Project summary-Final thoughts

The final picture: how well did the dots connect?

The intention of “Connecting the Dots” was to bring together two institutions in a collaborative effort to create EAC-CPF records. The focus was on Samuel Johnson and his Circle, broadly defined. Project staff deliberately pushed the boundaries of what might be considered a “circle” to facilitate the inclusion of records covering the three entities of corporate bodies, persons, and families. The project provided an opportunity to expand the chronological dimensions of Johnson’s circle to include modern collectors and scholars who were intimate with either Samuel Johnson or his biographer, James Boswell.

One of the first lessons learnt is that a circle is dynamic and difficult to contain. The original intention was to create fifty EAC-CPF records; the final count was seventy-eight. In a project with a three-month creative period, increasing the number of records by more than fifty percent did have an impact on the final product. In an endeavor that was primarily about testing the limits of EAC-CPF, a reduction in consistency across records was balanced out by the benefits of undertaking a thorough analysis of the standard. It is hoped that the imperfection inherent in experimentation will help others achieve more perfect results. With regard to connecting the dots, it soon became clear that we could make connections between nearly infinite numbers of points; we struggled to maintain a circle against the inclination to weave a web.

Caveat creator: Our best practices changed.

Throughout the project, we maintained best practices; however, as we went along we changed our definitions of best practices. Therefore, the final best practices are not the same as those we followed. In addition, since the creation of content-rich records is time consuming, some of our records are extensive, while others may be regarded as stubs.

Notes, observations, and conclusions regarding the creation of EAC-CPF records.

The rationale for choosing Samuel Johnson and his circle was twofold: several individuals in the circle were represented in the collections held at Harvard and Yale, and we presumed that the group would be interconnected and self-contained. The coherence of the group should have made it possible to improve efficiency, but we did not find a suitable way to do this. When we recognized that process of finding and inserting cpfRelations and resourceRelations into EAC-CPF records was time consuming, we considered creating a centralized database or other system for tracking and sharing relations. For a project where there is a discrete boundary, creating such a system at the beginning might save time; this approach would have made it easier to include a full range of relations for each record, but it was not feasible for us to act upon the idea.

Major Decisions.

Regarding specific points of discussion, the most persistent challenges came from our discussions of the elements: places, biogHist, relations.

<places>

The <places> element was a challenge from the beginning because of the stipulation that a place not be listed twice. This raised the question of which placeRole should take precedence: within a hierarchy that usually included birth, employment, and residence, we identified residence as the most inclusive description. Nonetheless, we considered that it would be desirable to assign multiple placeRoles to a given place to accompany the allowed element dateSet. We also discussed how granular we should be in describing specific regions within a city: Greater London, for instance, consists of thirty-two boroughs. Rather than spend the time identifying the specific borough in which an address lay, we chose to use London (England), except when a specific borough was identified.

Where we did opt for greater specificity was with institutional names represented in LCNAF: churches, schools, universities, etc. we identified by name rather than the town in which they existed. This decision not only allowed for greater specificity, but it also allowed us to apply a larger number of placeRoles: for example, “education” might be the role of the University of Edinburgh, while “residence” might define the city itself.

We considered omitting date from our list of places, but decided that this would reduce the relevance of the place element. The problems that listing a date posed are illustrated in the case of William Adams. As a cleric, he was assigned to several parishes at the same time. His associations with a given place did not necessarily mean a physical association; he might have spent a few days in Llandaff each year, but it is impossible to know how he parsed his time. In the case of Richard Savage, we know that he relocated to Richmond when he enjoyed some degree of financial success, but it was not clear precisely when that was or for how long he remained there. In other words, exact dates can be elusive and an association with a place can sometimes be important on paper (as with Adams’s appointments) without representing time actually spent in a place.

<biogHist>
The two enduring challenges we faced in populating biogHist were subjectivity and how to list creative works. While absolute objectivity may be impossible to achieve in any endeavor, the focus on Samuel Johnson made subjectivity a particular challenge in this project. Even written into the guidelines was the suggestion that the first paragraph mention the subject’s connection to Johnson, if any. In some cases, the relation to Johnson would have been tangential to the biogHist in most instances, but received primacy in our records because of our focus. This will no doubt be true of other projects in which EAC-CPF records are created around a particular local interest.

Questions about how to list creative works posed a practical problem. We initially listed these within a comprehensive chronList, and later created a chronList that was separate from that which outlined life events. We abandoned the two chronList approach because of the lack of `<span>` within the chronList, which we wanted in order to designate titles as “font-style:italic”. Within best practices documentation, we recommended providing a `<list>` of works in order to make use of `<span>`.

The categories of relation we used were cpfRelation and resourceRelation; functionRelation was not relevant to this project. While each presented unique challenges, the largest questions had some bearing on both.

One of the greatest challenges of the project came when we tried to grapple with the attribute xlink:arcrole. Creating a local vocabulary to describe arcrole for associative relations proved to be a struggle, especially in cases where there was no known correspondence or collaboration. After considerable time and debate, we prohibited arcrole because we lacked an RDF vocabulary with URIs for each term and creating such a vocabulary was not feasible within the scope of the project.

Even before we abandoned arcrole for practical reasons, it proved unworkable. We attempted to create a local vocabulary for cpfRelation and resourceRelation that was accurate but also broad enough to be widely applicable. In the process, we questioned the utility of our local vocabulary designations. The difficulty of achieving these aims is clear from the working list we compiled.

For "associative" alone we had: associatedWith, correspondedWith, collaboratedWith, colleagueOf, criticOf, stagedPerformancesFor, stagedWorksBy, supportedBy (financially), supporterOf, tutorOf, workedFor, workedWith, worksPublishedBy, writtenAboutBy.

The process of attributing an arcrole for associative relations revealed that some relationships were both associative and hierarchical; for instance, relations such as tutorOf and workedFor suggested a hierarchy.

In the end, we discarded “hierarchical” as a description of the relationType attribute because none of the relations was purely hierarchical and there was a strong case for each to be termed “associative”. We initially designated relations within and to a corporateBody as “hierarchical”, but decided that this would be more appropriate in corporate records where there is a departmental levels rather than in our case where the hierarchy itself was unclear, such as with the Club. In summary, our abandonment of “hierarchical” was not due to any limitations of it as a value; it was because “hierarchical” was not applicable to the records we created.

In the early stages, the local vocabulary we tested for “hierarchical” relations included: referencedIn (refers to manuscript collections), belongedTo (as in a person with relation to a club, or Francis Barber as a slave), collectorOf (referring to collectors of the manuscript materials), employedBy, memberOf, managedBy (i.e. Drury Lane), meetingPlaceOf, ownedBy (as a resourceRelation for collectors, but could apply to Francis Barber), proceededBy, precursorTo, succeededBy.

Beyond the dilemma of assigning an arcrole, the local requirement of including a descriptiveNote created its own complications. The description was supposed to indicate the link between the cpfRelation and the subject of the record, but at times that very relationship was imbalanced in a way that made phrasing awkward. The entry for Richard Savage in the Alexander Pope record reflects the awkwardness of passive construction: “He was the recipient of Pope’s generosity: Pope solicited funds to help Savage retire in Cardiff.”

With regard to the attribute resourceRelation, we designated each entity as as “creatorOf”, “subjectOf” or “other”, according to the rules of the EAC-CPF tag library. The resourceRelation was supposed to be “creatorOf” if the record subject was either the author/creator or recipient of an item. However, there are cases in which a collection contains numerous letters that merit the designation “subjectOf” and only one to which “creatorOf” applies; in this case, we still applied the “creatorOf” description since it has priority over “subjectOf”. Another challenge was whether there should be a threshold for inclusion: is the appearance of a cpf entity once within a collection enough to warrant inclusion as a resourceRelation? Similarly, what is a realistic level of description of the source? Should there be a descriptive note explaining inclusion beyond “other” in order to justify the value “creatorOf”, for instance when there are scattered letters to or from a cpf entity within a larger collection?

Time was a tremendous factor in limiting our work. We had hoped to include links to sources held at several repositories but in the interest of time focused primarily on local databases: OASIS and the Yale finding aid database. Even on a local level, this approach was limiting. For instance, Yale’s Lewis Walpole library has holdings that related directly to this project, but because their collections are not represented in the finding aid database, they were omitted from this set of records.

Finally, since the bibliographic information we included changed over the course of the project, here are some records for which resourceRelations are not in strict alphabetical order: they are in order according the name of the collection, but the addition of author/title information means that not all are in order according to bibliographic standards.

role:

The "role" sub-element did not pose the same problems of definition as arcRole, but we abandoned it at the same time as we abandoned arcRole, again because of a lack of RDF vocabulary.

Minor Decisions.

In addition to the many involved discussions that were an ongoing part of the project, we collaborated to answer many minor procedural questions that arose in the process of record creation. Among the elements we addressed were: sources, entityType, names, languagesUse, and structureOrGeneology.
sources:

Over the course of the project, we wondered how to handle non-perm links. Our final decision was that we would include current (but non-permanent) links in addition to full bibliographic information.

names:

The decisions we made concerned pseudonyms and which entries to include from LCNAF. For pseudonyms, we decided to create an entry according to AACR2 standards where an entry was lacking in LCNAF. For name entries generally, we first considered including only those 4xx names that were in Latin script but decided it was best to include all 4xx entries. For entries not listed in LCNAF, our future best practice would be to create an entry following AACR2. For instance, since the DNB provides a variation on the name of Edward Burney (Edward Francisco Burney rather than Edward Francis Burney), we would consider it best practice to acknowledge the DNB entry with an AACR2 name entry.

languagesUsed:

Initially, we took this to mean what the tag suggests: languages used by a person. Interpreted broadly, we listed languages that we knew an entity had learned. According to the EAC-CPF tag library, it should apply only to languages in which an entity was productive.

dentityType -- person as corporateBody:

Persons who are also corporate bodies may need another look: Dodsley and Millar were both people but published under their names as well. We found that this can complicate roles and relations, but did not have the opportunity to draw firm conclusions as to future approaches to take for entities that can be both person and corporateBody.

General context:

Initially, we intended to include general context notes on: book collecting, bluestockings, politics, publishing/printing, salon culture, and slavery. In the interest of time, we abandoned those entries and never did decide how general or tailored to make them. We had considered that they would be most useful if the notes were general to the time rather than specific to Samuel Johnson and his circle, since that would allow them to be used/exported for other purposes.