I. Introduction

The Social Network of Early Modern Collectors (SNEMC) Project is an effort to document and represent the exchange of curiosities in Early Modern England. It is the latest phase of the “Culture of Curiosity Project,” which is concerned with an examination of the cultural discourse and literary representation of early modern collections of rarities and curiosities in England and Scotland, 1580-1700. Central to this research is the Digital Ark, a web archive of some fifty XML-encoded catalogues, inventories, and records of these collections keyed to a project database of mentioned people, places, and bibliographic items. Our focus in phase two of this programme of research is the social network of collection: the people who facilitated the creation, manipulation, and exchange of objects and how they acquired them (gifts, purchases, discoveries, etc.). In tracing these events of exchange, we ask such questions as: who were the people involved in these collecting networks? What was their relationship to these objects and what were their roles in the production, acquisition, and exchange of these objects? And then, what does all this tell us about the function of collected objects in the cultural and intellectual practices of these social networks? To answer these questions, we are modelling and developing new applications of computer-assisted network visualization to the textual materials related to
these collections. In doing so, we want to bring the object to the fore of our analysis, while at the same time situating the object in its social context.

This paper details the process and reasoning by which our research team arrived at a data structure that serves the visualization needs of the project. We discuss SNEMC in the larger context of the Curiosity project, identify the aims and priorities of the project and how these priorities create certain needs and implications for our data structure unique to this project. We will then look at how a principle-driven methodology informed our deliberations between the various available tools, vocabularies, and perspectives in the project’s early development. Not only did this outlook indicate the broader aims of our work to our project members; the clarity afforded by this approach has helped us avoid potential pitfalls that could have arisen later in the pipeline. Finally, we conclude with a reflection upon how this process informed our largest shift to an event-driven relationship model and the implications of such a principles-driven methodology for DH projects that depend upon well-crafted data structures.

II. Aims and Priorities

The aims and priorities of the project derive from our goal of documenting and better understanding the social network of early modern collectors as represented in the various historical records of their transactions. Specifically, we are interested in two kinds of relationships: relationships between people and objects, and relationships between people involving objects. Given the nature and preparation of our data, the best method for this analysis seemed to be some sort of network visualization. In the first instance, our preparation of materials and our chosen methods and tools must serve our own objectives, but we will also want both our findings and our dataset to be useful to and adaptable by other scholars. Thus, we began with these as our informing principles.
A. Our methods must be object-oriented and object-centred

At first, an object-oriented dataset may seem counterintuitive given our aim to track the relationships among collectors; however, it is the personal relationship to the object that defines the particular nature of these interpersonal relationships. Moreover, the historical documents themselves, from which we identify these relationships, are object-oriented: they are catalogues, inventories, and narrative descriptions of collections and collected objects. In fact, sometimes the naming or description of an object is all a document explicitly states, and its implicit relationships to people need to be inferred. The object is also typically a catalyst for or even cause of relationships between the people mentioned in these documents. The donation of an object might be the only point of contact between a local citizen and collector. The act of donation might be incidental or opportunistic, or it could be part of a more deliberate act of cultivating and solidifying social connection. In many ways, the object influences if not determines the nature of the relationship and the interactivity between people mentioned in these documents. As an example:


Figure 1. Example from Nehemiah Grew’s Musaeum Regalis.

Here we have the mention of an object from the Royal Society’s collection recorded by Nehemiah Grew detailing the passing of a kidney stone—here taken to be a bone—by one Sir Throgmorton. That kidney stone was then acquired by Thomas Cox (perhaps directly from Throgmorton, or through his surgeon or some other intermediary), who gave the curiosity to the Royal Society, where Grew made this record while observing the collection. In this instance
then, we have evidence of four persons or corporate bodies connected through a single object from its production to Grew’s record. The object is integral to these interpersonal connections.

B. Our visualizations must feature and focus on the object

In order to incorporate objects as core entities in a network of relationships, we need to change how relationships are typically graphed. What we are after is not a simple indication of a relationship—a friend of a friend—but of actions and interactions involving objects as constitutive of those relationships. We need to be able to put the object into the network graphs along with, or rather, in between people. This is not how similar projects tend to represent their data. Shakeosphere, for instance, with its social network of Authors, Printers, Publishers, and Booksellers—whose relationships are defined by bibliographic objects they have in common—tells the user at a glance that between 1622-23 William Shakespeare had three printers: Nicholas Okes, Thomas Purfoot, and William Stansby. However, such a visualization relays no information as to the number of books between them and the author or what those books were. While this makes sense given their project’s priorities, our own project seeks to highlight more explicitly the role objects play in interpersonal relationships.
C. Accessibility and Linked Open Data (LOD)

Once these connections and their relationships are captured, however, how can they contribute to our scholarly understanding of networks of early modern collectors? A dataset as large as ours—currently thousands of relationships noted in some fifty documents comprising about 1M words of text—must both allow access in a way that facilitates the navigation of such a large amount of information. Just as a library is practically useless without a catalogue, our dataset will contribute little to scholarship without the tools to navigate it.

Beyond giving users the ability to investigate our data according to our parameters, we want them to have access to the raw data and its ontology so that they might either expand or incorporate it into their own dataset for their own purposes, or adapt our ontology to suit their own data. This level of interoperability has long been a focus of web developers:

Central to the development of the Semantic Web has been efforts to design and implement domain-specific ontologies to provide “a lexical or taxonomical framework for
knowledge representation ... [that] can be shared by different information systems communities” 
(Ramli and Noah 1154).

The key to this level of interoperability in our project is Linked Open Data (LOD). Prioritizing LOD allows other scholars to use the dataset(s) we create and adapt them to their own needs. It also provides opportunities for contributing to other projects and collaborating in a more seamless way than via mediated data transfers. In choosing our ontology, we strived for a balance between serving the project’s unique focus on the object while adhering to existing standards.

IV. Adapting and Creating A Principle-Driven Ontology

With our principles established, we could now evaluate the ontologies and models available to us according to criteria directly related to the project. These principles informed our selection but also any modifications we might need to make for a given ontology to fit the project. Thinking through how these models aligned with our principles led us from an ontology centred on objects to an object-oriented, event-driven model based in Web Ontology Language (OWL) and the N-ary relations model developed by the W3C’s Semantic Web Best Practices and Deployment Working Group.

A. Freely Available Ontologies

One of the first real challenges our project faced was determining the vocabulary that would structure our dataset, as these decisions would have long-lasting effects on the possibilities and affordances in visualization of our data as well as its accessibility and interoperability. While the preparation of our dataset was initially informed by our project’s own priorities and aims, it proved helpful to position our methodology within the larger fields of historical research and the Digital Humanities. Many of the studies we looked at came to the same conclusion: expanding an existing ontology with a local metadata schema unique to our dataset was the most efficient
approach and the most likely to encourage “application interoperability both on the syntactic and the semantic levels” (Ramli and Noah 1156).

Given the principle that making use of existing ontologies, when possible, is preferable to developing one’s own, we looked for similar projects in related fields to determine what tools and ontologies might already exist for our own use. However, available ontologies did not easily align with our project’s objectives and the nature of its material. An important consideration in looking at existing ontologies related to historical objects is that our focus is on representations of objects in historical documents rather than the objects themselves (most of which objects are now lost). This is distinct from many ontologies in use by museums, such as CIDOC’s Conceptual Reference Model for example, which are principally concerned with the present, real world object itself and its material properties. As a result, an ontology such as CIDOC-CRM, designed “to mediate between institutions such as museums, libraries, and archives,” has a sweeping ontology concerned with objects present and accounted for in existing collections; an ontology both too broad and still lacking certain important vocabulary we need concerning the social acts and activities of collecting.\(^1\) The Europeana Data Model (EDM) has distinct, but similar problems. We found EDM’s incorporation of digital resources enticing but the ontology is largely concerned with connecting digital surrogates (e.g. a scanned image) to an existing real-world object. This is again different from our own project, which aims to link mentions of often lost real world objects from multiple documents and prioritize those objects’ role as artifacts of social exchange. We concluded that EDM would not be an appropriate model for our own dataset.\(^2\)

---

2. See the Appendix for a fuller discussion of these and other object-oriented ontologies.
Given our commitment to adhering to the principles of Linked Open Data, the results of our literature survey and our own investigations all suggested that developing a new ontology and implementing it through a version of Web Ontology Language (OWL) using the Protégé ontology editor to create and test our ontology structure would be the best approach.³

B. Adapting Our Ontology – An Event-Driven Model

Perhaps the most significant revelation throughout the development of our dataset was the realization of what precisely it means to be object-oriented in our project and our consequent move toward an event-driven model. Our original intent had been to put the “GoPro” on the object—to follow and describe each object’s path and interactions as an artifact of social exchange—but there was some ambiguity about how exactly one might reflect such an intention in a data structure. In looking at other projects, we realized that an event-driven model would allow us to describe an object’s interactions and path from its perspective while clarifying the various roles of different entities in each event. Our conversations led to an N-ary event model as outlined by the W3C working group (2006). An N-ary model simply denotes the number of different entity types in that model; as defined by the working group “in some cases, the natural and convenient way to represent certain concepts is to use relations to link an individual to more than just one individual or value.” In our case, we use a ternary model with three entities: objects, people, and events. It is an “event model” because, with few exceptions, all other entities relate directly to an event as a means of articulating their relationships to one another.

One of the W3C Working Group’s use cases outlines a similar scenario to those we commonly encounter:

³ Protégé has the added benefit of being able to export to rdf to help with the LOD part of the equation.
The purchase event (Purchase 1) links buyer, seller, object, purpose, and amount to a single set of relationships rather than force the creation of many separate relationships that would make a cohesive expression difficult if not impossible. Similarly, as we have mentioned, our dataset often includes more than one person interacting with an object for multiple reasons. For example, as outlined in Figure 1., Grew, Throgmorten, and Cox, are involved in the exchange of an object through multiple means: “voided”, “given” (and ultimately received). The development of the final RDF via the OWL ontology remains a work in progress; we want to make sure the events we create address all of the research aims we have set out before we make it publicly available. However, the W3C’s work on N-ary relations has given us the ontology patterns for representing these types of relations or events in RDF and OWL. Defining the events, therefore, becomes vital to the success of the approach we are taking.

An “event” in this context refers to something that occurs over a period of time at certain places and acts upon or with other entities (Pustejovsky et al., 2003, p. 3). In the case of our project, developing events as an entity type allowed us to clear up certain ambiguities in the
relationships between other entities (i.e. objects, people) and squares with the n-ary relations framework. The most straightforward example is that one person might have multiple roles in relation to an object. For instance, a person might be the recipient of an object but then donate it to someone else. Rather than directly linking one entity (object) to another (person), our event-driven ontology allows us to describe two interactions between the entities that clarifies their relationship and allows us to load these distinct events with related information (e.g. the other donor and recipient, location, date, etc.).

Additionally, an event driven model became essential for the many cases where more than two individuals form links to an object: for example, an object might have not only a donor and recipient, but a courier between them as well. One of the key advantages of an event-driven model is its ability to represent sophisticated relationships among more than two individuals where the instances of properties in RDF can only directly link two individuals. But, in addition to allowing the expression of multiple relationships between the same entities and linking multiple individuals to the same object, events also allowed us to capture more granular attributes by loading them onto this new entity. We are able to clarify subclasses for acquisition (undefined, donation, sale), production (natural or artificial), manipulation, and observation that then have additional types that we need to define. For example, an acquisition could be a donation or a sale (or something else); the object’s production could be artificial or natural; and the manipulation or movement of the object could be related to transportation or extraction. The RDF expression would include, for example, the Acquisition event with one of several subclasses including Donation. And all these details are important to capture in order to record and visualize the richness of the connections between objects and people.
If we return to our example of the kidney stone from Nehemiah Grew’s catalogue, we can see the advantages in visualization through an event-driven ontology. First, there would really be no way to suggest these relationships in any meaningful way without the object as an entity in the network. Without the “bone,” we might be able to link Cox to Throgmorton and the Royal Society, but Grew’s relationship to those three entities becomes more difficult to articulate as he only ever really interacts with the object itself.

![Figure 4. Object-Oriented Representation](image)

An object-oriented model links each person to the object (i.e. the bone), but conveys limited information and does not allow for the representation of more than one type of relationship between two entities (figure 3). The event-driven model allows for a greater degree of depth (figure 4). For instance, we can see not only that Sir Throgmorton was involved in the production of the bone, but gave that object to Thomas Cox, who passed it along to The Royal Society. Where in the first model we have a static, nondescript association of individuals with an object that describes a single role, the event-driven visualization presents the object as the central entity in a series of interpersonal exchanges that allows for multiple relationships between persons and objects.
Figure 5. Event-Driven Representation

Implementing these events required careful consideration as to how to integrate the new “event” entity into our current dataset. The project was initially developed using a number of database tables for people, objects, locations, and, now via an editor we have built, yet another one for the events we have established.

Determining the events required much discussion and revision; we needed to be project-specific enough to properly describe different events while considering ways to minimize the number of terms and adhering to standards when possible. Too much customization and noise makes for poor visualizations and complicates the markup/event creation process. Conversely, the project has been encoding texts for years and a complete rework of the data was out of the question. There needed to be a balance between complete customization and full adoption of an existing ontology.
Our resulting ontology for event entities is relatively sparse. In order to keep our ontology focused and work-flow streamlined, we settled on a select number of event classes: Acquisition, Production, Manipulation, and Observation. For greater specificity, we also implemented a small number of subclasses (e.g. artificial vs. natural production). As we finalize this ontology, we feel that we have struck a good balance between a level of specificity that serves the needs of the project and the documents we represent and a generic enough vocabulary to adhere to existing standards and conventions. The next step is the processing of our data and creation of events. For this, we are developing an “Event Builder.” But this will have to wait for another presentation.

Figure 6. DigitalArk Event Builder Prototype

V. Conclusion

Codifying and clarifying our project principles and aims before structuring our dataset led to important revelations about the project’s needs and the consequent decisions our team made
benefited greatly from it. The most drastic change our principles-informed methodology led us to was the realization that an event-driven ontology suited our needs better than sticking to a vague notion of being object-oriented. In effect, we realized that the structure of our documents are already object-oriented and, rather than merely mimic that, we needed to implement a structure that would allow us the freedom to emphasize and clarify the interactions these objects played a role in. Treating objects as a primary entity would have, at best, necessitated loading our object-entities with a large amount of static data. Instead, the event-driven model carries the load of that data in a way that allows for greater specificity and flexibility, making the object an active agent that can participate in multiple interactions with other entities.

**CIDOC.** Conceptual Reference Model. [Cidoc-crm.org](https://cidoc-crm.org).


N-Ary Relations, W3C, [https://www.w3.org/TR/swbp-n-aryRelations/#Note](https://www.w3.org/TR/swbp-n-aryRelations/#Note)


Protégé. [https://protege.stanford.edu/](https://protege.stanford.edu/)


*Shakeosphere: Mapping Early Modern Social Networks*. Created by Blaine Greteman and David Eichmann. Iowa City: University of Iowa Libraries.  
https://shakeosphere.lib.uiowa.edu/networkAnalytics.jsp

W3C’s Semantic Web Best Practices and Deployment Working Group.  
https://www.w3.org/TR/swbp-n-aryRelations/#Note