The "Deceased" Preaches His Own Eulogy: Training Students to Provide Access Points on Discovery-Level Records

Felicia Piscitelli, Associate Professor, Special Collections Cataloger and Italian Resources Librarian, Texas A&M University Libraries

Lisa Furubotten, Associate Professor, Serials Cataloging. Texas A&M University Libraries

Anton duPlessis, Clinical Assistant Professor, Curator of the Mexican Colonial Collection; Director, Primeros Libros de las Américas, Texas A&M University Libraries

Alma Beatriz Rivera-Aguilera, Coordinadora de Automatización, Biblioteca Francisco Xavier Clavigero, Universidad Iberoamericana, Mexico

Ángel Villalba-Roldán, Jefe del Departamento do Catalogación, UNAM-IIB-Hemeroteca Nacional de México

Abstract

This paper is an outgrowth of Discovering a New World: Cataloging Old and Rare Imprints from Colonial and Early Independent Mexico, a project funded by the Council on Library and Information Resources (CLIR). The grant developed a methodology by which Hispanophone student workers, with minimal training, can create discovery-level records for collections by inputting bibliographic metadata into a web-based template.

In addition to descriptive elements, access points for corporate and personal names and some type of subject access are needed to truly make this collection discoverable by users. We asked: Can students with rudimentary training provide such access points? If so, are they limited to inputting uncontrolled terms, or can they be trained to use controlled vocabularies? We experimented with four input methods to evaluate which ones would effectively enable students to populate the access point elements:

- **Uncontrolled vocabulary:** for example, name headings entered in a [surname, forename] format, keywords, and summaries.

- **Controlled vocabulary lists embedded in the metadata collection tool:** for example, genre terms taken from the Rare Books and Manuscripts Section (RBMS) Controlled Vocabularies, and geographic subject headings taken from Library of Congress Subject Headings (LCSH).

- **Dynamic feedback list:** for example, an embedded list of controlled name headings for printers from the National Authority File (NAF), with a feedback mechanism permitting a cataloger update the list as needed.

- **Use of authorized thesaurus:** for example, the Virtual International Authority File (VIAF), from which students can select authorized name headings.

In this paper, we discuss the pros and cons of each method. We also note that while providing lists of controlled headings might be useful for small or homogeneous collections, this method proves unwieldy for a large heterogeneous collection, and it is worth training data inputters to use authority files.

The Mexican Colonial Collection at Cushing Memorial Library and Archives

The Mexican Colonial Collection at the Texas A&M University Cushing Memorial Library and Archives comprises more than 4,000 books, manuscripts, broadsides, newspapers, pamphlets, and religious works portraying practically all aspects of life in New Spain and Mexico from the sixteenth century to the second *Imperio Mexicano* (1867). This collection encompasses the fields of history, anthropology, linguistics, religion, political science, philosophy, civil and canon law, military science, the history of the book, the sciences, and arts. There are legal documents including items circulated by the *Santa Inquisición*, theological and religious works (many of which contain beautiful woodcuts and engravings), newspapers, discourses concerning Mexico's independence movement, and even cookbooks. Most items are in Spanish, but some pieces are in Latin, Nahuatl and other indigenous languages, French, English, or Italian. Many are unique as Texas A&M is the only U.S. institution owning copies.

Project Overview

Several hundred bound volumes containing multiple works, and more than 600 single-sheet items required cataloging to make them discoverable. In a 2012-2014 CLIR project, the investigators—three from Texas A&M and two from Mexican libraries—developed a methodology whereby Hispanophone students input bibliographic metadata into an easy-entry web-based template developed in SharePoint.[1] We hired 13 detail-oriented students with robust Spanish language skills—all the students were

native or second generation Hispanophones; half were Latin American international students or permanent residents.

The challenge was to identify strategies that would enable students to produce records similar to those created by professional catalogers, and which would integrate well with records already in the bibliographic catalog.

We divided the work into stages to determine how students would handle increasing degrees of difficulty, and to calculate the cumulative cost of generating more complex records. We prepared an instruction manual with illustrations and examples and conducted one-on-one student training sessions.

Basic Descriptive Elements		Additional Elements
Title	Author	Illustrations
Variant Title	Author 2	Illustration note
Subtitle	Author: Corporate Body	Ornaments
Variant Subtitle	Place: Country	Font Color
Statement of	Place: City	Printer's Device
Responsibility	Printer: Authorized Form	Binding
Publication: Place	Printer: Corporate Body	Marca de Fuego
Printer	Date: Date Form	Handwriting
Date		Bookplate/Ex Libris
Language		Bookstamp
Pages		

Fig. 1: Sample elements

These stages were as follows:

- **Basic descriptive elements.** The students were asked to input into the template values for the simple descriptive elements shown in Figure 1.
- **Keywords.** Students tagged the items, using natural language terms they thought appropriate.
- **Additional descriptive elements for rare books.** In this phase, additional descriptive element fields particular to the features of rare books were added to the template.

---

1   SharePoint is a web application framework and platform developed by Microsoft that facilitates collaborative projects. https://products.office.com/en-us/SharePoint/collaboration

- **Single-sheet items.** Because these formats require some unique elements, we designed a second template for the single-sheet and broadside items.

- **Authority file use.** Students were instructed in the use of Virtual International Authority File (VIAF) to test whether they could effectively select appropriate name headings for the item at hand.

Helping Students Populate the Record Elements

To enhance discoverability, besides descriptive elements, we wished to supply access points for persons and corporate bodies associated with an item, and some sort of subject access. We wondered if it were possible for students with only rudimentary training to provide such access. We assumed that our students could furnish uncontrolled access such as keywords, free-form notes (including summaries), corporate names in natural language order, and personal names in a simple [surname, forename(s)] format. We also inferred that the students' knowledge of Spanish and of naming practices in the Hispanophone world would help them correctly formulate names and deduce the substance of the items. Another question then arose: can students be trained to deploy controlled vocabularies such as the RBMS genre term list or Library of Congress Subject Headings, or copy headings from an authority file such as VIAF?

Because the records were destined for the bibliographic catalog, we wanted them to conform as much as possible to established bibliographic norms. At the same time, traditional cataloging requires mastering many complex standards such as rules for description, headings, and subject strings. This paper focuses on developing approaches that enable minimally trained students to populate bibliographic fields. We enjoyed the challenge of determining if we could enable students to produce records that would approximate those produced by trained catalogers, experimenting with the following strategies.

1. Uncontrolled Values

Some elements were satisfied with unrestricted uncontrolled values that students judged appropriate. Such values cannot later be indexed and are only useful for keyword searching of the records. These values provide imprecise inconsistent retrieval, faceting, and collation.

- **Subject keywords.** Students were instructed to consider terms *they* would use if searching for the item at hand, and to input those into an open-ended keyword field. They reported being comfortable with this task, and that information found on the item preliminaries[2] often adequately indicated the item's nature and contents. In other cases, they demonstrated resourceful ingenuity in searching the Internet for information about the item itself or the entities named on the item, to glean sufficient historical context to choose appropriate keywords.

- **Summaries.** For the proclamations and broadsides, students input summaries describing the nature of the item. Occasionally summaries could be copied verbatim from a vendor-supplied description.

- **Uncontrolled titles.** Students were asked to input the titles as found on the item. They were not required to identify the original title if the item was a translation or reprint; therefore, the resulting records will lack any required uniform titles and linking notes for

---

2   Title page, cover, colophon, etc.

the original publication, unless subsequently supplied by a cataloger. We considered these elements too difficult to require from students.

- **Uncontrolled author names.** In the first stage of record construction, students were asked to input the author's name in the Author field in [surname, forename(s)] format. Elsewhere in the record (e.g., the statement of responsibility, summaries), the students were to record names as found on the item. We thought this would be an easy task, but it was found to be difficult, as discussed below.

2. Controlled Values from Drop-down Lists

One of the reasons SharePoint was selected as a tool for hosting our record database is that it permits us to embed drop-down lists of preselected values in the fields. We used terminology from standard controlled vocabulary lists; in the case of a name, we used the authorized form found in the authority file. Database administrators may also implement rules controlling whether a selection is required, or if multiple values are permitted.

- **Static lists.** Lists of controlled values were embedded in these element fields: country, language, illustrations, ornaments, font color, binding types, author role, and title source. Students were given instructions and definitions to enable them to select appropriate values.
- **Static lists: genre.** A list of collection-appropriate genre terms selected from the RBMS Controlled Vocabularies was embedded in the genre element field.
- **Static lists: LCSH subjects.** Broad subject headings selected from the *Library of Congress Subject Headings* were used in the subject element field.
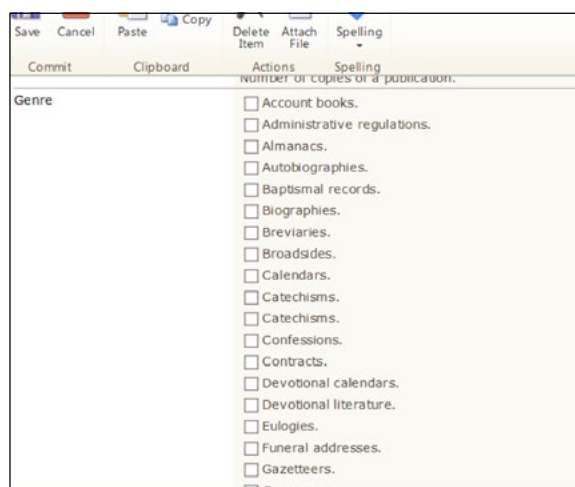

Fig. 2: Genre element with embedded list

Fig. 3: Subject element with embedded list

Controlling the possible values ensures compatibility with values used in existing cataloger-produced records. In turn, this promotes indexing and faceting when the new records are integrated into the catalog. The controlled vocabulary lists also reduced typographical errors.

Drop-down lists are most useful when working with smaller homogeneous collections with a limited number of applicable terms or headings. Conversely, a broad collection like the Mexican Colonial Collection requires more terms, yet this creates a problem—a larger list of specific terms that becomes unwieldy when embedded in a template field, and which requires more extensive student training. The only way to revert the list to manageable size is to use broader terms.

Furthermore, if the cataloger must examine a large number of items in a heterogeneous collection to develop a large list of precise terms, this is so labor intensive that he/she might as well assign the subject headings in the traditional manner.

To overcome these problems we attempted to modify this strategy and experimented with making lists dynamic, expanding them as needed through feedback provided by the students.

▪ **Dynamic lists. Example—City of publication:** For the city of publication element, we developed a list of city names in the form authorized by the Name Authority Cooperative Program (NACO) authority file, with cities likely to appear on the items in the collection. But if the students encountered a city of publication not on the embedded list, they could select the *Specify your own value* option and input the new city name as found on the piece. A cataloger would periodically review the database to find these new city names, look up the authorized form of the city, add the authorized form to the list as a value likely to occur on items in this collection, and correct that record by removing the *Specify your own value* choice, replacing that choice with the authorized form of the city now on the list.

▪ **Dynamic lists. Example—Authorized name of printers:** Similarly, we developed a list of authorized names of printers likely to be associated with this collection. Using the same process described above, new printer names could be added to the dynamic list as needed.

Using feedback generated by the student's examination of the items permitted us to provide larger numbers of controlled terms. Still, this strategy is more useful when working with a more homogeneous collection such as the *Primeros Libros*, which is limited to sixteenth-century imprints and thus associated with a limited number

Fig. 4: Drop-down list for Publication: City field

of printers. For the larger Colonial collection, because the number of printers grew exponentially throughout the eighteenth and nineteenth centuries, it became increasingly difficult to modify the list in a timely manner.

3. Student Use of Online Authority File

To avoid the impracticality and labor of embedding an extensive list in fields, we attempted to test whether students could learn to effectively search name authority files and discover for themselves the authorized form of a corporate or personal name.

For a subset of records that did not already have name headings provided by one of the previous strategies, the students searched the VIAF. They copied the 1XX field containing the authorized form of the name into the SharePoint template fields. This strategy required investing in training students to navigate the authority file and choose the correct record corresponding to the needed name. There are benefits to such training. However, students sometimes had difficulty identifying when they had the correct authority record corresponding to a name found on the item.

What Catalogers Know About Spanish Name Headings

Making access points for personal names can be daunting, even to people familiar with the language of the items being cataloged. Catalogers have the National Authority File and the rules set forth in Resource Description and Access (RDA) and the Anglo-American Cataloging Rules (AACR2) to guide them, but non-catalogers would be unaware of these resources and would require complicated training to use them. This project involved Spanish-language materials. Spanish-speaking students were chosen with the hope that their knowledge of Spanish would permit them, with only basic instruction, to formulate names correctly. That did not always happen.

Some Spanish names consist solely of forename(s)-surname. As is the case with names in English, a basic [surname, forename(s)] heading would suffice, as long as the design of a project does not require that each name is unique.

But most Spanish names are not so simple. Most people in the Hispanophone world use the surnames of both parents. The father's surname comes first and, thus, serves as the beginning of the access point. For example, in the case of Juan Antonio Pérez Vásquez, whose father's family was Pérez and whose mother's was Vásquez, his heading would be Pérez Vásquez, Juan Antonio. Prepositions such as *de*, articles

(*el, la, los, las*), the two in combination (*del, de la, de las, de los*), or conjunctives such as *y* may occur as part of a name. Often when a woman marries, she appends her husband's first surname to her family's surname(s), with *de*. Thus, if Isabel María Ortíz y Pino marries Juan Antonio Pérez Vásquez, her full name would be Isabel María Ortíz y Pino de Pérez. Her surname begins with Ortíz.

According to the rules for establishing Spanish names in both AACR2 and RDA, the heading should begin with the first element of the surname, excluding prepositions like *de*.

Before the seventeenth century, some people simply appended the name of a place or an attribute to their forenames. An example of this is Pedro de Alcántara, a sixteenth-century Spanish saint.

It is customary among Catholics to name their children after saints, so sometimes people are named with phrase-like names such as Pedro de Alcántara Martínez. It is easy to think that "de Alcántara" is part of this man's surname, but here it is part of his first name. He is named after the sixteenth-century saint. The correct input for his name is *Martínez, Pedro de Alcántara.* Inputting such names requires some expertise.

Variations in Spanish form or spelling often occur during the time period covered in this collection, for

instance, Yturriaga vs. Iturriaga, Leyva vs. Leyba, Florez vs. Flores, Gerónimo or Gerónymo or Jerónimo, Xavier vs. Javier, and so on. The popular first name José was usually rendered Joseph or Josep as late as 1800. Another potential difficulty is the versions of a name appearing in other languages. Pedro Martínez wrote in Latin, using the form Petrus instead of Pedro on the items. "San Juan Nepomuceno" is the Hispanicized form of a Bohemian-born saint, John of Nepomuk.

Titles of nobility add yet another layer of complexity, particularly if a person tended to use his or her title rather than their name on documents. The Colonial Mexican collection contains a document by General Francisco Xavier Venegas addressed to "Excmo. Sr. Duque del Infantado." The Duke's name does not appear in or on the piece, and the student entering the metadata only gave Venegas' name. It required research to reveal that the Duke's full name was Pedro de Alcántara Álvarez de Toledo y Salm-Salm.

Other categories of names, such as those represented by initials only, and pseudonyms, present problems in any language. For example, one would need to look in the authority file to know that Ignacio Tomay (also spelled Tomai or Thomay) was the pseudonym of José María Genovese, a Sicilian-born Jesuit living in New Spain.

Discussion

What Was Easy?

Students found the title element easy, because the students recorded information as it came on the item in hand. They also had no trouble with elements for which we were able to provide drop-down lists. For the investigators, the dynamic lists were easy to manage until the amount of feedback became overwhelming.

What Was Difficult?

Occasionally students had difficulty identifying the author(s). When several names were mentioned on an item, students were sometimes misled by the wording into selecting the wrong individual as the author. For example, in Spanish the word *por* can mean *for* or *by*, depending on the context. Usually the name of an author is introduced by *por*, as in *Don Quijote de la Mancha* por [by] Miguel de Cervantes. However, in funeral sermons, where the title statement appears in the form *Oraciones funerales por* [name of the deceased], students automatically assumed the individual named after the word *por* was the author. The word *del* can also be misunderstood in the same way, as in *"Solemnes exequias del illmo. señor dr. d. Jose Gregorio Alonso de Hortigosa, obispo que fue de la ciudad de Antequera …"* for which a student entered Hortigosa in the record as the author, when he is actually the deceased. This mistake happened often enough to provoke the humor in our paper's title.

Other errors occurred when students formulated name headings without help from the VIAF file. The following types of mistakes appeared frequently.

Type of Error	Entered	Correct
Incorrectly selecting an element as the first element of the surname	Antonio de Oviedo, Juan Jose De Urdanivia, Cárlos Joseph Diez	Oviedo, Juan José Antonio de Díaz de Urdanivia, Carlos Jose
Incorrectly including abbreviations of titles as part of the name	De Castro, D. Juan Francisco Br. D. Juan Gabriel de Contréras	Castro, Juan Francisco de Contréras, Juan Gabriel de
Beginning a name heading with *de*	de Contreras, Juan Martin de Jesus Soria, Rafael	Contreras, Juan Martín de Soria, Rafael de Jesús
Incorrect use or lack of diacritics	Perez Quixano, Agustin	Pérez Quixano, Agustín
Not recognizing when a person's forenames include the name of a saint	Nepomuceno Rivas, Juan	Rivas, Juan Nepomuceno (*Nepomuceno* is really part of the man's forename, as he was named after St. John of Nepomuk)
Not recognizing non-Spanish forms of names; i.e., names in Latin, Italian, or French, or even archaic spellings of Spanish	José Gregorio Alonso de Hortigosa Joseph Gregorii Alfonsi de Ortigosa (Latinized form of the name)	Alonso de Ortigosa, José Gregorio (heading from Biblioteca Nacional de España)
Not recognizing when an individual had only a forename (despite training)	Transfiguracion, Francisco de la Sales, Francisco de	Francisco de la Transfiguración Francis de Sales

Table 1: Most common errors, with examples and corrected forms.

What Was Impossible?

Three challenging areas were typically beyond students' abilities.

- **Works by corporate officials are entered under their official titles rather than their personal names.** This is not something we would expect a non-cataloger to know. For example, a document issued under the authority of the viceroy Martín de Mayorga should have as the author the form *New Spain. Viceroy (1779-1783: Mayorga)*, not *Mayorga, Martín de.* We accepted that records for official documents would incorrectly list personal name forms in the element fields.

- **Translations.** Without substantial training and extensive searching in OCLC, it would be difficult for students to determine when a publication was a translation or what its original title would be. We accepted that these records would not have translation uniform titles and would not collate with, or have linking notes to, the original editions.

- **Pseudonyms.** Although sometimes—with the help of the VIAF files or through an Internet search—the students were able to find the real name of an author appearing on the item with a pseudonym or initialism, this was a difficult task we did not require.

Conclusion

The purpose of this paper is not to argue for the strict use of controlled vocabularies and headings in bibliographic descriptions, either in traditional catalogs or in other databases using non-MARC based records. We understand that it is acceptable for record standards to vary depending on the type of access needed for a particular collection or project. And it is clear that there is a lack of trained personnel to create records at the highest level for the backlog of items that need to be processed. It

was not our intention to demonstrate the value of the work of the Library of Congress, OCLC (via VIAF), Dublin Core, ALA, and others. in promoting the sharing of authority records and of controlled vocabularies outside the traditional cataloging community. Yet, unexpectedly as a related issue, this project has reiterated to us the importance of these organizations' efforts as they facilitate the participation in projects (especially digital) of non-catalogers.

Such efforts allow project managers to shift focus away from the difficult, expensive, and time-consuming task of training non-catalogers (usually temporary personnel) to construct headings or use vocabularies according to standards. Instead, we can focus on the lesser task of training staff to navigate the authority record databases and select the correct record showing the authorized form of the topical or name heading needed. Users of the authority file records can see the correct authorized form of a heading, including diacritics, order of the elements of a name, correct spelling, and so on, so they do not need to know how to construct a heading. Professional catalogers—trained in the rules and standards for constructing headings and vocabularies—can dedicate themselves to building the authority records.

Using the online authority and vocabulary resources frees us from training non-catalogers how to construct a heading, but we still need to train them to identify the appropriate record. Experienced catalogers can identify non-obvious bits of information in authority records (including in the citations), which help them determine that a record corresponds to the individual named on the item at hand. For example, one student selected an individual as an author for an eighteenth-century item, while a cataloger would have known that, since the physics text cited

in the record's 670 *Source Data Found* field was published in the twentieth century, the individual represented by the authority record could not be the same individual as author on the older religious-themed item in hand. Non-catalogers, who lack the training and experience, cannot see these clues. The new RDA attributes[3] permit a non-cataloger to better identify which record/individual is the correct one, allowing staff to focus on inputting the proper information rather than debating among several possible records. As another unexpected outcome, this project demonstrated to us the merit of RDA authority record attributes in helping non-catalogers select correct individuals/corporate bodies and the authorized form of name for use in name and subject access.

References

Joint Steering Committee for Revision of AACR2. 2002. *Anglo-American Cataloguing Rules,* 2nd ed., 2002 rev. Chicago: American Library Association.

Joint Steering Committee for Development of RDA. 2010. *RDA: Resource Description & Access.* Chicago: American Library Association.

Name Authority Cooperative Program (NACO): http://www.loc.gov/aba/pcc/naco/.

Virtual International Authority File (VIAF): http://viaf.org/.

Additional Reading

Almuzara, Leticia Barrionuevo, M. Luisa Alvite Díez, and Blanca Rodrígue Bravo. 2012. A Study of Authority Control in Spanish University Repositories. *Knowledge Organization* 39 (2): 95-103.

DeZelar-Tiedman, Christine. 2011. Exploring User-Contributed Metadata's Potential to Enhance Access to Literary Works: Social Tagging in Academic Library Catalogs. *Library Resources & Technical Services* 55 (4): 221-233.

Lu, Caimei, Jung-ran Park, and Xiaohua Hu. 2010. User Tags Versus Expert-Assigned Subject Terms: A Comparison of LibraryThing Tags and Library of Congress Subject Headings. *Journal of Information Science* 36 (6): 763-779.

Maron, Deborah, et al. 2014. Lo-Fi to Hi-Fi: A New Way of Conceptualizing Metadata in Underserved Areas with eGranary Digital Library. *Proceedings of International Conference on Dublin Core and Metadata Applications,* Austin, Texas, October 8-11, 2014, 37-42. Available at dcevents.dublincore.org/IntConf/dc-2014/paper/download/259/228.

Yusuf, Felicia, and Christopher Nkiko. 2010. Involvement of Non-Professionals in Cataloguing: A Survey of Practices in Three Academic Libraries in Nigeria. *Malaysian Journal of Library & Information Science* 15 (1): 83-90.

---

3   Associated Dates, Associated Place, Address, Field of Activity, Associated Group, Occupation, Gender, Family Information, Associated Language, Fuller Form of Personal Name.