Collaboration and Education: Engaging High School Students with EAC-CPF Research*

Valerie Addonizio and Christopher Case, Sheridan Libraries, Johns Hopkins University

Abstract

The Special Collections Research Center at Johns Hopkins University received a CLIR Hidden Collections grant for processing the archives of the Roland Park Company from March 2013 to March 2014. The grant included partnerships between the university's Archives and its Technical Services department to create best practices for EAC-CPF records, as well as a partnership with a local high school history class to complete the required research. Rather than simplify ISAAR (CPF) and EAC-CPF for the students, project staff distilled them into discrete, easy-to-understand tasks that allowed for the production of controlled data in a high school environment.

Project Background and Grant Requirements

The Ferdinand Hamburger Archives is the official archival repository for Johns Hopkins University's Homewood campus divisions, the School of Education, the Carey Business School, and the Paul M. Nitze School of Advanced International Studies. Archival holdings include the business records of the university as well as a substantial body of manuscript collections documenting a variety of research areas including the history of science, literature, higher education, politics, and regional planning and development. In 2010 the archives received the Roland Park Company Records as a transfer from Cornell University. This collection, which focuses on the development of several important Baltimore neighborhoods, together with the Martin L. Millspaugh Archives, another prominent Baltimore collection gifted to us, was the impetus to write a CLIR Hidden Collections grant in 2011.

The project outlined here is a result of that grant, which included the development of a set of EAC-CPF (Encoded Archival Context-Corporate bodies, Persons, and Families) best practices through collaboration with the university's Archives and its Technical Services department; collaboration with a local high school for research and identification of biographical and related archival holdings information; and the adaptation of that information to create EAC-CPF records.

The success of this project was measured in three ways: first, against the terms laid out in the grant; second, in terms of the students' experience; and third, in terms of whether the quality of the records hewed to the best practices developed by the project team.

Formulating Local Best Practices

Because the vast majority of the EAC-CPF standard consists of optional elements, the team (consisting of the project archivist and the content management librarian representing technical services) knew that it would benefit all parties, and potentially the wider archival community, to

---

* This paper is an abridged version of Addonizio and Case 2014.

create a series of local best practices for implementing EAC-CPF. Background research included evaluating best practices and draft records in a number of EAC-CPF instances, including those created by Harvard and Yale (Diaz and Pyzynski 2012), Tufts (DAT-040), and the SNAC (Social Networks and Archival Context) project. For a full treatment of this process, see Addonizio and Case 2014.

Collaborating

After creating best practices, we began the collaboration between Johns Hopkins and the local high school. This involved the Johns Hopkins University archivist, the Roland Park Company records project archivist, the class's history teacher, and two school librarians.

The idea for this collaboration came out of a simpler question of whether a senior high school history class could come to campus for a project involving primary sources. Then we came up with the idea of doing EAC-CPF research. In conversation with the educators, we learned that they hoped to contextualize the assignment to be as much about research using archival material as about a real-life application of the work. In other words, the educators were interested in treating the assignment like a client-based work order with the client, Johns Hopkins, having a real-life request and the students having to follow strict instructions to deliver a usable product. This expectation was significant, because there were a number of times when Johns Hopkins and the school staff questioned whether the requirements of the project would overreach the normal workload or comfort zone of high school students.

Distilling

Because the educators were undaunted and, in fact, enthusiastic about the project's complexity, we moved forward with trying to distill our best

practices and the encoding realities of EAC-CPF into a project that could capture the complexity of ISAAR (CPF)/EAC-CPF and be managed outside our immediate supervision.

The initial idea was to provide the students with a single spreadsheet to fill out for each entity. Yet the reality was that we asked for more than the biographical and relationship information suggested by ISAAR (CPF). We also asked for controlled data required for encoding that reflected our best practices.

An example using <placeRole> helps demonstrate the challenge. For the actual encoding of EAC-CPF, our best practices limit the value of <placeRole> for a person to one of birth, residence, education, marriage, occupation, travel, death, or burial. We could not ask for only a list of predominant places added to a spreadsheet. We had to limit the places to those prescribed by our best practices, and the students needed to know that. This is the simplest example of the complicated interplay of content and requirements that we had to anticipate.

Given this complexity, oversimplification was a risk. Therefore, we decided not to simplify the requirements (i.e., fill out this single spreadsheet), but the method by which they could be fulfilled. As a result, we provided students with a suite of four documents, each of which helped explain what we were looking for (content) and how we needed to see it (controlled data, including authority work). The documents use the ISAAR (CPF) section numbering system like a primary key, allowing a student to cross-reference the documents. Our aim was that, following a general introduction to the project, we would only need to explain how to use the documents, then the documents would stand on their own. What follows is a description of each document, followed by how students used it.

The ISAAR Roadmap. This is a blank ISAAR (CPF) record repurposed as a roadmap that lists every requirement that needs to be met, and then points students to what that requirement means (content), and where and how to fill it in (controlled data). The fields are ISAAR (CPF), but the instructions combine our best practices and the requirements of EAC-CPF. The roadmap's secondary purpose is to help delineate between two major concepts—gathering information about corporate bodies, persons, or families (ISAAR 5); and gathering information about related resources on corporate bodies, persons, or families (ISAAR 6). For instance, it helps to point out that the bibliography the students write for the <biogHist> is different from the list of related resources. We created two roadmaps—one for persons and one for corporate bodies.

By referring to the ISAAR roadmap, the first thing the student sees is the ISAAR (CPF) section "5.1 Identity" (Figure 1). In that section the first task is to find the "5.1.2 Authorized form of name." That has two components—the authorized form of name, and a permalink. What does that mean

to the students? Further to the right, the roadmap points to two sets of instructions—"Where to Find Names" and "Permalinks" in another document (the project guide). After students read those instructions in the project guide, the roadmap shows that the data are to be entered on tab 5.1 on the project spreadsheet.

The Project Guide. At 10 pages, the guide is weighty, but it was written as a reference document rather than a long exposition. It provides definitions and context to certain usage, detailed instructions on how to get authorized names, places to look for related resources, and ways to determine permalinks and list relationships. Each of the sections is referenced from the ISAAR roadmap. We purposely used informal language and we aimed for relatable examples (see Figure 2).

The Spreadsheet. Ultimately, we provided a spreadsheet for all the data (except the narrative biography/history), but with multiple tabs, each of which represented a different ISAAR (CPF) section. Using a spreadsheet has two benefits: first, spreadsheet fields are easy to export

ISAAR 5. ELEMENTS OF AN AUTHORITY RECORD		
5.1 Identity		
5.1.2 Authorized form of name Use spreadsheet tab 5.1	Authorized form of name	See *Where to Find Names (CPF)* in your Project Guide.
	Permalink to source of name	See *Permalinks* in your Project Guide
5.2 Description Area		
5.2.2 History		Write a chronology using spreadsheet tab 5.2.2
		Write a narrative biography (800 words or less) and submit it using this link.
5.2.3 Places Use spreadsheet tab 5.2.3		See *Where to Find Place Names* in your Project Guide
5.2.5 Occupations Use spreadsheet tab 5.2.5		See *Where to Find Occupations* in your Project Guide

Fig. 1: ISAAR roadmap

Associative

This is the broadest relationship and the one you will use the most. Basically, if the relationship you're describing isn't one of the others in this list, use Associative. The important part is really in the description, which allows you to briefly state how the two entities are related.

Examples:

Mutual relationships	Directional relationships
If the relationship is mutual, like the ones listed below, the description does not need to imply the "direction" of the relationship.	Here are examples of when the direction of the relationship is described in the description.
Main topic: John Lennon **Associative relationship:** Paul McCartney **Description:** Band mates in The Beatles.	**Main topic:** Steve Jobs **Associative relationship:** Apple Computer, Inc. **Description:** Company founded by Steve Jobs.
Main topic: Ben Cohen **Associative relationship:** Jerry Greenfield **Description:** Co-founders of Ben & Jerry's Ice Cream.	**Main topic:** Pew Charitable Trust **Associative relationship:** National Public Radio **Description:** Receives financial support through charitable donations.

Fig. 2: Sample from the project guide

and manipulate, and would be advantageous in post-production; and second, drop-down fields allowed us to control vocabulary dictated by our best practices. We created two spreadsheets—one for persons and one for corporate bodies.

The narrative continues using the <placeRole> example from above. The roadmap indicates that the information for "5.2.3 Places" be added to tab "5.2.3 Places" on the spreadsheet. The cells in that tab limit students to the list of values for that element, allowing our best practices to guide student data entry (Figure 3).

The actual ISAAR (CPF) standard was used for conceptual context. We provided it just in case it helped the educators or students to pin down what we were looking for. Importantly, the numerical sections provide the vital framework for enabling all the documents to enter into a relative relationship.

Acquainting Students with the Project

After writing and testing the suite of documents in August 2013, we introduced students to the project in September. We developed two presentations. The first dealt with the nature of archival material and, significant to the project, examples of real finding aids. The second was an introduction to ISAAR (CPF), and consisted almost entirely of reading the ISAAR roadmap document together as a class. Students opened all four documents on their laptops and went from document to document, as the roadmap instructed.

We also discussed the difference between ISAAR 5 (the section that defines biographical information, familiar to students) and ISAAR 6 (the section that defines related collections, and was a new concept). At the end of the session, one student said, "But you can't have an ISAAR 5 without an ISAAR 6," meaning you can't have information about an entity without sources about

A	B	C
5.2.3 Place Name	**Source of Name**	**Role**
Westmoreland County (Va.)	LC Subject Headings	birth
Mount Vernon (Va.)	LC Subject Headings	residence

birth
residence
education
marriage
occupation
travel
death
burial

5.2.2 History | **5.2.3 Places** | 5.2.5 Occupations | 5.3 Relationships | 6. Related Resources | She

Fig. 3: Sample from the person spreadsheet

that entity. This statement demonstrated a fundamental understanding about the relationship between biographical content and the archival collections that contain it, and was a cause for team celebration.

Over the next few weeks, students e-mailed us some of their questions. Two examples reflect an increasing awareness of the complexities in creating archival authority records, suggesting that the students were participating in a valuable learning experience:

"When you have a person who has the same occupation for overlapping or several years, how do you differentiate? Example - Kessler was an urban planner, landscape architect, for several jobs, overlapping in time."

"I had a question regarding my ISAAR 6 on Charles Grasty. I have found some archival materials that are individual and not members of larger archives [*i.e., catalogued individually in Worldcat*]. They therefore don't have finding aids. Should I instead provide the Worldcat permalink of the source?"

Results

We reviewed the data twice before to the end of the project, in November and in late December 2013. In both cases, students provided sample records, which we reviewed in detail and found were quite solid. The students returned their data in January 2014, when the project archivist set to work adapting the information into EAC-CPF.

We immediately observed that the results

required careful proofreading before the data, especially the biographical or historical notes, could be adapted. Students used subjective language throughout. The only expectation was that the records be "Wikipedia-worthy," but nearly every note still needed our attention. Also, because of the complexity of the instructions, some students entered information in the wrong fields, requiring minor corrections and some additional research. Furthermore, we anticipated that almost all dates would need to be normalized, but mistakenly expected that we could automate that process.

Upon reflection, we realize that date normalization is something that can be easily explained to students in a future version of this project. Other challenges that arose with dates include incorrect dates, the recording of circa dates (something we had not anticipated and for which we take responsibility for not explaining), incomplete date ranges, and missing dates. Finally, in some cases there was missing and incomplete information, which ranged from not listing a person's

occupation to not submitting any biographical or historical notes. We do not find the students primarily at fault for these issues. All of these challenges can be better addressed in the future with careful instruction and ongoing evaluation by the archives team.

Analysis

As stated above, project success was measured in three ways. First, in relation to the terms laid out in the grant, we found that 15 EAC-CPF records were created in accord with EAC-CPF emerging best practices. Archives and Technical Services collaborated on a set of best practices. We were successful in translating the complexity of EAC-CPF requirements to high school students. And virtually all of the students demonstrated a fundamental understanding of what was required.

Second, we measured project success by responses from students and educators. Everyone at the school agreed that this was an exceptionally unique and meaningful project for students and a

Names	Chronology	Names	Sources	Bioghist	Occupations/ Functions	Places	Relationships	Related Resources
Atterbury, Grosvenor	X	X	N/A	X	X	X	X	X
Baltimore Country Club	X	X	X	X	X	X	X	X
Baltimore Sun	N/A	N/A	N/A	X	N/A	N/A	N/A	N/A
Bouton, Edward J.	X	X	N/A	X	X	X	X	N/A*
Fowler, Lawrence Hall	X	X	N/A	X	X	X	X	X
Grasty, Charles	X	X	N/A	X	X	X	X	X
Kessler, George Edward	X	X	X	X	X	X	X	X
LaMotte, Frank Linton	X	X	N/A	N/A	X	X	X	N/A
Maryland Transit Authority	X	X	N/A	X	X	X	X	X
Mowbray, John McConkey	X	X	X	X	X	X	X	N/A*
Olmsted Brothers	X	X	N/A	X	X	X	X	X
Olmsted, Frederick Law (Junior)	X	X	N/A	X	X	X	X	X
Palmer, Edward Livingston	X	X	X	X	X	X	X	X
Russell Sage Foundation	X	X	N/A	N/A	X	X	X	N/A
Urban Land Institute	X	X	N/A	N/A	X	X	X	N/A

*-Some missing information was discussed with the Project Archivist before the records were returned, and it was determined that this information truly did not exist and the students were not accountable for it.

Fig. 4: Analyzing the results

terrific example of experiential learning. They are eager to collaborate again. The Archives shares this sentiment.

The third measure is whether the quality of the EAC-CPF records hews to our best practices. Our assessment of this aspect of the project is mixed. Of the 15 EAC-CPF records created, only three were entirely complete according to Hopkins's best practices. Further, four could not be used because they would require more research to be complete. The remaining eight could be considered complete by a slightly less strenuous standard. If <sources> were not required, the number of complete records would increase to eleven.

We encountered a notable variation in record accuracy, quality, and comprehensiveness, because of the varied set of skills that each student brought to bear on the assignment. Figure 4 indicates which sections of the EAC-CPF record were provided. The extent to which these submissions meet our standards for publication was inconsistent. Over the course of the encoding, the project archivist found inaccurate dates, inconsistent dates, incorrectly formatted sections, and factual errors, suggesting that we should have had more direct collaboration with educators and students to better communicate our expectations. One of our most important lessons learned is that not every student should be expected to submit error-less work. In an actual work setting, professional staff do not just evaluate a product; supervisors review staff work and provide feedback until the product is worthy of submission. When partnering with students on a product, it should be the responsibility of the professional staff (here, the project archivist) to perform quality control and work closely with them to refine the product until it contains no errors.

Because very few of the records were entirely accurate or complete, we believe that in this third evaluation measure—the accuracy of EAC-CPF records vis-á-vis our best practices—the project was a complete success in terms of being a proof of concept, though the proof was in the negative. However, we also believe we can significantly address critical concerns in a future project.

These concerns fall into three areas. First, we believe that some of our quality control challenges raise interesting questions. For instance, does the quality have more to do with the age and experience of the students than with the method? Some of the work was excellent and complete, but was asking for this level of detail from high school students too much, even with educators' encouragement? It is possible that a similar project would be more successful with college-age students and a virtual environment made for populating EAC-CPF records.

Second, because this was a high school research project, students were not allowed to simply mine Wikipedia for the answers to their questions. We wonder if doing that for especially well-known entities such as, say, Frederick Law Olmsted Jr., would have been more successful than conducting a short-lived and limited-resourced research project on that individual. Allowing for repurposed information might have yielded fewer inaccuracies and gaps.

Third, we learned much about how and when to review data during a project. We conducted two reviews prior to final submission, and both times we deemed the data accurate. Unfortunately, those samples ended up being the most accurate records the students had produced. One recommendation is to review all the work being done, not only a few sample sets. Also, we would recommend there be two submission dates rather than having the project be the final assignment. This would allow for iterative grading,

something the educators might not usually do. Those students whose work did not meet our standards in the first submission would be given the chance to complete the records before final submission. We need to better account for the discrepancy between expectations within a learning environment and those within a professional environment.

As for the students, we believe they digested a seriously complex assignment, became familiar with archives and archival holdings, did some original historical research, tested the limits of their capacity for following instruction, and gained a greater appreciation for the real-life application of historical research. The variations in accuracy of their final data should not diminish the accomplishment of undertaking this complex assignment.

Conclusions and Observations

The success of this project depended heavily on the active participation of two school librarians. These individuals were essential to both the teacher and the students in interpreting and understanding the ISAAR (CPF) standard. One of the librarians became the principal contact after the project began, and asked questions that could only be articulated, understood, and applied by another information professional. Her fundamental understanding of how to apply the standard allowed her to be on the frontline of student questions, and helped assure us that the interpretation would be accurate. Without these two librarians, whose participation we accounted for in the project design, the work would likely not have gone so smoothly. The "do not try this at home" disclaimer can be qualified with "unless under the direct supervision of an information professional."

A second observation coalesced as we thought about the relationship between ISAAR 6 (relationships to resources) and the real-life advantages of EAC-CPF. In this project we tasked students with manually gathering biographical or historical information for an individual or corporate body, then seeking out and finding related collections. Clearly, this was a time-intensive approach. Like others interested in EAC-CPF, we contemplated the pros and cons of both manual input and automation models. At the same time, unrelated to that debate, we framed a lot of project discussion around the differences in the need for accuracy between the information required for ISAAR 5 and that required for ISAAR 6. We found that we preferred that the accuracy lie on the ISAAR 5 side of the record (accurate birth dates, lists of related corporate bodies, persons and families, occupations), but we could never expect an exhaustive list of every related archival collection for every entity.

Finally, we hope that this case study of collaboration and EAC-CPF will be useful as the profession moves forward with the standard. Although the future of EAC-CPF will likely include automated data harvesting such as that done by SNAC, this project demonstrates that well-designed and monitored cooperation may also play a significant role. Although automation offers the prospect of generating mass numbers of records, the reality remains that many older EAD finding aids do not contain the level of detail and nuance that EAC-CPF allows for, and archivists' workloads can only benefit from well-organized collaboration. We also hope that the outreach component of this collaboration provided a meaningful experience to the students and faculty that extends beyond the project outcomes produced.

We anticipate that all related project material, including the best practices and three student documents, will be available on the CLIR Hidden

Collections website at http://www.clir.org/hidden collections/registry/hc.0880.

References

Addonizio, Valerie, and Christopher Case. 2014. "Collaboration and Education: Engaging High School Students with EAC-CPF Research." *Journal of Archival Organization* 12, 1-2.

DAT-040 DCA Best Practices for EAC. Available at http://sites.tufts.edu/dca/files/2011/03/DAT-040 DCABestPracticesforEAC.pdf.

Diaz, Enrique A. and Susan C. Pyzynski. 2012. Connecting the Dots: Using EAC-CPF to Reunite Samuel Johnson and His Circle, Harvard University and Yale University. Available at https://wiki.harvard.edu/confluence/display/connectingdots/.

SNAC Prototype, The Social Networks and Archival Context Project. Available at http://soci-alarchive.iath.virginia.edu/xtf/search.