

**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-March 2014. The grant included partnerships between the Archives and Technical Services to create a Best Practices for EAC-CPF records, as well as 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 focused on the development of a number of important Baltimore neighborhoods, as well as the gift of the Martin L. Millspaugh Archives, another prominent Baltimore collection, was the impetus to write a CLIR Hidden Collections grant in 2011.

* This paper is an abridged version of an article of the same name appearing in the forthcoming issue of the *Journal of Archival Organization*.

The project outlined in this paper was a result of that grant, which included the following:

1) the development of a set of EAC-CPF Best Practices as a collaboration of both the Archives and Technical Services; 2) collaboration with a local high school for research, and identification of biographical and related archival holdings information; and 3) the adaptation of that information for the creation of EAC-CPF records.

The success of this project was measured in three ways. First, the project was measured against the terms laid out in the grant. Second, the success was measured in terms of the experience of the students. And third, it was measured in terms of the quality of the records in accordance to the Best Practices developed by the Project Team. The evaluation of the project according to these metrics concludes the paper.

Formulating Local Best Practices

Since 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 involved, and potentially the wider archival community, to create a series of local Best Practices for implementing EAC-CPF. Background research included evaluation of the Best Practices and draft records in a number of EAC-CPF instances, including those created by Harvard and Yale (*Connecting the Dots*)¹, Tufts², and the SNAC project³. For a full treatment of this process, please see the article of the same name in the forthcoming issue of the *Journal of Archival Organization*.

Collaboration

After the creation of Best Practices, the collaboration between Johns Hopkins and the local high school could begin. This involved the University Archivist for Johns Hopkins, the Project Archivist for the Roland Park Company Records, the history teacher whose class was participating, and two school librarians.

The idea for this collaboration came out of a much simpler question of whether a senior History class could come to campus for a project having to do with primary sources, but once the idea of doing EAC-CPF research materialized, the philosophy of collaboration evolved. In conversation with the educators it emerged that they hoped to contextualize the assignment as being 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: the client, Johns Hopkins, has a real-life request and they, the students have to follow strict instructions in order to deliver a usable product. This is 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 a high school student.

Distillation

Since the educators were undaunted and in fact enthusiastic about the complexity of the project, we moved forward with trying to distill both our Best Practices and the encoding realities of EAC-CPF into a project that could a) capture the complexity of ISAAR (CPF)/EAC-CPF and b) be managed outside our immediate supervision.

The initial idea was to provide the students with a single spreadsheet they could fill out about each entity (i.e., ISAAR (CPF) translated into a single spreadsheet). Yet, the reality was that we were asking for more than the biographical and relationship information suggested by ISAAR (CPF); we were also asking for controlled data required for encoding that was reflective of our Best Practices.

An example using <placeRole> helps to 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. So we could not ask for only a list of predominant places to be 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 the complexity, over-simplification was a risk. Therefore the team decided to simplify not the requirements (i.e., fill out this single spreadsheet), but the method by which they could be fulfilled. As a result, the students were provided with a suite of four documents (detailed below), each of which helped to explain what we were looking for (content) with 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 between them. Our aim was that after a general introduction to the project, all we would need to explain was *how* to use the documents, and then they

would stand on their own. What follows is a description of each document, followed by exactly how it is used by the students.

1) Begin with 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 the students to a) what that requirement means (content) and b) where and how to fill it in (controlled data). The fields are ISAAR (CPF), but the instructions are our Best Practices and the requirements of EAC-CPF combined. Its 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 than the list of related resources. Two roadmaps were created: one for Persons and one for Corporate Bodies.

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 <i>Where to Find Names (CPF)</i> in your Project Guide.
	Permalink to source of name	See <i>Permalinks</i> 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 <i>Where to Find Place Names</i> in your Project Guide
5.2.5 Occupations Use spreadsheet tab 5.2.5		See <i>Where to Find Occupations</i> in your Project Guide

Figure 1. ISAAR Roadmap

By referring to the ISAAR Roadmap, the first thing the student sees is the ISAAR (CPF) section *5.1 Identity*. In that section, the first task is to find *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 titled “Where to Find Names” and “Permalinks” in another document (the Project Guide). After reading those instructions in the Project Guide, the Roadmap shows that the data is to be entered on tab 5.1 on the project spreadsheet.

2) Reference the Project Guide. The Guide is a weighty document at ten pages, 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, how to determine permalinks, and how to list relationships. Each one of the sections is pointed to from the ISAAR Roadmap document. The language was purposefully informal and the examples aimed to be relatable. An example follows:

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

Figure 2. Sample from the Project Guide

3) Add data to the spreadsheet. Ultimately there was a spreadsheet for all the data (except the narrative biography/history), but with multiple tabs that each represented a different ISAAR (CPF) section. There are two benefits: first, spreadsheet fields are easy to export and manipulate, and would be an advantage in post-production; and second, drop-down fields allowed us to control vocabulary dictated by our Best Practices. Two spreadsheets were created: 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 the students to the list of values for that element, allowing our Best Practices to guide the student data entry.

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 instructed by the Roadmap.

There was also a necessary discussion on the difference between ISAAR 5 (the section that defines biographical information, familiar to students) and ISAAR 6 (the section that defines related collections, new concept). At the end of the session a student stated, “But you can’t have an ISAAR 5 without an ISAAR 6;” or, you can’t have information about an entity without sources *about* that entity. This statement demonstrated a fundamental understanding between biographical content and the archival collections that contain it, and was a cause for celebration for the team.

Over the next few weeks some questions came through e-mail. Two significant examples:

“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?”

These questions reflected an increasing awareness of the complexities in creating archival authority records, suggesting that the students were participating in a valuable learning experience.

Results

There were two reviews of the data prior to the end of the project, one in November and one in late December 2013. In both cases sample records were provided and the progress was reviewed in detail and found to be quite solid. The students returned their data in January 2014, at which time the Project Archivist set to work adapting the information into EAC-CPF.

It was immediately observed that copy editing required careful proofreading before the data (especially the biographical or historical notes) could be adapted. Subjective language was also used throughout; the only standard applied when editing was “Wikipedia-worthy,” but nearly every note still needed attention. Also, due to 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 erroneously expected that we could automate that process. Upon reflection, date normalization is something that could be easily explained to the students in a future version of this project. Other challenges arose with dates, including incorrect dates, the recording 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 occupations to not submitting a biographical or historical note of any kind. We do not find the students primarily at fault for these issues; all of these challenges could be better addressed in the future with careful instruction and ongoing evaluation by the Archives team.

Analysis

As stated above, because of the grant requirements, the success of this project was measured in three ways. First, when considering if the project was a success according to the grant, fifteen EAC-CPF records were created in accord with EAC-CPF emerging best practice. The Archives and Technical Services did collaborate on a set of Best Practices; the complexity of translating the requirements of EAC-CPF to a project for high school students was a success; and virtually all of the students demonstrated a fundamental understanding of what was required.

Second, we measured the success of this project according to the response from the students and educators. Everyone at the school agreed that this was an exceptionally unique and meaningful project for the students, a terrific example of experiential learning, and they are eager to collaborate again. The Archives share this sentiment.

The third vital question considers the quality of the EAC-CPF records according to our Best Practices. Our assessment of this aspect of the project is mixed. Of the fifteen EAC-CPF records created, only three were entirely complete according to Hopkins' 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> was not required, the number of complete records would increase to eleven.

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.

Figure 4. Analyzing the results

We encountered a notable variation in accuracy, quality, and comprehensiveness of these records, owing to the varied set of skills that each student brought to bear on the assignment. While the table above indicates whether a section of the EAC-CPF record was 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 there should have been more direct collaboration with the educators and students during the course of the project in order to better communicate our expectations. One of our most important revelations in overseeing the project (and thereby a lesson learned) was simply not accounting for a very basic fact about the education environment: not every student should be expected to submit work that is completely without error. In contrast, the work of professional staff does not just involve evaluating a product, but also constitutes a process of refinement, whereby supervisors review the work of staff and provide feedback

until the product is worthy of submission. As a result, we believe that 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 to work closely with the students on the process of refinement until the product is without errors.

Since very few of the records were entirely accurate or complete, the authors feel that in this third aspect of project evaluation—the accuracy of EAC-CPF records commensurate with our Best Practices—this project was a complete success in terms of being a proof of concept, but that the proof was in the negative. However, the authors also feel that critical concerns could be significantly improved in a future project.

First, the authors believe that some of our quality control challenges raise some interesting questions, namely, does this have more to do with the age and experience of the students than with the method? While some of the work that was provided was in fact excellent and complete, was asking for this level of detail from high school students asking too much, even with the educators' encouragement to do so? It is possible that a similar project would be more successful with college-aged students and a virtual environment made for populating EAC-CPF records.

Second, owed to the fact that this was a high school research project, the students were not allowed to simply mine Wikipedia for the answers to their questions. Frustratingly for the students, the team wonders if doing exactly that for especially well-known entities, for instance, Frederick Law Olmsted Jr., would have been much more successful than a short-

lived and limited-resourced research project on that individual. Given the same project, fully allowing for re-purposed information might have yielded fewer inaccuracies and gaps.

Third, much was learned about how and when to review the data during the project. There were two reviews prior to the final submissions, and both times the data was deemed quite accurate. Unfortunately, those samples ended up being the most accurate records, and so a recommendation would be to review *all* the work being done, not just a few sample sets. Additionally, the authors would recommend that there be two submission dates rather than having the project be a final assignment. This could allow for iterative grading, something that the educators might not usually do. Those students that did not produce work that met our standards in the first submission would be given the chance to complete the records before final submission. This discrepancy between the expectations of a learning environment and the needs of a professional environment should be better accounted for.

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

A significant conclusion is that the success of this project depended very heavily on the active participation of the 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 us 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 front lines of student questions, and helped assure us that the interpretation would be accurate. The participation of these two librarians was accounted for in the design of this project, without which we could not have expected it to go so smoothly. This '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 the team thought about the relationship of ISAAR 6 (relationships to resources) and the real-life advantages of EAC-CPF. In this project, students were tasked with manually gathering biographical or historical information for an individual or corporate body and then seeking out and finding related collections. Clearly, this was a time-intensive approach, and like others interested in EAC-CPF, the authors contemplated the pros and cons of both manual input and automation models. At the same time, but unrelated to the debate, the team framed a lot of its discussion for the project around the differences in the need for accuracy between the information required for ISAAR 5 and 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, the authors hope that this case study of collaboration and EAC-CPF will serve some function as the profession moves forward with the standard. The authors posit that 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. While 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. It is also our hope that the outreach component of this collaboration ultimately provided a meaningful experience to the students and faculty that extends beyond the project outcomes that were produced.

Post-Script

The team anticipates 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/hiddencollections/registry/hc.0880>

Citations

¹ *Connecting the Dots: Using EAC-CPF to Reunite Samuel Johnson and His Circle*, Harvard University and Yale University, accessed November 26, 2013,

<https://wiki.harvard.edu/confluence/display/connectingdots/>

² *DAT-040 DCA Best Practices for EAC*, Tufts University, accessed November 26, 2013,

<http://sites.tufts.edu/dca/files/2011/03/DAT-040DCABestPracticesforEAC.pdf>

³ *SNAC Prototype*, The Social Networks and Archival Context Project, accessed November 26, 2013,

<http://socialarchive.iath.virginia.edu/xtf/search>