 Goethe- und Schiller-Archiv, Weimar, Germany ([GSA](#))

Database:

Verzeichnis der Bestände (guide to the collections)

<http://www.weimar-klassik.de/gsa/index.html>

This guide provides basic information on the creators (name, dates, profession) and the physical extent of the collection.

At present, various methods of record storage are used for the different research and indexing projects at the Goethe- and Schiller-Archive: allegro, TUSTEP, client/server applications (Delphi 5 / ORACLE 8). Data on persons and institutions are at present administrated by the respective projects themselves. The letter data base provides, for example, data fields for "author" and "addressee", where the respective names are stored (cf. papers for the MALVINE project). It is planned to make use of authority files for names and institutions after the transition to the client/server solution.

5.3.2.6 Institut Mémoires de L'Édition Contemporaine (Institute of Contemporary Publishing Archives), Paris, France ([IMEC](#))

Lists of creators and collections on <http://www.imec-archives.com/fonds/fond1.asp>

5.3.2.7 Narodna in Univerzitetna Knjižnica (National and University Library), Ljubljana, Slovenia ([NUK](#))

Database:

NUK – Normativna in bibliografska baza (NIBB)

URL: <http://enuka/nibb/System/Index.asp>

(only for known IP addresses, due to firewall security)

Rules:

PPIAK - Verona, E., *Pravilnik i prirucnik za izradbu abecednih kataloga.*

AACR2R

UNIMARC/Authorities

Number of records: Persons and corporate bodies: 315

The NUK database NIBB was created for the LEAF project on the basis of the previous data file. It is an internet application that replaces unstructured word files.

(The data base was not available on line during the preparation of Ch. 5 and 6.)

5.3.2.8 Österreichische Nationalbibliothek (Austrian National Library), Vienna, Austria ([ÖNB](#))

Database:

NAK (*Nachlaß- und Autographenkatalog der Österreichischen Nationalbibliothek*)
<http://nak.onb.ac.at/cgi-bin/allegro/nak/nak.pl>

Number of records:

Persons: 4819
Corporate bodies: 763

Rules:

RNA - *Regeln zur Erschließung von Nachlässen und Autographen*
<http://zka.staatsbibliothek-berlin.de/rna/>

Bernd Reifenberg, *HANS Anwender-Handbuch*. 1-3. Lieferung (2001).

Local procedures:

Max Kaiser, "Ansetzung von Personennamen und Körperschaften in Nachlaß- und Autographenkatalog (NAK) der Österreichischen Nationalbibliothek: Richtlinien, Geschäftsgänge und Datenfluß" (2002; workflow _authorities_onb-nak_02.pdf, report to RA 2002-04-25)

NAK contains references to manuscript collections (HAN), music collections (MUS), and the Austrian Literary Archives (ÖLA) with exceptions stated on the information page on <http://nak.onb.ac.at/cgi-bin/allegro/nak/nak.pl>

ÖNB uses the system [Allegro-HANS](#) based on the exchange format MAB. The main entries are kept on two authority files - the PND (*Personennamendatei*) for personal names, the GKD (*Gemeinsame Körperschaftsdtei*) for names of corporate bodies that are linked to the descriptions of the objects (collections, documents, audio-tapes, images). NAK is exchanging, both importing and exporting, authority records with the central Austrian authority files of the Austrian Library Consortium (*Österreichische Bibliothekenverbund*; Acc 10 for persons and Acc 11 for corporate bodies), and the German authority files (PND and GKD. Since 2001 all records must have references to PND and GKD.

5.3.2.9 Riksarkivet (National Archives) ([RA](#))

The National Archival Database
<http://www.nad.ra.se/NAD.html>

Number of records:

Persons and corporate bodies: c. 170 000

Rules:

Rules for authority records of the National Archives (*Riksarkivets auktoriseringsregler*) on
http://www.ra.se/nad/auktor/index_ab.html

The rules comply with ISAD(G) and ISAAR(CPF), but there is a lot of legacy data in different standards, such as the large number of records from the first version of the National Registry of Private Archives (REA) 1987-1993. In that system the name authority and collection information formed one record, and there was no support for many to many relations between persons/corporate bodies and archival records. The national and regional archives have since 1993 had a system (ARKIS) in which authority records could be linked to descriptions on the top (fonds/collection) level. This system is now replaced by ARKIS II that fully supports multi-level description with links to authority records on all levels. With the exception of Lund University Library, the manuscript departments of the research libraries have no systems for automatic updates to the National Archival Database (NAD). NAD has so far only been available on CD-ROM, latest edition 1998(referred here to as RA-NAD98). The NAD on the web page above represents only the national and regional state archives.

ARKIS II is not completely loaded yet. It is therefore not certain that all the examples discussed below will be available for the LEAF prototype.

5.3.2.10 Staatsbibliothek zu Berlin (Berlin State Library), Berlin, Germany ([SBB](#))

Database: Kalliope

URL: <http://kalliope.staatsbibliothek-berlin.de/>

Number of records:

Persons: 239 605

Corporate bodies: 21 272

Rules:

Regeln zur Erschließung von Nachlässen und Autographen (RNA)

<http://zka.staatsbibliothek-berlin.de/rna/>

The records are maintained by the "Zentralkartei der Autographen (ZKA)", and the authority records follow the common German authority file [Personennamendatei \(PND\)](#).

The German Union Catalogue for Modern Manuscripts and Letters (Zentralkartei der Autographen, ZKA) was established in 1966 in SBB. More than 150 partner institutions in Germany provide their descriptions of modern manuscripts and letters regularly to the ZKA. In preparation of the conversion from the card catalogue to the database cataloguing rules for manuscripts and letters were established. The *Regeln für die Erschließung von Nachlässen und Autographen - RNA* (i.e. Rules for the cataloguing of literary archives and autographs) has now been generally accepted by manuscript departments in Germany and Austria.

The database Kalliope now contains records for 239 605 persons with references to more than 470 000 items ((May 2002) from more than 150 German repositories (libraries and archives). On the basis of the name authority file DIANA (*Deutscher Index zu Autographen und Nachlässen*, i.e. German Index of modern manuscripts, letters and literary archives) was created. DIANA offers overall information about the locations where documents related to a specific person are stored.

Authority information is regularly exchanged and updated with the common German authority file for personal names [Personennamendatei \(PND\)](#) and corporate names Gemeinsamen Körperschafts-Datei ([GKD](#)) of the Deutsche Bibliothek.

On the database Kalliope see also above under [DLA](#)

5.3.2.11 Schweizerische Landesbibliothek (Swiss National Library), Bern, Switzerland ([SNL](#))

Database:

Repertorium der handschriftlichen Nachlässe (SNL-RHN) <http://www.snل.ch/repertoi/alsrep.htm>

Number of records:

Persons and corporate bodies: 7 200

The *Repertorium* includes individual and family archival collections held in around 260 archives, libraries, and museums.

This is an old static version latest updated 1997. A new version is now tested and expected to be available the second half of 2002.

It is not an authority file in the proper sense, but the records contain fields with the same function: Name and first name (for individual collections)/Name, possibly place of origin (for family collections), Dates (birth and death, or of the collection), and Profession or activity

5.4.3.12 Biblioteca de Universidad Complutense, Madrid, Spain ([UCM](#))

Database:

CISNE (MARC)

<http://cisne.sim.ucm.es/>

Number of records:

Persons: 3003

Corporate bodies: 1323

Rules:

Autoridades de la Biblioteca Nacional: recurso electrónico
(Madrid: Biblioteca Nacional: Chadwick-Healey España, 1996-)
Reglas de catalogación (Madrid: Ministerio de Cultura, 2000)
Directories para entradas de autoridad y referencia, ANABAD (1993)
Anglo-American Cataloguing Rules

5.3.3 National Registries of Private Archives

There are a number of national registers of private archives and related projects that are implementing authority files. Many of them started with manual registers or card indexes, where references to holdings were noted on records with a person or a corporate body as an entry.

The starting points for such registers were hence the entities--persons or corporate bodies--not the material (see also **National Register of Archives 1995**)

5.3.3.1 Europe

5.3.3.1.1 UK: Archival Networks

The National Register of Archives →

<http://www.hmc.gov.uk/nra/nra2.htm>

The indexes to the NRA contain references to papers of c.150,000 corporate bodies, persons and families relating to British history with a further 100,000 connected records. These are held in a network of national and local record offices, university libraries and specialist repositories in the UK and abroad. The system is based on the concept of authority files, and there are detailed rules for creating the records: *Rules for the Construction of Personal, Place and Corporate Names* (National Council of Archives, 1997)

Archive Hub: access to descriptions of archives held in UK universities and colleges →

<http://www.archiveshub.ac.uk/introduction.shtml>

AIM25 Archives. University of London Computing Centre → <http://www.aim25.ac.uk/>

For more information on British networks see:

A2A--Access to Archives → <http://www.a2a.pro.gov.uk/> <http://www.ariadne.ac.uk/issue30/archives>,
and the description on <http://www.ariadne.ac.uk/issue30/archives>

HMC/Archives Research Projects Register → <http://www.hmc.gov.uk/arpr/keyword.asp>

The U.K. projects are all aiming to be ISAD(G) and ISAAR(CPF) compliant

5.3.3.1.2 Finland: National Archives

The Archival Database and Online Finding Aids → <http://www.narc.fi/atengl.html>

Interface in Finnish, Swedish, and English. It follows the standards ISAD(G) and ISAAR(CPF), and includes altogether 19 946 records creators (Jan. 2002).

5.3.3.1.3 Norway: Union Catalogue of Private archives

Norske privatarkiver - Samkatalogen → <http://www.riksarkivet.no/arkivsok/index.html>

Interface only in Norwegian. Probably ISAAR(CPF) compatible.

5.3.3.1.4 Denmark: National Register of Private Archives

DANPA → <http://danpa.dda.dk/index.asp>

Search interface in Danish, information also in English and German. The data structure is unknown.

5.3.3.2 Australia and New Zealand

National Archives of Australia: Records Search → http://www.naa.gov.au/the_collection/recordsearch.html

Bright Sparcs, the database of the Australian Science and Technology Heritage Centre → <http://www.asap.unimelb.edu.au/bsparcs/bsparcshome.htm>

RAAM (Register of Australian Archives and Manuscripts) → <http://www.nla.gov.au/raam/>

RAAM is a guide to collections of personal papers and non-governmental organizational records held by Australian libraries and archives.

Database description → <http://www.nla.gov.au/raam/raamdb.html>.

ASAW (Australian Science at Work) → <http://www.austehc.unimelb.edu.au/asaw/asawhome.htm>

ASAW and Bright Sparcs are both managed using the Online Heritage Resource Manager (OHRM) - tool designed specifically to manage archival authority records and information and resources associated with them. Information about the OHRM and some of its implementations can be found at: <http://www.austehc.unimelb.edu.au/ohrm/>

NRAM (New Zealand National Register of Archives and Manuscripts / Te Raarangi Puuranga, Tuhinga Ake o te Motu) → <http://www.nram.org.nz/>

5.3.3.3 U S A

NUCMC (National Union Catalog of Manuscript Collections) → <http://lcweb.loc.gov/coll/nucmc/nucmc.html>.

There are a several projects with authority information included in EAD files and facilities for name search, but few are actually using authority files.

5.3.3.4 Canada

CAIN (Canadian Archival Information Network | Réseau canadien d'information archivistique) <http://www.cain-rcia.ca/>.

CAIN has also links to the provincial and territorial archival networks →

http://www.cain-rcia.ca/cain/network.html?sessionKey=1013508129015_62_20_57_206

5.4 Rules and Principles

5.4.1 Mandatory and Essential Elements

5.4.1.1 IFLA: UBCIM

IFLA UBCIM Working Group on Minimal Level Authority Records gives priorities to the following elements in *Mandatory Data Elements for Internationally Shared Resource Authority Records* <http://www.ifla.org/VI/3/p1996-2/mlar.htm> :

1. Record Identification

Includes coded information on Record Status, Type of Record, Unique record identification, Date and time of transactions

2. Language

Language and script of cataloguing

3. Authority

Code for the descriptive cataloguing rules followed in establishing the heading and references.

Source Library/Agency for the Record

Code for the library or agency responsible for the record content

4. Entity

Differentiated or undifferentiated personal name [recommended addition to UNIMARC]

Authorized Heading

Nationality of Entity

5. References

Variant Forms of the Authorized Heading

Related Authorized Heading ('See Also')

6. Notes

Citation for a consulted source or the item catalogued which provided information about the authorized heading and/or variant forms

The following data elements are highly recommended:

Biographical, Historical, or Other Information about the Entity (to be a recommended addition to UNIMARC)

Source Data Not Found, that is Citation for sources examined but heading not found

General notes

5.4.1.2 ICA: ISAAR (CPF)

The present version of ISAAR(CPF) follows the recommendations of IFLA concerning the elements that are essential for exchange of information and the identity of the entity. To these are added a number of elements for the purpose of providing more detailed information on the functions and activities of the entity, which play an important role when authority records are linked to archival descriptions.

During the present review of ISAAR(CPF) the committee is now discussing the following structure (references to present ISAAR(CPF) within brackets):

1. Identity Area

The purpose of this area is to provide elements which identify uniquely the corporate body, the person, or the family in accordance with national and international standards, and to define a standardised access point.

2. Description Area

The purpose of this area is to describe the corporate body, person or family named in the Identity Area. This can consist of a free text biographical, corporate or family history and/or a structured text and/or links to other relevant sources. All elements should include date qualifiers.

3. Relationships Area (1.6, 2.1.8, 2.2.8, 2.3.8)

The purpose of this area is to record relationships with other corporate bodies, persons and /or families. All relationships should include date qualifiers.

4. Related Archival Materials and Other Resources Area

The purpose of this area is to provide links to archival descriptions, bibliographical and other relevant resources. All links should provide date qualifiers.

5. Control Area

The purpose of this area is to document and record sources used to compile the authority record, its creator(s) and status.

In these areas only the following elements are regarded as mandatory/essential:

- **Type of entity** (i.e. corporate body, person, family) (1.2)
- **Authorised form of name** (1.3)

- **Dates of existence**
- **International Record ID** (identity code(s) of the authority record) (1.1) (in countries like the UK where a single agency, the Historical Manuscripts Commission, will maintain a national name authority file, the ID will be a national one prefixed by 'GB', elsewhere archival institutions will identify themselves as the creator of authority records by their institutional ID.)

A draft of the new version of ISAAR(CPF) will be available to the public 2003 on the home page of the committee (<http://www.hmc.gov.uk/icacds/icacds.htm>)

5.4.1.3 LEAF Data Providers

The data providers implementing PND are thereby also compliant with the recommendations of IFLA's working group in their priorities. A difference is the high priority given to life dates by all data providers. It is generally agreed that this element is essential for differentiating entities.

Archival systems, such as RANAD, focussing more on corporate bodies, give also priorities to elements such as place and functions.

When it comes to practice there is a long tradition—from manual indexes to database systems—giving priorities to a very limited number of elements:

- **Name(s)**
- **Dates**
- **Description**

These elements have been supplemented with fields for control of processes and exchange of data in information systems.

5.4.2 Local Procedures

Most of the LEAF data providers do not have documented routines for the entering, updating, and quality control of authority records that have been available. The responsibility for quality control lies by the individual staff.

DLA has very detailed rules for the creation of authority records in the rules referred to above. These rules also include lists of sources.

RA has developed rules for authority records in cooperation with the Royal Library. Within the organization of National and Regional Archives there a system for authorizing staff for updating different parts of the system. Routines for central control of authority records for public agencies are emerging, but they have not yet been adopted for private corporate bodies and persons. The procedures for control of data provided from other institutions to NAD is generally agreed to be insufficient, and are limited to control of the data transfer format.

SBB is a special case in that all authority records in Kalliope are "copies" of authority records in PND. Whenever SBB or any of its partners finds and uses a person name in PND which still does not exist in Kalliope, a provisional "Kalliope-record" containing only the name, gender and, possibly, additional information and the PND ID number is created. The staff responsible for PND relations at SBB just adds this Kalliope record ID number in the PND record itself and gives it a "z" mark. This provides enough information so that with the next weekly update this very record is automatically matched to the Kalliope record. The procedure is almost the same for entering new records, only that all information is sent from SBB to PND.

ONB has provided a detailed description of the local procedures and the communications with the general Austrian and German authority files (KAISER 2002). In the NAK of ONB the elements name, life dates, gender (m/f), and country code are mandatory. Only authority records for differentiated persons are created in accordance with guidelines established by the Austrian library association (*Zentralredaktion der Österreichischer Bibliothekenverbund*). If there is a record for an entity in (1) the Austrian authority files Acc 10/11 or (2) the German PND/GKD in PICA/ILTIS that record will be imported to Allegro-Hans. If that is not the case, a new record will be created in Allegro-Hans and communicated with the general authority files.

5.4.3 Names and Entities: Differentiated and Undifferentiated Records

The foremost function of authority files in libraries has been to serve as tools for cataloguers to ensure that the proper forms of names are used. Systems have usually been designed in such a way that each name or variant name appearing in publications has constituted one authority record. An author publishing scientific works under his real name and novels under a pseudonym will in such systems be represented by two authority records. In the management of archives and manuscripts it is, on the other hand, more natural to see the person and his all his works as a whole. In that environment it is natural to come to the conclusion that the record should be constituted not by a name but by the entity *per se*. (It would be possible to combine these points of view; a system might have several authority name records referring to one individual, if relations are created between these records stating this fact. EAC has taken this possibility into consideration.)

Among the LEAF data providers it appears to be a generally agreed that an entity should be represented by only one record including all the variant forms of names (tracings). In RNA (http://zka.staatsbibliothek-berlin.de/rna/rna_03.htm) it is clearly stated, §12, that a person always should be entered under the same form of name, and in § 13, if he person appears under different name usually the most common form should be used as an entry.

In the Report of the IFLA UBCIM Working Group on Minimal Level Authority Records *Mandatory Data Elements for Internationally Shared Resource Authority Records* (<http://www.ifla.org/VI/3/p1996-2/mlar.htm>) an element "Differentiated or undifferentiated personal name" is proposed as a recommended addition to UNIMARC.

Differentiated personal name A unique personal name heading which is distinguished from all others by the addition of qualifying data. Qualifying data may be added to the authority heading or included as qualifying information in the authority record in accordance with the standards followed by individual agencies

Undifferentiated personal name A personal name heading which, because it cannot be uniquely distinguished, conflicts with or normalizes to match another personal name heading. When distinguishing information becomes available, differentiated headings are established for each distinct bibliographic identity.

In National Council of Archives [UK], *Rules for the Construction of Personal, Place and Corporate Names* (1997), 2.1.2B, there is the following declaration: "A personal name is constructed by combining mandatory and optional components of the name so that the person concerned can be identified with certainty and distinguished from others bearing similar names".

The distinction between records for "persons" and "names" is essential in the PND and the rules developed by *Die Deutsche Bibliothek*. These rules are adopted by LEAF partners in Germany and Austria.

In the rules developed by DLA—*Gebrauchsanweisung für die Namennormdatei (Version 1.5)*—this distinction is put forward as the essential concept in the process of creating authority records:

Individualisiert = Personensatz Differentiated record = entity record

Nicht individualisiert = Namensatz Undifferentiated record = name record

These distinctions give actually quite different ontological status of the records. A differentiated record refers to a person or a corporate body as such, as an entity, while the undifferentiated record remains a mere name record. In the latter case we do not even know if a person with that name has existed, it might be a pseudonym for one or more persons.

In the Austrian PND (Acc 10) of the Austrian Library Consortium only differentiated person records are created, and in contrast to the German rules it is not allowed to link to undifferentiated authority records. These rules are applied in the specific rules for NAK (KAISER 2002)

5.4.4 Dates and Date Formats

In MARC based systems life dates for persons is not an element but a qualifier to name elements, for example:

x00\$a **Elizabeth** \$b I,\$c **Queen of England** \$d **1533-1603** (source: LOC)

When authority records now are regarded as being constituted by the individual, a consequence would be to elevate this qualifier to an element. In the case of corporate bodies there is no corresponding qualifier at all in MARC with the exception of records for congresses

The LEAF data providers who have data base system have anyhow dates as a separate element. Also in the case of corporate bodies eight of the data providers have a special field for standardized Dates of Corporate Body (WP 3.2, *Leaf Metadata Tables: Person Names And Corporate Bodies*, 4, element 19).

The dates are generally stored in a format that can be converted into the **ISO 8601:1988 International Date Format**: "year-month-day".

In the case of MARC records, as the example above, the format is generally **yyyy-yyy**, sometimes combined with qualifiers such as "c.". German speaking data providers have usually the form **dd.mm.yyyy** in special fields for standardized dates. Sweden is one of the few countries in the world where the ISO format is also a national standard for dates in general use. In RANAD the ISO date format might be followed by qualifiers such as "?", "c:a" (*circa*), and a decade such as 1910 to 1919 might be expressed "*1910-tal*".

It is likely that the LEAF export function will be able to deal with all these variants, as long as the dates are distinguishable from other information, and consistent.

A note on German and Swedish field names

Note that words which etymologically are the same are defined quite differently in the German and Swedish systems.

SBB, DLA: "Gültigkeit, von/zu" the dates of existence for a corporate body

RANAD: "Giltighetstid, från/till" the dates of use for alternative names

SBB, DLA: "Wirkungszeit, von/zu" the flourishing dates of a person or corporate body

RANAD: "Verksamhetstid, från/till" the dates of existence for a corporate body

5.5 Problems and Samples

5.5.1 Language Dependent Name Conventions

The IFLA Section on Cataloguing has pointed out a number of problems when linking multi-language authority files (MURTOMAA 1994, TILLET 2001):

- differences due to language dependent qualifiers or geographical jurisdictions prescribed by cataloguing rules
- different practices with regard to abbreviations
- different transliteration and romanization schemas
- different practices for word division from Romanized forms
- differences in spelling practices even for the same language (such as US versus most other English speaking countries that cause retrieval problems).

A particular problem is the different national rules for normalizing names of sovereigns, such as all the variants of the Latin form **Carolus**: Carl, Karl, Charles, Carolus, Carolo, Karol &c.

It can also be the case that there are different rules for spelling of the same name within a language. In Sweden there exists no general agreement on whether names of historic kings should be normalized or not in the case of for instance Carl/Karl, Gustav/Gustaf. Authority files should, of course, contain all variants, but that is not the case in practice. In RANAD98 you need to search on both variants.

The Swedish king who got killed at Lützen 1632 is in English usually referred to as **Gustavus Adolphus** or **Gustavus II**, in Swedish **Gustav II Adolf**, and in Library of Congress the coded form is

\$a Gustaf \$b II Adolf,\$c King of Sweden \$d 1594-1632

In Kalliope there is an impressive number of variants of the name from PND, but it is still not certain that it would be possible to create a link with corresponding authority records from data providers of other languages.

Name	Gustaf Adolf <Sverige, Konung, II.>
Verweisung	Gustav Adolf <Sverige, Konung, II.> Gustav Adolph Gustav Adolph <Schweden, König, II.> Gustav Adolph <Sverige, Konung, II.> Gustaf Adolf <Schweden, König, II.> Gustave Adolphe <Schweden, König, II.> Gustav Adolphe <Schweden, König, II.> Gustavus <Magnus> Gustavus <Suecia, Rex> Gustavus Adolphus <Schweden, König, II.> Gustavus <Sverige, Kuning, II.> Gustavus <Schweden, König, II.> Gustavus-Adolphus <Schweden, König, II.> Gustavus-Adolphus <Finnland, Großfürst> Gustaf Adolf <Finnland, Großfürst> Gustaf Adolf <Etonien, Herzog> Gustavus-Adolphus <Etonien, Herzog> Gustavus-Adolphus <Karelien, Herzog> Gustaf Adolf <Karelien, Herzog> Gustaf Adolf <Carelien, Herzog> Gustavus-Adolphus <Carelien, Herzog> Gustaf Adolf <zu Ingermanland> Gustavus-Adolphus <zu Ingermanland> Ingermanland, Gustaf Adolf zu Ingermanland, Gustavus Adolphus zu Gustaf Adolf <Estland, Herzog> Gustavus Adolphus <Estland, Herzog> Gustaf Adolf <Sverige, Konung, II.> Gustav Adolf <Schweden, König, II.>

In the following example concerning the Duke of Alba, PND has, as it seems, all possible variants:

Fernando <Alba, Duque, III.>	
Verweisung	Álvarez de Toledo, Fernando Toledo d'Alba, Fernando A. de Alba, Fernando A. de Alba, Ferdinand von Toledo D'Alba, Fernando A. de Toledo, Fernando Á. de Alvarez de Toledo d'Alba, Fernando Alvárez de Toledo, Fernando Alba, Fernando d' Alvárez de Toledo d'Alba, Fernando Alvárez von Toledo, Ferdinand Alvárez de Tolède, Ferdinand Alvárez de Toledo y Pimentel, Fernando Toledo d'Alba, Fernando A. Toledo y Pimentel, Fernando A. Pimentel, Fernando A. de

Toledo y Pimentel, Fernando A. de
 Toledo d'Alba, Fernando Alvarez de
 Alba, Fernando Alvarez de
 Alvarez d'Alba, Fernando
 Ferdinandus <Alba, Herzog, III.>
 Fernando <Alba, Duque, III.>
 Ferdinand <Alba, Herzog, III.>

Indiv.kennzeichen individualisiert
 Lebensdaten 1507 - 1582 (29.10.1507-11.12.1582)

Still there is no complete and exact literal match to the authorized Spanish name in the record of UCM:

```
008 020311nn aznnaabn |a aaa cnz n
100 1 Alba, Fernando Álvarez de Toledo, |cDuque de, |d1507-1582
400 1 Alba, Fernando Álvarez de Toledo y Pimentel |bIII, |cDuque de
400 1 Álvarez de Toledo, Fernando |bIII, |cDuque de Alba
670 Autoridades BNE
```

In cases where data providers only give the name normalized according to national rules automatic linking appears to be impossible, and it is obvious that the name qualifiers cannot be included in such a procedure. In the following cases it is not even sure that a user not knowing French or Italian would be able to see that the records refer to the same individual:

Elizabeth I (1533-1603) Queen of England and Ireland (source: NAR)

\$a Elizabeth \$b I,\$c Queen of England \$d 1533-1603 (source: LOC)

Elisabetta I d'Inghilterra

Elizabeth 1, reine d'Angleterre

Elizabeth, Königin v. England

Elisabeth I, drottning av England

Up to the 18th Century it is common that an author appears both under a Latin and a vernacular form of name. The conventions concerning which form of name that should be authorized are likely to differ between different data providers due to language conventions and the information available. Furthermore, if the person was awarded a knighthood it can become even more confusing.

The naturalist **Carl Linnaeus**, for instance, was knighted **Carl von Linné** in 1761. His papers and correspondence are very dispersed, most in the Linnean Society of London. In English his name is usually authorized as **Linnaeus, Carl**, while the authorized name in Swedish, French and German usually is the latter form **Carl von Linné**, rarely Karl, as a writer he usually appeared in the Latin form **Carolus Linnæus**

Search on "Linné" in NRA gives no result, In BL-MOLCAT we get the see reference:

"Linné (Carl) . see Linnaeus.

Linnaeus (Carl)" gives then 14 references to manuscripts

In NRA Carl Linnaeus (1707-1778) Naturalist, 6 references to repositories

In RA-NAD98 only one of five data providers has recorded Linnaeus, Carl as alternative name, the rest only providing Linné, Carl von

KALLIOPE (PND) provides all possible alternative forms of names in different languages:

Name	Linné, Carl von (Biologe, 1707-1778)
Sortiername	Linné, Carl –von– (Biologe, 1707-1778)
Verweisung	Linné, C... F... Linnaeus, Carl Linnaeus, C.

Linnaeus, Carolus
 Linné, Cararolus à
 Linné, ...
 Linné, Karl von
 Linnaeus, ...
 Linnaeus, Carl von
 Linnaeus, Carolus von
 Linné, Carolus à
 Linné, Carolus a
 Linné, Carolus
 Linnaeus, Carolus
 Linnaeus, Karl
 Linné, Carl a
 Linnaeus, Carolus a
 Linné, Carolus C. a
 Linné, Charles
 Linneo, Carlos
 Linne, Karl von
 Linnäus, Carl

The sample **Linné** is a case that most users would be able to deal with, but less well known authors from the early modern period would be more problematic. If at least one data provider has recorded the variant forms of names and all have the life dates, links between these records can be established. For the user it would then be sufficient to have knowledge of one variant to get all the references.

The problem with different standards for transliteration of Cyrillic are well-known. A record for the famous 19th Century Russian composer has all the following variants in Kalliope (PND):

Verweisung	Cajkovskij, Pëtr I. Tschaikowski, Pjotr I. Cajkovskij, Pëtr Il'ic Cajkovskij, Pjotr Il'ic Cajkovskij, Petr I. Caikovskij, Petr I. Chaikovsky, P. Csajkovszkij, P. Tchaikovsky, P. I. Tchaikovsky, Peter I. Tciaikowski, Pietro Tschaikowsk, Peter Tschaikowsky, P. Tschaikowski, Peter Tschaikowsky, Peter Tschaikowsky, Peter I. Tschaikowsky, Peter Iljitsch Tschaikowski, Peter Iljitsch Tschaikowsky, Peter Ilych Tchaikovsky, Piotr I. Tschaikowski, P. Tschaikowsky, Pjotr I. Tschaikowski, Peter I. Tschajkowski, Peter I. Tchaikovsky, Peter Tschaikowsky, Peter J. Tschaikowski, Peter J. Tschaikowsky, Peter I. Tchaikovsky, Petr I. Tschaikowsky, Peter Tschaikowsky, P. I. Tchaikowski, P. Tchaikovsky, P. Tchaikovsky, Peter J. Ciaikovsky, Piotr I. Csajkovszkij, Péter Cajkovskij, Petr
------------	--

Tchaikovsky, Pyotr I.
 Tschaikowsky, ...
 Tchaikovsky, Petr
 Tchaikovsky, Pyotr
 Chaikovskii, Petr I.
 Tschaikowsky, Petr I.
 Tschaikowski, P. I.
 Tchaikovsky, Pyotr Il'yich
 Tchaikovsky, Piotr Illitch
 Tchaikovski, Piotr I.

But still the most used form in e.g. Denmark and Sweden "Tjajkovskij, Peter" is not represented. The now recommended transcription form "Čajkovskij, Pëtr Il'ic" is probably unknown to the general public in Europe.

5.5.2 Matching Names and Dates

For the purpose of testing the possibilities of automatic and manual linking of records a large number of searches has been done in the databases available on line, both those of the LEAF partners and the national registers listed above. It has, however, been difficult to find out a method for systematic comparisons of samples. Although many attempts were made in the pre study for this report, few results were significant enough to be included here. It has, of course, also been a problem that only a couple of authority files of the data providers have been available on line.

An interesting selection was obtained by searching for exiled Germans in the Swedish national archival data base (5.2.1); a comparison of names and dates in SNL-RHN (5.2.3) demonstrates the feasibility of creating automatic links by matching names and dates.

5.5.2.1 Example: Records of German exiles

In RA-NAD98 there are records for exile German collections kept by *The Labour Movement Archives and Library* (<http://www.arbarkiv.nu/english.htm>) in Stockholm.

Of 45 records in RA-NAD98 12 could be identified in Kalliope. DLA could provide three matches from Kallias. In seven cases these records of Kalliope only had flourishing dates (*Wirkungszeit*). In such cases the identification is dependent by context information which hardly can be automated. An informed user would be able to create links between the records.

Even if no links are created the LEAF server would still be of great value in this case for a German cataloguer or researcher whose attention will be drawn to a Swedish repository which probably only is known to specialists in the history of labour movements.

The output below is generated by copying and pasting from the public interfaces. The name on the first row is the name authorized, below variant names from the alternative names/see references.

Data Base	Name	Life Dates	Fl. Dates	Comment
NAD	Blachstein, Peter	1911 - 1977		
KALLIOPE	Blachstein, Peter Greiff, Will [Pseud.]	1911 - 1977		
NAD	Brandt, Willy	1913 - 1992		
KALLIOPE	Brandt, Willy Frahm, Herbert Ernst Karl [Wirkl. Name:]	1913 - 1992		
KALLIAS	Brandt, Willy Frahm, Herbert Ernst Karl	19131218 – 19921008		
NAD	Friedländer, Otto	1897 - 1954		
KALLIOPE	Friedlaender, Otto Friedrich, Otto [Pseud.]	1897 - 1954		
KALLIAS	Friedlaender, Otto	1897 - 1954		

Data Base	Name	Life Dates	Fl. Dates	Comment
NAD	Greid, Hermann Markus Grabscheid Hans Dirk	1892 - 1975		
KALLIOPE	Greid, Hermann Wiener [Pseud.]	1893 - 1975		
NAD	Harthern, Ernst Ludwig Jacobson Niels Hoyer [= pseudonym]	1884 - 1969		
KALLIOPE	Harthern, Ernst		1910 - 1911	The context information of the DIANA record allows the conclusion that the records concern the same individual, but the LEAF server would not be able to draw the same conclusion by an automated procedure.
KALLIAS	Harthern, Ernst Harthern, Ernst Harthern-Jacobson, Ernst Ludwig Jacobson, Ernst Ludwig Harthern-	1884 1969		
NAD	Heinig, Kurt	1886 - 1956		
KALLIOPE	Heinig, Kurt		1956-	See the note to Harthern, Ernst
NAD	Mockrauer, Franz	1889 - 1962		
KALLIOPE	Mockrauer, Franz		1914 - 1933	See the note to Harthern, Ernst
NAD	Philipp, Rudolf	1895 -		
KALLIOPE	Philipp, Rudolf		1950-	See also the note to Harthern, Ernst. The records might concern the same individual, but one cannot be sure
NAD	Reinowski, Hans J. Hans Reinow	1900 -		
KALLIOPE	Reinowski, Hans Johann Reinowski, Johannes Reinowski, Johann Reinowski, Hans Reinow, Hans [Pseud.] Rastlos, Hannes [Pseud.]			
NAD	Sonntag, Wolfgang	1910 - 1970		
KALLIOPE	Sonntag, Wolfgang		1948-	Might be the same, but cannot be certified from the record
NAD	Stechert, Kurt			
KALLIOPE	Stechert, Kurt		1957-	The context "Exilarchiv" makes an identification likely
NAD	Trepte, Curt	1902 -		
KALLIOPE	Trepte, Curt Trepte, Kurt		1938 - 1966	The same?

5.5.2.2 Test: Names and Dates in SNL-RHN

A study made by Pierre Clavel on records in the Swiss national registry of manuscripts SNL-RHN has given the following result (communications by e-mail to the present author 21/03/02, 24/06/02).

Contents of the data base SNL-RHN: 7 100 records of corporate bodies, families, and persons.

Selection of data: A population of all records of individual persons with dates of birth and death, total 5200 records.

Distribution of years of birth and death in the selected population:

First year of birth:	1417
Last year of birth:	1957
First year of death:	1461
Last year of death:	1992

Numbers of birth years per century:

15th C.	16th C.	17th C.	18th C.	19th C.	20th C.
13	48	116	930	3557	534

Extreme set of years of birth and death shared by at least two individuals:

First: 1695-1756

Last: 1921-1989

Assuming that the distribution is linear, and allowing 80 possible death years for each birth year (the rounded difference between the shortest and longest lives), there would be over 43000 possible combinations; a number large enough for the percentage of people sharing the same birth and death years to be 0%.

Results:

Surname, first name, and years of birth and death were compared.

48% of these records had at least one other records sharing the same year of birth and date. These were spread in c. 1000 groups of combinations of birth and death years containing 2 to 7 people. Of these only three groups of individuals had the same surname and the same years of birth and death:

De Candolle	Augustin-Pyramus	1778-1841
De Candolle	Jacob-Michel-François	1778-1841
Meyer	Karl Alfons	1883-1969
Meyer	Eugen	1883-1969

Solier	Jean	1717-1789
Solier	Marc	1717-1789

Out of 250 persons sharing the same surname and first name with someone else of the sample (including a group of 5 homonyms), there is only one case where two persons also are born the same year, and no case of people also having died the same year.

Usteri	Paulus	1768-1795	Zeichner, Kaufmann
Usteri	Paulus	1768-1831	Arzt, Politiker, Journalist

Concerning the first names there were 4400 individuals not carrying unique first names, and only 17 were born and dead the same year as someone having the same first name.

It can also be noted that one example was found on three individuals sharing the same names and professions, and distinguished by the life dates only:

Amstein	Johann Georg	1744-1794	Arzt, Naturforscher
Amstein	Johann Georg	1778-1818	Arzt, Naturforscher
Amstein	Johann Georg	1819-1892	Arzt, Naturforscher

(Arzt, Naturforscher = Physician, naturalist)

Conclusion:

From this test it can be assumed that the combination of names and dates are sufficient enough for the purpose of LEAF to distinguish individuals. However, as it will appear below (5.3), SNL-RHN is exceptional with its high percentage of records with life dates.

5.5.3 Contents of Essential Elements

A questionnaire was sent (23/1/02) to the LEAF data providers concerning their local practices and the actual use of essential fields:

- Which are your local rules on mandatory and essential fields?
- What percentage of records for persons has data in the following elements:
 - life dates with dates of birth and death distinguished from other descriptive text
 - nationality/nation of residence, entered with country code or controlled vocabulary
 - place (according to national standards)
 - profession/function/activity or corresponding descriptive field

5.5.3.1 Frequency Table

The table below gives the percentage of records containing data in the fields specified.

	DLA	FDOP	ONB	RA	SBB	SNL	UCM
--	-----	------	-----	----	-----	-----	-----

Life dates				40,9	36	80	53
Date of birth	35,5	19,8	67				
Date of death	15,4	13,5	52				
Flourishing dates			41	-	44	-	
begin	0,1						
End	0,1						
nationality/nation	0,3	0,2	99	-	80	0	
Place				79,6	0,1	0	
place of birth	12	16,9	51				
place of death	5,8	9,5	39				
place of activity	0,8	12,2					
profession/function/activity	41,9	33,5	72	(93,5)	80	75	

5.5.3.2 Specifications as Reported and Comments

DLA:

111642 100,00% Total records in NND

39590 35,46% with date of birth < 2003*
 17180 15,39% with date of death < 2003*
 130 0,12% with start of activity period < 2003*
 79 0,07% with end of activity period < 2003*
 276 0,25% with country code
 13454 12,05% with place of birth
 6435 5,76% with place of death
 857 0,77% with place of activity
 336 0,30% with place as part of a postal address
 6953 6,23% with profession/function (verbal)
 46781 41,90% with encoded classification as: writer, creator of collections ("Bestandsbildner"), artist, composer, photographer ...

FDÖP

date of birth: 19,8 %
 date of death: 13,5 %
 nationality: 0,2 %, but neither country code nor controlled vocabulary
 place of birth: 16,9 %
 place of death: 9,5 %
 place of flourishing: 12,2 %
 profession: 33,5 %

ONB

Dates of Birth/Death are obligatory and are distinguished from other descriptive text. If no Dates of Birth/Death can be given, Date of Flourishing is obligatory.

Frequency:

Date of Birth (#808): given in 3221 of 4819 records (person names authorities)
 Date of Death (#809): given in 2483 of 4819 records (person names authorities)
 Date of Flourishing (#806w): given in 1985 of 4819 records (person names authorities)
 Nationality (#806n): given in 4778 of 4819 records (person names authorities)
 Place of Birth (#808A): given in 2445 of 4819 records (person names authorities)
 Place of Death (#809A): given in 1901 of 4819 records (person names authorities)
 Profession (#805): given in 3483 of 4819 records (person names authorities)
 Affiliation (#807...807Z): given in 1985 of 4819 records (person names authorities)
 Also mandatory: Gender (#806g)

RA NAD98

Number of records: 169762

Life dates: 69412/169762=0.409

Nationality: can not be distinguished from other places in NAD1998

Place: 135180/169762=0.796

Classification/function: 158654/169762=0.935

in most cases on a very high level with codes just for distinguishing public and private corporate bodies

5.5.4 Complementary Authority Records

The LEAF data providers have in one respect two types of authority files: on the one hand there are the conventional authority files containing a large number of names but not much of other information, on the other hand specialist institutions with a limited number of records, but instead providing more information in depth, and serving as biographical dictionaries. When such records can be brought together, the result will be very rewarding for the users, both in terms of access points and detailed information. Even if the user only is interested in the material referred to by SBB or ÖNB, the FDÖP record is still very valuable for the context information it provides.

LEAF element	FDÖP	SBB	ÖNB
Leaf.Local_ID_number	FDÖP-NrAutor: 130	ZKA-Nummer 00038680	
National Authority Record ID		PND-Nummer 118529277	
Leaf.Main_heading	Ehrenfels, Christian Freiherr von	Ehrenfels, Christian von	Ehrenfels, Christian von
Leaf.Other_Name (See-Reference)		Ehrenfels, Maria Christian Julius Leopold Karl von	
Leaf.Other_Name		Ehrenfels, Christian	
Leaf.Date_of_birth	20.06.1859	1859	20.06.1859
Leaf.Date_of_birth_standardised	20.06.1859		
Leaf.Date_of_death	08.09.1932	1932	08.09.1932
Leaf.Date_of_death_standardised	08.09.1932		
Leaf.Place_of_birth	Rodaun (Wien)		Rodaun (Wien)
Leaf.Place_of_death	Lichtenau		Lichtenau (NO)
Leaf.Place_of_flourishing	Wien, Prag		
Leaf.Nationality		AT	
Leaf.Profession		Philosoph	
Leaf.Profession, non standardised	Philosoph	Freiherr, Professor; Prag, Wien-Rodaun, Lichtenau/NÖ (Wirkungsorte) Österr. Philosoph, Psychologe und Dramatiker; Begründer der Gestaltpsychologie, auch Philosophie der Mathematik	Philosoph, Schriftsteller Freiherr
Leaf.Gender	m		
Leaf.Reference_books	Personalakt der Universität Wien (Lebenslauf, verfaßt von Robert Zimmermann; Unterlagen zum Habilitationsverfahren), Vorlesungsverzeichnisse der deutschen Universität Prag; Personalstandsverzeichnisse der deutschen Universität Prag.		
Leaf.Date_of_capturing	31.12.1995		
Leaf.Curriculum	Curriculum: Sohn des Leopold von Ehrenfels und der Klothilde von Coith; besuchte von 1870 - 1876 die Realschule in Krems a. d. Donau; heiratete 1894 Emma v. Hartmann; aus der Ehe gingen zwei Kinder hervor. 1877: Studium an der Hochschule für Bodenkultur in Wien; inskribierte im WS 1879/80 an der Philosophischen Fakultät der Universität Wien Philosophie, Germanistik und		

	<p>Geschichte, daneben auch juristische Studien; Herbst 1882 - September 1883: Militärdienst als Freiwilliger, den er mit der Reserve-Offiziersprüfung beim 4. Dragonerregiment beendete; folgte Meinong im WS 1884/85 nach Graz und promovierte daselbst im März 1885 mit der Dissertation "Über Größenrelationen und Zahlen. Eine psychologische Studie"; zugleich Publikation zweier dramatischer Dichtungen: "Die Brüder von Hartenstein" und das Gedicht "Melusine". Herbst 1887: Reise nach Zürich: Zusammentreffen mit Richard Avenarius, dem Psychiater Forel und Gottfried Keller; 1888: Habilitation an der Wiener Universität mit der psychologischen Studie "Über Fühlen und Wollen" (bereits im Februar 1887 fertiggestellt), Tod des Vaters; 1889/90: Aufenthalt in Berlin; 1891 - 1896: Vorlesungen an der Wiener Universität, jeweils nur in den Sommersemestern, Reisen nach Deutschland und Italien ; Sommer 1896: Ruf als Extra-Ordinarius an die deutsche Universität in Prag, wo durch den Weggang Friedrich Jodls eine Stelle frei geworden war; 1899: Ernennung zum o. Univ.-Prof. in Prag, 1929: Emeritierung; er hielt aber weiterhin Vorlesungen ab bis zum WS 1931/32.</p>		
Leaf.Lehr_und_Forschungsschwerpunkt	Lehre von den Gestaltqualitäten, Werttheorie.		
Leaf.Bequest	Nachlaß: Privatarchiv der Familie von Ehrenfels in Lichtenau, Niederösterreich (Originale), Kopien von nachgelassenen Manuskripten und Briefen (1370) in der FÖP, Ein Inventarverzeichnis der Manuskripte, sowie ein Katalog der Korrespondenz befinden sich in der FÖP		

5.6 Conclusions

5.6.1 Identification and Automatic Linking

From the samples available for this study, it can be concluded that there generally is enough information for an informed user to identify the entities. The obvious exceptions are the undifferentiated authority records that the cataloguer has not been able to identify.

Only a few variables in the data now available can be used for the creation of automatic links:

- National or international authority numbers
- Name
- Dates

Other candidates could have been values such as nationality and place of birth/death. These elements are, however, not very well covered. Nationality is a required element in MLAR and there is an international standard for coding nationality. However, the problem with historic nationalities remains to be solved.

The data providers SBB, DLA, and ONB have, or, have the intention, to record a common ID, the PND number. When that is the case, automatic links between these data providers can be created with the highest level of certainty. With the present selection of data providers for the prototype, that would probably deal with most records. PND appears also to provide a significant number of non German authors, and if available on the web, it would be possible for cataloguers in other countries to cross reference to it.

Concerning the name element a number of difficulties have been pointed out. It remains to see how frequent these problems will be in practice.

It is obvious that life dates for persons and dates of existence for corporate bodies are essential both for automatic and manual linking. Life dates are in fact the only data about a person that will be true, regardless of country or language of cataloguing (cf. TILLET 1996). The data providers are already giving this element a high priority, but still only around 30-50 % of the records have such data. Although it might be so that the records that do not have life dates concern less well known individuals who are not very likely to appear in different systems, the element must be given the highest priority. This is a question that should be possible to answer when the LEAF server is implemented and loaded with a sufficient number of records from different data providers. It is also recommended that tests are carried out with the same methods as used here with the SNL-RHN data.

5.6.2 Selection of Data Providers

The present analysis and the forthcoming prototype have a certain bias to records from the German speaking cultural sphere (in which Sweden and Slovenia might be included). Since the BL records probably cannot be included, it is necessary to find another representative for English language records by including the U.K. National Registry of Archives that already has a compatible format.

There are three data providers representing Romanic languages. However, IMEC does not have data ready for testing. BN and UCM have general authority files that only can be used for test purposes, and not be included in the permanent LEAF/Malvine server.

5.6.3 Completeness

At present there is a big problem for users to judge the completeness of the data provided by the partners. All projects are in different stages of development. Some have authority records for all creators of collections, some have authority records for all authors (sender/receivers) of documents within a limited number of collections, but do not cover all creators of collections.

The LEAF server must provide declarations on what users can expect from the service and its data providers (see 6.5)

5.6.4 The Concept of Authority Record

It can be questioned whether it is reasonable to maintain the concept of authority records. Barbara Tillett has written and spoken for many years about the idea of moving away from the concept of "authority control" to "access control," and by that she means providing more flexibility in user views of which form of a name, title, or subject they want to see for an entity. In the case of authority records for persons and corporate bodies, the obvious consequence of this view is, that it is person or corporate body as such that it is the person or corporate body as such that should be represented by the record, not a particular form of name. We can achieve much more agreement on what an entity is and what its characteristics are, than on what single form of its name everybody in the world has to call it. We have seen how the same person might be known under different forms of names even within the same language. Various names are linked to the entity—person—and we can implement displays in various ways, once we have organized the information about the entities in a clear manner (TILLET 1996). The exchange format chosen by the LEAF project, EAC, is well designed to meet this point of view. Each EAC record can contain several "authorized names" and there are attributes for information on the authorizing agencies for each variant.

Archivists and manuscript librarians are in their daily work primarily focussed on identifying individuals in order to contextualize the unpublished material they are dealing with. First after this descriptive work comes the process of creating the access points. The result of this process might then rather be called: **Records of persons and corporate bodies with standardized access points and context information.**

5.6.5 Minimal Requirements for Authority Records in LEAF

In the report **MRAD2** 3.3.6 it is recommended that the minimal requirements for authority records to be part of the LEAF system are indications of **name** and **life dates**. Also the present analysis of practice leads to the same recommendation. However, an acceptance of this recommendation would result in the exclusion of about half the number of records of the LEAF data providers, which is not acceptable from a user perspective. The LEAF system must accommodate also records that are not authority records in the proper sense. By an automated procedures such incomplete records could be marked, and excluded from automatic linking procedures.

Taking the fact into consideration that present and potential data providers are in different stages in the development of local systems, the following requirements are proposed:

Each data provider to LEAF should

- implement local rules for authority records that are compliant with the professional standards of IFLA or ICA;
- assume a plan for the work to fulfil the minimum level requirements adopted by LEAF within a limited time;
- document the local procedures for quality assurance;

document the completeness and coverage of data provided for publication on the LEAF server site

6 Bibliography

[ABATE & al. 1998]: Marcey L. Abate, Kathleen V. Diegert, Heather W. Allen, "A Hierarchical Approach to Improving Data Quality", DATA QUALITY (September 1998 Vol. 4, Nr. 1). URL: <http://www.dataquality.com/998abate.htm>

[ACIG 2002]: LITA/ALCTS CCS Authority Control in the Online Environment Interest Group: Bibliography of Current Issues in Authority Control (covering January 2000 to July 2001); <http://www.lita.org/igs/Acig/2002bibliography.pdf>.

[BIBCONTROL 2002]: Bibliographic Control of Web Resources: A Library of Congress Action Plan; Update; November 20, 2002; <http://lcweb.loc.gov/catdir/bibcontrol/actionplan.pdf>

[Bide 1999]: Mark Bide: <in decs> Directory of Parties: Outline Specification; October 25, 1999; <http://www.indecs.org/pdf/DirectoryofParties.pdf>.

[Bide 2002]: The InterParty Project. A Framework for the Interoperable Unique Identification of Parties in E-Commerce. An Introduction to Interparty. InterParty Workshop, London, September 30, 2002.

[Bourdon 2001]: Françoise Bourdon: Modéliser les données d'autorité. In: Bulletin des bibliothèques de France 46:5 (2001) p. 117-119, http://bbf.enssib.fr/bbf/html/2001_46_5/2001_5-p117-bourdon.xml.asp.

[Chan/Hu/Lo 2000]: A Collaborative Project on Chinese Name Authority Control: The HKCAN Model; February 2000; <http://hkcan.ln.edu.hk/papers/authenpa.pdf>.

[Delsey 2002]: Functional Requirements and Numbering of Authority Records. A Conceptual Model. Prepared by Tom Delsey for the IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Draft; June 07, 2002 (not available online).

[Dixon 2002]: Larry E. Dixon: Bath Profile. Functional Area D: Authority Record Search and Retrieval in Online Library Catalogs. Presentation for LITA (Atlanta), June 16, 2002; <http://www.lita.org/igs/Acig/authcontrol2.htm>.

[FORUM 2002]: Library of Congress Conference 2000 Action Plan Forum, American Library Association 2002 Annual Conference, Discussion Highlights; <http://www.loc.gov/catdir/bibcontrol/highlights061602.pdf>.

[FRBR 1998]: IFLA Study Group on the Functional Requirements for Bibliographic Records / [International Federation of Library Associations and Institutions. IFLA Universal Bibliographic Control and International MARC Programme, Deutsche Bibliothek, Frankfurt am Main]: Functional Requirements for Bibliographic Records. Final Report. München : Saur 1998 (= UBCIM publications ; N.S., 19); <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.

[Hodge 2000]: Gail Hodge "Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files," CLIR Publications & Resources, pub91 (April 2000). URL: <http://www.clir.org/pubs/reports/pub91>

[Hu/Lo/Kuo/Lau 2000]: Lily Hu / Patrick Lo / Evelyn Kuo / Maria Lau: The Implementation stage of the Hong Kong Chinese Authority Name (HKCAN) Database; 2000; http://hkcan.in.edu.hk/papers/hkcan_pa1.doc.

[Guidelines for Authority Records 2001]: *Guidelines for Authority Records and references*. 2. ed. / rev. by the IFLA Working Group on GARE Revision. UBCIM publications. New series ; 23 (München : Saur, 2001)

[ICACDS 2001]: International Council on Archives – Committee on Descriptive Standards: Minutes of First Plenary Meeting, Belgian State Archives, Brussels, Belgium, 3–6 October, 2001; <http://www.hmc.gov.uk/icacds/eng/minutesoct2001.htm>.

[ICACDS 2002a]: International Council on Archives – Committee on Descriptive Standards: Minutes, Madrid Meeting, June 12–15, 2002; <http://www.hmc.gov.uk/icacds/eng/minutesjune2002.pdf>

[Indecs 2000a]: Putting Metadata to Rights. Summary Final Report. <Indecs> Interoperability of Data in E-commerce Systems; June 2000, <http://www.indecs.org/pdf/SummaryReport.pdf>.

[ISAAR (CPF) 1995]: ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families. Final ICA approved version. Prepared by the Ad Hoc Commission on Descriptive Standards. Paris, France, 15–20 November 1995. Ottawa: The Secretariat of the ICA Ad Hoc Commission on Descriptive Standards 1996; http://www.ica.org/biblio/cds/isaar_eng.pdf
<http://www.hmc.gov.uk/icacds/eng/standards.htm> and
<http://www.ica.org/eng/mb/com/cds/descriptivestandards.html>

[ISAD(G)]: *ISAD(G): General International Standard Archival Description - 2nd Edition*. Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999 (2000) URL: as ISAAR(CPF) above

[KAISER 2002]: Max Kaiser, "Ansetzung von Personennamen und Körperschaften in Nachlaß- und Autographenkatalog (NAK) der Österreichischen Nationalbibliothek: Richtlinien, Geschäftsgänge und Datenfluß" (2002; workflow _authorities_onb-nak_02.pdf, reported to RA 2002-04-25)

[Kowok 2002]: Lolita Kwok: HKCAN Project: an overview. Presentation at the JULAC-HKCAN / OCLC Meeting; May 7, 2002; http://hkcan.in.edu.hk/presents/0205_hkcan71.ppt.

[Lam 2001]: XML and global name access control. In: OCLC Systems & Services 18:2 (2002), p 88–96.

[Lam/Kwok 2002]: Ki-Tat Lam / Louisa Kwok: XML Name Access Control Metadata Repository. An Experiment at HKUST Library; Oct. 2, 2002; <http://library.ust.hk/info/reports/xmlnac.html>

[Lau 2001]: Maria Lau: HKCAN 7XX Model. Presentation at the 2nd Hong Kong Innovative Users Group (HKIUG) Meeting; Dec 11, 2001; http://hkcan.in.edu.hk/presents/0112_cuhk_maria.ppt.

[Lunau 2002]: Carrol Lunau: Profiles: what are they and why do I care? Real World Steps to Interoperability ALA, Atlanta, Georgia, June 16, 2002; <http://www.lita.org/igs/Acig/authcontrol4.htm>.

[MARBI 2001]: MARBI Multilingual Record Task Force: Multilingual Authority Records in the MARC 21 Authority Format, June 8, 2001; <http://lcweb.loc.gov/marc/marbi/2001/2001-dp05.html>.

[MARD-3]: Per-Gunnar Ottosson: LEAF – Linking and Exploring Authority Files. Model Application Requirements Document, Part 3: Analysis of Practice. Final Draft, Vers. 0.4, Nov. 6, 2002.

- [MLAR]: IFLA UBCIM Working Group on Minimal Level Authority Records, *Mandatory Data Elements for Internationally Shared Resource Authority Records*. URL: <http://www.ifla.org/VI/3/p1996-2/mlar.htm>
- [NACO 1996]: Library of Congress: NACO Participants' Manual. 2nd Edition. Washington, D.C.: Library of Congress: Cataloging Distribution Service 1996; <http://www.loc.gov/catdir/pcc/naco/npm2ed.pdf>
- [MRAD1]: Model Requirements Analysis Document - 1: Analysis and Evaluation of Input Data (=LEAF Report, D 5.1, part 1)
- [MRAD2]: *Model Requirements Analysis Document - 2: User Groups Statements and Special User Scenarios*. (=LEAF Report, D 5.1, part 2)
- [MURTOMAA 1994]: Murtomaa, Eeva and Eugenie Greig with help of Joan Aliprand. "Problems and Prospects of linking Various Single-Language and/or Multi-language name Authority Files," *International Cataloguing and Bibliographic Control*, v. 23, no. 3 (July/Sept. 1994), p. 55-58
- [NACO 7XX]: Use of 7XX fields in NACO Name Authority Records; Jan. 4, 2002; <http://www.loc.gov/catdir/pcc/naco/7xx.html>.
- [NACOPARA]: NACO Program Parameters; <http://www.loc.gov/catdir/pcc/nacopara.html>.
- [Naito 2001]: Eisuke Naito: Why Authority Control among Chinese, Korean and Japanese Scripts?; Presentation at Workshop on Authority Control among Chinese, Korean and Japanese Languages; Jan. 10, 2001; <http://www.nii.ac.jp/CAT-ILL/PUB/CJK-WS/1-1Naito.pdf>.
- [National Register of Archives 1995]: *The National Register of Archives : an international perspective : essays in celebration of the fiftieth anniversary of the NRA*, ed. Dick Sargent. Historical research. Special supplement ; 13. (London : Univ., 1995)
- [NDL Newsletter 2001]: National Diet Library Newsletter, 119, May 2001; http://www.ndl.go.jp/en/publication/ndl_newsletter/119/192.html
- [Ottosson 2002a]: Per-Gunnar Ottosson: EAC Crosswalk. ISAAR(CPF)-2 / Madrid Draft to EAC / Alpha Version; Draft, Vers. 0.2, Sept. 04, 2002; <http://www.iath.virginia.edu/eac/documentation/eacxwalkisaar.html>.
- [Ottosson 2002b]: Per-Gunnar Ottosson: Encoded Archival Context (EAC) – Recent Developments. In: LEAF Newsletter, No. 1, March 2002; http://www.crxnet.com/leaf/news_online.html#eac2.
- [Patton 2002]: Glenn Patton: International Efforts to Improve Interoperability. ALCTS/LITA Authority Control in the Online Environment, June 16, 2002; <http://www.lita.org/igs/Acig/authcontrol3.htm>.
- [Pitti 2001]: Daniel V. Pitti: Creator Description: Encoded Archival Context. Paper presented at Computing Arts 2001: Digital Resources for Research in the Humanities. University of Sydney September, 26-28, 2001; <http://setis.library.usyd.edu.au/drrh2001/papers/pitti.pdf>
- [Plassard/Ratthei 1999]: Marie-France Plassard / Stephanie Ratthei: Universal Bibliographic Control and International MARC (UBCIM). Annual Report 1999; <http://www.ifla.org/VI/3/annual/ann99.htm>.
- [RNA]: Deutsche Forschungsgemeinschaft, *Regeln zur Erschließung von Nachlässen und Autographen (RNA)*. Deutsches Bibliotheksinstitut, Berlin, 1997 mit Ergänzungen - Stand Dezember 1998. URL: http://zka.staatsbibliothek-berlin.de/rna/rna_00.htm
- [RND]: Gunnar Karlsten: LEAF – Linking and Exploring Authority Files. Recommended Name DTD, 2nd Draft, Sept. 18, 2001.
- [Rules 1997]: *Rules for the Construction of Personal, Place and Corporate Names* (National Council of Archives, 1997)
- [Rust/Bide 2000]: Godfrey Rust and Mark Bide: The <indec> metadata framework. Principles, model and data dictionary; June 2000; <http://www.indec.org/pdf/framework.pdf>.
- [Suleman et al. 2001]: Hussein Suleman, Anthony Atkins, Marcos A. Gonçalves, Robert K. France, Edward A. Fox, Vinod Chachra, Murray Crowder, Jeff Young: Networked Digital Library of Theses and Dissertations. Bridging the Gaps for Global Access. Part 1: Mission and Progress, Part 2: Services and Research. In: D-Lib

Magazine 7:9 (September 2001), <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html> (DOI: 10.1045/september2001-suleman-pt1) and <http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html> (DOI: 10.1045/september2001-suleman-pt2).

[Tillett 1996]: Tillett, Barbara B., "International Shared Resource Records for Controlled Access", *Authority Control in the 21st Century: An Invitational Conference*. URL: <http://www.oclc.org/oclc/man/authconf/tillett.htm>

[Tillett 2001]: Barbara B. Tillett: Authority Control on the Web. Final Version; Jan. 23, 2001; Bicentennial Conference on Bibliographic Control for the New Millennium; http://lcweb.loc.gov/catd/ir/bibcontrol/tillett_paper.html.

[Tillett 2001a]: Barbara B. Tillett: A Virtual International Authority File; 67th IFLA Council and General Conference, Boston, August 16-25, 2001; <http://www.ifla.org.sg/IV/ifla67/papers/094-152ae.pdf>.

[Tillett 2002]: Barbara B. Tillett: A Virtual International Authority File. Presentation for the Council on East Asian Libraries, Association for Asian Studies, Committee on Technical Processing, 2002 Annual Meeting, Washington, D.C., April 4, 2002; http://kcan.in.edu.hk/presents/0204_ceal2002.ppt.

[Tillett 2002a]: Barbara B. Tillett: Work Plan Form for Action Item 2.2; May 30, 2002; <http://www.loc.gov/catdir/bibcontrol/2-2workplan.pdf>.

[Young 2001]: Jeffrey A. Young: NDLTD: Authority Linking Proposal. DRAFT, 12 April, 2001; <http://alcme.oclc.org/ndltd/AuthLink.html>.

[Zillhardt / Bourdon 1998]: Sonia Zillhardt / Françoise Bourdon: AUTHOR Project. Transnational Application of National Name Authority Files. Final Report, 5 June, 1998, <http://portico.bl.uk/gabriel/projects/pages/cobra/author.pdf>.

Definitions, Acronyms and Abbreviations

The following list provides the definitions and abbreviations used within this document.

- CNAF** Central Name Authority File
- CNAR** Central Name Authority Record
- EAC** Encoded Archival Context.
URL: <http://www.library.yale.edu/eac/> and <http://jefferson.village.virginia.edu/eac/>
- ICA/CDS** International Council of Archives/Committee on Descriptive Standards
- IST** Information Society Technologies
- GKD** [*Gemeinsame Körperschaftsdatei*](#) Name authority file for corporate bodies, refers here to the GKD of the German Library if not otherwise specified
- LAF** LEAF Authority File. Collection of all LARs.
- LAL** LEAF Automatic Link
- LAR** LEAF Authority Record. Any single authority record in the LEAF system
- LEAF** Linking and Exploring Authority Files
- LML** LEAF Manual Link
- LNL** LEAF Negative Link
- NAK** Nachlaß- und Autographenkatalog (NAK) [Catalogue of Collections and Manuscripts of the National Library of Austria]. URL: <http://nak.onb.ac.at/cgi-bin/allegro/nak/nak.pl>
- NRA** The UK National Register of Archives (NRA) maintained by the Historical Manuscripts Commission. URL: <http://www.hmc.gov.uk/nra/>
- PICA/ILTIS** The union catalogue system of the German Library
- RA-NAD98** Riksarkivet: the National Archival Database on CD-ROM (1998)
- PND** [*Personnamendatei*](#) Name authority file for persons, refers here to the PND or the German Library if not otherwise specified
- SLAF** Shared LEAF Authority File: Collection of all SLARs
- SLAR** Shared LEAF Authority Record: Any group of two or more LARs linked as a result of a rule
- WP** Work Package

Partner Acronyms:

- BL** British Library
- BN** Biblioteca Nacional
- CNS** Crossnet Systems Ltd.
- DLA** Deutsches Literaturarchiv
- FDÖP** Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie

- GSA** Goethe- und Schiller-Archiv
- IMEC** Institut Mémoires de l'Edition Contemporaine
- JRS** Joanneum Research
- NUK** National and University Library, Ljubljana, Slovenia
- ÖNB** Österreichische Nationalbibliothek
- RA** Riksarkivet
- SBB** Staatsbibliothek zu Berlin
- SNL** Swiss National Library
- UCM** Biblioteca de Universidad Complutense de Madrid
- UoB** University of Bergen