LEAF - Linking and Exploring Authority Files

IST-2000-26323

RECOMMENDED NAME DTD

DATE	15.12.2001
AUTHOR	Gunnar Karlsen
QUALITY ASSURANCE	Robert Bull
WORK PACKAGE	4.0
KEYWORDS	DTD, XML, Schema, data format
DELIVERABLE NUMBER	D 4.1

DOCUMENT HISTORY

Version	Date	Comments	Status	Distribution
0.1	08.08.2001	Within this document the recommended name DTD for LEAF is described	Draft	UoB's LEAF team
0.2	18.09.2001	Version for commenting by partners	Draft	LEAF consortium
0.3	12.10.2001	Final version	Draft	LEAF consortuim
1.0	15.12.2001	Final version, incorporating corrections by R.Bull, QA Acceptance	Final version	LEAF Consortium and European Commission

Table of Contents

1 Introduction

1.1 Scope

This document is the report on a recommended name Document Type Definition (DTD), prepared for the Project LEAF (Linking and Exploring Authority Files; project number IST-2000-26323). This document was done under Work Package 4.0 (*Data Representation Study*) of the current project.

The purpose of this document is to establish a name XML DTD for the project and show how it integrates with the metadata set.

The markup permitted in a particular XML application can be documented in a DTD. The DTD lists all legal markup and specifies where and how the markup may be included in a document. Particular document instances can then be compared to the DTD. Documents that match the DTD are said to be *valid*.

The following document presupposes that the reader is familiar with the basic principles of markup-languages and concepts of XML (Extended Markup Language).

1.2 Document Overview

Chapter 1 is this *introduction* of the Recommended Name DTD. In **Chapter 2** an introduction is given of the already existing standards for the description of content in libraries, archives and museums. Together with the presentation of other relavant projects in **Chapter 3**, this should present the needs for a seperate dataformat for authority files to fit the LEAF project. **Chapter 4** presents the new XML Schema standard and compares this to DTDs. Finally the work of the EAC group with the new DTD is presented in **Chapter 5**, together with examples of LEAF data providers' formats encoded in EAC in **Chapter 6** The work on a recommended name DTD for LEAF is concluded in **Chapter 7**.

2 Existing standards

This section describes the use of person and corporate bodies information in different existing standards, such as UNIMARC, CIMI DTD, EAD and related standards, as well as projects intended to convert data between these.

In the archives and libraries sector there are already several existing standards for description of content, some of them also including authorities information. An important distinction lies in whether the dataprovider/repository uses a separate authority file for the creator of the material or if this information is included in the bibliographical description of it.

2.1 MARC/UNIMARC

The original MARC (Machine Readable Catalogue) was developed to serve the needs of libraries as a convenient way of storing and exchanging bibliographic data. Despite cooperation in the development of the standard, there has emerged several versions, e.g. UKMARC, INTERMARC and USMARC, whose paths diverged owing to different national cataloguing practices and requirements. A solution to this problem of

incompatibility was to create an international MARC format (UNIMARC, http://www.ifla.org/VI/3/p1996-1/unimarc.htm) which would accept records created in any MARC format.

The other major MARC format is the USMARC (also known as MARC21 after merging with CAN/MARC), maintained by the Library of Congress and available at http://www.loc.gov/marc/.

2.1.1 UNIMARC AUTHORITIES

In 1991 a new authorities format, UNIMARC Authorities, was published. Previously an author's name was entered into the bibliographic format as many times as there were documents associated with him or her. With this format, a new single authoritative form of the name (with references) can be created in the authorities file and the record control number for this name is the only item included in the bibliographic file.

The UNIMARC Authorities format is designed to allow an agency to hold in one place the authoritative name of the author, corporate body name, etc., together with references to other forms of the name. Such data is linked to a bibliographic record by subfield $3 (Authority record number) in fields in the 7-- block of the bibliographic format. The data can then be included in the bibliographic record either at the time of creation or when a user views that record. See http://www.ifla.org/VI/3/p1996-1/uniafull.htm for documentation of the format.

2.1.2 Examples, MARC-records

Below is a selection of MARC records in 'native' format. The same records are encoded using the proposed EAC DTD later in this report (see page 18)

Example, USMARC from Universidad Complutense de Madrid (UCM), Spain:

Authority file – personal name

```
A10000471    Actualizado el: 21-08-00  Creado el: 21-08-00  Revisión: 1TIPO: -
FUNCION: -         ASUPRIMIR: -
001     000000014x
008                                   0        nz  a
100  1  Pi Sunyer, Carles,|d1888-1971
400  1  Pi Suñer, Carlos
400  1  Pi i Sunyer, Carles
400  1  Pi Sunyer, Carlos
```

Description of fields:

100 1 Heading – Personal Name (indicator 1 denotes that the name is a surname formatted in inverted order or a name without forenames that is known to be a surname), $d are dates associated with the name.
400 1 See From Tracing – Personal Name

Authority file – corporate name

```
 A10000288    Actualizado el: 21-08-00  Creado el: 21-08-00  Revisión: 1TIPO: -
FUNCION: -         ASUPRIMIR: -
001     000000001r
008                                   0        nz  a
110  2  Instituto Nacional de Fomento de la Exportación (España)
510  2  Instituto Español de Comercio Exterior (Madrid)
510  2  Centro de Documentación e Información del Comercio Exterior (España)
```

```
663     |iVéase además
```

Description of fields:

110 2 Heading – Corporate Name (indicator 2 denotes that the name is in direct order).
510 2 See Also From Tracing – Corporate Name. Is used in an established heading or established heading and subdivision record to trace a see also from reference from a corporate name to a related established heading.
663 Reserved for local use

Both the examples are simple, containing little other information besides the name of the entry and other or related name forms.

2.1.3 Examples, UNIMARC Authorities

Example, UNIMARC Authorities from Biblioteca Nacional (BN), Portugal):

Authority file – corporate body

```
001:1
005:19980922162000.0
100:  ^a19890804apora0103    ba
152:  ^aRPC
210:01^aBraga.^bBispo,^c1932-1963 (Ant-onio Bento Martins J-unior)
305:0 ^aAs obras deste autor encontram-se sob os seguintes cabe-calhos de acordo com o
exerc-icio do cargo eclesi-astico ou n-ao:^bBraga. Bispo, 1932-1963(Ant-onio Bento
Martins J-unior);^bMartins J-unior, Ant-onio Bento, 1881-1963
500: 1^aMartins J-unior,^bAnt-onio Bento,^f1881-1963
801: 0^aPT^bBN^c19890804
810:  ^aGr. Enc. Port. Bras.
830:  ^9PT^aReligioso
```

Short description of fields:

001 record identifier
005 version identifier
100 general processing data
152 rules
210 heading, corporate name
305 textual see also
500 see also reference tracing personal name
801 originating source
810 source data found
830 general cataloguers note

And a corresponding file for the personal name

```
001:2
005:20010209113700.0
100:  ^a19890804apora0103    ba
152:  ^aRPC
200: 1^aMartins J-unior,^bAnt-onio Bento,^f1881-1963
305:0 ^aObras deste autor encontram-se sob os seguintes cabe-calhos conforme desempenho
ou n-ao do cargo eclesi-astico:^bBraga. Bispo, 1932-1963(Ant-onio Bento Martins J-
unior);^bMartins J-unior, Ant-onio Bento, 1881-1963.
510:01^aBraga.^bBispo,^c1932-1963 (Ant-onio Bento Martins J-unior)
801: 0^aPT^bBN^c19890804
810:  ^aGrd. Enc. Port. Bras.
830:  ^9PT^aAutor de pastorais. Colab. de jornais e revistas
```

Short description of fields:

Same as above, except:

200 heading, personal name

510 see also reference tracing corporate name

2.1.4 Conversion tools for MARC formats

For both USMARC and UNIMARC, there have been several projects attempting to create a standard DTD to support the conversion of catalogue data from the MARC families of formats (and back) without the loss of data. The following are of notable interest.

MARC DTD Project

On http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html there is an introduction to the MARC DTD project. Two PERL programs which convert from MARC to MARC DTD and back have been developed. See http://lcweb.loc.gov/marc/marcsgml.html. The MARC-to-MARC DTD program is supposed to take different MARC versions, like USMARC, UNIMARC, etc., as input. It incorporates MARC data elements from five different separate formats into two DTDs: A DTD for the *USMARC Format for Bibliographical Data* includes additional elements from the *USMARC Format for Community Information* and the *USMARC Format for Holdings Data* to permit the creation of three record types within the framework of a single DTD. Similarly, a second DTD for the *USMARC Format for Authority Data* includes additional data elements from the *USMARC Format for Classification Data* to permit the creation of authority and classification records within a single 'authority' DTD.

The output file will be well-formed, but not necessarily valid XML. The names chosen for the elements (i.e. the XML tags) in the MARC DTD are number-based, derived from the variable field tags and subfield codes in the MARC formats themselves, often resulting in a hard to read and large file.

Medlane XMLMARC

Medlane XMLMARC provides an alternative to convert a MARC record to an XML document. The record format should be USMARC, Medlane have not experimented with other formats. Users can design a DTD file and specify any mapping that suits their need.

The Medlane XMLMARC software consists of two components: Client and Server. The Client program provides a graphical user interface and interaction with the server, while the Server program is responsible for converting a MARC record to an XML document according to the users' specifications.

The conversion tool supports USMARC only, uses a special designed DTD that is a subset of the MARC format and is available as freeware licence under GNU. See the project's official homepage http://xmlmarc.stanford.edu/ for more information.

BiblioML

The BiblioML project has published XML DTDs that defines the structures for bibliographic and authorities records based on the UNIMARC bibliographic and authorities formats. The conversion is accomplished in two steps; first the records are scanned and converted to a simple XML syntax by a Java-program, then the XML records are converted to BiblioML or AuthoritiesML, using XSLT.

The conversion tool supports UNIMARC, and requires Java 2 environment and Saxon v.5.3.1 or later for the stylesheet conversion. See the project's official home page http://www.culture.fr/BiblioML/en/index.html for more information.

MARC.pm – MARC::XML module

The MARC::XML module is a part of the MARC.pm PERL 5 module for reading, manipulating, outputting and converting bibliographic records in the MARC format. The MARC::XML program can output both single and

batches of MARC records, where the number of records in a batch is determined by the memory capacity of the hardware. The conversion tool supports USMARC only and requires PERL 5.004 or later. See the offical homepage http://marcpm.sourceforge.net/ for more information.

2.2 EAD

The Encoded Archival Description (EAD) DTD is a standard for encoding archival finding aids using SGML/XML. The standard is maintained by the Network Development and MARC Standards Office at the Library of Congress(LC) in partnership with the Society of American Archivists. The current version is 1.0 which has been available since 1998. An upgrade is in progress. See http://www.loc.gov/ead/ for the official home page.

In MALVINE, EAD was also used to encode authority files from some of the data providers, like the SBB [Bruvik99]. Some archives propably do this too to a certain extent, since the authority information then is included in the biographical records. When the personal name authority files consist mainly of a series of biographical records, all information will fit into the element <biohist> in EAD. The series of <biohist> elements might be encoded inside the <archdesc> like this, without any component <c> element:

```
<ead>
                                                    <eadheader>
                                                        ...
                                                    <archdesc>
      <bioghist>

                                                    </bioghist>
                                                    </archdesc>
</ead>
```

This is valid encoding in EAD, and it will emphasize that this is not a catalogue of a collection, but an encoding of biographical information.

Several other projects and standards are related to EAD, like the MUS-EAD. This is a project of the Museum Computer Network (MCN) Special Interest Group to test implementation of the EAD DTD in a museum environment. It is designed to test whether the EAD DTD, which was designed primarily for archival inventories and registers, is truly applicable to the many types of collections held in museums. The goal is to encourage as many museums as possible to encode at least one of their collection guides in EAD and to provide feedback to the MCN, which will report the group's findings to the EAD developers and to the Society of American Archivists EAD Working Group. See the projects home page http://www.mcn.edu/standards/mus-ead.html for more information.

2.3 TEI

The Text Encoding Initiative (TEI) has made comprehensive guidelines on how to encode text mainly for scholarly purposes, but also for the language industries more generally. TEI covers a broad range of texts and is well documented [TEI 1999]. See http://www.tei-c.org/ for the official home page. The current version is P3, but there is ongoing work to improve and extend TEI. Based on TEI guidelines it is possible to make different DTDs to cover a wide range of purposes.

As TEI is intended for the texts and archival material itself, it does not contain a wide series of elements for archival description and context.

2.4 CIMI

The Consortium for the Computer Interchange of Museums Information, CIMI, has developed the CIMI DTD for use in this part of the cultural heritage community. Version 4.0 of the CIMI DTD is basically a subset of TEI, with some new features to make it easier to produce exhibition catalogues and exhibit wall texts. The CIMI DTD was developed as part of the project CHIO (Cultural Heritage Information Online). According to [Light97], the CIMI DTD is not used in any other project. The current version of it, v.4.0, is not compatible with the CIMI profile for Z39.50 or Dublin Core. There are suggestions from 1997 to make a version 5.0, but this work appears not to have been continued. The development of the CIMI DTD seems to have stopped and the use of it is limited.

On behalf of the international museum community, CIMI has also been contributing to the Museum Documentation Association's (MDA) development of a DTD for SPECTRUM, MDA's international museum standard for describing objects, a well-known standard widely used in the museum world. CIMI will aim to ensure that the resulting XML standard is tested within the international user community and lead an XML-DTD testbed. See the home page http://www.cimi.org/wg/xml_spectrum/xml_testbed_desc.html for more information on this project.

2.5 ISAAR(CPF)

ISAAR(CPF) is an international standard for archival authority records for corporate bodies, persons and families. It is developed and maintained by the International Council on Archives, see http://www.ica.org/ for home page and a complete description.

The standard consists of a set of numbered information elements which may be used to control the creation and use of access points in descriptive records and/or to describe a corporate body, person, or family as units within an archival descriptive system, and it determines the types of information that could be included in an archival authority record. In contrast to MARC, EAD, etc, the ISAAR(CPF) is not an encoding scheme, but rather a description and set of rules of good archival practice.

2.6 Conclusion

MARC/UNIMARC Authorities are the only excisting standards for authorities information. These are designed for the exchange of information in the libraries, and are not suited for archival description. The tools for conversion to XML are not very robust.

Other standards, like the EAD, are designed for description of archival content, and not archival context, i.e. information about the creators.

3 Current work and projects

This section contains a description of related work done internationally in the libraries, the archives and the museums sector and the work done in relevant projects, like AUTHOR, ONE-2, COVAX, FRANAR, INDECS and MACS.

There are several ongoing or completed projects that are related to the use of authority files or information in archives and libraries. Note that only one of these explicit works on authority files, while the common factor for the other ones is the attempt to bring together different collections of related information or materials, often by the use of XML.

3.1 AUTHOR

The European Project AUTHOR, running from 1995-97, was the first real attempt to bring together 5 national libraries (Belguim, France, Portugal, Spain and the United Kingdom), each responsible for authority files with different rules, formats and languages. National records in various formats were mapped into UNIMARC formatted records and then exchanged using a program called 'UseMARCON'. Each of the partners created a sample set of authority records to test. The samples allowed comparison of some of the same authority records from the different dataprovoders, including all records for headings starting with O and T, records for persons with name Martin, Smith and Garcia, and some international corporate bodies. A final report on the project is available on http://www.bl.uk/gabriel/cobra/author.pdf.

This is the only project that actually has tried to share authority files across different national libraries. Since the tests were done on a small and limited amount of data, all of them within the MARC-family of formats, the results has little implication on the LEAF data format, that should cover a wider selection of local formats. The UseMARCON software could be of interest, and have been extensively used by the British Library national bibliographic service, by the Finnish union catalogue (converted 15 million records) and it is integrated in the Endeavour Voyager library management system. The software has also been considerably improved by Crossnet under contract to the British Library and is available to download from their web site http://www.bl.uk/ .

3.2 ONE-2

The ONE-2 extends the results of the preceding libraries project OPAC Network in Europe (ONE) by integrating additional library services and involving new application areas. The previous ONE project brought together several major European bibliographic institutions and libraries in implementing common software tools, profiles and service agreements for Z39.50 and On-line Public Access Catalogues (OPAC). The partners in the project will extend their Z39.50 based Search and Retrieval services to integrate other related functions, such as item ordering, inter library lending, copy cataloguing and update and electronic document delivery. There will also be a focus on searching across domains and the emerging use of Z39.50 and XML. The project started in 1999 and was finished on 18 October 2001. See the official home page http://www.one-2.org/ for project repots and more information.

3.3 COVAX

The pupose of COVAX (Contemporary Culture and Virtual Archives in XML) is to analyse and draw up the technical solutions required to provide access through the Internet to homogeneously encoded document descriptions of archive, library and museum collections based on the application of SGML/XML. The project will demonstrate its feasibility through a prototype containing a meaningful sample of all different types of documents to build a global system for search and retrieval. It is based on the assumption that in libraries, archives and museums an enormous number of descriptions could be made available over the Internet by converting existing records or by creating new ones to specific SGML/XML DTDs.

COVAX intends to combine document descriptions from libraries, museums and archives with digitised surrogates of their materials, in order to build a global system for search and retrieval. It will allow the widely distributed primary documents from these institutions to be accessed regardless of their location. The basic

operational core is the application of SGML/XML and the various DTDs currently being defined in libraries (MARC DTD), archives (EAD) and museums (MUS-EAD), as well as the DTD defined by the Text Encoding Initiative (TEI) for the mark-up of electronic versions of cultural texts. The use of SGML/XML will enable standardization, interoperability and interconnectivity between libraries, archives and museums in the processes of handling, browsing, searching and retrieval of all kinds of descriptions and documents.

The project started 01.01.00, runs for 24 months, and at the current stage there is a public report delivered on available software (editors, parsers, browsers) of interest to the COVAX project on the official homepage http://www.covax.org

3.4 FRANAR

As a follow-up to the IFLA Working Group on Minimal Level Authority Records and the ISADN, which had completed its task in 1998, the FRANAR (Functional Requirements And Numbering of Authority Records) working group was created in June 1999 under the auspices of the Division of Bibliographic Control and the IFLA UBCIM Programme. The Working Group comprises ten members representing various interests and institutions.

The working areas of the group are to define functional requirements of authority records, to study the feasibility of an International Standard Authority Data Number (ISADN and to to serve as the official IFLA liaison to, and work with, other interested groups concerning authority files.

Information about the initative is available at the official home page http://www.ala.org/alcts/alcts_news/v10n1/gateway.html

3.5 INDECS

The Indecs (Interoperability of Data in E-Commerce Systems) project is an international initiative of rights owners creating metadata standards for e-commerce. Indecs was designed as a fast track, infrastrukture project aimed at finding practical solutions to interoperability affecting all types of right-holders in a network, e-commerce environment. The project focussed on the practical interoperability of digital content identification systems and related rights metadata within multimedia 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.

A generic metadata model is the foundation work of Indecs. The model is best described as a data model with a formal structure for describing and uniquely identifying intellectual property. The basis of the model is the assumption that it is possible to produce generic systems to handle all creation types. The project was completed in March 2000 and a public summary report is available at the project home page http://www.indecs.org.

3.6 MACS

The MACS (Multilingual Access to Subjects) project's goal is to develop a management system to join various European Subject Heading Languages (thesauri) into one, virtual search space. The MACS Project is related to work in the Cobra+ Project (**CO**mputerised **B**ibliographic **R**ecord **A**ctions, homepage http://www.bl.uk/gabriel/cobra/finrap3.html).

A key issue of the project is to maintain the independence of the various contributing Subject Heading Languages. In order to achieve this, there is established a *federated SHL/thesaurus management system*. It does not physically or logically combine the contributing SHLs into one merged, consolidated system. Instead, it acknowledges the independence of the contributing SHLs and uses explicit links between them which are maintained outside of the SHL data bases. The maintenance of the Link Database is decentralized and done by the participating partners in a federated organization. A prototype which contains a small subset of data from the indexing languages and libraries, has been developed for testing. The project's official home page is http://infolab.kub.nl/prj/macs/

3.7 Conclusion

Except for the AUTHOR project, little work has been done to explore the exchange of authority files in archives and national libraries. This confirms the statement in the LEAF technical annex, saying that 'there is no commonly accepted model for authority data in Europe' and no possibilities for the users to search across different holders of such data, although some parallell problems are discussed in working groups like FRANAR.

4 XML DTD or Schema?

4.1 XML Schema

With the approval in May 2001 of the XML Schema as a W3C Recommendation (see www.w3.org), it is necessary to take into account the future use of this specification as a data format. The XML Schema language is intended to supplement the basic DTD mechanism included in XML version 1.0 with a much more rigorous framework for declaring the structure and contents of XML documents. Besides basic elements and attribute relationship facilities built into XML 1.0, schemas allow application writers to enforce specific datatype restrictions on their documents' contents. It also provides support for the construction of user-defined complex datatypes, data ranges, and masks.

The XML Document Type Definitions are a part of the XML 1.0 W3C Recommendation Feb-98, used to define constraints on the logical structure of the XML document. XML documents are usually validated against an associated DTD.

By definition an XML document is *valid* if it has an associated document type definition and if the document complies with the constraints expressed in it. XML schemas work in the same way as a DTD by providing the constraints by which the document is validated against. There is still a difference, since a schema allows us to test for two kinds of validity; the validity of the content models and the validity of the specific units of data. Content model validity tests whether the order and nesting of tags is correct, similar to a DTD, while datatype validity is the ability to test whether specific units of information are of the correct type and fall within the specified legal values. If we look at the differences between a DTD and a Schema, this datatyping is a major improvement in Schema for certain purposes.

4.1.1 Improvements in XML Schema

The main distinctions between DTDs and Schema are:

XML-syntax

XML Schema documents are XML documents. This means that they use elements and attributes to express the semantics of the schema and that they can be edited and processed with the same tools that you use to process other XML documents. Based on XML syntax, a Schema can also easily be rewritten to a more or less formal structure, elements can be added or removed and all this can be done by XSLT. The XML syntax also gives a higher degree of human readability to Schema, compared to DTDs.

Datatypes

In XML Schema there is a basic difference between complex types which allow elements in their content and may carry attributes, and simple types which cannot have element content and cannot carry attributes, but contain numbers, strings, dates, etc. Attributes always have simple types. There is also a major distinction between definitions that create new types (both simple and complex) and declarations which enable elements and attributes with specific names and types (both simple and complex) to appear in documents instances.

The simple types are predefined in a similar way to programming languages and can be further constrained through the application of one or more facets. The following table shows some examples of the simple types and a selection of applicable facets (see http://www.w3.org/TR/xmlschema-0/, Appendix 3 for a complete list):

Simple types	Facets				
	length	minlength	maxlength	pattern	enumeration
string	y	y	y	y	y
integer				y	y
boolean				Y	
language	y	y	y	y	y
IDREF	y	y	y	y	y

Table 1: Schema Datatypes

Derived simple types and complex types can be constructed from the list of simple types, allowing an almost unlimited degree of constraining the data.

Support for regular expressions

Schema allows you to specify a *pattern* on elements and attributes which uses the familiar *regular expression* syntax to describe the sorts of things that are allowed in a given field. To specify an element with a pattern, for example, you would create a new type based on an existing type and define the pattern like this:

```
<simpleType name=”PostalCode” base=”string”>
      <pattern value=”\d{4}” />
</simpleType>
```

This pattern would define a four-digit postal code, as used in Norway.

Use of XML namespaces

Namespaces have two functions in XML

1. To distinguish between elements and attributes from different XML applications that share the same name.

2. To group all related elements and attributes from a single XML application together so software can recognize them easily.

Namespaces are implemented in XML by attaching a prefix to each element and attribute. Each prefix is mapped to a URI. Schema itself is connected to a required nemespace which is always defined in the header of any Schema:

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
```

4.2 Why still use a DTD?

4.2.1 General arguments

Since XML Schema uses XML syntax, it parses with ordinary XML parsers, but there is a lack of reliable software and problems with existing Schema validators as they are still only beta versions (pr. 08.08.01). An example is the Xerces validator (http://xml.apache.org/xerces-c/schema.html), which only implements a subset of the XML Schema Recommendation. The target is to implement all the features of the current XML Schema Recommendation before the end of this year. This date is still tentative and subject to change.

Every DTD can at a later stage be converted to a Schema specification. Software that performs the first stage of this process exists, and it is expected that these will offer a more detailed conversion. Any DTD can be converted to a Schema specification at the same level of constraint, but further refinement requires more accurate knowledge of the data in the model and is likely to be done manually.

EAC (and other complex DTDs like EAD, DocBook and TEI) make extensive use of such parameter entities in their designs. Parameter entities do not map directly into Schema structures, and thus to accomplish the same or similar semantic and structural objectives in Schema requires some rethinking and "re-engineering". For large structures, like EAD, this will require a major effort.

On http://www.w3.org/2000/04/schema_hack/ there is a free conversion tool from simple DTDs to XML Schema. The tool is written as a Perl script, and although it does not reflect the last changes in the recommendation (it is based on the previous, 20000407 version, of XML Schema), it still works for a first stage conversion.

The script dtd2xsd.pl takes a DTD and does a simple conversion to a schema, where the local namespace by default is named 't' and set to http://www.w3.org/namespace/ in the header.

If we look at the date-element in the proposed EAC DTD:

```
<!ELEMENT date
     (%later; | altdate)*
>
<!ATTLIST date
          type          (begin | end | active | begin-end)  #IMPLIED
          form          (single | rangeclosed | rangeopen)  #IMPLIED
          era           CDATA                               "ce"
          calendar      CDATA                               "gregorian"
          normal        CDATA                               #IMPLIED
>
```

Note the parameter entity:

```
<!ENTITY % later
     '#PCDATA'
>
```

The conversion tool uses the XML Schema elements from the XML Schema namespace (http://www.w3.org/2001/XMLSchema) and the converted part of this section, after being processed by the dtd2xsd.pl script, is:

```
<element name='date'>
  <complexType mixed='true'>
   <sequence minOccurs='0' maxOccurs='unbounded'>
    <element ref='t:altdate'/>
   </sequence>
   <attribute name='type' use='optional'>
    <simpleType>
     <restriction base='string'>
      <enumeration value='begin'/>
      <enumeration value='end'/>
      <enumeration value='active'/>
      <enumeration value='begin-end'/>
     </restriction>
    </simpleType>
   </attribute>
   <attribute name='form' use='optional'>
    <simpleType>
     <restriction base='string'>
      <enumeration value='single'/>
      <enumeration value='rangeclosed'/>
      <enumeration value='rangeopen'/>
     </restriction>
    </simpleType>
   </attribute>
   <attribute name='era' type='string' use='default' value='ce'/>
   <attribute name='calendar' type='string' use='default' value='gregorian'/>
   <attribute name='normal' type='string' use='optional'/>
  </complexType>
 </element>
```

Note that the conversion tool does not handle the use of parameter entities and that this entity is lost in the conversion, while the `altdate` element is kept with its attributes.

After this first stage in the conversion, XML files that were valid according to the initial DTD will still be valid according to the Schema. The data types can now be modified to draw full value of the possibilities provided by Schema, but, as a result of this, the content of the XML files may have to be corrected.

The whole date element could also be changed by using one or more of the simple types in Schema for time and dates, like timeDuration or recurringDuration. These are ISO 8601 formatted (recurring) time intervals and they have the maxInclusive, maxExclusive, minInclusive, minExclusive, duration, period, pattern and enumeration constraing facets.

4.3 Conclusion

Looking at the scope and functionality that schemas will provide, they seem like a great improvement over DTDs. Certain kinds of applications, exchanging information between databases, for example, and ecommerce are certainly going to be made simpler and more interoperable by XML Schema.

Still, the primary virtue of DTDs today are that they are well understood and they do offer a good way to describe the structure of a document for exchange. It will take some time before XML Schema are as well understood and supported. Looking to the EAD working group, they have deferred making a decision concerning Schema until there is a major support for it as demonstrated by significant software developments and implementations, and there are clear benefits associated with having an EAD Schema.

When we also look at the big gaps and differences between the different data providers' formats, the use of a more specified and granular standard for data representation, like Schema, would make it even harder to find an accurate mapping against the LEAF metadata set. These arguments are all based on the assumption that the Schema will be more resticted than the result of the conversion shown above, assumed that the conversion can be done without loss, to make full use of the improvements in Schema. A proper use of Schema would also complicate the conversion process as the LEAF conversion tool would have to convert the local formats to a much more constrained common format.

The conclusion seems that a DTD still is the choice as data format for the LEAF system, and that use of Schema should occur at a later stage when it is properly supported by software.

5 The EAC (Encoded Archival Context) DTD

It has for a while been recognized that there is a need for an archival standard for describing information about the creators of archival materials – individuals and families, organisations and corporate bodies. With the exceeding use of EAD to describe archival content, it has been suggested [Pitti99] that this standard should be extended with a new DTD for authority control and that this should be based on ICA's International Standard Archival Record for Corporate Bodies, Persons, and Families (ISAAR(CPF)). With financial support from the Delmas Foundation to the Research Library Group (RLG), there has been two meetings in 2001 to advance work on this standard, which has been named Encoded Archival Context (EAC). During this process, the LEAF project has been represented in the working group to ensure that the needs of LEAF are implemented in the standard. Special attention has been paid to ensure that the LEAF metadata set matches the EAC DTD.

At the current stage (July 2001), there is a test version of the DTD available at http://bembo.village.virginia.edu/eac.html, together with some examples of manually created records. There is also released a tag library, but no further documentation. XML/SGML DTDs are by nature hard to read and use of the test version requires familiarity with these standards.

The next stage in the development will consist in encoding a broad range of records and evaluating the functionality of the structure. An official alpha version is expected around September 2001.

5.1 Structure of the EAC

The main structure of the EAC is a division at the top level into the two elements <eacheader> and <condesc> (see figure).

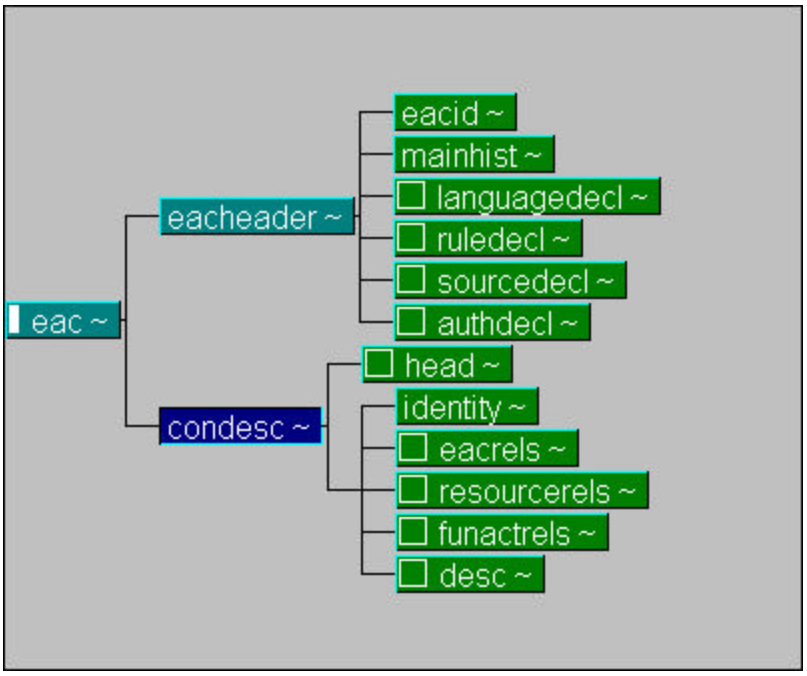

Figure 1: Main structure of the EAC DTD

The <eacheader>-element is used in a similar way as the header in EAD and TEI, and contains (meta)-information for editorial control of the EAC-instance itself.

The <condesc>-element (contextual descriptions) contains the authority file and is divided into the following elements:

- <identity> : contains information required for unique identification of the instance, including authorized and other forms of name, dates and additional information.
- <eacrels> : references to related entities described in other EAC instances
- <resourcerels> : references to other archival materials, books, museum objects etc.
- <funactrels> : references to functions and activities held or maintained by the instance.
- <desc> : for controlled and/or prose description of the entity. It may contain highly formalized descriptions (e.g. sex, occupation, function) or relatively unstructured prose descriptions.

5.2 The encoding analog attribute and backward conversion

A special detail in EAC, that is inherited from the EAD DTD, is the encoding analog attribute. This is used in the following way:

In the header element, there is an (optional) attribute named encodinganalogsys that allows for declaring the name of the local format, like in the following example:

```
<eacheader type="p" status="draft" langencoding="ISO 639-2B" scriptencoding="ISO 15294" dateencoding="ISO 8601"
countryencoding="ISO 3166-1 a2" ownerencoding="ISO 11551" encodinganalogsys="UNIMARC Authorities">
```

Each element containing information converted from a local format can now be mapped to the original format by keeping the name of the local field in the ea attribute:

```
<part ea="200: 1 $a" type="surname">Martins J-unior</part>
```

```
<part ea="200: 1 $b" type="forename">Ant-onio Bento</part>
```

This way of keeping the internal field number in the EAC record can reduce the loss of information that could occur in a conversion process, and, if used correctly, may also serve to reconstruct the original record.

5.3 EAC and LEAF metadata

The candidate DTD for LEAF must be mapped against the LEAF metadata set. This has been done through the whole process with EAC, as the metadata were available during the development of EAC.

It is important to note that XML DTDs, like the EAD and now the EAC, are loose structures, often allowing a field to be encoded in several different ways. The value of an element is often relative to the surrounding element(s) in the XML structure, where the actual semantic content of the field is determined by the way the element is used and its attributes. Note the following example with dates of birth and death:

```
<identity>
        <pershead rule="RPC" authorized="BN" languagecode="por" scriptcode="latn" type="identity">
                <part ea="200: 1 $a" type="surname">Martins J-unior</part>
                <part ea="200: 1 $b" type="forename">Ant-onio Bento</part>
        </pershead>
        <nameadds ea="200: 1 $f">
                <date form="rangeclosed" era="ce" scope="begin-end" calendar="gregorian"
                normal="1881/1963">1881-1963</date>
        </nameadds>
</identity>
```

EAC has only one date element for all dates, but inside the identity and nameadds elements we know that this is a date associated with a person, and therefore dates og birth and death.

The result of this is that there is no exact match between the elements in the LEAF metadata set and the elements in EAC, but the flexibility of the EAC DTD and the encoding practice will ensure that the metadata are covered.

6 Examples of encoding in EAC

As an important part of the testing of the EAC model, samples of data from different LEAF data-providers have to be encoded to ensure that the model covers all relevant information from the original data.

Below is a selection of records in EAC, each with comments on the encoding and possible conversion problems.

The following formats are used as examples of encoding in EAC:

Data provider/LEAF-partner	Local format
Biblioteca Nacional (BN), Portugal	UNIMARC Authorities
Universidad Complutense de Madrid (UCM), Spain	USMARC
Forschungsstelle und Dokumentationszentrum für Österreichishe Philosophie (FDÖP), Austria	Local format
Österreichische Nationalbibliothek (ÖNB), Austria	Allegro-Hans

Table 2: Dataproviders and formats tested in EAC

More examples can be found at the temporary EAC website http://bembo.village.virginia.edu/eac.html

General comments:

A conversion program for making all local date formats into ISO 8601 may be necessary. See the use of the `normal` attribute in all `date`-elements.

6.1 Example 1, UNIMARC Authorities from BN

```
<?xml version="1.0"?>
<!DOCTYPE eac SYSTEM "C:\Documents and Settings\gunnar\Mine Dokumenter\EAC\eac\eac.dtd">
<eac>
	<eacheader type="p" status="draft" langencoding="ISO 639-2B" scriptencoding="ISO 15294" dateencoding="ISO 8601"
	countryencoding="ISO 3166-1 a2" ownerencoding="ISO 11551" encodinganalogsys="UNIMARC Authorities">
		<eacid syskey="example07" countrycode="NO" ownercode="UoB">NO::UoB::example07
		</eacid>
		<mainhist>
			<mainevent maintype="create">
				<name>Gunnar Karlsen</name>
				<maindate calendar="gregorian" normal="20010801">August 01,2001</maindate>
				<maindesc>Manually created from BN testdata</maindesc>
			</mainevent>
		</mainhist>
		<languagedecl>
			<language languagecode="eng" scriptcode="latn" ea="100">English in Latin Script</language>
		</languagedecl>
		<ruledecl ea="152 $a">
			<rule id="RPC">Regras Portuguesas de Catalogação</rule>
		</ruledecl>
		<sourcedecl>
			<source id="bn" syskey="2">
				<title type="simple">BN test-data for LEAF, record id=1</title>
				<sourceinfo ea="005">
					<date form="single" normal="20010209">20010209113700.0</date>
					<note audience="internal">version identifier</note>
				</sourceinfo>
				<sourceinfo ea="100">
					<date form="single" normal="19890804" ea="100">19890804</date>
					<archref/>
				</sourceinfo>
				<sourceinfo ea="810 $a">
					<bibref>Grd. Enc. Port. Bras</bibref>
				</sourceinfo>
			</source>
		</sourcedecl>
	</eacheader>
	<condesc>
		<identity>
			<pershead rule="RPC" authorized="BN" languagecode="por" scriptcode="latn" type="identity">
				<part ea="200: 1 $a" type="surname">Martins J-unior</part>
				<part ea="200: 1 $b" type="forename">Ant-onio Bento</part>
			</pershead>
			<nameadds ea="200: 1 $f">
				<date form="rangeclosed" era="ce" scope="begin-end" calendar="gregorian"
				normal="1881/1963">1881-1963</date>
			</nameadds>
		</identity>
		<eacrels>
			<eacrel ea="510:01" reltype="associative" syskey="1">
				<corpname>
					<part ea="210:01 $a" type="Entry element">Braga.</part>
					<part ea="210:01 $b" type="Subdivision">Bispo,</part>
					<date normal="1932/1963">1932-1963</date>
```

```
                                        </corpname>
                                </eacrel>
                        </eacrels>
                        <desc/>
                </condesc>
</eac>
```

Problems and questions regarding the encoding:

- There is redundant information in the original fields 100, 801 and 200, 305, 500

- How to encode field 305, Textual See Also Reference?

6.2 Example 2, USMARC from UCM

```
<?xml version="1.0"?>
<!DOCTYPE eac SYSTEM "C:\Documents and Settings\gunnar\Mine Dokumenter\EAC\eac\eac.dtd">
<eac>
        <eacheader type="p" status="draft" langencoding="ISO 639-2" scriptencoding="ISO 15924" dateencoding="ISO 8601"
        encodinganalogsys="USMARC">
                <eacid syskey="Example05" countrycode="NO" ownercode="UoB">NO::UoB::Example05</eacid>
                <mainhist>
                        <mainevent maintype="create">
                                <name>Gunnar Karlsen</name>
                                <maindate calendar="gregorian" normal="20010607"> June 07,2001</maindate>
                                <maindesc>Manually created record from UCM testdata</maindesc>
                        </mainevent>
                </mainhist>
                <languagedecl>
                        <language languagecode="en" scriptcode="latn">English in Latin Script.</language>
                </languagedecl>
                <ruledecl>
                        <rule id="aacr">Anglo-American Cataloguing Rules.</rule>
                        <rule id="aacr2">Anglo-American Cataloguing Rules, Second Edition.</rule>
                </ruledecl>
                <sourcedecl>
                        <source  syskey="a10000471" id="UCM-authority">
                        <title>Universidad Complutense de Madrid Name Authority File, a10000471.</title>
                        <sourceinfo>No sources provided in import record.</sourceinfo>
                        </source>
                </sourcedecl>
        </eacheader>
        <condesc>
                <identity>
                        <pershead rule="aacr2" authorized="UCM" languagecode="spa" scriptcode="latn" ea="100">
                                <part type="surname">Pi Sunyer</part>
                                <part type="forename">Charles</part>
                        </pershead>
                        <pershead rule="aacr2" languagecode="spa" scriptcode="latn" ea="400">
                                <part type="surname">Pi Su</part>
                                <part type="forename">Carlos</part>
                        </pershead>
                        <pershead rule="aacr2" languagecode="spa" scriptcode="latn" ea="400">
                                <part type="surname">Pi i Sunyer</part>
                                <part type="forename">Charles</part>
                        </pershead>
                        <pershead rule="aacr2" languagecode="spa" scriptcode="latn" ea="400">
                                <part type="surname">Pi Sunyer</part>
                                <part type="forename">Carlos</part>
                        </pershead>
                        <nameadds>
                                <date scope="active" calendar="gregorian" normal="1888/1971" form="rangeclosed"
                                ea="100$d">1888-1971</date>
                        </nameadds>
                </identity>
```

```
        </condesc>
</eac>
```

Problems and questions regarding the encoding:

- What rules are used by UCM? Default is set to AACR2, since this is a USMARC record.

6.3 Example 3, Authority file from FDÖP

```
<?xml version="1.0"?>
<!DOCTYPE eac SYSTEM "C:\Documents and Settings\gunnar\Mine Dokumenter\EAC\eac\eac.dtd">
<eac>
    <eacheader type="p" status="draft" langencoding="ISO 639-2B" scriptencoding="ISO 15924" dateencoding="ISO 8601"
    countryencoding="ISO 3166-1 a2" detaillevel="1" encodinganalogsys="local format">
        <eacid syskey="example02" countrycode="NO" ownercode="UoB">NO::UoB::example02</eacid>
        <mainhist>
            <mainevent maintype="create">
                <name>Gunnar Karlsen</name>
                <maindate calendar="gregorian" normal="20010710">July 10,2001</maindate>
                <maindesc>Manually created record from FDÖEP testdata</maindesc>
            </mainevent>
        </mainhist>
        <languagedecl>
            <language languagecode="eng" scriptcode="latn">English in Latin Script.</language>
        </languagedecl>
        <ruledecl>
            <rule id="local"/>
        </ruledecl>
        <sourcedecl>
            <source id="fdoep" syskey="FDOEP-NrAutor: 310">
                <title type="simple">FDÖEP test-data for LEAF, FDÖEP-NrAutor: 310</title>
                <sourceinfo>
                    <date form="single" era="ce" calendar="gregorian" normal="19951231"
                    ea="Leaf.Date_of_capturing">31.12.1995</date>
                    <date form="single" era="ce" calendar="gregorian" normal="20010426"
                    ea="Leaf.Last_Correction">26.04.2001</date>
                    <persname ea="Leaf.Author_of_the_data_file">
                        <part>uh</part>
                    </persname>
                </sourceinfo>
            </source>
        </sourcedecl>
</eacheader>
<condesc>
        <identity>
            <pershead rule="local" authorized="FDOEP" languagecode="ger" scriptcode="latn" ea="Leaf.Main.Heading">
                <part type="surname">Kröner</part>
                <part type="forename">Franz</part>
            </pershead>
            <nameadds>
                <date scope="begin" form="single" era="ce" calendar="gregorian" normal="18891212"
                ea="Leaf.Date_of_birth">12.12.1889</date>
                <date scope="begin" form="single" era="ce" calendar="gregorian" normal="18891212"
                ea="Leaf.Date_of_birth_standardised">12.12.1889</date>
                <place>Schönwald b. Bautch (Nordmähren)</place>
            </nameadds>
            <nameadds>
                <date scope="end" form="single" era="ce" calendar="gregorian" normal="19580424"
                ea="Leaf.Date_of_death">24.04.1958</date>
                <date scope="end" form="single" era="ce" calendar="gregorian" normal="19580424"
                ea="Leaf.Date_of_death_standardised">24.04.1958</date>
                <place>Innsbruck</place>
            </nameadds>
        </identity>
        <eacrels>
            <eacrel ea="Leaf.Relationship" reltype="subordinate" type="akadem Lehrer">
                <persname>
                    <part>R.Reininger</part>
                </persname>
```

```
                <!-- how to know which type of relation (pers, corp, fam)? -->
            </eacrel>
        </eacrels>
        <desc>
            <persdesc>
                <sex ea="Leaf.Gender">
                    <value>m</value>
                </sex>
                <funactdesc ea="Leaf.Profession">
                    <p>Philosoph</p>
                </funactdesc>
                <legalstatus ea="Leaf.Academic_Title">
                    <value>Dozent</value>
                    <date scope="begin" form="single" era="ce" calendar="gregorian" normal="1940"
                    ea="Leaf.Date_of_first_flourishing">1940</date>
                    <date scope="end" form="single" era="ce" calendar="gregorian" normal="1945"
                    ea="Leaf.Date_of_last_flourishing">1945</date>
                    <place ea="Leaf.Place_of_flourishing">Graz,Z&#252;rich</place>
                </legalstatus> -->
            </persdesc>
        </desc>
        <resourcerels>
            <resourcerel>
                <bib ea="Leaf.Reference_books"> PA Archiv der Univ. Graz; Schuch,  Sch&#246;nafinger (=Klappentext,
                2.Aufl. "Anarchie...", dort Foto</bib>
            </resourcerel>
        </resourcerels>
</condesc>
</eac>
```

Problems and questions regarding the encoding:

- Rules for encoding practice is mandatory in EAC. This is set to ‘local’ here.

- How to decide which type of relation (corporate, person or family) between the actual record and the information pointed to in ‘Leaf.Realtionship’, here encoded using the eacrels element?

6.4 Example 4, Allegro-Hans record from ÖNB

```
<?xml version="1.0"?>
<!DOCTYPE eac SYSTEM "C:\Documents and Settings\gunnar\Mine Dokumenter\EAC\eac\eac.dtd">
<eac>
<eacheader type="p" status="draft" detaillevel="1" langencoding="ISO 639-2B" scriptencoding="ISO 15924" dateencoding="ISO 8601"
countryencoding="ISO 3166-1 a2" encodinganalogsys="Allegro-HANS">
        <eacid syskey="example01" countrycode="NO" ownercode="UoB">NO:UoB:example01</eacid>
        <mainhist>
            <mainevent maintype="create">
                <name>Gunnar Karlsen</name>
                <maindate calendar="gregorian" normal="20010711">July 11,2001</maindate>
                <maindesc>Manually created from LEAF testdata provided by &#214;NB</maindesc>
            </mainevent>
        </mainhist>
        <languagedecl>
            <language languagecode="eng" scriptcode="latn">English in Latin script.</language>
        </languagedecl>
        <ruledecl ea="005">
            <rule id="PND">Personennamendatei</rule>
            <rule id="other"/>
        </ruledecl>
                <!-- where to put local id= p2603 ea="003" sourcedecl is not finished here    -->
        <sourcedecl>
            <source syskey="118500775" id="onb" countrycode="AT" ownercode="12-7-SWB">
                <title type="simple">ÖNB authority record</title>
                <sourceinfo> <!-- is this correct for time and agent for last correction? -->
                    <date form="single" era="ce" calendar="gregorian" normal="20010313/09:30:57"
                    ea="007">20010313/09:30:57</date>
                    <persname ea="007">
                        <part>HESUL</part>
                    </persname>
```

```
                    </sourceinfo>
                </source>
        </sourcedecl>
</eacheader>
<condesc>
        <identity>
                <pershead rule="PND" ea="800" authorized="ÖNB" languagecode="ger" scriptcode="latn" type="identity">
                        <part type="surname">Adorno</part>
                        <part type="forename">Theodor W.</part>
                </pershead>
                <pershead rule="other" ea="830" scriptcode="latn" type="identity">
                        <part type="surname">Wiesengrund</part>
                        <part type="forename">Theodor</part>
                </pershead>
                <pershead rule="other" ea="830A" scriptcode="latn" type="identity">
                        <part type="surname">Wiesengrund-Adorno</part>
                        <part type="forename">Theodor</part>
                </pershead>
                <pershead rule="other" ea="830B" scriptcode="latn" type="identity">
                        <part type="surname">Wiesengrund Adorno</part>
                        <part type="forename">Theodor</part>
                </pershead>
                <pershead rule="other" ea="830C" scriptcode="latn" type="identity">
                        <part type="surname">Adorno</part>
                        <part type="forename">Theodor Wiesengrund-</part>
                </pershead>
                <pershead rule="other" ea="830D" scriptcode="latn" type="identity">
                        <part type="surname">Adôrnô</part>
                        <part type="forename">T.W.</part>
                </pershead>
                <pershead rule="other" ea="830E" scriptcode="latn" type="identity">
                        <part type="surname">Adorno</part>
                        <part type="forename">Teodor</part>
                </pershead>
                <pershead rule="other" ea="830F" scriptcode="latn" type="identity">
                        <part type="surname">W.-Adorno</part>
                        <part type="forename">Theodor</part>
                </pershead>
                <pershead rule="other" ea="830G" scriptcode="latn" type="identity">
                        <part type="surname">Adorno</part>
                        <part type="forename">Teodor V.</part>
                </pershead>
                <nameadds>
                         <date scope="begin" form="single" era="ce" calendar="gregorian" normal="19030911"
                        ea="808">11.09.1903</date>
                        <place ea="808a">Frankfurt am Main</place>
                </nameadds>
                <nameadds>
                        <date scope="end" form="single" era="ce" calendar="gregorian" normal="1969" ea="809">1958</date>
                        <place ea="809a">Visp</place>
                </nameadds>
        </identity>
        <eacrels> <!-- can you encode a relation to a seperate record for a pseudonym like this?-->
                <eacrel reltype="identity" syskey="115058060" ea="856">
                        <persname>
                                <part type="surname">Zwieback</part>
                                <part type="forename">Castor</part>
                        </persname>
                        <note>Pseud.</note>
                </eacrel>
        </eacrels>
        <desc>
                <persdesc>
                        <sex ea="806g">
                                <value>m</value>
                        </sex>
                <funactdesc ea="805">
                        <list>
                                <item>Philosoph</item>
                                <item>Soziologe</item>
                                <item>Musiktheoretiker</item>
                                <item>Komponist</item>
                                <item>Schriftsteller</item>
                        </list>
                </funactdesc>
```

```
        </persdesc>
        </desc>
</condesc>
</eac>
```

7 Concluding remarks

There is no other suitable DTD available for LEAF than the EAC working draft. With the participation of LEAF in the EAC working group the needs of the LEAF project should also be well covered and the LEAF project is considered an important test-bed for the DTD.

EAC will develop into a tested first version in time for the LEAF project.

There are some problems that are likely to occur during the conversion from some local formats:

- not all data providers maintain a distinction between the 3 types of authority records that we find as mandatory in EAC (corporate, personal or family).

- for every relation to another entity you also have to know the type of entity referred to.

Still these are only minor problems that should be solved, either by modifying the draft for the DTD before its final version or in the technical solutions in LEAF.

8 Definitions, Acronyms and Abbreviations

8.1 Partner Acronyms

Partner Name	Acronym
Staatsbibliothek zu Berlin	SBB
Joanneum Research	JRS
University of Bergen	UoB
Crossnet Systems Ltd.	CNS
Österreichische Nationalbibliothek	ÖNB
Biblioteca Nacional	BN
Biblioteca de Universidad Complutense de Madrid	UCM
Swiss National Library	SNL
National and University Library, Ljubljana, Slovenia	NUK
Institut Mémoire de l'Edition Contemporaine	IMEC
Riksarkivet	RA
Deutsches Literaturarchiv	DLA
British Library	BL
Goethe- und Schiller-Archiv	GSA
Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie	FDÖP

8.2 Other definitions and abbreviations

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

AUTHOR See http://www.bl.uk/gabriel/cobra/author.pdf

CHIO Cultural Heritage Information Online, http://www.cimi.org/old_site/projects/chio.html

CIMI Consortium for the Computer Interchange of Museums Information, http://www.cimi.org

COBRA+ Computerised Bibliographic Records Actions

COVAX Contemporary Culture and Virtual Archives in XML

DTD Document Type Definition

EAC Encoded Archival Context, http://bembo.village.virginia.edu/eac.html

EAD Encoded Archival Description, http://www.loc.gov/ead/

FRANAR Functional Requirements And Numbering of Authority Records, http://www.ala.org/alcts/alcts_news/v10n1/gateway.html

GNU Recursive acronym for "GNU's Not Unix", http://www.gnu.org/

ICA Internation Council of Archives, http://www.ica.org/

IFLA International Federation of Library Associations and Institutions

INDECS Interoperability of Data in E-Commerce Systems, http://www.indecs.org/

IP Intellectual Property

ISAAR(CPF) International Standard Archival Authority Record for Corporate Bodies, Persons and Families, http://www.ica.org/

ISADN International Standard Authority Data Number, http://www.ifla.org/VI/3/p1996-2/mlar.htm

ISO International Organization for Standardization, http://www.iso.org/

IST Information Society Technologies, http://www.cordis.lu/ist/home.html

LEAF Linking and Exploring Authority Files, http://www.leaf-eu.org

MACS Multilingual Access to Subjects, http://infolab.kub.nl/prj/macs/

MARC Machine Readable Cataloguing, http://www.loc.gov/marc/

MCN Museum Computer Network, http://www.mcn.edu/

MDA Museum Documentation Association, http://www.mda.org.uk/

ONE OPAC Network in Europe, see ONE-2

ONE-2 OPAC Network in Europe-2, http://www.one-2.org/

OPAC Online Public Access Catalog

RLG Research Libraries Group, http://www.rlg.org

SGML Standard Generalized Markup Language

TEI Text Encoding Initiative, http://www.tei-c.org/

URI Uniform Resource Identifier, http://www.w3.org/Addressing/

WP Work Package

W3C World Wide Web Consortium, http://www.w3.org/

XML Extended Markup Language, http://www.w3.org/XML/

XSLT Extensible Stylesheet Language Transformations, http://www.w3.org/Style/XSL/

8.3 References

The following list shows references that are made within this document.

[Bruvik99] T.M.Bruvik, *Technical Note, ver.0.1*, Nov.1999, prepared for the MALVINE project under W.P. 2.3.4 - 2.3.5.

[Light97] Richard Light: *CIMI DTD v.5 Design Principles*, 23 September 1997, http://www.cimi.org/documents/dtd5design.html

[Pitti99] Daniel Pitti: Encoded Archival Description, An Introduction and Overview, D-Lib Magazine, Nov.1999, http://www.dlib.org/dlib/november99/11pitti.html

[TEI99] C.M.Sperberg-McQueen and Lou Burnard: Guidelines for Eelctronic Text Encoding and Interchange, May 1999, http://www.tei-c.org/Guidelines/index.htm