Survey of Four META-DATA Candidates

You would have a difficult time finding a resource in a digital library without consulting some type of database that contains descriptions of each resource. The information used to describe a book, journal, article or other resource in such a database is called meta-data. There are many formats for encoding meta-data. This document presents one set of bibliographic data describing a single resource in four formats used in various projects related to digital libraries. The MAchine Readable Catalog (MARC) record format is used by hundreds of online catalog systems around the world to encode bibliographic data about books, journals, video and audio materials for libraries. RFC-1807 specifies a bibliographic file format that was intended primarily to describe computer science technical reports. It is used in conjunction with the Dienst handle server, to uniquely identify a document. The Envision DTD for bibliographies, developed by the Computer Science Department at Virginia Tech, describes an SGML application for tagging bibliographic information such as articles, research papers, etc. The Text Encoding Initiative (TEI) header is part of an SGML application used to tag electronic texts in the humanities. When compared to a MARC record, only TEI comes close to supporting enough elements needed for encoding the amount of data a professional cataloger can amass for an individual item. RFC-1807 and Envision support a significant subset of the types of elements required but are designed to describe articles and reports rather than documents of types as varied as books, journals, audio or video segments, software, digital images, texts in a series, etc. While they each support a note field which can contain anything a user wishes, this data is unstructured and essentially inaccessible unless the full text of the record is searchable. In the end, the MARC record is still the strongest candidate for a meta-data template for describing objects in a digital library, although it might be expanded or paired with one or more of the other formats described below to describe items at different levels.

Let us start by examining the MARC record for the Journal of Technology Education, from which examples in each of the other formats mentioned above were derived.

MARC:
MARC's primary weakness (and one shared with TEI) is that it is not really intended to describe items such as journal articles or book chapters. This has not been an issue in traditional libraries, except in the case of electronic reserve services. A digital library meta-data record must describe some resources at that level because the boundaries of physical journals and texts will increasingly become irrelevant in a digital library. Pressures such as ownership of materials reverting to authors and users requesting the ability to perform searches that combine articles, videos, books and journals into one result set will mean that items might need to be described at multiple levels.

This record was captured from VTLS. It describes a print journal that is also available as an electronic journal on the World-Wide Web. Title information is stored in 2xx fields, such as the 245 which contains the main title. Keywords are stored in multiple occurrences of the 650 field, and electronic access information such as the document URL (Uniform Resource Locator) are stored in the 856 field:

```
Local lvl: 0        Analyzed: 0 Operator: 00         Edit:
CNTL:   20001913   Rec stat: n Entrd: 900125        Used: 951110
Type: a Bib lvl: s Govt pub: s Lang: eng  Source: d S/L ent: 0
Repr:   Enc lvl: I Conf pub: 0 Ctry: vau  Ser tp: p Alphabt: a
Indx: u Mod rec:   Phys med:   Cont: ^    Frequn: f Pub st: c
Desc: a Cum ind: u Titl pag: u ISDS:    1 Regulr: x Dates: 1989-9999
010     sn89-6813
012      \l 1
022 0  1045-1064
035     0903-28960
040     NSD \c NSD \d VPI
042     nsdp
049     VPI$
090     T61 \b .J687
222 00 Journal of technology education
245 00 Journal of technology education.
246 33 JTE
260 00 Blacksburg, Va. : \b Technology Education Program. Virginia
        Polytechnic Institute and State University, \c 1989-
263     8911
265     Mark Sanders, 144 Smyth Hall, Virginia Tech, Blacksburg, VA
        24061
300     v. ; \c 23 cm.
350     $8.00 (individual) \a $15.00 (library)
362 0  Vol. 1, no. 1 (fall 1989)-
500     This journal is also available electronically. Internet
        work-stations are available in the University Libraries.
550 0  Cosponsored by International Technology Education
        Association and Council on Technology Teacher Education.
650  0 Vocational education \x Periodicals.
650  0 Industrial arts \x Education \x Periodicals.
650  0 Technology \x Study and teaching \x Periodicals.
710 20 Virginia Polytechnic Institute and State University.
710 20 Council on Technology Teacher Education (U.S.)
710 20 International Technology Education Association.
856 7  \u http://scholar.lib.vt.edu/ejournals/JTE/jte.html
856 7  \u gopher://scholar.lib.vt.edu/
949     GMc/jep 1/19/90 ; GMc 1/18/95(rev.); mwo 4/04/95 (rev) ; GMc
        5/8/95 (rev.)
```

RFC-1807:

This is how the above MARC record might look as a Dienst-compatible bibliographic record. Most of the data from the MARC record is present, although there are no obvious places to put information such as ISSN, physical size, cost or starting issue, or to specify that this item is a journal rather than a book or "report". It is a stretch to

adapt this format to other material such as books or journals. One advantage of this format is that it is compact and easy to understand.

```
BIB-VERSION:: CS-TR-v2.1
ID:: VPI.LIB//JTE
ENTRY:: January 19, 1990
ORGANIZATION:: Technology Education Program, Virginia Polytechnic
               Institute and State University
TITLE:: Journal of technology education
TYPE:: Journal
CORP-AUTHOR:: International Technology Education
CORP-AUTHOR:: Association and Council on Technology Teacher Education
CONTACT:: Mark Sanders, 144 Smyth Hall, Virginia Tech, Blacksburg, VA
          24061 <msanders@vt.edu>
COPYRIGHT:: Copyright Journal of Technology Education ISSN 1045-1064.
            Permission is given to copy any article or graphic provided
            credit is given and the copies are not intended for sale.
HANDLE:: hdl:vpi.lib/JTE
OTHER_ACCESS:: URL:<http://scholar.lib.vt.edu/ejournals/JTE/jte.html>
KEYWORDS:: Vocational education
KEYWORDS:: Industrial arts
KEYWORDS:: Technology
LANGUAGE:: English
NOTES:: Journal published quarterly
END:: VPI.LIB//JTE
```

Envision (TRBIB) DTD:
This is how the above MARC record might look as an Envision bib record. Some data that could not be encoded with RFC-1807 is supported with unique markup elements here, such as frequency, editor and ISSN. There is no way to include the absolute address (URL) of an electronic document, but the Envision software constructs a URL that points to each document using the document id and an internally encoded URL stub. Envision is well suited for describing items such as journal articles. It has entry types for describing articles in books, journals, periodicals, proceedings, as well as entries for books, booklets, broadcast programs, charts, computer software, video, letters, maps, master's theses and a pair of generic entries for items that do not fit into any of these categories.

```
<bib id="Education96jte">
<b.misc>
       <au><cnm>International Technology Education</cnm></au>
       <au><cnm>Association and Council on Technology Teacher
           Education</cnm></au>
       <title>Journal of technology education</title>
       <pub><name>Technology Education Program</name><address><onm>Virginia
           Polytechnic Institute and State University</onm></address></pub>
       <editor><fnm>Mark</fnm><snm>Sanders</snm></editor>
       <frequency>quarterly</frequency>
       <note>Accessible on the World-Wide Web at
             http://scholar.lib.vt.edu/ejournals/JTE/jte.html</note>
```

```
        <issn>1045-1064</issn><lang>English</lang>
        <crt>Copyright Journal of Technology Education ISSN 1045-1064.
            Permission is given to copy any article or graphic provided
            credit is given and the copies are not intended for sale.</crt>
        <keywords>Vocational education</keywords>
        <keywords>Industrial arts</keywords>
        <keywords>Technology</keywords>
</b.misc></bib>
```

TEI DTD Header:

This is how the above MARC record might look as a TEI header. A TEI header can be used to encode virtually all of the information in a MARC record. Unfortunately, encoding can be subject to the tagger's interpretation of the markup elements. For example, I chose to place information about the editor in the <titleStmt> element of the <fileDesc> portion of the header. I instead could have placed this information in the <sourceDesc> portion of the header. TEI headers are time-consuming to develop but like all of the formats presented here, they could be automatically generated from MARC records.

```
<teiHeader type=text>
<fileDesc>
  <titleStmt>
    <title>Journal of technology education</title>
    <author>International Technology Education</author>
    <author Association and Council on Technology
    Teacher Education</author>
    <sponsor>Technology Education Program, Virginia Polytechnic
    Institute and State University</sponsor>
    <respStmt><name>Mark Sanders</name><resp>Editor</resp>
    </respStmt>
  </titleStmt>

  <extent>approximately 200 to 500K an issue</extent>

  <publicationStmt>
    <publisher>Technology Education Program, Virginia Polytechnic
    Institute and State University</publisher>
    <pubPlace>Blacksburg, Virginia</pubPlace>
    <authority>Scholarly Communications Project</authority>
    <idno type=issn>1045-1064</idno>
    <availability status=free>http://scholar.lib.vt.edu/ejournals/JTE.html
    </availability>
  </publicationStmt>

  <sourceDesc>Original files supplied in Microsoft Word format</sourceDesc>
</fileDesc>
<encodingDesc>
 <classDecl>
    <taxonomy id='LCSH'>
      <bibl>Library of Congress Subject Headings</></>
    <taxonomy id='LC'>
```

```
            <bibl>Library of Congress Classification</></>
    </classDecl>
</encodingDesc>

 <profileDesc>
  <langUsage>
     <language id=EN wsd='english usage=100>English</language>
  </langUsage>
  <textClass>
     <keywords scheme='LCSH'>
       <list>
          <item>Vocational education</item>
          <item>Industrial arts -- education</item>
          <item>Technology -- Study and teaching</item>
       </list>
     </keywords>
     <classCode scheme='LC'>JC 177</>
  </textClass>
</profileDesc>
</teiHeader>
```

James Powell
Library Automation
University Libraries, Virginia Tech
April 9, 1996