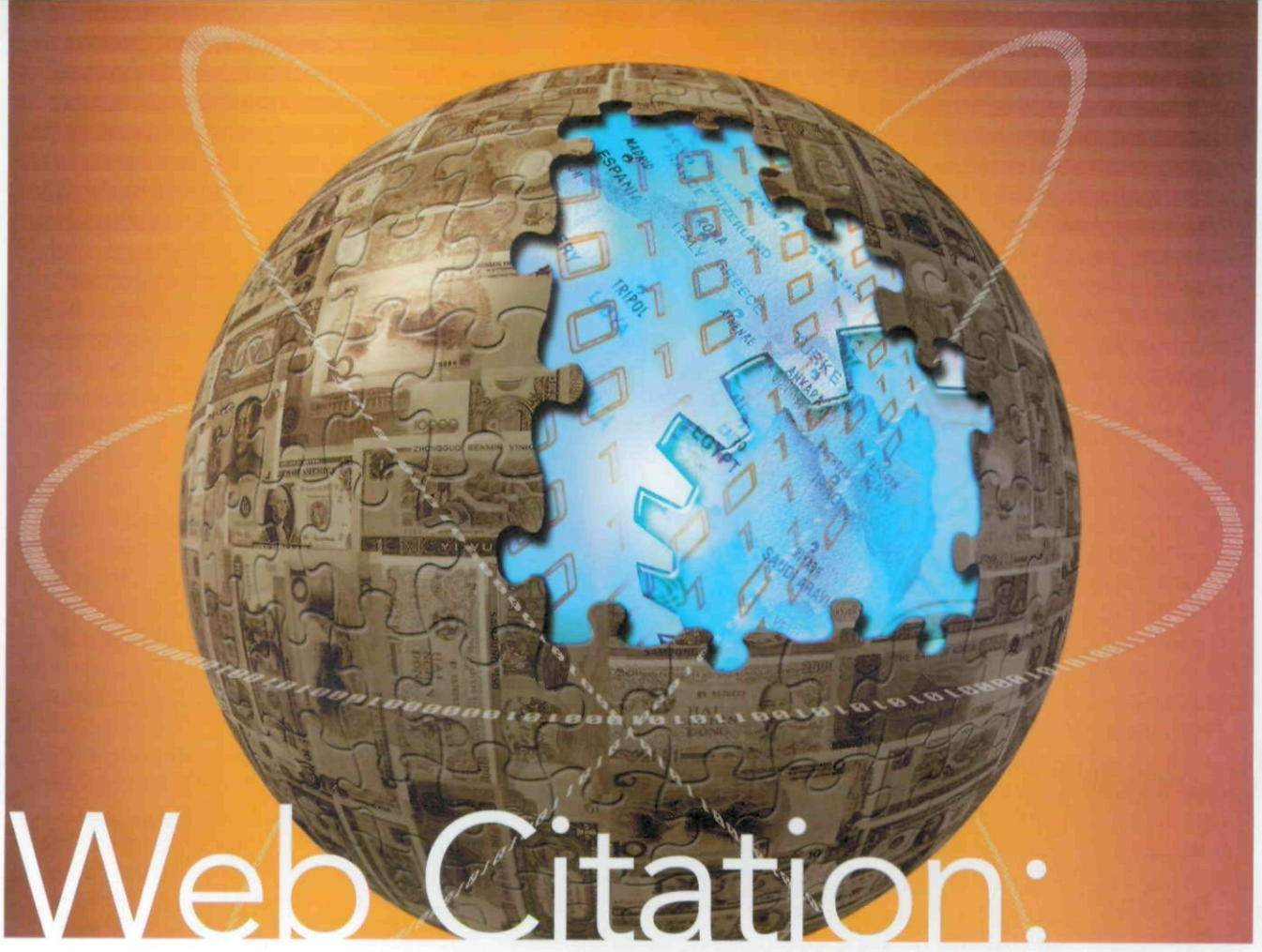

Web Citation:

A Proposal for Standardized Specification

by SHEILA BECK AND RICHARD BECK

This paper proposes a new standard in HTML/XHTML for creating accurate citations. This is an original proposal by the authors, who are requesting comments for review.

HyperText Markup Language, or HTML, is the lingua franca in which Web pages are expressed.

As every librarian knows, it is difficult for students and scholars alike to cite Web pages accurately in their footnotes and bibliographies. It is not for lack of trying. The multitude of citation styles—AP, MLA, Chicago, and Turabian—and the unique details of each make the task next to impossible. Plus, the citation guides were originally written for printed materials consulted during the research process, not for the electronic sources used by today's students and scholars. Accurate references remain crucial to researchers. As Janice Walker points out in her article "Citing Online: Online Serial Publications and Citations Systems" (*The Serials Librarian*, v. 33, no. 3/4, 1998: pp. 343–356), "Many of the existing formats for scholarly citations fail to take into account the unique structure of electronic publications, or fail to adequately address some of the specific issues prompted by online publications."

OTHER APPROACHES

Other people are working on ways to create automatic citations. Peter Jörgensen ("Citations in Hypermedia:

XHTML, a relatively new standard, **clears up contradictions** and confusions created by HTML's loose syntactic standard. **XML**, with its origins in document markup, is more syntactically formal than HTML and has become **the de facto standard** for distributing information between systems.

Implementation Issues," *Information Technology and Libraries*, v. 24, no. 4, December 2005: pp. 186–191) discusses implementation issues with extracting embedded bibliographic information such as metadata tags. Groups such as NISO (National Information Standards Organization) [www.niso.org/standards/resources/MI-Citation_Elements_v1.xls], IMS Global Learning Consortium [www.imsglobal.org/rli/rliv1p0imsrli_infov1p0.html], and the California Digital Library [www.cdlib.org:8081/inside/diglib/guidelines/mdreqs.htm#mdguidelines] are among those making proposals.

HyperText Markup Language, or HTML, is the lingua franca in which Web pages are expressed. HTML, a standard formalized by the World Wide Web Consortium, details how content is to be displayed. The specific browser, such as Internet Explorer or Mozilla Firefox, decodes the HTML expressions and displays the page according to its instructions.

HTML version 4.1 provides the current specification. This version includes mechanisms for style sheets, scripting, embedding objects, improved support for right to left and mixed direction text, and enhancements to forms for improved accessibility for people with disabilities [www.w3.org/TR/xhtml1]. It provides no support to assist in the area of citation.

XHTML, a relatively new standard, clears up contradictions and confusions created by HTML's loose syntactic standard. XML, with its origins in document markup, is more syntactically formal than HTML and has become the de facto standard for distributing information between systems. A well-formed XHTML document eliminates the ambiguity that can lead to format confusion. Modern Web browsers all support XHTML and XML in addition to HTML. Billed as the successor of HTML, the current XHTML effort makes no attempt to address issues of citation [www.w3.org/MarkUp].

Web pages are divided into two main parts—the head and the body. The body contains information to be displayed, while the head contains information about the page, such as the title. A later addition to the head portion is the <META> tag, which was designed to provide search terms to be encoded with the page. As search engines locate and analyze the page, the <META> tag information helps them properly index the page. The <META> tag can, for example, contain synonyms for terms on the page so that a searcher can find the page by terms that are not actually in the body text.

INTRODUCING BIBLIO

We propose that a new capability be introduced to the <HEAD> portion of an HTML page. The new, optional <BIBLIO> tag would provide bibliographic information to assist researchers in referencing the specific page and its contents.

The <BIBLIO> tag would treat the page as a container of one or more items of information, each expressed with enough information to render a full and proper bibliographic entry for the page and for the information on the page. Browsers could be programmed to display the bibliography in full and correct form. Word processors and other such programs could intercept the page, extract the tag, and thus fill in their own bibliographic information.

The remainder of this paper discusses the <BIBLIO> tag in a simplified form. The tag meets the criteria for standard HTML, as well as the criteria for XHTML pages.

The schema (formal specification) for the tag is discussed below. Examples are rendered in XPATH notation [www.w3.org/TR/xpath].

TAG LEVELS

There are two levels to the <BIBLIO> tag. The first describes the page itself as a container of information; the second describes information contained in the page. The attributes of the container would be the title of the page, the creator, the title of the information on the page, the date of generation, and the URL. Some of this information is (optionally) already available on the page, but it would be formalized in the <BIBLIO> section. The attributes of what is contained include author, date published, title, where published, publisher, and version. Since the elements of a citation are the same regardless of style, these components could be used to create accurate citations. The following examples parallel the discussion in the *MLA Handbook for Writers of Research Papers*, *Publication Manual of the American Psychological Association*, and *The Chicago Manual of Style*.

Consider the case of multiple items contained on the page. Clearly, the scholar would need to identify the specific item of interest. In cases such as the example on page 34 of an article on the CNN home page, the software can

offer a list to the researcher of each story for selection into the bibliography.

In some cases, information to be included in the bibliographic entries would not be in the <BIBLIO> section itself. For example, the date of retrieval would clearly be known to the browser, but could not be encoded onto the page. This is shown in the examples in brackets as the predefined replacement tag [currentdate]. The square brackets indicate the browser is expected to do a direct replacement with the identified information. Similarly, the <BIBLIO> section provides a holder for the URL of the formal page identification; this would not necessarily be the URL through which the page was accessed. The URL tag would be helpful in situations such as a page that had been located though redirection or cached in a search engine.

The <BIBLIO> tag proposal is in development. It will be formally submitted to the W3C for consideration as a proposed standard. See Figure 1 on page 34 to view the current schema layout.

More formally, here are some examples of citations using the specific stylebook configurations and their <BIBLIO> schema counterparts.

1. Internet article based on print source

MLA

"Milius, Susan. "Meat-Eating Caterpillar: It Hunts Snails and Ties Them Down." Science News Online 23 July 2005. 29 July 2005 www.sciencenews.org/articles/20050723/fob1.asp.

<biblio/contains/author>."<biblio/contains/title>." <u><biblio/contain></u> <biblio/contains/date>[currentdate]<biblio/origin>.

APA

Milius, S. (2005). Man eating caterpillar: It hunts snails and ties them down. *Science News, 68,* 51.Retrieved July 29, 2005, from <www.sciencenews.org/articles/20050723/fob1.asp>

<biblio/contains/author>.<biblio/contain/date>.<biblio/contains/title>.<biblio/contains> <biblio/contains/version>.Retrieved [currentdate], from <biblio/origin>

Chicago (Bibiliography)

Milius, Susan. "Man Eating Caterpillar: It Hunts Snails and Ties Them Down." *Science News* July 23,2005. www.sciencenews.org/articles/20050723/fob1.asp (accessed July 29, 2005).

<biblio/contains/author>."<biblio/contains/title>."<biblio/contains><biblio/contains/date>.<biblio/origin>(accessed [currentdate].)

2. Article in an Internet-only journal

MLA

Nelson, Geoffrey, Anne Westhues, and Jennifer Macleod. "A Meta-Analysis of Longitudinal Research on Preschool Prevention Programs for Children." Prevention & Treatment 6. Article31 (2003). 25 July 2005 < http://journals.apa.org/prevention/volume6/pre0060031a.html>.

<biblio/contains/author>."<biblio/contains/title>".<u> <biblio/contains></u><biblio/contains/version><biblio/contains/date>[currentdate]<biblio/origin>

APA

Nelson, G, Westhues, A. & Macleod, J. (2003, December 18). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment,6,* Article 31. Retrieved July 25, 2005, from http://journals.apa.org/prevention/volume6/pre0060031a.html

<biblio/contains/author>.<biblio/contains/date>.<biblio/contains/title>.<biblio/contains>,<biblio/contains/version>.Retrieved [currentdate], from <biblio/origin>

Chicago (Bibliography)

Nelson, Geoffrey, Anne Westhues, and Jennifer Macleod. "A Meta-Analysis of Longitudinal Research on Preschool Prevention Programs for Children." *Prevention & Treatment* 6, Article 31 (December 18, 2003), http://journals.apa.org/prevention/volume6/pre0060031a.html (accessed July 25, 2005).

<biblio/contains/author>."<biblio/contains/title>." <biblio/contains><biblio/contains/version> (<biblio/contains/date>),<biblio/contains/origin> (accessed [current date]).

3. Document on Web site

MLA

"Third Arrest Over July 21 Bombs."CNN.com. 2005. Cable News Network.29 July 2005 < www.cnn.com>.

"<biblio/contains/title>".<u> <biblio/contains ></u>. <biblio/date>. <biblio/sitetitle >.[currentdate]<biblio/origin>.

APA

Third arrest over July 21 bombs. Retrieved July 29, 2005, from <http//www.cnn.com>

<biblio/contains/title>.Retrieved [currentdate],from <biblio/origin>

Chicago (Bibliography)

"Third Arrest Over July 21 Bombs." *CNN.com,* July 29, 2005.www.cnn.com (accessed July 29, 2005).

"<biblio/contains/title>." <biblio/contains>,<biblio/date>. <biblio/origin>(accessed [current date]).

Sheila Beck *[sbeck@qcc.cuny.edu] is a librarian at Queensborough Community College, City University of New York.* ***Richard Beck*** *[dickbeck@optonline.net] is a member of the technical team at Information Builders, Inc., N.Y.*

REFERENCES

Gibaldi, Joseph. *MLA Handbook for Writers of Research Papers.* New York: Modern Language Association, 2003.

Publication Manual of the American Psychological Association. Washington, D.C.: American Psychological Association, 2001.

The Chicago Manual of Style. Chicago. Chicago: University of Chicago Press, 2003.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="www.w3.org/2001/XMLSchema"
elementFormDefault="qualified" attributeFormDefault="unqualified">
 <xs:element name="biblio">
  <xs:annotation>
   <xs:documentation>Standard bibliographic information
</xs:documentation>
  </xs:annotation>
  <xs:complexType>
   <xs:sequence>
    <xs:element name="sitetitle">
     <xs:annotation>
      <xs:documentation>This is the title of the page itself
</xs:documentation>
     </xs:annotation>
    </xs:element>
    <xs:element name="creator" maxOccurs="unbounded">
     <xs:annotation>
      <xs:documentation>Creator of the page itself
</xs:documentation>
     </xs:annotation>
     <xs:complexType>
      <xs:attribute name="type" use="optional">
       <xs:simpleType>
        <xs:restriction base="xs:string">
         <xs:enumeration value="compiler"/>
         <xs:enumeration value="editor"/>
         <xs:enumeration value="creator"/>
        </xs:restriction>
       </xs:simpleType>
      </xs:attribute>
     </xs:complexType>
    </xs:element>
    <xs:element name="contains" minOccurs="0"
maxOccurs="unbounded">
     <xs:annotation>
      <xs:documentation>Title of information on the page.
</xs:documentation>
     </xs:annotation>
     <xs:complexType>
      <xs:sequence>
       <xs:element name="author"/>
       <xs:element name="date">
        <xs:annotation>
         <xs:documentation>Date of the object in std format.
</xs:documentation>
        </xs:annotation>
        <xs:complexType>
         <xs:attribute name="format">
          <xs:simpleType>
           <xs:restriction base="xs:string">
            <xs:enumeration value="yyyymmdd"/>
            <xs:enumeration value=""/>
           </xs:restriction>
          </xs:simpleType>
         </xs:attribute>
        </xs:complexType>
       </xs:element>
       <xs:element name="version">
        <xs:annotation>
         <xs:documentation>Info. regarding the item itself.
</xs:documentation>
        </xs:annotation>
        <xs:complexType>
         <xs:attribute name="volume" use="optional">
          <xs:simpleType>
           <xs:restriction base="xs:string">
            <xs:enumeration value="pagination"/>
            <xs:enumeration value="volume"/>
           </xs:restriction>
          </xs:simpleType>
         </xs:attribute>
         <xs:attribute name="page" type="xs:string"
use="optional"/>
         <xs:attribute name="issue" type="xs:string"
use="optional"/>
        </xs:complexType>
       </xs:element>
       <xs:element name="publisher"/>
       <xs:element name="title">
        <xs:complexType>
         <xs:attribute name="type" use="optional">
          <xs:simpleType>
           <xs:restriction base="xs:string">
            <xs:enumeration value="journal"/>
            <xs:enumeration value="magazine"/>
            <xs:enumeration value="poem"/>
            <xs:enumeration value="anthology"/>
           </xs:restriction>
          </xs:simpleType>
         </xs:attribute>
        </xs:complexType>
       </xs:element>
      </xs:sequence>
      <xs:attribute name="type" use="optional">
       <xs:simpleType>
        <xs:restriction base="xs:string">
         <xs:enumeration value="website">
          <xs:annotation>
           <xs:documentation>Actual title of the web page
</xs:documentation>
          </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="poem">
          <xs:annotation>
           <xs:documentation>Titile of information on the page.
For example, an online journal carrying a poem.</xs:documentation>
          </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="photo"/>
         <xs:enumeration value="article"/>
         <xs:enumeration value="story"/>
         <xs:enumeration value=""/>
        </xs:restriction>
       </xs:simpleType>
      </xs:attribute>
     </xs:complexType>
    </xs:element>
    <xs:element name="date">
     <xs:annotation>
      <xs:documentation>Date of generation of the page itself
</xs:documentation>
     </xs:annotation>
    </xs:element>
    <xs:element name="origin" minOccurs="0">
     <xs:annotation>
      <xs:documentation>Formal URL of this page.  Not the access
URL</xs:documentation>
     </xs:annotation>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

Figure 1. *Schema layout*