**Connecting the Dots: Using EAC-CPF to Reunite Samuel Johnson and His Circle**

Submitted by:
Houghton Library, Harvard University: Susan Pyzynski, Krista Ferrante, Melanie Wisner
Beinecke Rare Book and Manuscript Library, Yale University: Ellen Doon, Michael Rush

Date: September 30, 2011, rev. October 18, 2011

Note:  To address the feedback from the Library Lab Review Committee we have added a final paragraph to the Background section to further explain how this project can influence future projects using EAC-CPF.

**Summary**
We propose a project to demonstrate the benefits of using Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) to describe creators of manuscript collections and encode meaningful semantic links between those creators themselves and the primary sources that document their lives and work.  The project will focus on lexicographer Samuel Johnson (1709-1784) and his circle of fellow writers, artists, political thinkers and friends, as well as collectors thereof.  EAC-CPF records will describe Johnson and members of his circle; document and make navigable their relationships; and provide links to the rich Johnsonian archival resources held by the Beinecke and Houghton Libraries.  The project will attempt to answer the following questions:

1) How can two separate institutions (Beinecke/Yale and Houghton/Harvard) collaborate to create, share, maintain EAC-CPF records and develop a set of metadata best practices for this new archival standard?

2) What is a content-rich EAC-CPF record?  What contextual information and linking will be most beneficial to our users?

**Background**
Adopted by the Society of American Archivists (SAA) as an official standard in January 2011, Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) is a new metadata transmission standard for encoding information about archival creators.  EAC-CPF was developed by an international committee of archivists and is maintained by SAA and hosted by the State Library of Berlin.  EAC-CPF records can include three basic types of information about creators of archival records: an entity's identity as expressed by the authorized form(s) of their name; description, narrative or otherwise, of the entity; and relationships to other entities or to resources by or about the entity being described, including but not limited to descriptions of archival resources and digital surrogates of those resources.

 EAC-CPF is an XML schema and was developed for use in an online environment; it is Linked Data-friendly in its ability to link to other online resources in semantically meaningful ways.  The standard was released to the archival community for use in March 2010.  Since its release

questions about implementation of the standard have concerned the archival community. Several stand-alone projects are happening on the international front, such as the *People Australia* project at the National Library of Australia. The most notable project in the U.S. is SNAC: The Social Networks and Archival Context Project, sponsored by the National Endowment to the Humanities (NEH). This project is investigating the "the feasibility of using existing archival descriptions in new ways, in order to enhance access and understanding of cultural resources in archives, libraries, and museums." The goals of SNAC are two-fold: 1) Create open-source tools to separate the process of describing people from that of describing records; and 2) Create a prototype system that will link descriptions of people to one another and to descriptions of resources in archives, libraries and museums and other diverse resources.

What SNAC is not doing is creating any best practices for what an EAC-CPF record should contain for optimal user discovery, nor is it investigating how EAC-CPF records are best created, maintained and shared long-term and in a collaborative environment. This is what the Beinecke and Houghton Libraries wish to test in this project.

The work that results from this project should illustrate what a highly-curated EAC-CPF record can contain, and the extent to which a set of interconnected EAC records can enhance research in a particular area. As North American repositories consider implementing a new mode of access to archival materials, examples – and accompanying best practices – will be a critical part of encouraging experimentation and adoption of the new standard.


**Project Plan**
Project members have determined that building a test database of 25-50 EAC-CPF records would be enough to investigate using EAC-CPF in a collaborative environment and develop a set of best practices. Samuel Johnson and his circle were chosen as the topic since Houghton holds the Hyde Collection of Samuel Johnson and the Beinecke owns the papers of Johnson's biographer and close associate James Boswell. There is a rich group of related contemporaries (David Garrick, Hester Thrale Piozzi, etc) from which to select entities to be described in our set of EAC-CPF records, as well as a common collector who donated much of the material to both institutions. The Beinecke and Houghton curators, Kathryn James and John Overholt, will use their specialized knowledge to work with the project team to define the group of people for whom we will make records, and help identify the archival resources to which we will link the EAC-CPF records.

We will use ICA-AtoM (http://ica-atom.org) as our database software. ICA-AtoM is web-based archival description software that is based on International Council on Archives (ICA) standards. ICA-AtoM is free, open-source software which has the ability to both import and export EAC-CPF records (the Archivists' Toolkit is not designed to export EAC-CPF). We still have to determine where the software should live and be administered from, Harvard or Yale.

Once we have selected our creator names for the EAC-CPF records we will start the encoding. To do this we will hire two 3-month temporary (casual) employees, one at Houghton and one at the Beinecke. These two project members will do the actual creation of the records, working with each other, plus the Houghton and Beinecke staff in creating consistent, standard EAC-CPF

records.  We are setting the minimal standard for an EAC-CPF record as having name authority information and at least on link to another EAC-CPF record and one link to an archival resource, but expect many of the records will be much richer.  The two temporary project members will also be documenting each decision we make (possibly using a wiki?).  We think that by working through all the issues in this collaborative fashion we will come up with a set of best practices that can be shared and used by others in the archival community.

Kathy Wisser, Assistant Professor at Simmons College, Graduate School of Library and Information Science, will be an outside consultant for the project.  Wisser was the chair of the EAC Working Group that developed the standard and now chairs SAA's Technical Subcommittee for Encoded Archival Context, the committee responsible for the standard's maintenance.  We expect to invite her to meet in person with the project members once or twice during the duration of the project and she will be available for email and telephone consultation on an as-needed basis.

After the EAC-CPF records are created they will be exported and sent for inclusion in the SNAC project.  Daniel Pitti, Project Director of SNAC, is very supportive of this potential Beinecke-Houghton collaboration and is interested in what we produce.

If practical, our final step would be to install/host an instance of the SNAC's prototype historical resource and access system (http://socialarchive.iath.virginia.edu/xtf/search), the code for which is built on XTF and freely available.  We could then, perhaps as a later Library Lab proposal, test and assess the functionality of our test EAC records with the research community.


**Timeline**
1st month - install and test ICA AtoM software, write job descriptions and post jobs for 3-month temporary/casual staff, determine names/persons for EAC-CPF records, select a source or sources of name authority information; plan the work for temporary/casual staff

2nd month - hire temporary/casual staff, begin training staff, set up documentation wiki, 1st in-person meeting of Houghton/Beinecke staff

3rd-4th months - create EAC-CPF records, midway in-person meeting with Kathy Wisser to review progress, document decisions and workflows

5th month – finish EAC-CPF record creation, compile and review documentation and begin creating a best practices document

6th -7th months  – final in-person meeting to review and refine best practices documentation, export EAC-CPF records and send to SNAC project, possible installation of a SNAC XTF instance for public access to our EAC-CPF records.

**Budget**

? hours of technical support and the ability to host ICA AtoM software on a web server/database server (MySQL)

? hours of technical support and the ability to host SNAC/XTF software on a web server

| | |
|---|---|
| 3 month (13 weeks) temporary full-time person at Houghton Library- | $13,000 |
| 3 month (13 weeks) temporary full-time person at Beinecke Library | $13,000 |
| Honorarium from Kathy Wisser's consultation/meeting time | $1,000 |
| Travel/transportation for 3 in-person meetings of Beinecke/Houghton staff | $500 |
| Contingency (supplies/XML software, $4^{th}$ in-person meeting, etc) | $2,500 |
| Total: | $30,000 |