

Panning for Gold

Utility of the World Wide Web for Metadata and Authority Control in Special Collections

Nadine P. Ellero

This article describes the use of the World Wide Web as a valuable name authority resource and tool for special collections analytic-level cataloging and the specific goal of “fully discovering” the names of people who lived in the past as well as those from the present. Current tools and initiatives such as the Name Authority Component of the Program for Cooperative Cataloging (NACO) and the Library of Congress Name Authority File have a specific mission and are partially helpful. Web resources encompassing special collections are often intricate and require global and enhanced resources to continue what have been the guiding principles, tradition, and value of cataloging: to discover works via many points of entry; to find works by or about the same person, topic, or title; and to continue the great cataloging legacies of standards and cooperation.

From August 2000 through January 2002, the Historical Collections and Services Department of The Claude Moore Health Sciences Library (TCMHSL), along with the guidance of the Head of Intellectual Access at TCMHSL, used the Web as a tool for name authority work with the digitization of The Philip S. Hench Walter Reed Yellow Fever Collection. Specifically, authority data for names of persons were created and supplemented with information obtained via resources found on the Web.

Russell and Spillane (2001) presented a synthesis of the function of authority control in general cataloging practice and the utility of the Web in obtaining information on authors. They focused on the benefits of the NACO initiative and presented a case for using the Web for contact information for authors. This Web information would supplement the authority record in the 670 field. Author and company Web sites and catalogs of the world's national libraries are two examples of sources. In addition, they discussed the evaluation of this contact information mined from the Web. Web sites created by the author or institution are considered as authoritative as corresponding print reference sources, whereas Web pages or sites created by “fans” need to “be treated with a healthy degree of skepticism” (78).

While Russell and Spillane concentrated on contemporary authors, this article will show how the use of the Web, in a very focused and specialized way, can be extremely valuable for special collections cataloging and metadata creation for fully discovering the names of people who lived in the past as well as those from the present.

While the NACO initiative and the Library of Congress Name Authority File (commonly known as the NAF) are unsurpassed, they were designed to include

Nadine P. Ellero (npe6f@virginia.edu) is Head of Intellectual Access, The Claude Moore Health Sciences Library, University of Virginia Health System in Charlottesville.

Manuscript received April 1, 2002; accepted May 17, 2002.

names based on items of “literary warrant” as they are added to library collections and therefore serve a very specific purpose. Our project, while involving both famous and little known people derived from analyzing a manuscript collection, fell outside the mission of the NAF. As more and more institutions (i.e., libraries, archives, and museums) in the United States and around the world process special collections of unpublished materials on an analytic level and make these resources available on the Web, an enhanced and global system for authority records will become essential. It is interesting to note that in almost ten years the NAF has increased an astonishing 667%. In 1992, the NAF contained 500,000 name authority records (*Library of Congress Information Bulletin* 1992); as of April 13, 2002, the total was 3,835,384 (Sturtevant 2002). Additionally, in 1994 the Library of Congress Name Authority File became the Anglo-American Authority File (*Library of Congress Information Bulletin* 1992) and in 2001 19.43% of the new Name Authority Records (NARs) were contributed by international sites.¹ We were surprised to find that even with this amazing expansion of the NAF, many names of prominent persons were not found, including United States Congressional Gold Medal honoree Aristides Agramonte, Surgeon General of the United States Army Raymond W. Bliss, President of the National Academy of Medicine in Columbia Roberto Franco, and U.S. Secretary of War James W. Good. An enhancement to the NAF for facilitating faster personal name identification would be to include a qualifier field indicating the person’s profession (e.g., actor, author, historian, lawyer, physician, etc.). While there is a 678 biographical or historical data field in the MARC option of data fields for authority records, it was often not present from our experience. In a sampling of 264 NAF records, 5 contained a 678 field, 141 did not contain a 678 field or other field to indicate profession, and 108 contained a 670 field that indicated profession. The balance of NAF records that contained fields indicating profession were 8 with 510 fields and 2 with 400 fields. In addition, projects such as this one could report to NACO on data obtained for death dates of persons. Some examples from this project include: Edgar Erskine Hume, Foster Kennedy, and William Dosite Postell.² The Web implies an international scope; therefore international and enhanced resources will be required to continue what have been the guiding principles, tradition, and value of cataloging: to discover “works” via many points of entry; to find works by or about the same person, topic, or title (such as all the versions of the Bible); and to continue the great cataloging legacies of standards and cooperation.

Background

The Philip S. Hench Walter Reed Yellow Fever Collection, held by Historical Collections and Services of TCMHSL, is

an archive of largely primary source documents focusing on the discovery of the transmission of yellow fever. Walter Reed and his assistants James Carroll, Aristides Agramonte, and Jesse Lazear proved that the *Aedes aegypti* mosquito was the transmission vehicle for the yellow fever virus. Philip S. Hench’s passion for this subject led him to collect books, articles, correspondence, photographs, and artifacts from the Yellow Fever Commission of 1900. The archive also includes military artifacts and photographs of American troops during the Spanish-American War and the First American Occupation of Cuba (1898–1902), and is therefore international in scope. (For more information, see the Web exhibit *This Most Dreadful Pest of Humanity: Yellow Fever and the Reed Commission, 1898–1901* at www.med.virginia.edu/hs-library/historical/yelfev/tabcon.html and the Philip S. Hench Walter Reed Yellow Fever Collection Web site at: <http://yellowfever.lib.virginia.edu/>.)

TCMHSL’s curator, Joan Echtenkamp Klein, realizing the value and importance of digitizing and preserving this collection, assembled several library staff to complete an application for the Institute of Museum and Library Services (IMLS) National Leadership Grant in December 1998. On September 24, 1999, the library received notice of an award of \$250,041. The goal of this project was to digitize a large selection of the primary resources (i.e., the written correspondence, photographs, artifacts, and maps) both as images and searchable text to create a Web site for displaying, searching, and learning about the Yellow Fever Commission and discovery of the transmission of yellow fever. The grant writing and work on the project itself was a successful experience of cross-departmental cooperation requiring expertise from all areas of library operations, especially cataloging for the metadata aspects. The alignment of technical services and special collections in this creation of digital objects and Web sites will undoubtedly be a continuing trend (Crosby 2000; Bradshaw and Wagner 2000). The work at TCMHSL on building the Yellow Fever Web site is a testimonial to the richness of cataloger and curator collaboration and communication.

Primary Documents and “Metadata”

The project team along with David Seaman, director of the Electronic Text Center at the University of Virginia, decided to encode the text using the Text Encoding Initiative (TEI) Guidelines, which are a subset of the Extensible Mark-up Language (XML). Robust flexibility was desired for searching as well as the ability to create pre-determined searches by subject, personal name, etc. XML was also a promising archival choice and one that would allow for future applications as yet unknown. Seaman designed a template for the insertion of tagged text and the

project team designed a data grid to capture the essential descriptive metadata for senders (i.e., a sender of a letter), recipients (i.e., a receiver of a letter), broad subjects, and significant geographic names and places.

The popular definition of metadata as “data about data” is pervasive and poor. Hearn (1999, 7) described metadata structures as “the development of metadata [that] can be thought of as a loosely consolidated effort to create a standard structure for differing communities to use for the description and retrieval of records.” Glogoff and Forger (2001, 9) used metadata “as the indexing that is applied to electronic information.” In other words, there is the structure (i.e., XML language or MARC) and there are the standards in the form of content (i.e., the Library of Congress Name Authority File and the prescribed form of entry and standard conventions set forth in the Anglo-American Cataloguing Rules (AACR2) for the creation of entries. The goal of metadata, as stated by Milstead and Feldman (1999, 25), is to “improve matching by standardizing the structure and content of indexing or cataloging information.” Vellucci makes several important observations that we aimed for in our metadata approach and design:

. . . the successful use of multiple metadata schemes in the library environment will depend on authority control . . . (2000, 33)

Metadata are data that describe the attributes of a resource; characterize its relationships; support its discovery, management, and effective use; and exist in an electronic environment. (34)

For purposes of this article, the phrase “descriptive metadata” will be used. Descriptive metadata include types of data that, by their nature, are of an intellectual content and include such elements as author (in the Yellow Fever Collection, mostly known as senders of letters), recipient (the receiver of a letter), personal name as subject, topical subjects, geographic locations, etc. These descriptive metadata (as the entry in a 100 or 600 MARC field, for example) require consistency and completeness. The consistency aspect serves to control input of the data in a tag or field and better guarantees resource discovery. Completeness of name-entry—that is, having both a surname and forename whenever possible—also facilitates resource and information discovery and retrieval. In the language of catalogers, this is authority control. In addition, anyone familiar with working with primary source documents knows that names of people and places are often incomplete, have variations in spelling, and frequently contain nicknames. These unique materials, especially the letters, require the creation of titles for XML headers, and dates often have to be supplied or

approximated. One of the most challenging aspects of these materials is the high degree of personal name variation, which demands authority control for insuring accurate transcription and effective discovery of the descriptive or intellectual content contained in these primary resources.

At the time of this project, a full authority system—that is, one allowing for automatic mapping of variant names to a complete or controlled name—was not developed for this Web site. A collection level MARC record was added to the library’s online catalog, but not the 5,562 individual items that comprise the Web site collection. Likewise, the authority controlled personal names from the Web site were not added to the library’s online catalog authority file. The personal names list (or the Who’s Who on the Web site) contains *see* references and lists alternate forms of names following the controlled name and functions as an interim measure for collocating name variations as well as providing brief biographical data. A future enhancement would include an active authority system applied to the Web site as is available in many automated library catalogs. DiLauro et al. (2001) describe an automated name authority control system for the creation of controlled metadata name entries, employing an indexing scheme able to locate and “learn” specific patterns and facilitate collocation of related names and/or concepts. This automated name authority system still requires an added controlled name element or field. Other possibilities that utilize an authority number may allow faster computer indexing and minimize the size of records that in current models and practice require adding controlled name entries in addition to names as they appear in text.

Authority Control and the Web as Authority Source Information

The head of Intellectual Access (a member of the project team), who functioned as the metadata consultant, created a personal names list. For completeness and consistency, the personal names list was used by production staff to enter the correct (i.e., controlled or authoritative form) and complete forms of personal names. Later, the personal names list was renamed “Who’s Who” (<http://yellowfever.lib.virginia.edu/reed/whoswho.html>). A list of geographic names was also created and became the “Places” component (<http://yellowfever.lib.virginia.edu/reed/places.html>). Both the Who’s Who and Places lists reside on the “Collection” side of the Philip S. Hench Walter Reed Yellow Fever Collection Web site.

Personal names appeared in many forms: first and last name; last name only; first name only; nicknames, etc. Often these names were of significant personages, such as presidents of countries, surgeon generals, or key players in

the yellow fever story. The example of George Sternberg illustrates the common problem with the collection material. Sternberg was surgeon general of the United States Army from May 30, 1893, to June 8, 1902, and a member of the Yellow Fever Commission. In the primary documents, his name can be found in the following forms:

Dr. Sternberg
G. M. Sternberg
General Sternberg
Geo. Sternberg
Sternberg
Uncle George Miller

The descriptive metadata (located in the header of the XML record) used the inverted form of “Sternberg, George Miller” and the header and title sections of the XML record used the direct form of “George Miller Sternberg.” Sternberg is an example of an easy or regular case, and his name was quickly and easily found in the NAF. Other challenging examples are described in the next section.

The Process

Name-authority procedures implemented in common cataloging workflow are often done with the work in hand. For this project, work-in-hand procedures were not an option. The required speed of production needed to meet the grant deadlines, combined with the fact that the digitized documents in the beginning were not loaded on a server and easily viewed or consulted, prevented work-in-hand processing for name authority. Accuracy and complete identification of people and places remained important to this project and efforts were made to maintain goals of effective information discovery and retrieval. In the early stages of processing, names were established without having full authority control applied and a different form of personal name was often chosen by the metadata analysts. Again, due to grant deadlines, the usual routine of consulting the NAF with the work in hand was not possible. In later processing, when a personal name authority was discovered for a locally created name (e.g., Carlos Finlay), it was noted with the phrase “Partial LC Authority [10/3/2000] Full LC is: Finlay, Carlos Juan, 1833–1915” and in that way flagged for a later processing cleanup routine. Whenever a name was added to the personal names list, the creation date was entered in brackets and corresponded to the date marked on worksheets that the metadata analysts gave to the head of Intellectual Access. The date served as a reference for backtracking to the original worksheets. A decision was made to not include birth and death dates with personal name entries.

As a result of the constraints discussed above, a hybrid workflow was developed. The workflow is referred to as

hybrid because it was an adaptation to facilitate meeting the grant requirements and to allow the employment of several part-time staff working simultaneously. The workflow was hybrid and not an ideal. An ideal workflow would have included the metadata analysts searching the NAF as documents were examined and data grids completed, with problem names set aside for later in-depth investigation. The hybrid workflow that was instituted required the compilation of lists of names by the metadata analysts. These lists of names or worksheets included bits of information that would enable the discovery and determination of proper or complete forms of personal names, names that were later worked on by the head of Intellectual Access as she created the authoritative personal names list and handled problem names. The salient bits of information included years, countries, activities, and so on—essentially anything that would assist in identifying a particular person. In addition, “Input Guidelines” for creating entries for incomplete names were devised. These guidelines were placed on an Intranet page available to all project staff (see figure 1). The personal names list was a simple html tagged file with a refresh command, which enabled the list to be built as project staff were completing the data grids and ensured that project staff would see the most current additions and corrections to the list (see figure 2).

The example of Hagedorn illustrates the hybrid process. On one analyst’s list appeared:

Hagedorn, [s.n.] author of a book on Leonard Wood

(Note: [s.n.] was a convention used when the first name was unknown.)

When “Hagedorn” was first searched in the NAF (via the Online Computer Library Center, Inc. [OCLC]) with the first name unknown, the search produced 93 entries. The examination of 93 entries was quickly deemed an ineffective use of time. Following a hunch, the head of Intellectual Access decided to use the Web and the Google search engine to find Hagedorn’s first name. The simple search phrase of “Hagedorn Leonard Wood” was used and almost immediately the full name of “Hermann Hagedorn” was discovered. These Web searches were often not complicated, as many names were unique and, when coordinated with the bits of information supplied by the metadata analysts, resulted in short hit lists. Since this collection was processed in a health sciences library and not a general academic library, there were few immediately accessible print world biographies, etc. The NAF was partially helpful, but not very useful when only the last name, such as “Hagedorn” or a common name such as “Henderson,” was supplied. Due to the nature of the collection material, many governmental Web sites were fre-

Transcribe only top level names in the personal name metadata field. Top level names include all persons associated with the yellow fever experiments, all members of the Reed family, and close relatives and friends. Many names appear in the letters that are nicknames or terms of endearment that will not be transcribed for subject access of personal names. **When names appear incomplete*, the following conventions will apply:**

- **Lawrence, Mrs.** (when only the title is known and not the first name).
- **Fort Thomas [Arizona Territory?]** (when the name of fort is given and you think you might know the location), see also agreed upon geographic and place terms.
- **[Beech...?], [s.n.]** (when transcribing what you think the last name is, and no first name is given)
- **Methren, [s.n.]** (when the first name was not given and we can not surmise the first name)
- **Rawley Springs [s.l.]** (when you know the name but not the location)
- **[s.n.], Jacob** (when the surname was not given and we can not surmise the surname)
- **[s.n.], Ellen** (same as above and we were given the information "Sister Ellen", see also agreed upon personal name terms)

* *The above conventions were adopted from the Anglo American Cataloging Rules, 2d edition, 1998 revision, section 22.20 "Undifferentiated Names" (418–19)*

Do not transcribe names for which only the words "brother/brother-in-law" or "sister/sister-in-law," etc. appear and the specific person can not be identified.

Figure 1. Input Guidelines

quently used and bookmarked. The Web became a gold mine of information (see tables 1 and 2). The instability of URLs, however, continues to be problematic. During this project and the final months of writing this article, the information sources consulted in table 1 continued to manifest URL changes. For purposes of documentation in this project, paper copies were made of all Web pages that supported the discovery and establishment of persons' names.

Seaman encouraged the work and development of the personal names list for purposes not only of standardization in the entry of descriptive metadata but also as a self-help resource for visitors of the Yellow Fever Web site and as a reference aid for e-mail inquiries. With this encouragement, the decision was made to enhance the list with as much information about each person as possible. A minor but interesting person on the Yellow Fever Web site was Mabel L. Conat, to whom Hench wrote in 1940 at the Public Library of Detroit. Conat's signature was not clear on her letter and while her name was typed on the responding letter from Hench, it was often the case that the spelling of a name from a respondent letter was misspelled. In this case, it turned out that the spelling was correct and a Web search provided verification. Conat was not in the NAF, but was mentioned in the ACRL's Guide to Policies and Procedures (www.ala.org/acrl/policy/polyindx.html). Her name was verified as Mabel L. Conat in Chapter 15. It was interesting to discover that she was ACRL president in 1942–43.

The Yellow Fever Web site features many Hispanic names, as the site involved relations with Cuba. One worksheet highlighted the name Estanislao Pardo Figueroa along with the notation of "President of the Academy of Medicine, Lima, Peru." Due to the hybrid workflow procedures, much of the personal name work was done in isola-

tion from the primary documents and name entry and identification was given scrupulous attention. As expected, Figueroa was not in the NAF. A string search of "Estanislao Pardo Figueroa" on the Web brought up the document "Imágenes Históricas de la Medicina Peruana" (http://200.48.26.79/bibvirtual/libros/Medicina/Ima_Histo_Med_Per/cap_27.htm). Chapter 27 noted that Estanislao Pardo Figueroa was president of the Academy of Medicine from 1919 to 1921.

The Web also proved to be helpful in the case of Francisco Dominguez Roldan. The metadata analysts struggled with this name, as it appeared in every possible form. A Web search led to the page entitled HISA (www.fiocruz.br/coc/hisa/itiC.htm), a sublink of the Web site of Casa de Oswaldo Cruz, a foundation that provides research and special projects and is part of the Brazilian medical institution known as Fundação Oswaldo Cruz. The HISA page contained the entry "Centenario del Nacimiento del Dr. Francisco Dominguez Roldan: 1864–1942," which led to a short biographical entry (www.fiocruz.br/coc/hisa/CTIC.HTM#1696). This Web search provided both name verification and the additional facts of Dominguez Roldan's birth and death dates.

Names that did not bear identification clues on the worksheets, such as Harold W. Jones, became prime candidates for employing the Web after NAF searches failed to produce sought-after names. Interestingly, Jones appeared on the National Library of Medicine's Web site at www.nlm.nih.gov/exhibition/tour/portraits3.html.

In our final analysis of May 2002, the personal names list (Who's Who on the Yellow Fever Web site) contained 1,692 unique name entries (not including the *see* references for other forms of names) and 265 places. Of the 1,692 names, 272 (16%) were verified and constructed using the NAF, leaving 1,421 (84%) names established by

```

<HTML>
<HEAD>
<TITLE>
PERSONAL NAMES MASTER LIST:
"WHO'S WHO" ON THE PHILIP S. HENCH WALTER REED
YELLOW FEVER COLLECTION WEBSITE
</TITLE>
</HEAD>
<BODY>
<meta http-equiv="Refresh" content="300;url=http://knowledgeweb.lib.virginia.edu/docs/cmas/intellect/wrnamaster.html">
<H1> MASTER LIST OF PERSONAL NAMES: "WHO'S WHO" ON THE PHILIP S. HENCH WALTER REED YELLOW FEVER COLLECTION
WEBSITE
</H1>
&nbsp;

<P><B>List Created and Maintained by Nadine P. Ellero *** LAST UPDATED 11/19/2001 AT 11:42 AM *** </B>

<P><B>Each category of information is delimited by a pipe symbol (|) according to the following format:</B>

<P><B>Last Name, First Name Middle Name | Additional information; Other names; etc. | Authority Source </B>

<P><B>NOTE: NAMES THAT HAVE AN ASTERISK (*) BEFORE THEM AND/OR ARE BOLDED CONTAIN INSTRUCTIONS TO BE
REPLACED WITH A FULLER FORM OF THE NAME. AS AN INTERVENING STAGE,AND FOR THE PUBLIC, THESE NAMES WILL
BECOME SEE REFERENCES TO THE PROPER FORM.</B>
<P>&nbsp;

<P><B>KEY</B>
<P><B>(date) = Reference Date(s) to place the person and their position/function in time</B>
<P><B>Local Source = Form of name created from original document(s)</B>
<P><B>LC = Library of Congress </B>
<P><B> [date] = Date the name was added to the list </B>
&nbsp;

<P>Abbot, William Richardson |Opened and taught a Charlottesville Institute attended by Reed; Was born 1839 -; not sure if the same as Prof. Abbott |
Local Source [8/21/2001]

<P>Abbott, Gen.|General Abbott (some military person) | Local Source [9/12/2000]

<P>Abbott, Prof. | Professor at the University of Virginia | Local Source [10/23/2000]

<P>Abbott, [s.n.] | See document 00916034; not sure if this is any of the other listed Abbotts | Local Source [9/11/2001]

<P>Abel, Dr. | Doctor in Louisville (1939) | Local Source [10/13/2000]

<P>Abercrombie, John William | (1866-1940) Alabama State Senator 1896-1898; University of Alabama President 1902-1911; United States Congressman
1913-1917; Superintendant of Education 1920-1926 | Local Source [4/10/2001]

<P><B>*Abercrombie, [s.n.] | REPLACE WITH ABERCROMBIE, JOHN WILLIAM |

<P>Acheson, Dean | Secretary of State, United States (January 21, 1949-January 20, 1953); see: www.spartacus.schoolnet.co.uk/USAacheson.htm | Partial
LC Authority [8/21/2001] Full LC is: Acheson, Dean, 1893-1971.

<P>Acker, Mollie Flint |South Bend, Indiana (1924); markup has Mollie "Heink" Acker which is not right |Local Source [11/16/2001]

<P>Acton, H. W. |Author; Physician? |Local Source [11/14/2001]

<P>Adams, E.S. | Major General, Adjutant General, The Adjutant General's Office, War Department, United States; | Local Source [8/29/2001]

```

Figure 2. Sample of the Personal Names List

<P>Adamson, Estelle |Student, Tulsa, Oklahoma (1927) | Local Source [10/22/2001]

<P>Adamson, William C. |Congressman from Georgia, United States; served from March 4, 1897, until December 18, 1917; see: <http://bioguide.congress.gov/scripts/biodisplay.pl?index=A000051> |Partial LC Authority [11/6/2001] Full LC is: Adamson, William Charles, 1854-1929.

<P>Agostini, Dr. | Physician, Cuban sanitation, 1908 | [3/20/2001]

<P>Agramonte, Aristides |A doctor; born June 3, 1866 and died August 17, 1931; congressional gold medal recipient | Local Source [10/2/2000]

<P>Agramonte, Eduardo | Father of Aristides Agramonte, died 1872? | [3/20/2001]

<P>Agramonte, Frances F. | Wife of Aristides Agramonte | Local Source [4/26/2001]

<P>*Agramonte, Mrs. |REPLACE WITH AGRAMONTE, FRANCES F. Wife of Aristides Agramonte | Local Source [10/17/2000]

<P>Ahrendts, J. L. |Representative, John Wyeth & Brother Inc., Philadelphia, Pennsylvania (1942) |Local Source [9/3/2001]

<P>Ainsworth, Frederick Crayton | United States Major General |Partial LC Authority [10/23/2000] Full LC is: Ainsworth, Frederick Crayton, 1852-1934.

<P>*Ainsworth, [s.n.] | MIGHT BE THE SAME AS FREDERICK CRAYTON AINSWORTH |1905 U.S. Army Medical Officer | 3/20/2001

<P>Alberti, Steward | Worked with Reed at Fort Apache | Local Source [8/20/2001]

<P>Albertini, A. Diaz | Cuban Physician; Director of Finlay Institute in Havana; full name is Antonio Diaz Albertini, however he preferred signing his name as "A. Diaz Albertini [8/17/2001] | Local Source [4/26/2001]

<P>*Albertini, Antonio Diaz | REPLACE WITH ALBERTINI, A. DIAZ |

<P>*Albertini, Dr. | REPLACE WITH ALBERTINI, A. DIAZ | Local Source [10/2/2000]

<P>Alderman, Edwin Anderson | Former University of Virginia President |Partial LC Authority [10/5/2000] Full LC is: Alderman, Edwin Anderson, 1861-1931.

<P>Alexander, Martha | Worked for the Journal of the History Medicine and Allied Sciences (1951) | Local Source [9/6/2001]

<P>*Allery, P. | WRONGLY TRANSCRIBED. SHOULD BE TILLERY, P. A. |

<P>*Allery, P. | WRONGLY TRANSCRIBED. SHOULD BE TILLERY, P. A. |

<P>Allison, William B. | United States Representative to Congress 1863-1871 and Senator 1873-1908 from Iowa |Partial LC Authority [10/23/2000] Full LC is: Allison, William B. (William Boyd), 1829-1908.

<P>Alspaugh, Edna |Student, Tulsa, Oklahoma (1927) | Local Source [10/22/2001]

<P>*Alvare, Dr. | REPLACE WITH ALVARE, IGNACIO |

<P>Alvare, Ignacio | Cuban physician, Havana; helped Hench with research on Camp Lazear, 1940 | Local Source [4/27/2001]

<P>Alvarez, Jacinto Mendez | Volunteer in the Yellow Fever Experiments, 1900, 1901 | [3/20/2001]

<P>Alvarez, Joaquin Maria | Cuban Physician; identified site of cemetery | Local Source [5/10/2001]

Figure 2 (cont'd). Sample of the Personal Names List

project staff. Of the 1,692 names, 107 were found to have authoritative biographies located on the Web and the corresponding URLs were added to the Who's Who list. In May of 2002, we rechecked the names not found originally

in the NAF and discovered that two (William Abbot and Domingos Freire) had been added to the NAF. The two names represent a percentage of less than 1% of the unfound personal names added in the intervening time

Table 1. Web Sources Frequently Used for People

Web Source: People	Web Address	Comments
Biographies of the U.S. Chiefs of the Army Corps of Engineers	www.hq.usace.army.mil/history/coe.htm	The title of this Web site as reviewed July 25, 2002, is "Commanders: Portraits and Profiles."
Famous West Point Graduates	www.dmi.usma.edu/Milresources/Generals/famgrads.htm	The title of "Famous West Point Graduates" changed to "Notable USMA Graduates," as viewed July 25, 2002.
Political Graveyard	http://politicalgraveyard.com/index.html	
Principle Officers of the Department of State	www.state.gov/www/about_state/history/officers.html	
Surgeon Generals of the Public Health Service	www.surgeongeneral.gov/library/history/sylist.htm	
Surgeon Generals of the United States Army	www.armymedicine.army.mil/history/tsgs/default.htm	
United States Congressmen	http://bioguide.congress.gov/biosearch/biosearch.asp	
United States State Senators	www.senate.gov/search/index.html	
United States Department of Interior's Secretary of the Interior	www.doi.gov/anniversary/secretaries.html	The United States Department of Interior's Web page had the following notice: "Access to the DOI website has been restricted in compliance with a court order. Select DOI Web pages will be made available to the public through a private internet service provider." (As viewed February 20, 2002.)
United States Party Leaders in Congress 1789–1997: Vital Statistics	www.house.gov/rules/97-136.htm	
United States Presidents	www.americanpresidents.org/ www.whitehouse.gov/history/presidents/index.html http://lcweb2.loc.gov/ammem/ndlpedu/orientation/preslist.html	
United States Secretaries of State	www.state.gov/www/about_state/history/sectravels2.html#tenure	
United States Secretaries of the Treasury	www.treas.gov/Architext/AT-allquery.html	
[United States] Secretaries of War and Secretaries of the Army: Portraits and Biographical Sketches by William Gardner Bell	www.army.mil/cmh-pg/books/sw-sa/SWSA-Fm.htm	
Yellow Fever Experimentations Congressional Gold Medal Awardees	http://dallaslibrary.org/CGI/goldmedals/yellowfever.html http://clerkweb.house.gov/histHigh/Congressional_History/goldMedal.php	

Table 2. Web Sources Frequently Used for Events

Web Source: Events	Web Address	Comments
<i>The Army Medical Department 1865–1917</i> by Mary C. Gillett	www.armymedicine.army.mil/history/booksdocs/spanam/gillett3/	
Public Health in Cuba 1865–1917	www.armymedicine.army.mil/history/booksdocs/spanam/gillett3/ch9.htm	This is Chapter 9 of a larger work: <i>The Army Medical Department 1865–1917</i> by Mary C. Gillett.
Records of the Military Government of Cuba	www.archives.gov/research_rom/federal_records_guide/military_government_of_cuba_rg14.html	Contains significant collection of Major General Ludlow's papers.

(i.e., time between the name was created for the Web site and the time the name appeared in the NAF) and in these instances, less than one month and eight months, respectively. Refinement work continues on the Web site and the Who's Who list is expected to grow over the next year.

Impact/Conclusion

The Web has the potential for tremendous impact on the ability of archivists, catalogers, and historians to flesh out the description of collections, bibliographic records, and the stories behind and surrounding primary materials. The

Web saved time and allowed us to work more effectively and meet project deadlines. We locally “cooperated” along departmental lines at TCMHSL to share expertise and created a Web resource that satisfied preservation goals, digitized primary materials, and enriched access to the yellow fever materials with the Who’s Who name list and biographical links. This local and internal cooperation became vital when grant deadlines necessitated a quick processing speed and the relinquishment of an ideal authority control procedure. Many personal names were mined from the Web and used for authority establishment, enhancement, and control, due to the size and continuing growth of the Web and associated metadata. In particular, government and educational sites are getting better and becoming more complete with information, easier to search and browse, and more prevalent. Schreiner and Somers (2002) have also recognized the Web as a gold mine of sources and compiled a listing of biographical Web resources.

Tillett (2001, 169–70) has proposed the construction and application of an International Name Authority File System as part of IFLA’s (International Federation of Library Associations and Institutions) effort of Universal Bibliographic Control (see: www.ifla.org/VI/3/ubcim.htm):

We could shift our attention from a single authorized form that everyone in the world had to accept and could instead share parallel or complementary records through the Internet—moving more into what I’ve called for years “access control” . . . Rather than exchanging authority records with the overhead of locally maintaining such a file, we would instead create a virtual database on the Internet that allowed simultaneous searching of multiple national authority files.

Tillett (2001, 3) describes a situation where not only is there an international authority file, but also the ability for searchers to customize their view of names or for library catalogs or search engines to customize a view for a certain audience(s):

We want to have the authorized form preferred by a library as the default offered to most users, but we can also envision offering user-selected preferences, through client software, or “cookies” that let the user specify once what their preferred language, script, or cultural preference is—for example for spelling preferences when cultures have variations, like American English and spelling preferences in the United Kingdom: labor and labour.

This kind of customization can be extremely useful for Web sites that involve international collaboration or cross international boundaries contentwise, such as the Yellow Fever Web site. A by-product of the Yellow Fever Web site was the Who’s Who names list that could become a source, albeit small, of names to aid in cooperatively adding to an international personal name depository. As OCLC has led the way for sharing bibliographic records, another system, whether it be one gigantic personal names depository (Barrueco Cruz et al. 2002) or a Z39.50 link is needed to facilitate order and increase searching precision and recall on the Web depositories (Tillett 2000). Tillett (2001, 5) states:

Authority control will help users of the Web to benefit from collocation and search precision that authority control enables and do it in ways that are meaningful to users in their preferred language and script.

It has been recognized that before such an international name authority system can be facilitated, professionals need to discuss and address the following questions as outlined by Françoise Bourdon (2001, 8–9):

[I]n May 2001, FRANAR decided to move its work toward the definition of functional requirements for authority records. It seems more pertinent first to know more about nature and functions of authority data we want to manage before defining an international numbering system supposed to identify them . . . When the concerned entities, the elements of data which constitute authority records, and the real or possible users of these records have been defined, how should we organize them? What are the characteristics of each entity, of each element of data, of each user? How are these components linked together to finally create an information system?

Work involving a standard authority number such as the International Standard Authority Number (ISAN), with author numbers as the authority or controlled form, has been proposed and is in early stages of design and implementation (Snyman and van Rensburg 2000).

The Philip S. Hench Walter Reed Yellow Fever Collection Web site is but one Web site that has used the NAF and the Web to flesh out its story and provide complete names of persons. It would have been easier and better if there were an international and enhanced name authority file, files, or system from which the many Hispanic names, such as Francisco Dominguez Roldan, could have been searched and verified. Would that an auto-

matic name authority control system existed for creation and searching of Web site materials. Work needs to be done to create new tools and depositories for the future. The Web as a communication tool, data exchange tool, and information warehouse or gateway offers us (i.e., catalogers, descriptive metadata creators, et al.) the opportunity to fly forward into the digital age. Vellucci (2001, 42) strongly urges catalogers to “expand their concepts of authority control, for although the underlying goals will remain the same, the authority control process will change. . . . And most critically, catalogers must actively participate in the development of system architectures and data registries.” Setting standards, working cooperatively, describing documents, and enhancing searching are the key characteristics in the cataloger’s hall of fame. OCLC, Cooperative Online Resource Catalog (COCR), RLIN, and NACO stand in this hall of fame as well as all the local initiatives not written about but seen in library catalogs, Web pages, and Web sites that have been enhanced with descriptive metadata. We need to seize these opportunities to create new systems and sources of information as well as continue to share new ways of working that get our work done faster and more efficiently (as Russell and Spillane [2001] concluded) and cooperatively create and contribute to products (such as an international Web name authority system) that can be consulted and shared by all. While panning for personal names on the Web was a highly useful and fun application for creating controlled name entries, the process can be made easier and more effective with an international and enhanced Web name authority system.

Looking even further into the future, Berners-Lee, Hendler, and Lassila (2001, 41-42) describe how the application of XML and RDF (Resource Description Framework) can encode meaning and relationships between concepts to enable the following scenario:

Suppose you wish to find the Ms. Cook that you met at a trade conference last year. You don’t remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is “Cook” (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that’s on your list of clients and follow links to Web pages of their children to track down if any are in school at the right place.

Here is another type of “name-hunt,” not too unlike the hunting experienced in establishing full names of persons on the Yellow Fever Web site and the dreams that exist in creating a global name depository or international

name authority file system. Information retrieval and precision (i.e., resource discovery) can be enhanced by harnessing the power of an international name authority system. We can employ techniques such as XML and RDF to the infrastructure of a new and improved Internet. A fulfilled dream would be to employ an automatic name authority system, a system that does not mandate a particular “right form” of name, yet recognizes related forms. Behind the scenes there may be an ISAN, ISADN, etc., pulling these variants together powered by some yet to be determined software and hardware. Cataloging principles and especially those of authority control will light and lead the way for a future built on standards, cooperation, and more effective resource description and discovery.

Notes

1. See www.loc.gov/catdir/pcc/intco/graphs01.html. Accessed Aug. 5, 2002.
2. Edgar Erskine Hume, death date of 1952. Accessed Aug. 5, 2002, www.arlingtoncemetery.com/eehume.htm; Foster Kennedy, death date of 1952. Accessed Aug. 5, 2002, www.whonamedit.com/doctor.cfm/1181.html; William Dosit Postell, death date of 1982. Accessed Aug. 5, 2002, www.tulane.edu/~matus/postell.html.

Works Cited

- Barrueco Cruz, José Manuel, et al. 2000. Personal data in a large digital library. Paper presented at Fourth European Conference on Research and Advanced Technology for Digital Libraries, September 18–20, 2000. Accessed March 11, 2002, <http://openlib.org/home/krichel/phoenix.html>.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific American* 284, no. 5: 35–43.
- Bourdon, Françoise. 2001. Functional requirements and numbering of authority record (FRANAR): To what extent authority control can be supported by technical means? Paper presented at sixty-seventh IFLA Council and General Conference, August 16–25, 2001. Accessed March 12, 2002, www.ifla.org/IV/ifla67/papers/096-152ae.pdf.
- Bradshaw, Elaine Beckley, and Stephen C. Wagner. 2000. A common ground: Communication and alliance between cataloger and curator for improved access to rare books and special collections. *College and Research Libraries* 61, no. 6: 525–34.
- Crosby, Ellen. 2000. Technical services in twenty-first century special collections. *Cataloging and Classification Quarterly* 30, no. 2/3: 167–76.
- DiLauro, Tim, et al. 2001. Automated name authority control and enhanced searching in the Levy Collection. *D-Lib Magazine* 7, no. 4: 1–8. Accessed January 11, 2002, www.dlib.org/dlib/april01/dilauro/04dilauro.html.
- Glogoff, Stuart J., and Garry J. Forger. 2001. Metadata protocols and standards: Bringing order to our digital objects. *Internet Reference Services Quarterly* 5, no. 4: 5–14.

- Hearn, Stephen S. 1999. Metadata structures and authority control. *Technicalities* 19, no. 6: 7–9.
- LC achieves cataloging milestones. *Library of Congress Information Bulletin* 51 (June 29, 1992): 290.
- LC name authority file expanded: Becomes Anglo-American authority file. *Library of Congress Information Bulletin* 53 (July 11, 1994): 293.
- Library of Congress. 2002. E-mail to the author.
- Milstead, Jessica, and Susan Feldman. 1999. Metadata: Cataloging by any other name. *Online* 23, no. 1: 25–31.
- Russell, Beth M., and Jodi Lynn Spillane. 2001. Using the Web for name authority work. *Library Resources and Technical Services* 45, no. 2: 73–79.
- Schreiner, Susan A., and Michael A. Somers. 2002. Biography resources: Finding information on the famous, infamous, and obscure. *College and Research Libraries News* 63, no. 1: 32–35, 39.
- Snyman, M. M. M., and M. Jansen van Rensburg. 2000. NACO versus ISAN: Prospects for name authority control. *The Electronic Library* 18, no. 1: 63–68.
- . 1999. Reengineering name authority control. *The Electronic Library* 17, no. 5: 313–22.
- Sturtevant, Carolyn (NACO Coordinator, Library of Congress). 2002. Further comments from Sturtevant, May 29.
- Tillett, Barbara B. 2000. Authority control at the international level. *Library Resources and Technical Services* 44, no. 3: 168–72.
- . January 2001. “Authority control on the Web.” Paper presented at bicentennial Conference on Bibliographic Control for the New Millennium, Final Version, January 23, 2001. Accessed March 11, 2002, http://lcweb.loc.gov/catdir/bibcontrol/tillett_paper.html.
- . August 2001. “A virtual international authority file.” Paper presented at sixty-seventh IFLA Council and General Conference, August 16–25, 2001. Accessed March 11, 2002, www.ifla.org/IV/ifla67/papers/094-152ae.pdf.
- Topper, Joby, and Historical Collections and Services. The Claude Moore Health Sciences Library. University of Virginia. 1997. “This most dreadful pest of humanity”: Yellow fever and the Reed Commission, 1898–1901. Accessed July 18, 2001, www.med.virginia.edu/hslibrary/historical/yelfev/tabcon.htm.
- Velluci, Sherry L. 2000. Metadata and authority control. *Library Resources and Technical Services* 44 (1): 33–43.