2000

MARC to XML : an enhanced name authority record

Antonio M. Calvo
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Calvo, Antonio M., "MARC to XML : an enhanced name authority record" (2000). *Master's Theses*. 1975.
DOI: <https://doi.org/10.31979/etd.zhgv-wf6a>
https://scholarworks.sjsu.edu/etd_theses/1975

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

MARC to XML:
AN ENHANCED NAME AUTHORITY RECORD

A PROJECT REPORT

Presented to
The Faculty of the School of Library & Information Science
San Jose State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Library and Information Science

by
Antonio M. Calvo

May 2000

UMI Number: 1399779

Copyright 2000 by
Calvo, Antonio M.

All rights reserved.

UMI[®]

UMI Microform 1399779

Copyright 2000 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

copyright 2000
Antonio M. Calvo
ALL RIGHTS RESERVED

**APPROVED FOR THE
SCHOOL OF LIBRARY & INFORMATION SCIENCE**

Linda Main

Dr. Linda Main

Judy Weedman

Dr. Judy Weedman

Patricia Elliott Stroh

**Patricia Elliott Stroh, MA, MLIS,
Curator, The Beethoven Center**

APPROVED FOR THE UNIVERSITY

William Fish

ABSTRACT

MARC TO XML: An Enhanced Name Authority Record

by Antonio M. Calvo

This report describes a project that creates an enhanced name authority record using XML (Extensible Markup Language). The enhanced record provides a finer granularity of biographical detail than that which exists in the current MARC (Machine Readable Cataloging) name authority record. Because XML takes full advantage of the interoperability of the World Wide Web, records structured with XML may be viewed and manipulated without the previous requirement of proprietary library catalog software.

Sections include: histories of the MARC format and markup languages; the process undertaken in the creation of the enhanced record including the markup of sample records from the Ira F. Brilliant Center for Beethoven Studies at San Jose State University; and a discussion of the Nomen DTD (document type definition). Appendices include: the Nomen DTD; the Nomen Tag Library; the Nomen Worksheet; and screen shots from the project web site.

This project report is dedicated to the memory of

My grandmother
Mary Wilhelmena Hamilton
1917-1999

ACKNOWLEDGEMENTS

This project report marks the culmination of two years of study and would not have been possible without the support numerous individuals.

I had a fabulous committee and wish to thank each member. Dr. Linda Main, my project advisor, boldly took our class into XML, which provided the technical foundation for the project. Her encouragement from the inception of my idea through its fruition was invaluable. Dr. Judy Weedman, of the SJSU SLIS campus in Fullerton California, gave careful readings and offered useful comments and suggestions. Patricia Elliott Stroh, curator of the Beethoven Center, in the midst of the Beethoven 2000 Festival and moving the Center to temporary quarters across campus, somehow found the time to thoughtfully read and critique the project records, website and report. Her comprehensive knowledge of music librarianship and the world of Beethoven never ceases to amaze me.

Also at the Beethoven Center, I would like to thank Dr. Elizabeth Fleming, for her help on some of the finer points of musicological research, and Dr. William Meredith for his calm reassurances.

Very special thanks must go to my dear sister Luz Calvo and Dr. Catrióna Rueda Esquibel. I was most fortunate to have them agree to edit large portions of the manuscript. My

sister's advice, especially in academic matters, has always meant a great deal.

I would also like to thank my brother, Andrew Calvo, for planting the seed for the idea of doing a biographical record and Dr. Geoffery Liu for explaining to me what an authority record was during our summer cataloging course.

Indeed, I am grateful to the whole of the School of Library & Information Science. The faculty, staff and students -- especially my "buddies" in ALASC. I must mention my longtime friend and fellow Mac-fanatic Albert Schulze for his assistance with the emulation software used in the earliest stages of my work, and Graciela Hernandez for her diligent and timely formatting of the report. I'd also like to thank my friends in San Jose, Victor Bubbett and Lance Cleveland for being around to listen to and feed me through all this.

Finally, my utmost thanks and appreciation goes to my parents, Tony H. Calvo and Beverley Calvo. Their unrelenting support was felt throughout this endeavor -- as it has been throughout my life.

It is cold in the scriptorium, my thumb
aches. I leave this manuscript, I do not
know for whom; I no longer know what it is
about: stat rosa pristina nomine, nomina
nuda tenemus.

Umberto Eco, *The Name of the Rose*

TABLE OF CONTENTS

INTRODUCTION	1
REVIEW OF LITERATURE	5
Thirty Years of MARC	5
Authority Work and Authority Control	9
XML - Extensible Markup Language	11
MARC to XML	14
PROJECT DESIGN AND METHODOLOGY	19
Overview	19
Enhanced Name Authority Record	20
XML Document Type Definition (DTD)	20
Markup of Sample Records from the Beethoven Center ...	21
Display of Records using XSL Style Sheet	23
RESULTS	24
The Nomen DTD	24
Nomen Tag Library	30
Applying the DTD to Beethoven Center Records	30
Display of XML MARC and Enhanced Records on the Web ..	32
CONCLUSION	34
REFERENCES	37
APPENDIX A: Nomen Document Type Definition	41
APPENDIX B: Nomen Tag Library	44
APPENDIX C: Enhanced Name Authority Worksheets	52
APPENDIX D: Screen Shots of MARC and Enhanced Records	55

INTRODUCTION

If the role of the librarian is to facilitate access to recorded information, then an integral component of this process must be bibliographic control. The need for the continued development of systems that describe, organize and thus make accessible collections of information becomes increasingly apparent as the volume of human knowledge expands over time. During the short period encompassing the development of electronic digital technology, we have seen what might be called a "revolution" in bibliographic control: the widespread implementation of the MARC format--MARCII, LCMARC, USMARC, and UNIMARC, among others--as the standard for digital electronic cataloging.

MARC (Machine Readable Cataloging) has become the standard library format to express bibliographic data. Its use has stimulated the development of library automation and information networks nationally and internationally (Avram, 1975; Crawford, 1989). As a result of this shared bibliographic format, records can be easily transferred among libraries. This sharing of information has contributed to the development of cooperative cataloging, (the sharing of catalog records), which has greatly reduced duplication of effort.

The MARC format was based on bibliographic data previously found on Library of Congress cataloging cards. The initial purpose of the format was the distribution and

maintenance of the same information in machine-readable form (Avram, 1975). The development of MARC indicates a significant shift, from print form to electronic form, in the representation and distribution of bibliographic data. However, its design was profoundly influenced by cataloging practices developed for the printed card (Crawford, 1989). In today's purely digital cataloging environment, the opportunity exists to develop bibliographic data that contains a depth and richness that was not possible on a 125 x 75 millimeter index card.

Since the first use of MARC thirty years ago, technologies have emerged that allow the platform-independent transfer of data over networks. Much in the same way that MARC allowed the sharing of bibliographic data among libraries, these new technologies have allowed the sharing of digital media worldwide. The Internet in general, and the World Wide Web in particular, are evidence of this interoperability. The World Wide Web--based on open, platform independent TCP/IP and HTTP network and transfer protocols--has experienced dramatic growth since its early deployment in 1989. This is due in part to its use of the text based Hyper-text Markup Language (HTML). HTML is an application of the Standardized General Markup Language (SGML), and allows the presentational mark-up and display of text and graphics transferred over the Internet via easily obtained "browser" software.

The Extensible Markup Language (XML), a subset of SGML, appears perfectly suited to transfer MARC data to any computer with Web browsing capabilities. Like HTML, XML is designed to work on any computer with mainstream browsing software. However, unlike HTML, XML provides descriptive, logical markup of information, leaving the display to a separate style sheet. Using the interoperability of XML, the transfer of bibliographic data could be achieved without the limitations of proprietary MARC compatible software or complex server-side conversions.

In addition to streamlining the transfer process, breaking out of the MARC format with XML should allow the expansion and enhancement of bibliographic data. For example, personal name authority records could include detailed biographical information well beyond the typical name, birth and death dates, and name variants.

Throughout the transitions from book catalogs, to card catalogs, to electronic catalogs using the MARC format, the underlying principle of bibliographic control has remained the same: improved access to recorded information. This report will describe a project that contributes to the development of a small but vital piece of the cataloging process: personal and corporate name authority, through the conversion of MARC authority records to an enhanced structure developed with XML. The project will be conducted in the

Center for Beethoven Studies library at San Jose State
University.

REVIEW OF LITERATURE

Thirty Years of MARC

MARC (Machine Readable Cataloging) was developed in the 1960s by the Library of Congress. MARC was created to enable the electronic creation, storage, manipulation, transfer and display of bibliographic information based on the United States Library of Congress catalog card format. Since its creation, MARC has become the *de facto* format for electronic cataloging systems allowing the transfer of bibliographic data from cataloger to user with reasonable success.

The Library of Congress (LC) began investigations into the possibility of using automated techniques for its internal operations in the late 1950s (Avram, 1975). Subsequently, the Librarian of Congress requested a grant from the Council on Library Resources (CLR) to perform a feasibility study of applying automated techniques to the operations of LC (Avram, 1975; Crawford, 1989). The study, published in 1963, recommended that a group within LC design and implement automated procedures for cataloging, searching, indexing and document retrieval. A systems staff was assembled in the Office of the Librarian to proceed with the recommendations (Avram, 1975).

Concurrently, another study was commissioned by the CLR to consider more specifically methods and implications of converting LC catalog card data to machine-readable form for the purpose of printing bibliographical products by computer.

LC, CLR, and the Association of Research Libraries (ARL) co-sponsored a conference in January 1965 to consider both of the studies and their significance. The conferees concluded that:

1. Availability of machine-readable catalog records produced and distributed by LC would help those libraries that have automated systems.

2. The machine-readable record should include all the data presently available on LC's printed card, plus additional information to produce a multipurpose record.

3. Agreement by a broad segment of the library community on the elements to be included in the record was most desirable, and the design of the record at LC was probably the best means of achieving standardization. (Avram, 1975)

Henriette Avram, Ruth Freitag and Kay Guiles, working as the new systems staff that had been assembled in the Office of the Librarian, analyzed LC cataloging data from a machine processing point of view (Crawford, 1989). In June 1965, a report was issued for a proposed record format. After extensive comments from LC staff and the greater library community, (some of which were published as a supplement to the report), it became clear that there was sufficient enthusiasm and support for LC to become the distribution source of machine-readable records (Avram, 1975).

Accordingly, LC sought and was granted further funding from CLR to conduct a pilot project to test the feasibility and utility of the distribution of cataloging data in machine-readable form from LC to user libraries. The project was named MARC, for *machine-readable cataloging* (Avram, 1975).

The pilot project to distribute bibliographic records in machine readable form was announced by Henriette Avram in February 1966 and began in September 1966, with the first test tape being mailed to the 16 participating libraries in October 1966 (Pattie, 1998). While the project was originally slated to end in June 1967, LC announced at the midwinter meeting of the American Library Association (ALA) in January 1967 that it would be extended through June 1968 (Avram, 1975). At the June 1967 ALA/MARC meeting, LC announced that a full-scale, permanently operational MARC distribution service was in the planning stages (Avram, 1975).

The MARC pilot project provided a prototype on which to build the full-scale system, MARC II. In addition to experience gained through application and evaluation of the MARC format in the pilot project, an important event occurred which exerted a strong influence on the design of MARC II. The British National Bibliography (BNB) expressed interest in mounting a UK/MARC project of its own. Their interest, combined with visits from a number of foreign librarians, directed thinking toward a standard communications format suitable for interchanging bibliographic data not from one organization (LC) to many, but among organizations, perhaps crossing national boundaries (Avram, 1975).

The structure of MARC II has remained essentially fixed since the original draft was issued in January 1968

(Crawford, 1989). The philosophy behind MARC II was the design of one format structure capable of containing bibliographic information for all forms of material and related records (name and subject authority records). The three components of the format are the *structure*, or "empty container," the *content designators* (tags, indicators and subfield codes) and the *content*, the data itself (Avram, 1975).

The changes that have taken place to the format have been in the content designator and content components. The library community gives input through an ALA committee, established prior to 1974, called the Committee on Representation in Machine-Readable Form of Bibliographic Information (MARBI). MARBI has typically met at the ALA's midwinter meeting and annual conference. As changes are proposed to the Library of Congress, meetings are held at LC to further discuss the proposed changes. The Library of Congress publishes USMARC formats and has the final authority over them. MARC II was adopted as a national standard in 1971 by the American National Standards Institute as ANSI Z39.2 and by the International Standards Organization in 1973 as ISO 2709 (Avram, 1975; Crawford 1989). For the purposes of this project report, the use of the initials "MARC," will refer to this evolving standard as implemented by the U.S. Library of Congress and commonly referred to as, "USMARC."

Incorporated into MARC II was a specially designed extended character set for Roman alphabet languages, which included special characters and diacritics for languages other than English. This multilingual capability was further expanded to include nonroman text with the LC publication of USMARC Character Set, Chinese, Japanese, Korean in 1986. This system, jointly developed by the Research Library Group (RLG), the Library of Congress and MARBI, provided a mapping system of Chinese, Japanese, and Korean (CJK) characters to ASCII equivalents (Crawford, 1989).

For most libraries, the computer has become the only technology used to compile or search bibliographic information, primarily due to the implementation of the MARC record. Users now query a library's integrated user interface, often from home or other remote locations, to find what is located at libraries worldwide (Hagler, 1997).

Authority Work and Authority Control

Bibliographic control is dependent on accurate, and in some cases, uniform, data. Bibliographic records consist primarily of names, titles, and subject headings, in addition to various descriptive data. Of these, names and subject headings have the distinction of being managed by the authority system. The library cataloging practice that deals with organizing these names and subject headings is known as authority control or more generally, authority work.

Authority work involves the establishment of standard forms of headings to be used in the catalog, providing references to possible variants of these headings, and, in computerized systems, links to other items in the catalog where the heading occurs.

Authority work consists of the creation, use, maintenance, and evaluation of records, which are gathered together in an *authority file*. This file is then linked to a bibliographic file; together these form an *authority system*. Authority systems support the creation of accurate, or authoritative, bibliographic records (Burger, 1985).

Prior to the implementation of electronic authority systems, authority control was accomplished by printing subject-heading and name-heading records and maintaining these lists in book or card-file form. As new bibliographic records were added into the catalog, corresponding subject headings and name headings would be verified against the established forms in the authority files or, if need be, added to it. In public card catalogs, users would be directed from variant versions of terms to established terms via "see" references.

When the digital MARC format was implemented, an additional MARC format, distinct from the one used to describe bibliographic data, was created to describe authority data. Known as a MARC authority record, this

unique structure is used for both *name* authority and *subject heading* authority records (Burger, 1985).

LC published the first version of a MARC format for authority information in 1976. A comprehensive revision was completed by Josephine S. Pulsifer and Margaret Patterson culminating with the Library of Congress' MARC Development Office's, Authorities: A MARC Format in 1981. This publication was replaced in June of 1987 by USMARC Format for Authority Data Including Guidelines for Content Designation (Crawford, 1989). Authority files have traditionally been used to exercise control of the headings used in bibliographic records and often as a source of reference for user and librarian alike.

XML - Extensible Markup Language

Broadly defined, "markup" involves the use of symbols or characters used with text to indicate significance or action (Flynn, 1998). A markup language is a defined set of computer readable symbols or characters that indicate significance or action. A markup language may exist *inline*--markup inserted within the text at points of significance, or *out-of-line*--where the markup is stored separately from the text to which it applies (Flynn, 1998).

Flynn, in his book, Understanding SGML and XML Tools: Practical Programs for Handling Sturctured Text argues that the history of markup can be traced to the early days of the

book, before the printing press, when scribes and authors wrote explanations and comments in between lines or in the margins of the manuscripts, and used special colored ink that stood out from the rest of the text, as a way of indicating to their readers the importance of certain letters, words and phrases (Flynn, 1998). Also viewing markup in a broad sense, Coombs, Renear, and DeRose include punctuation as a type of markup in their article, *Markup Systems and the Future of Scholarly Text Processing*, in what they refer to as "Punctuational markup...the use of a closed set of marks to provide syntactic information about written utterances" (Coombs, Renear, and DeRose, 1987, p. 935). Instances of punctuational markup include the use of periods, question marks and exclamation marks to signal the end of sentences and shades of meaning.

Furthered by the advent of the printing press in the fifteenth century, the use of markup has evolved to include the use of italics, bold face type, underlined type, centered text, indentation and special spacings. Common use of these features has led readers to immediately understand the meaning implied for each. For example, in English language formatting, italics signify the title of a book, words in a foreign language, or emphasized terms, while bold-faced type or underlined type indicates a heading. Word processing software made this type of markup accessible to a large body of users that had previously relied on professional printers

to produce fully marked-up textual materials. HyperText Markup Language (HTML), used to build documents on the World Wide Web, has further increased the accessibility and reach of markup. However, markup as used by traditional moveable type printers, word processors, and HTML documents is purely presentational and relies on the reader to infer meaning from the formatting. For elements in a marked up document to have specific meaning that computers can understand and users manipulate, *descriptive* or *logical* markup languages such as Standardized General Markup Language (SGML) and more recently, Extensible Markup Language (XML) must be used.

SGML is a metalanguage that allows the creation of markup languages. HTML was created using SGML. Other examples of SGML application include the Encoded Archival Description (EAD) project which began as the Berkeley Finding Aid Project in the early 1990s (Pitti, 1998); and the Text Encoding Initiative (TEI) which began as an effort to facilitate the interchange of data among humanities scholars. SGML, however, is a complex technology, hard to learn and too difficult for mainstream browsers to handle (Bosak & Bray, 1999).

Created for the technology of today, XML is a streamlined metalanguage. XML allows the creation of Document Type Definitions (DTDs) which enable the logical markup of data which can then be displayed on a mainstream browser. (It should be noted, however, that at the time of

this writing, Microsoft's Internet Explorer 5.0 for Windows or Macintosh (IE5) is the only XML compliant browser available.) Another of XML's strengths is its use of a new standard called Unicode, a character-encoding system which supports the intermingling of text in all the world's languages. Unicode utilizes a multi-byte character set allowing more than one byte per character therefore increasing the range of characters possible to represent (Ridley, 1997).

The XML version 1.0 specification was completed and recognized by the World Wide Web Consortium (W3C) in early 1998. Much "hype" has surrounded its release, although to date few examples of useful implementation can be found on the web. This may be attributed to two factors: the year-long wait for the first mainstream XML compliant browser to appear, and the fact that major programs had already begun constructing DTDs using SGML, thus causing reluctance to change course and implement the same projects using XML.

MARC to XML

The Library of Congress began work on an SGML MARC DTD in December 1995. The primary purpose of the project was to create standard SGML Document Type Definitions to support the conversion of cataloging data to and from the MARC data structure without loss of data (Library of Congress Network Development Office, 1998). While the complexity of this

project is immense--especially with the two-way conversion--some useful and interesting decisions were made during the design process that are worth elaborating.

The LC SGML MARC DTDs were made general rather than being designed for any specific MARC formats. The five MARC formats were accommodated in two DTDs: one for Bibliographic Data--expanded to include community and holdings information--and another for Authority data--used for both name and subject authority records. Names of the element tags in the DTDs were derived from variable field tags and subfield codes in the MARC formats themselves. This naming convention included field-level elements, subfield-level elements, and indicator names.

This monumental effort to replicate the MARC structure in SGML shows the potential of the concept. The LC Background and Development document states that, "The driving force behind this project was the desire for a standardized non-proprietary conversion between MARC encoded data and SGML data" (Library of Congress Network Development Office, 1998). LC may have succeeded in nearly taming the unwieldy SGML MARC DTD. However, without easily accessible browsers to view it, its contribution may ultimately be useful only as a model for future development, not for widespread implementation.

The Lane Medical Library, at Stanford University Medical Center, has undertaken another, purely XML, approach called the MEDLANE project. This effort to create a flexible

retrieval and display mechanism for bibliographic, authority, and other library information using XML was presented at the 1999 Medical Library Association Conference in a talk entitled "Restructuring serial, circulation, and traditional bibliographic data for deployment in changing digital environments" (Miller et al., 1999).

The MEDLANE approach differs from LC's not only because the markup language is XML, but because it does not make an attempt to replicate every last detail of the existing MARC format. Moreover, Lane has "somewhat aggressively" enhanced their bibliographic and authority records (Miller et al., 1999). This is clearly an experimental project, and the information provided on their website suggests that revision of the DTDs will continue.

One particularly interesting feature that is planned for the MEDLANE project is an implementation of an XML technology called XLink. Using XLink, MEDLANE hopes to provide links to authors within bibliographic records so that users can access biographical information from an authority record, view a portrait linked to the authority record, retrieve works by the author, or limit works to full text-linked records only. At present, Xlink technology is not yet supported by Internet Explorer 5.0 (Miller et al., 1999).

The MEDLANE group acknowledges its study of LC's SGML MARC DTD but found that the "immensity and literal mapping of MARC really did not address the fundamental problem of

complexity" (Miller et al., 1999). While the intent of the MEDLANE experiment is evidently an attempt to demonstrate feasibility, not to establish a standard, it is clearly useful to watch their progress.

With regard to the potential significance of logical markup languages (in this case SGML as applied in the EAD project) to the practices of bibliographic control, Jackie M. Dooley of the University of California at Irvine asserts that, "the development of EAD has generated more interest in both the U.S. and international archival communities than any other technological development in the thirty years since the automation revolution began to change the way in which cultural repositories of all kinds conduct their business" (Dooley, 1998, p. 1).

Repeated queries of indexes and databases such as Dissertations Abstracts Online, Library Literature and Library and Information Science Abstracts, using the string "XML and MARC", and variations on that theme, yielded zero results. We may surmise that due to the newness of XML, there has been little chance for sufficient inquiry and documentation in traditional research publications. Searching the Internet with various search and metasearch engines did uncover one additional MARC to XML development.

K.T. Lam of Hong Kong University has constructed and posted to the Web an example of both bibliographic and authority MARC records in XML along with XSL style sheets for

display. Minimal documentation is provided. Upon preliminary examination, Lam's records appear to differ from those of the MEDLANE group in that they carefully replicate existing MARC structure. Lam refers to and gives examples of the LC SGML MARC DTD. Lam's XML MARC examples are fully functional and viewable using IE5. In addition, Lam has taken advantage of XML's Unicode functionality which enables translation to Chinese characters using the BIG5 character set (Lam, 1999).

The March 1999 issue of *American Libraries* contains additional reference to MARC and XML. David Dorman's "Technically Speaking" column discusses the conversion of MARC to XML and the tools necessary for new XML based systems. Dorman concludes that by the beginning of the millennium, the library community will be asking the same questions about the XML compatibility of integrated library systems that they were asking about MARC compatibility twenty years ago. Moreover, he predicts that the movement to XML will "accelerate the convergence of the library-automation marketplace with the larger information-management marketplace" (Dorman, 1999, p. 76).

PROJECT DESIGN AND METHODOLOGY

"I met Neuromancer. He talked about your mother. I think he's like a giant ROM construct, for recording personality, only it's full RAM. The constructs think they're there, like its real, but it just goes on forever."

William Gibson, *Neuromancer*

Overview

The objective of the project is the design of an enhanced name authority record using XML. A careful examination will be made of the existing elements in the current specifications for the MARC Name Authority Record. From this analysis, an enhanced record structure will be developed and expressed as an XML Document Type Definition (DTD). Using this DTD, a sample set of personal and corporate name authority records from the Center for Beethoven Studies will be marked up. Finally, using the Extensible Stylesheet Language (XSL), these records will be displayed via IE5 browsing software. This project will not deal with automated conversion of existing MARC records into the newly designed structure. However, efforts will be made to design the new records structure in such a way to allow such conversion, should it be desired at a later date.

Enhanced Name Authority Record

The complete and current specification for US MARC authority records is published in USMARC Concise Format for Authority Data (Library of Congress, 1997). A critical analysis will be undertaken of this format with an eye toward elements that can be expanded to include a finer detail of biographical information that may be of use to catalogers and scholars. Based on this analysis, decisions will be made as to the precise set of elements to be included in the enhanced record. In addition, the progress of the MEDLANE project at Stanford University (Miller et al., 1999) and K.T. Lam at Hong Kong University (Lam, 1999) and their decisions with regard to the enhancement and structure of authority records will be noted.

XML Document Type Definition (DTD)

An XML DTD is a coded document that defines the structure for specific type of information. In comparing the DTD with the three parts of the MARC format, it could be said that the MARC DTD would contain *the record structure* and *content designator* portions of the record.

DTDs are constructed in a hierarchical tree structure with branches of information elements emanating from the root. In addition, each element can be assigned attributes, which further define it. These attributes could be seen as equivalent to the indicators and subfields in a MARC record.

As starting points in the development of the Name Authority XML DTD record for this project, the three projects discussed above--the SGML MARC DTD (Library of Congress, 1998), the MEDLANE Project (Miller et al., 1999), and the K.T. Lam example (Lam, 1999)--will be studied as models. Once the name authority record DTD is established, actual records can be marked up using this structure.

Markup of Sample Records from the Beethoven Center

The Ira F. Brilliant Center for Beethoven Studies located at San Jose State University is a special collection of material devoted to the composer Ludwig van Beethoven. The material in the collection has been extensively cataloged using MARC records that are stored in an alternate part of the University Library's catalog: the Beethoven Bibliography Database.

The Beethoven Bibliography Database Thesaurus was created in response to the domain specific nature of the collection: the life and works of Ludwig van Beethoven (Elliott & Bakke, 1992). The authority for these specialized subject headings is maintained in a separate thesaurus database, which is periodically printed and stored in binders.

Name authority is maintained through a separate file on cards. As monographic, musical score, and analytical MARC records are added to the database, all names are checked

against this authority file. If an authority record exists, a check is made to ensure that the authorized form is used. If no card exists, online searches of the Library of Congress authority file are conducted. At times it is obvious that a correct match has been made between the name entry on the new MARC record and the LC name authority record; however there are cases when it is difficult to confirm a match due to inadequate data on the LC authority record.

Indeed, the limited information given on traditional MARC name authority--name, dates, and name variants--creates a problem in the identification of persons and groups. This project will attempt to remedy this problem through the design of a MARC authority record that allows a finer detail of biographical information.

The goal of this project is to test the concept and usefulness of an enhanced authority record. The Beethoven Center was chosen as the test site because of its tightly focused historical collection in which names play an important role in accessing information. In addition, the author has become intimately familiar with the Center's authority file having been employed to work on the Beethoven Bibliography Database for the past six months.

In order to evaluate the effectiveness of the enhanced name authority record, a set of sample records will be marked up using the enhanced structure. These records will be selected from the Beethoven Center name authority file.

Using the extensive Beethoven-related resources available in the Center, the records will be augmented to include the detail of biographical information allowed by the enhanced record structure.

Display of Records using XSL Style Sheet

The final portion of the project will involve the design of XSL Style Sheets to display the records. While the DTD describes the data, the Style Sheet functions to present the data. As is common in many online public access catalogs (OPAC), users will have the option of viewing data in full MARC view, or in a more user-friendly fashion. This will be accomplished through the use of two distinct XSL documents. Considerations derived from graphic design and page layout will be taken into account in order to obtain a clear, easy-to-read display of each of the sample authority records. These will be exhibited using mainstream browsing software that supports XML and XSL.

RESULTS

The project can be viewed at the following web address:
<http://senna.sjsu.edu/lmain/Nomen>.

The results of this project are two-fold. First is the construction, display and manipulation of the MARC data using XML. The second is the enhanced record structure, as expressed in the Nomen DTD and as demonstrated through the creation of XML records based on selected name authority records from the Beethoven Center. The enhanced records show the finer granularity of biographical information allowed by the new record structure.

The Nomen DTD

The starting point in the development of the enhanced name authority record was the construction of an XML DTD. DTDs begin with a "root" element which contains all the other elements. The root element, and hence the name of the DTD for this project is *Nomen*, the Latin word for *name*. The complete Nomen DTD is included in Appendix A.

As stated previously, both the Library of Congress' SGML MARC DTD and K.T. Lam's XML MARC DTD faithfully represented the complete MARC format, while Stanford University's Medlane project began anew with unique XML bibliographic and authority record structures.

Because the intent of the Nomen DTD was to amplify the detail of authoritative biographical information contained in

the record, it was decided to approach the design of the record structure with a clean slate. Instead of replicating each detail of the MARC format, the underlying function of the record was maintained and expanded while constructing the XML DTD. The flexibility of XML allows for the display of re-created pseudo-MARC records, as is evident on the MARC-View portion of the Web site.

The Nomen DTD is intended to fulfill the function of the USMARC Name Authority format, which is "to be a carrier for authoritative information concerning the standard forms of names [and subjects] that should be used as references to the standard forms, and the interrelationships among these forms" (Library of Congress Network Development Office, 1997). Therefore, the first sub-element, after the obligatory unique identifier element, was for the authorized form of the personal or corporate name.

The dissimilarity of the Nomen DTD to the MARC authority format becomes readily apparent in the authorized name field (equivalent to the 100 or 110 field in MARC). The MARC format specifies that the full authorized form of the name be in the initial subfield (\$a), and allows inclusion of various additional information in subfields like: birth and date years (\$d); fuller form of name (\$q); titles (\$c); a number of subfields to indicate works associated with the name; and general chronological and geographical subdivisions (\$y and \$z). In contrast, the Nomen DTD subdivides this

information differently, and deals with some data elements in other parts of the record structure. For example, the authorized form of the name is handled by four elements one of which, <Additional>, has three sub-elements of its own. Chronological dates and geographical information are dealt with in detail in a separate section of the record.

The <Source> element is included as a sub-element to <Authorized>. Its function is to provide a mechanism with which to indicate the particular reference source where the authorized form of the name was obtained. The <Source> element is used a total of four times in the Nomen DTD, each time to specify the reference source for particular informational items.

The authorized name itself is handled by the following elements: <Family>; <Given>; <Additional>, with its three sub-elements <Add1>, <Add2>, and <Add3>; and <Title> (in this case "Title" refers to a title bestowed upon someone, not the name of a creative work). Description of appropriate content for each of these elements is contained in the Nomen Tag Library (Appendix B).

The decision to divide the name into separate elements was made in an effort to allow for non-western cultures which normally present names with family names preceding given names. This flexibility was accomplished although it should be noted that the extra fields add a degree of complexity to the display of the data.

The variant names (MARC fields 400 and 410) were tagged using a suite of elements essentially identical to the authorized name elements with the addition of the attribute: *number*. The *number* attribute allowed the <Variant> element to be repeated up to five times and the data from each extracted and displayed.

The final element of the Nomen DTD, <Descriptio> and its sub-elements <Inceptio>, <Conclusio>, <Regnum> and <Affiliatio>, depart significantly from traditional MARC format. These elements do, however, contain some of the same data found in USMARC fields which, as demonstrated on the web site, can be extracted and displayed in MARC-like fashion.

The <Descriptio> element provides a detailed record structure for identifying the most important chronological dates, locations, knowledge domains and affiliations associated with the person or group that is the subject of the record. While basic descriptions of these elements are contained in the Nomen Tag Library (Appendix B), what follows is a discussion of some of their distinct features and how they relate to traditional MARC.

MARC field 100 subfield \$d allows for the years of birth and death to be linked to the authorized name. The Nomen DTD contains this information in the elements <Inceptio> and <Conclusio> which further define the chronological information to include the month, day and location of the

birth or inception, and death or conclusion, of persons or groups.

It should be noted that MARC field 045 (Time Period of Heading) allows for time period codes and/or formatted time period associated with the 1XX. While helpful, the 045 field lacks the precision of the enhanced record's ability to specify exact dates and dates of multiple events.

The <Locatio> element (a sub-element of <Inceptio> and <Conclusio>) provides sub-elements: <Republic>, <Regio>, <City>, <Street>, <StreetNumber>, and <PostalCode>, effectively offering the capability to include a complete mailing address of the place of birth and death--if that information is available or deemed important. In traditional MARC, geographical information can be included in the 052 and 043 fields using geographical codes. However, because they are not associated with an event, these codes provide only a general sense of location.

Finally, both <Inceptio> and <Conclusio> offer the <Source> element for referencing the source where the information was obtained.

The Library of Congress Network development office states that the MARC Notes fields, 66X - 68X, "contain information that is primarily intended to guide catalogers. These notes often are not written in a form adequate for public user display. Field 680 (Subject Scope Note) is an exception to this general practice" (Library of Congress

Network Development Office, 1997)). These MARC fields, while not widely used, do allow for some of the detailed biographical information that is included in the <Descriptio> element of the Nomen DTD. For example, MARC field 670 (Source Data Found) contains the citation for a consulted source in which information is found about the 1XX field. This is equivalent to the enhanced record's use of <Source> as a sub-element of <Authorized>.

MARC field 678 (Epitome) is meant to contain biographical, historical, or other information about the 1XX heading. While rarely seen in the Library of Congress Authority file, this information could be useful to catalogers and scholars in determining the identity of an individual or group. The Nomen DTD addresses this need through the application of the last two sub-elements of the <Descriptio> element, <Regnum> and <Affiliatio>.

The <Regnum> element is intended to capture work and knowledge domains that help to describe the subject of the record. For example, this repeatable element would ordinarily contain the occupation(s) and/or specialization(s) of the person or group.

The <Affiliatio> element describes the subject through its significant affiliations. This element has the attribute: *relationship*, which can have the following values: employer; corporation; patron; associate; teacher; or student. Sub-elements include versions of the <Inceptio> and

<Conclusio> elements which allow documentation of the beginning and end of the affiliation as well as the locations in which these events occurred. The <Source> element is also included for works referenced.

Nomen Tag Library

The Encoded Archival Description (EAD) Tag Library Version 1.0 states: "Because DTDs are intended to be read chiefly by computers, they are often accompanied by documentation designed for human comprehension. This tag library or master list of elements is one such piece of documentation" (Encoded Archival Description Working Group of the Society of American Archivists, & Network Development and MARC Standards Office of the Library of Congress, 1998, p. vii).

The Nomen Tag Library (Appendix B) was created to fulfill the same purpose as the EAD Tag Library. Its design shows the complete Nomen DTD in a human readable form as well as indicating the appropriate content for each element.

Applying the DTD to Beethoven Center Records

Once the Nomen DTD for the enhanced name authority record was established, detailed biographical information on selected sample authority records was collected at the Beethoven Center. This was accomplished by gathering preliminary data from MARC records from the Library of

Congress Authority file and the Beethoven Center index card authority file. This information was written on an Enhanced Name Authority Worksheet (Appendix C) developed for this purpose. The worksheet followed the structure of the Nomen DTD. Tag names are followed by blank spaces where data relevant to the various elements in the DTD may be written. In order to obtain information that was beyond the scope of traditional authority records, other resources were examined.

In most cases basic music reference works such as Baker's Biographical Dictionary of Musicians, The New Grove Dictionary of Music and Musicians, and Die Musik in Geschichte und Gegenwart were sufficient to gather the data. In one instance, a secondary source was consulted.

With the data collected on the worksheets, the final step in the creation of the XML files was to markup the data using the elements from the Nomen DTD. This was done with text editing software. The files were saved using the subject's name followed by the .xml file extension. Ultimately, these records would be part of one long file of all the authority records for a given collection. However, in an effort to simplify display and file management processes, the decision was made to save each as a separate file.

Display of XML MARC and Enhanced Records on the Web

With the authoritative biographical data marked up using the Nomen DTD tags, it became possible to display the data

using XSL stylesheets. XSL works by selecting or "matching" data contained by a specific element, formatting it, and displaying it using web-browsing software. The three files: the .dtd, the .xml and the .xsl work together in creating the final output. This differs significantly from HTML, where just one file is used. The three layers add significant complexity but also a higher level of flexibility.

For example, by using different stylesheets, identical XML files can be displayed in different ways. The project web site demonstrates this by allowing a MARC view and Enhanced view of each sample record (Appendix D). In addition, elements can be displayed more than once, (the birth and death years in the MARC view) or not at all, as was the case in the source information for each of the informational items.

Although the Nomen DTD contains four <Source> elements to cite the source where information was found, only one was displayed in each enhanced record. This was because it was difficult to format the display in a way that would not detract from the readability of the record. A related issue arose in determining the amount of bibliographical information that should be included in the <Source> element. In both of these cases it is expected that once full hyper-linking capability is reached within XML and XSL, the displayed content of an element could be hyper-linked to more detail about that particular item. For example, a title

cited as a source could be hyper-linked to its full bibliographic record. Or, a particular form of a name could be linked to the title of the document where it was found, which could in turn be linked to the bibliographic record.

The pseudo-MARC record display was accomplished by inserting the three-digit MARC tags as part of the stylesheet and then extracting and formatting the essential data elements from the XML file to recreate a MARC like display. The display of the enhanced records includes the additional biographical information contained in the XML files. Examples of the enhancements include the day and month of birth, the place of birth, the day and month of death, the place of death, the Occupation field, and the Affiliation field.

In comparing the traditional MARC name authority format to the Nomen DTD, it becomes apparent that the Nomen DTD is missing a way to relate creative works associated with the subject. In MARC format this is accomplished through the use of subfields of the 1XX field. For example, it would have been useful to allow for the inclusion of significant musical compositions in the Beethoven record. This is something that should be remedied in future versions of the Nomen DTD.

CONCLUSION

The implementation of the MARC format thirty years ago indicated a significant shift in bibliographic control. This project examined the prospect of increasing the level of detail in one form of MARC--the Name Authority record. XML was chosen as the technology with which to create a new record structure because XML enables data to be transferred over shared networks like the World Wide Web. Structuring records in this platform-independent metalanguage demonstrates the potential to expand their usefulness by increasing their accessibility. The magnification of detail in the enhanced name authority records shows their potential to offer amplified precision in the identification of individuals and groups.

The project methodology included analysis of the MARC Authority format and other current MARC/XML projects that had similar intentions. Based on these findings, an XML DTD, named the Nomen DTD was constructed. Research at the Ira F. Brilliant Center for Beethoven Studies at San Jose State University resulted in the creation of six enhanced records. These records were marked-up using the tags from the Nomen DTD and made accessible via the World Wide Web using XSL stylesheets.

To the viewer of the MARC name authorities and enhanced name authorities on the project website, the XML coding that went into the project is difficult to detect. The pages

appear as though they were coded in HTML. In fact, the data being presented comes from XML files that reference the Nomen DTD and are displayed with XSL. Some of the data elements presented, for example, the year of birth and death, are extracted and displayed repeatedly in a single MARC view record though they appear just once in the XML document. Another instance of "behind the scenes" XML functionality is the use of identical XML files for both the MARC and enhanced records.

This project validates the work of Miller et al (1999) and K.T. Lam (1999). MARC records structured with XML can be adequately displayed on any computer with sufficient Web browsing capabilities--without the use of proprietary library software. While beyond the scope of this project, it is conceivable that further development of client-side processing scripts could increase users' abilities to search and sort records. Additional utilities for the input of records and the conversion of existing records will also be necessary.

We are approaching the time when a purely XML cataloging environment will be viable. By structuring records in a platform-independent manner and increasing the level of detail in records--the ability of librarians to facilitate access to recorded information through bibliographic control is heightened. This project confirmed the feasibility of using XML for bibliographic and related data. Finally, this

project demonstrated that the enhanced name authority record structure--because of its increased detail--would provide catalogers and scholars with more and better information.

REFERENCES

Avram, H. D. (1975). MARC its history and implications. Washington: Library of Congress.

Bosak, J., & Bray, T. (1999). XML and the second-generation Web. Scientific American, 280(5), 89-93.

Burger, R. H. (1985). Authority work: The creation, use, maintenance, and evaluation of authority records and files. Littleton, CO: Libraries Unlimited.

Chesnutt, D. R. (1998). SGML and the digital libraries of tomorrow. The Journal of Academic Librarianship, 24(2), 232-236.

Coombs, J. H., Renear, A. H. & DeRose, S. J. (1987). Markup systems and the future of scholarly text processing. Communications of the ACM, 30(11), 933-947.

Crawford, W. (1989). MARC for library use. (2nd ed.). Boston: G.K. Hall.

Die Musik in Geschichte und Gegenwart: allgemeine Encyclopädie der Musik (1979). Kassel: Bärenreiter-Verlag.

Dooley, J. M. (Ed.) (1998). Encoded archival description: Context, theory, and case studies Chicago: The Society of American Archivists.

Dorman, D. (1999). Technically speaking: Midwinter exhibits: Sifting through the shifts. American Libraries, 30(3), 75-77.

Elliott, P., & Bakke, C. (1992). Special collections and cataloging standards: Issues and compromises at the Steinbeck Research Center and the Center for Beethoven Studies. In B. H. Weinberg (Ed.), Cataloging heresy: challenging the standard bibliographic product (pp. 135-147). Medford, NJ: Learned Information.

Encoded Archival Description Working Group of the Society of American Archivists, & Network Development and MARC Standards Office of the Library of Congress (1998). Encoded archival description tag library, version 1.0. Chicago: The Society of American Archivists.

Flynn, P. (1998). Understanding SGML and XML tools: Practical programs for handling structured text. Boston: Kluwer Academic Publishers.

Hagler, R. (1997). The bibliographic record and information technology. (3rd ed.) Chicago: American Library Association.

Lam, L.T. (1999, May 25). Moving from MARC to XML. [WWW document]. URL <http://home.ust.hk/~lblkt/xml/marc2xml.html>

Library of Congress Network Development and MARC Standards Office (1998, May 22). MARC DTDs document type definitions: Background and development [WWW document]. URL <http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html>

Library of Congress Network Development and MARC Standards Office (1997, August 29). USMARC concise format for authority data (1993 ed. Update No. 2 March 1997) [WWW document]. URL <http://www.loc.gov/marc/authority/ecadhome.html>

Miller, D.R., Buttner, M., Murnane, P.M., Li, Y., Feng, M., Woelfel, R., Dharma, P., & Grappone, T.C. (1999, September 30). Restructuring serial, circulation, and traditional bibliographic data for deployment in changing digital environments. [WWW document]. URL <http://krypton.stanford.edu:8080/~dmiller/>

Pattie, L.-Y. W. M. (1998). Henriette Davidson Avram, the Great Legacy. In C. Myall & R. C. Carter (Eds.), Portraits in cataloging and classification : Theorists, educators, and practitioners of the late twentieth century (pp. 67-81). New York: The Haworth Press.

Pitti, D. V. (1998). Encoded archival description: The development of an encoding standard for archival finding aids. In J. M. Dooley (Ed.), Encoded archival description: Context, theory, and case studies (pp. 7-22). Chicago: The Society of American Archivists.

Ridley, M. (1997). Beyond MARC. Paper presented at the International Conference on the Principles and Future Development of AACR, Toronto, Canada. [WWW document]. URL <http://www.nlc-bnc.ca/jsc/r-beyond.htm>

Sadie, S. (Ed.) (1980). The New Grove dictionary of music and musicians. Washington, DC: Grove's Dictionaries of Music.

Slonimsky, N. (1992). Baker's biographical dictionary of musicians (8th ed.). New York: Schirmer Books.

Sperberg-McQueen, C. M. (1998). XML and the future of digital libraries. The Journal of Academic Librarianship, 24(4), 314-317.

APPENDIX A: Nomen DTD

```

<!ELEMENT Nomen (RecordNumber, Authorized+, Variant*, Descriptio)>
<!ATTLIST Nomen type (personal|corporate) #REQUIRED >

<!ELEMENT RecordNumber (#PCDATA)>

<!ELEMENT Authorized (Family, Given, Additional, Title, Source)>
  <!ELEMENT Family (#PCDATA)>
  <!ELEMENT Given (#PCDATA)>
  <!ELEMENT Additional (Add1, Add2, Add3)>
    <!ELEMENT Add1 (#PCDATA)>
    <!ELEMENT Add2 (#PCDATA)>
    <!ELEMENT Add3 (#PCDATA)>
  <!ELEMENT Title (#PCDATA)>

  <!ELEMENT Source (Work)>
    <!ELEMENT Work (Titilus, Subtitle)>
      <!ELEMENT Titilus (#PCDATA)>
      <!ELEMENT Subtitle (#PCDATA)>

<!ELEMENT Variant (vFamily, vGiven, vAdditional, vTitle)>
<!ATTLIST Variant number (one|two|three|four|five) #REQUIRED >

<!ELEMENT vFamily (#PCDATA)>
  <!ELEMENT vGiven (#PCDATA)>
  <!ELEMENT vAdditional (vAdd1, vAdd2, vAdd3)>
    <!ELEMENT vAdd1 (#PCDATA)>
    <!ELEMENT vAdd2 (#PCDATA)>
    <!ELEMENT vAdd3 (#PCDATA)>
  <!ELEMENT vTitle (#PCDATA)>

<!ELEMENT Descriptio (Inceptio?, Conclusio?, Regnum*, Affiliatio*)>
  <!ELEMENT Inceptio (Year, Month, Day, Locatio, Source*)>
    <!ELEMENT Year (#PCDATA)>
    <!ELEMENT Month (#PCDATA)>
    <!ELEMENT Day (#PCDATA)>
    <!ELEMENT Locatio (Republic, Regio, City, Street,
StreetNumber, PostalCode)>
      <!ELEMENT Republic (#PCDATA)>
      <!ELEMENT Regio (#PCDATA)>
      <!ELEMENT City (#PCDATA)>
      <!ELEMENT Street (#PCDATA)>
      <!ELEMENT StreetNumber (#PCDATA)>
      <!ELEMENT PostalCode (#PCDATA)>

    <!ELEMENT diSource (diWork)>
      <!ELEMENT diWork (diTitilus, diSubtitle)>
        <!ELEMENT diTitilus (#PCDATA)>
        <!ELEMENT diSubtitle (#PCDATA)>

```

```

      <!--ELEMENT Conclusio (dcYear, dcMonth, dcDay, dcLocatio*,
dcSource*)>
        <!--ELEMENT dcYear (#PCDATA)>
        <!--ELEMENT dcMonth (#PCDATA)>
        <!--ELEMENT dcDay (#PCDATA)>
        <!--ELEMENT dcLocatio (dcRepublic, dcRegio, dcCity,
dcStreet, dcStreetNumber, dcPostalCode)>
          <!--ELEMENT dcRepublic (#PCDATA)>
          <!--ELEMENT dcRegio (#PCDATA)>
          <!--ELEMENT dcCity (#PCDATA)>
          <!--ELEMENT dcStreet (#PCDATA)>
          <!--ELEMENT dcStreetNumber (#PCDATA)>
          <!--ELEMENT dcPostalCode (#PCDATA)>
        <!--ELEMENT dcSource (dcWork)>
        <!--ELEMENT dcWork (dcTitilus, dcSubtitle)>
          <!--ELEMENT dcTitilus (#PCDATA)>
          <!--ELEMENT dcSubtitle (#PCDATA)>

      <!--ELEMENT Regnum (#PCDATA)>
      <!--ATTLIST Regnum number (one|two|three|four|five) #REQUIRED >

      <!--ELEMENT Affiliatio (aNomen, aiInceptio, acConclusio)>
      <!--ATTLIST Affiliatio relationship
(employer|institutio|organization|corporatio|patron|associate|teacher|st
udent) #REQUIRED>

        <!--ELEMENT aNomen (#PCDATA)>
        <!--ELEMENT aiInceptio (aiYear, aiMonth, aiDay, aiLocatio,
aiSource)>
          <!--ELEMENT aiYear (#PCDATA)>
          <!--ELEMENT aiMonth (#PCDATA)>
          <!--ELEMENT aiDay (#PCDATA)>

          <!--ELEMENT aiLocatio (aiRepublic, aiRegio, aiCity,
aiStreet, aiStreetNumber, aiPostalCode)>
            <!--ELEMENT aiRepublic (#PCDATA)>
            <!--ELEMENT aiRegio (#PCDATA)>
            <!--ELEMENT aiCity (#PCDATA)>
            <!--ELEMENT aiStreet (#PCDATA)>
            <!--ELEMENT aiStreetNumber (#PCDATA)>
            <!--ELEMENT aiPostalCode (#PCDATA)>
          <!--ELEMENT aiSource (aiWork)>
          <!--ELEMENT aiWork (aiTitilus, aiSubtitle)>
            <!--ELEMENT aiTitilus (#PCDATA)>
            <!--ELEMENT aiSubtitle (#PCDATA)>

```

```

        <!--ELEMENT    acConclusio (acYear, acMonth, acDay,
acLocatio, acSource)-->
        <!--ELEMENT    acYear      (#PCDATA)-->
        <!--ELEMENT    acMonth     (#PCDATA)-->
        <!--ELEMENT    acDay      (#PCDATA)-->

        <!--ELEMENT    acLocatio   (acRepublic, acRegio,
acCity, acStreet, acStreetNumber, acPostalCode)-->
        <!--ELEMENT    acRepublic  (#PCDATA)-->
        <!--ELEMENT    acRegio     (#PCDATA)-->
        <!--ELEMENT    acCity      (#PCDATA)-->
        <!--ELEMENT    acStreet    (#PCDATA)-->
        <!--ELEMENT    acStreetNumber  (#PCDATA)-->
        <!--ELEMENT    acPostalCode  (#PCDATA)-->
        <!--ELEMENT    acSource     (acWork)-->
        <!--ELEMENT    acWork       (acTitilus,
acSubtitle)-->
        <!--ELEMENT    acTitilus  (#PCDATA)-->
        <!--ELEMENT    acSubtitle (#PCDATA)-->

```


APPENDIX B: Nomen Tag Library

<Nomen>

Description: The root element for the enhanced name authority DTD.

Attributes: type Personal|Corporate

<RecordNumber>

Description: Contains a unique identifying number for each record.

<Authorized>

Description: This element's children contain the authorized form of the personal or corporate name. Equivalent to MARC 100 or 110 field.

<Family>

Description: Contains the family name in personal records. In corporate records this contains the full authorized group name.

<Given>

Description: Contains person or group's primary given name.

<Additional>

Description: Contains the additional name or parts of names of the person or group.

<Add1>

Description: Contains the additional name or parts of names of the person or group.

<Add2>

Description: Contains the additional name or parts of names of the person or group.

<Add3>

Description: Contains the additional name or parts of names of the person or group.

<Title>

Description: Contains any titles bestowed upon person or group. Equivalent to subfield c of MARC 100 field.

<Source>

Description: Contains reference for the source of authorized name information (This could be a link to full bibliographic record if available)

<Work>

Description: This element's children contains the full title and subtitle of referenced source for inception information.

<Titlus>

Description: Contains title of referenced work.

<Subtitle>

Description: Contains subtitle of referenced work.

<Variant>

Description: This element's children contain the variant forms of the personal or corporate identity. Equivalent to MARC 400 or 410 field.

Attributes: *number* : one|two|three|four|five

<vFamily>

Description: Contains the variations of person or groups family name in personal records. In corporate record this contains variations of the full group name.

<vGiven>

Description: Contains variants of primary given name.

<vAdditional>**<vAdd1>**

Description: Contains the additional variant name or parts of names of the person or group.

<vAdd2>

Description: Contains the additional variant name or parts of names of the person or group.

<vAdd3>

Description: Contains the additional variant name or parts of names of the person or group.

<vTitle>

Description: Contains any titles bestowed upon person or group. Equivalent to subfield c of MARC 400 field.

<Descriptio>

Description: This element's children contain detailed descriptive information about the person or group or group.

<Inceptio>

Description: This element's children contain the complete date and geographical information about the individual's birth or group's inception.

<Year>

Description: This element contains the four digit *common epoch* year number of birth or inception.

<Month>

Description: This element contains the month of birth or inception.

<Day>

Description: This element contains the calendar day of birth or inception.

<Locatio>

Description: This element's children contains the geographical location of birth or inception.

<Republic>

Description: This element contains the name of the country or republic of birth or inception.

<Regio>

Description: This element contains the name of the state, province, or region of birth or inception.

<City>

Description: This element contains the name of the city or town of birth or inception.

<Street>

Description: This element contains the name of the street, road, or boulevard of birth or inception.

<StreetNumber>

Description: This element contains the street number of the location of birth or inception.

<PostalCode>

Description: This element contains the postal code of the location of birth or inception.

<diSource>

Description: This element's children contain the citation for the source of inception information (This could be a link to full bibliographic record if available)

<diWork>

Description: This element's children contains the full title and subtitle of referenced source for inception information.

<diTitlus>

Description: Contains title of referenced work.

<diSubtitle>

Description: Contains subtitle of referenced work.

<Conclusio>

Description: This element's children contain the complete date and geographical information about the individual's death or group's conclusion.

<dcYear>

Description: This element contains the four digit *common* epoch year number of death or conclusion.

<dcMonth>

Description: This element contains the month of death or conclusion.

<dcDay>

Description: This element contains the calendar day of death or conclusion.

<dcLocatio>

Description: This element's children contains the geographical location of death or conclusion.

<dcRepublic>

Description: This element contains the name of the country or republic of death or conclusion.

<dcRegio>

Description: This element contains the name of the state, province, or region of death or conclusion.

<dcCity>

Description: This element contains the name of the city or town of death or conclusion.

<dcStreet>

Description: This element contains the name of the street, road, or boulevard of death or conclusion.

<dcStreetNumber>

Description: This element contains the street number of the location of death or conclusion.

<dcPostalCode>

Description: This element contains the postal code of the location of death or conclusion.

<dcSource>

Description: Contains reference for the source of conclusion information (This could be a link to full bibliographic record if available)

<dcWork>

Description: This element's children contains the full title and subtitle of referenced source for conclusion information.

<dcTitulus>

Description: Contains title of referenced work.

<dcSubtitle>

Description: Contains subtitle of referenced work.

<Regnum>

Attributes: *number* : one|two|three|four|five

Description: This repeatable element contains the field of specialization or occupation of the person or group.

<Affiliatio>

Attributes: *relationship* : employer | institutio | organization | corporatio | associate | patron | teacher | student

<aNomen>

Description: This element contains the complete name of the person, place or organization of which the person or group is affiliated.

<aiInceptio>

Description: This element's children contain the complete date and geographical information about the person or group's affiliation with the person or place.

<aiYear>

Description: This element contains the four digit *common* epoch year number of birth or inception.

<aiMonth>

Description: This element contains the month of birth or inception.

<aiDay>

Description: This element contains the calendar day of birth or inception.

<aiLocatio>

Description: This element's children contains the geographical location of birth or inception.

<aiRepublic>

Description: This element contains the name of the country or republic of birth or inception.

<aiRegio>

Description: This element contains the name of the state, province, or region of birth or inception.

<aiCity>

Description: This element contains the name of the city or town of birth or inception.

<aiStreet>

Description: This element contains the name of the street, road, or boulevard of birth or inception.

<aiStreetNumber>

Description: This element contains the street number of the location of birth or inception.

<aiPostalCode>

Description: This element contains the postal code of the location of birth or inception.

<aiSource>

Description: Contains reference for the source of inception information (This could be a link to full bibliographic record if available)

<aiWork>

Description: This element's children contains the full title and subtitle of referenced source for inception information.

<aiTitlus>

Description: Contains title of referenced work.

<aiSubtitle>

Description: Contains subtitle of referenced work.

<acConclusio>

Description: This element's children contain the complete date and geographical information about the affiliation's conclusion.

<acYear>

Description: This element contains the four digit *common epoch* year number of affiliation's conclusion.

<acMonth>

Description: This element contains the month of affiliations's conclusion.

<acDay>

Description: This element contains the calendar day of affiliation's conclusion.

<acLocatio>

Description: This element's children contains the geographical location of affiliation's conclusion.

<acRepublic>

Description: This element contains the four digit *common epoch* year number of affiliation's conclusion.

<acRegio>

Description: This element's children contains the geographical location of affiliation's conclusion.

<acCity>

Description: This element contains the name of the city or town of affiliation's conclusion.

<acStreet>

Description: This element contains the name of the street, road, or boulevard of affiliation's conclusion.

<acStreetNumber>

Description: This element contains the street number of the location of affiliation's conclusion.

<acPostalCode>

Description: This element contains the postal code of the location of affiliation's conclusion.

<acSource>

Description: Contains reference for the source of affiliation's conclusion information (This could be a link to full bibliographic record if available)

<acWork>

Description: This element's children contains the full title and subtitle of referenced source for affiliation's conclusion information.

<acTitulus>

Description: Contains title of referenced work.

<acSubtitle>

Description: Contains subtitle of referenced work.

APPENDIX C: Enhanced Name Authority Worksheets

NOMEN

Marc to XML: An Enhanced Name Authority Record

WORKSHEET

Record Number: _____

Personal: ☐

Corporate: ☐

AUTHORIZED

Family _____

Given _____

Additional

Add1 _____

Add2 _____

Add3 _____

Title _____

Source

Work

Title _____

Subtitle _____

VARIANT

one ☐ two ☐ three ☐ four ☐ five ☐

Family _____

Given _____

Additional

Add1 _____

Add2 _____

Add3 _____

Title _____

VARIANT

one ☐ two ☐ three ☐ four ☐ five ☐

Family _____

Given _____

Additional

Add1 _____

Add2 _____

Add3 _____

Title _____

NOMEN

Marc to XML: An Enhanced Name Authority Record

WORKSHEET

DESCRIPTIO

Inceptio

Year _____

Month _____

Day _____

Locatio

Republic _____

Regio _____

City _____

StreetNumber _____

PostalCode _____

Source

Work

Title _____

Subtitle _____

Conclusio

Year _____

Month _____

Day _____

Locatio

Republic _____

Regio _____

City _____

StreetNumber _____

PostalCode _____

Source

Work

Title _____

Subtitle _____

REGNUM

NOMEN

Marc to XML: An Enhanced Name Authority Record

WORKSHEET

AFFILIATIO

employer ☐

institutio ☐

organization ☐

student ☐

corporatio ☐

patron ☐

associate ☐

teacher ☐

DESCRIPTIO

Inceptio

Year _____

Month _____

Day _____

Locatio

Republic _____

Regio _____

City _____

StreetNumber _____

PostalCode _____

Conclusio

Year _____

Month _____

Day _____

Locatio

Republic _____

Regio _____

City _____

StreetNumber _____

PostalCode _____

Figure 3

Figure 4