Curating Object-Oriented Collections Using the TEI

Brent Nelson

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.
This paper aims to contribute to the ongoing discussions of the Ontologies SIG about development of the TEI for treating objects identified in historical documents. In this paper, I provide a case study of a digitization project that is object-focused and poses challenges to any attempt at a standard data structure for identifying and describing objects and the circumstances surrounding them. The project is The Digital Ark, a web-delivered virtual museum of collections of rarities and curiosities in England and Scotland from 1580 to 1700, comprising documentary and graphical representation of up to 10,000 specimens and artifacts collected in that period, some of them surviving in museums in England today. These collections were heterogeneous, typically containing both natural specimens and artifacts, as well as objects that complicated this binary distinction: anything that aroused curiosity. Many objects were of interest for their strange or exotic qualities; others for their ingenuity or singularity. Objects were both foreign and domestic: Greek and Roman as well as Anglo-Saxon antiquities; naturalia and artifacts from the Far East or the New World; but equally, strange fish or animal remains found locally. These collections thus
represented the sum of early modern European knowledge and experience of the world in a time of rapid scientific and geographical expansion and reflected fundamental epistemological shifts in attitude toward curiosity, wonder, and credulity on the cusp of the modern age.

The documentary representation of these collections is similarly heterogeneous. Some of these objects are extant in modern museums, such as the Ashmolean, which was founded on the collection of John Tradescant the Elder (1570s–1638) and Younger (1608–1662). The cabinets belonging to John Bargrave (1610–1680) are also largely intact at Canterbury Cathedral. In a few cases, then, we have modern documentation in a structured form produced by museum specialists. Most of the collections, however, have been long since dispersed and are represented to us only in the seventeenth-century documentary record. These materials take the form of diary entries; inventories, lists, and catalogues; personal correspondence; and in a few cases, pictorial illustrations, such as drawings or prints.

Thus far, efforts to represent references to the object world in the historical record have focused on the question of ontologies and on mapping the TEI to the CIDOC standard to facilitate the sort of object descriptions used in the museum world (Ore and Eide 2009; Eide 2014–15; Ciula and Eide 2014), sometimes in explicit application to museum catalogues (Eide and Holmen 2006). Less attention has been paid to the way in which we might deal with objects mentioned in historical documents. In the wake of thing theory (Brown 2001), one might imagine a scholar wanting to mark objects in fictional or historical literature in a way that is analogous to entities already well supported by the TEI Guidelines—people and places. In what follows I use examples from the Digital Ark to model some key functionality for curating textual representations of material objects on the web. Focusing on metadata, Øyvind Eide posits two approaches to modeling ontologies: using existing models to integrate data from different sources; or studying and analyzing a domain of interest to infer a new ontology (2014–15, paras. 9–10). My approach is somewhere between these two: I am beginning with an existing model (the TEI) to see how far it can go in describing the sorts of representations one encounters in a particular (but, I think, illustrative) domain of interest related to the object world.
1. Web Curation

The Digital Ark project bridges the object world of the museum and the documentary world of the humanities scholar. Because we are dealing closely with the world of objects, we need to shift our point of reference somewhat from thinking about textual materials. A natural point of reference is the museum and museum practice. Therefore, I would like to frame this process by analogy as “curation.” The activities of curation (from the Latin cura, to care, to manage) apply aptly to the web, where digital objects are proliferating. The semantic limits of curation have been stretched in recent years in the context of Web 2.0 social practice, where personal content is communicated, selected, collected, and packaged for sharing with others on Facebook, Pinterest, or any number of other social media platforms. The language of curation has also crept into textual studies and editing as we produce, communicate, select, and present texts in strategic ways within a digital archive. But of course, the traditional home of curatorial practices lies in the Galleries, Libraries, and Museums (GLAM) sectors. The digital age has facilitated a rethinking of what constitutes the work of curation in this context. As in the humanities, digital tools have revolutionized museum theory and practice. While the museum is a convenient, and proper, place to begin a consideration of what constitutes curation in the web environment, I am doing so particularly from a humanities scholar’s perspective. There is a large and growing body of literature on web curation from the museological perspective, but the needs and interests of the museum curator and the humanities scholar do not entirely coincide.

Many of the activities involved in material curation in the physical space of a museum also apply in the context of building a digital archive of materials related to early modern collections. The editor/curator is tasked with managing the collection by locating, collecting, documenting, preserving, and conducting research on the collected materials. These activities apply as well to digital objects (in the case of the Digital Ark, images, metadata, and XML-encoded transcriptions of historical documents, as well as images and metadata on extant objects) as they do to the management of material collections. These are the in-house activities of collections management. Then there is the public face of the museum exhibiting or presenting the material to its potential users. This act of curation involves selection of materials, contextualization, and presentation. In the pre-Internet days, and still today in physical collection, selection was and is a function of limitation. Most collections cannot display all their materials at once in a way that can be processed
and appreciated by their public users, so careful choices must be made. In the case of a special exhibition, a theme becomes a principle for selection. On the basis of the theme (or other principle of selection), a curator makes meaningful groupings of objects and articulates the relationships between them based on that theme. These groupings, together with other, framing materials, provide a context that helps the user make sense of these objects.

In the digital environment, limitations of space and (to some degree) resources do not apply with the same force. In a digital space, all objects (i.e., surrogates of objects) of a large museum can, in theory, be stored, located, selected, retrieved, and even packaged for presentation with relative ease. These are functions that computers are very good at, if the materials are appropriately structured and prepared. In what follows I focus on the function of selection. Indeed, I want to argue that the ability to select is the core function of curating a digital archive, facilitating the ability to identify, locate, collect, and then manipulate objects for study. A necessary corollary in the context of historical materials is the ability to correlate selected objects with their documentary contexts.

2. The Text Encoding Initiative and a Document-centered Treatment of Objects

In the context of the Digital Ark, from a humanities perspective, our interest in documents is defined by their object-orientation: these are documents that are explicitly and in a dedicated way about objects. That said, from the perspective of the study of material culture, our approach to objects is decidedly document-oriented, in a way that the approaches of museum, gallery, and even archaeological professionals are not. Our primary interest is in the representation of objects, rather than the objects themselves. Indeed, in the case of the first museums, the number of objects represented in the historical record far exceeds the number of objects that are extant: the vast majority of these contents have been lost in the intervening centuries. We are also interested in the “subject–object relation”: how an object-world was perceived and received by the historical actors associated with it, as represented in the historical record (Brown 2001, 4 and passim). In the museum context, information related to objects typically comes as metadata in a structured form, such as a catalogue entry. The materials of the humanities scholar (letters, diaries, travel
narratives) are typically unstructured data, commonly discursive, often in continuous prose, but sometimes also in fragments. For this reason, the TEI is a good place to begin looking for metadata solutions for curating object-oriented documents and texts.

8 At this point, I need to return to the ongoing consideration of what the TEI might do to support representation and documentation of material objects. In recent years, the direction adopted by the Ontologies SIG has been to explore how far the current guidelines for representing the manuscript as material object could be extended to handle other kinds of objects: in the first instance, text-bearing objects (for example, a vase), but also other kinds of objects. This means that the first efforts are being applied to the header module of the TEI Guidelines (P5). What follows is an attempt to use an object-oriented case study to see how far the TEI as it presently stands could be generalized to apply certain Manuscript Description (<msDesc>) elements more broadly in a document to account for references to non-text-bearing objects. Although most of these elements belong to the Manuscript Description module and are intended for header metadata, some are also allowed within <body> in the context of elements that are commonly used for structured prose: <p>, <ab>, <seg>, or <list>.

9 The first consideration, then, is how generalizable the manuscript object is and how a TEI treatment of it maps onto other kinds of objects. There is a great deal that does not apply to other kinds of objects: catchwords, signatures, and watermarks, for instance, are all specific to manuscript and printed books. The entire sections on heraldry (10.3.8) and rubrics, incipits, explicits, and other quotations from the text (10.6.3) apply only to manuscripts. There are, however, several elements that could be used to treat objects of any kind, and others could be employed if the rules governing their application could be modified:

- <material> “contains a word or phrase describing the material of which the object being described is composed.” Of course, all material objects are composed of something.
- <desc> “contains a brief description of the object documented by its parent element.” This is a better, more generalizable option than <objectDesc>, which is limited to text-bearing objects and must be contained by <physDesc> within a header.
• <dimensions> contains a dimensional specification. Dimensions can be further particularized with <height>, <width>, and <depth>, or more generally represented with <dim>. Similarly, the @quantity attribute applied to <dim> would be useful in cases where more than one object is concerned (e.g., “ten shells”). Another option for recording quantity of objects is <num> with @value.

• <objectType> “contains a word or phrase describing the type of object being referred to.”

  This element could be useful for simple statements of a type or class of object: “a dog,” “a necklace,” “a monstrous fish.”

• <origDate> “contains any form of date, used to identify the date of origin for a manuscript or manuscript part.” A better treatment of dates associated with objects, however, might be the more general <date> element with an array of values for @type indicating, for examples, date of “origin”, “exchange”, or “creation”.

• <trait> “contains a description of some status or quality attributed to a person, place, or organization typically, but not necessarily, independent of the volition or action of the holder and usually not at some specific time or for a specific date range.” Of course, this could apply equally well to a material object, but as suggested above, the more general <desc> is preferable regardless.

This much applies to the object itself, but there are also elements that pertain to relevant context and circumstances: not just the what (the object itself), but also the who (<persName>), where (<placeName>), and when (<date>). It is also possible to associate @role with <persName> as with <placeName>. All this adds up to a structuring of an <event> in the life of an object; however, <event> is intended principally for use in the header within <listEvent>, and therefore its application is awkward within the body of a document. It must be nested in <listEvent>, but in the sort of documents I am modeling here, there is no list, only a mention of a single event. It also must contain a <label>, and therefore its application is not as nimble as <name>, for example. An alternative to <event> is <incident>, which can be applied almost anywhere within the <body>, but the semantics of this element are not a good fit: “any phenomenon or occurrence, not
necessarily vocalized or communicative, for example incidental noises or other events affecting communication.” For documenting an event, then, one might best resort to a stand-off solution, either in the header of the TEI document, or in RDF, or in a database.

An important difference needs to be observed in how these elements might be applied in the body of a document rather than in the metadata of the header. In the header, the values for these elements can be supplied by the encoder. In the body of a document, there must be strings of text already present to which these elements can be applied. In keeping with the general principles of the TEI Guidelines, my intention for the body of the document is, as much as possible, to mark the content as presented in the original document, to let the document’s structure speak for itself. I want to capture, in the first instance, the way in which the document represents, expresses, articulates, and categorizes the objects it references. As we will see, this distinction leads to some complications in how we model the markup of mentioned objects in a document.

Now let us see how well these apply and to what degree they account for the qualities and circumstances of a mentioned object. The following is a fairly complex example of a documented object from an early modern collection of curiosities. It comes from a seventeenth-century collector’s manuscript catalogue of his own collection. The collector was John Bargrave (1610–1680), a canon of Canterbury Cathedral, who amassed a modest and very personal collection, much of it acquired during his several trips to Europe, especially Italy. This is one of the more complicated object descriptions in his manuscript:

50. Item, a cravat, a shass or girdle, and a small pair of gaiters of curious work, by the inhabitants of the north-west (whether passage or no passage) of America, in the West Indies, made of porcupine quills very artificiously. In Italy there are butchers’ shops particularly for venison, in which shops are every week hanged up store of these porcupines; but we foreigners did not much approve of the meat. The cravat, &c., with divers other things, were sent me by one Mr. Tymothy Couley, now a marchant in London, by way of gratuity, he being one of the 162 slaves that I redeemed from Argeers, when I went thither by King Charles 2 commission and 10,000lb of hierarchical money, 1662, for that purpose. Amongst the chains of the redeemed I kept only this man’s, which I have now by me, and intend to have it hanged up over my grave in memorandum.

(Canterbury Cathedral Lit. Mss. E16a, f. 70)
There are various names mentioned, referring to the actors in the event, and there are roles implied—Bargrave as rescuer of Couley and recipient of Couley’s gift; Couley as gift-giver; King Charles II as financier—that could be represented with @role in association with <persName>. Still on the theme of <event>, we have a <date> for the rescue and, implicitly, for Bargrave’s acquisition of the objects. There is no date of origin for the objects, but we could approximate one for this set of artifacts (mid-seventeenth century), though there is no text to that effect. With respect to place of origin, we could use <placeName> to mark “the north-west (whether passage or no passage) of America, in the West Indies” with a specification of @type or @role as <origPlace>. There is clearly an <event> alluded to here—in fact, two events: the freeing of slaves and Couley’s gift of this set of artifacts to John Bargrave in gratitude. In this case, one might nest these two events in a <listEvent>, which makes some sense, given that they are related events; but a list implies parallel items, while in this case, the depiction of the event of gift-giving encompasses the mention of the rescue, so that there is at once a contained and a subordinate relationship, where the rescue is mentioned as a cause to explain an effect (the gift-giving). One can also imagine cases where an event occurs in the context of another, encompassing event, which would imply a nesting of <event>s.

With respect to the objects themselves, <objectType> could be useful and appropriate for marking the “shass,” “girdle,” and “gaiters,” as well as the “chains”; however, in early modern documents concerning curiosities, object type is not always stated. In many cases, it is not clear whether a mention of an object should be understood as a type of object. Often an object is mentioned by a name, such as “Nautilus,” where the type would properly be understood as something like a “shell.” In such cases, the type of object would have to be indicated as an attribute. We also have an indication of <material> in the porcupine quills that were used to make the “shass,” “girdle,” and “gaiters” as well as a <trait> in the adjective “curious.”

A full accounting of the application of current TEI elements and attributes would require more extensive modeling than this paper can allow, but this is enough to demonstrate that much can be generalized and applied to mentions of or reference to objects, although there are some limitations in applicability. There is, however, one much more fundamental need in our adaptation of the TEI.
This need is for something analogous to the naming function of `<name>` and its variants as it is used to identify mentions of other real-world entities, but in fact, the need in this context extends far beyond simple naming: it involves the very function of selection described above.

3. Identifying the Span of a Mention of an Object

At the most basic level, for documents that refer to, mention, or describe real-world objects, we must be able to identify what constitutes a reference to an object, identify a string of text that can function as a verbal identifier (e.g., a “name” in the loosest sense), and then correlate with that named object all the relevant context. This is the first step in treating historical documents with a curatorial sensibility. This identification is crucial for the function of selection. We must be able to identify what constitutes an object (i.e., a mention of a real-world object) in a text so we can select it and then do things with it: correlate and compare it with other mentions of the same real-world object or with other mentions of similar objects; build events involving that object in relation to other entities (named people and places) and temporal indicators (dates); attach annotations to it; or simply represent it in some curated way for an imagined user.

The context of collections of curiosities highlights the challenges posed by diversity of materials and forms in this respect. The case of a simple mention of an object analogous to the mention of a person or place name in a novel, for example, is easily managed. A mention of the Eiffel Tower can be treated much like the mention of a named person with a `<persName>` as `<name type="object">`, but the forms of reference represented in the documents related to early modern collections are not so straightforward. Here we are not typically dealing with proper nouns. Objects are not typically named or even namable, but rather described. The closest we come to a name is the identification of a type or class of object: a loadstone; a canoe; a unicorn horn. A more flexible solution, given the semantic restriction of `<name>`, is the `<rs>` (referencing string) element, which will form the core of what I describe below.

The simplest case can serve as a starting point. An inventory of objects typically consists of little more than the mention of an object in a list. The TEI has semantics to handle this documentary form.
In each case the mention is discrete and self-contained; each item in the list constitutes a reference to a real-world object. The text string contained in each item is small enough that the user can quickly process and understand the sort of object that is being referred to. But, as noted above, lacking @type, there is no way to indicate this item is an object. So, we need some other way to identify what constitutes a treatment of or a reference to an object.

An issue arising from this example is the question of when an inventory becomes a catalogue, and whether a catalogue should be considered a list. The question is pertinent in the context of seventeenth-century collections. According to Paula Findlen, the emergence of the catalogue in the early modern period went hand in hand with the development of collections of curiosities, marking a significant generic and epistemological departure from the inventory:

Inventories record the contents of a museum. They quantify its reality, listing the objects without attaching analytical meaning to them. Catalogues purport to interpret. Their appearance in the late sixteenth century further suggests how novel the practices of Renaissance collectors were.

The practical application of this distinction is difficult. In the case cited above, are we dealing with an inventory or a catalogue? Does the identification of a coin as gilded silver and as minted under the authority of Charles I constitute analysis? A related question is how to classify the structure of this document in TEI terms. Given Findlen’s definition, an inventory is most certainly a list. But can a catalogue be a list? The *Oxford English Dictionary* considers the identification of “catalogue” with a “list, register, or complete enumeration” now obsolete or archaic, noting that a catalogue is “[n]ow usually distinguished from a mere list” (OED 2). This being the case, how do we identify what constitutes an entry in a catalogue, even if entries are numbered like items in a list? Unlike more familiar prose forms (e.g., the novel or the monograph), these documents tend to be fragmented and segmented. Sometimes paragraphs are clearly identifiable in catalogues, but...
more commonly we are dealing with chunks of text for which <ab>—an “anonymous block” of text “analogous to, but without the semantic baggage of, a paragraph” (TEI 2013)—is more appropriate, or perhaps <seg>, and “arbitrary segment” which “represents any segmentation of text below the ‘chunk’ level” (TEI 2013). In this case, we do not have even a complete grammatical sentence, much less a proper paragraph:

4. Item, a brass flat piece, with the figure of a man drawing an ox by the horns; very ancient, being dug out of another temple of Hercules that stood upon the Aventine Hill, on the place where he killed the thief Cacus, where now standeth a church dedicated to St. Stephen, which by its title beareth the memory of the old story of Cacus, it being still called Sto Stefano nel Caco.

(Canterbury Cathedral Libraries and Archives 1676, ff. 4–5)

18

This entry, despite its length, seems to identify itself as an <item> belonging to a <list>, at least in TEI terms. In other cases, the scope of the text span treating an “item” in a catalogue might be so expansive as to constitute a full paragraph or even multiple paragraphs comprising an entire <div>.

Often the mention and description of an object, whether extensive or brief, are embedded in the context of continuous prose. In these cases the specific treatment of an object needs to be distinguished and separated from its broader context in order to define the limits of its full mention. So then, the first step is to identify the parameters of what constitutes a mention of an object, where a mention might include not only a reference to an object (analogous to naming) but also what is said about that object, whether as an item in a list (<item>) or in a full paragraph (<p>), an arbitrary, discrete chunk of text that is not exactly a paragraph (<ab>), or a segment within one of these larger elements (<seg>) or perhaps across more than one paragraph or anonymous block of text. We then need to indicate that each mention pertains to an object. We can use @type=“object” to say, in effect, that this portion of text (as defined by one of the above elements) constitutes a complete mention of an object, but here we again run into limits in the TEI because the <p> and <item> elements cannot take @type. Moreover, semantically it might not make sense to give a type value of “object” to a paragraph. It might not make sense either to apply @type
to an `<ab>` if it is understood to be a defined chunk of text “analogous to a paragraph.” It does make semantic sense, however, to say that an “item” is an “object,” and perhaps also that a `<div>` represents, structurally, the treatment of an object.

The most semantically neutral option for identifying the full context comprising a treatment of an object is the `<seg>` element with an `@type` attribute of "object". Another, less semantically appropriate, option would be `<rs>`, which seems to apply at the phrase level (as appropriate for a noun phrase referencing an object) rather than to a “chunk” of text as `<seg>` does. The `<seg>` element could be embedded in any of the structural elements identified above.

```xml
<list>
  <item>
    <seg type="object">1 Siluer peece guilded of Charles I K of England.</seg>
  </item>
</list>
```

or

```xml
<div>
  <p>
    <seg type="object">A Jubile peece of siluer, with the porta Santa.</seg>
  </p>
</div>
```

One might attach the `@type` element to those structural elements that allow it, but for consistency’s sake, it makes some sense to use an embedded `<seg>` in all instances. There are two advantages to doing so. First, it would allow distinctions between more than one object treated within a single `<p>`, `<item>`, or `<ab>`. Second, it would also be possible to link `<seg>`s that are part of a continuous or even discontinuous treatment of an object across document structures such as `<p>` or `<div>` (see TEI 2013, sect. 16. “Linking, Segmentation, and Alignment”).

Another challenge is, again, in the semantics of the TEI. Exactly what might be meant by an “object” entity is somewhat ambiguous. In the clearest TEI manner, we are interested only in mentions of a real-world entity, as when “name” is used to identify a mention of a person. The work of correlating that mention with a real-world entity is managed either by a name list in the TEI header which provides commentary or, as in my case, by an authority database that correlates name (using `@ref`
or, alternatively, @key) with biographical information about the referent. But such correlation is a challenge when we do not have named entities. And yet it is crucial to express, in some way, that the linguistic act in question is one of referencing.

4. Creating a Handle

We need, then, to be able to identify the core element of the reference, the object as distinguished from the things that are said about it in context. We need to give it a “handle” by which it can be known and by which we can grab it (i.e., select it). Creating a handle is crucial in cases such as item 22 in John Bargrave’s manuscript catalogue, which begins with a full-paragraph description of his visit to the catacombs in Rome (i.e., context), and it is not until the second paragraph that Bargrave finally references the objects he collected there:

22. Ten miles, almost, round about Rome, under the vineyards and cornfields, are hollow caves, streets, rooms, chappells, finely paynted, &., which is called Rome underground, or the Catacombe, wherein to the poor Christians in the times of persecution fled to hide themselves, to perform the Christian duties of preaching and prayer and sacraments. I going down a pair of stayre, and so walked some streets in Rome underground, a second story deep, until we came to water, which made us return. I taking all along on the one side, and my companions on the other, we could meete with never an O rate pro anima — praying for the souls of the dead not being then known, in the primitive times, there being no such thing as purgatory then known in the world, — that being of a later invention, to bring a vast revenue to the Pope or Camera Apostolica.

From this Rome underground I brought a very fair small ancient lamp, and a small bottle with a long neck—both of them of a very fine red earth; which, by Dr. Plott, I sent as a present to the cabinet of Oxford Library. One other earthen lamp, and a glass bottle with such a long neck, and a broken one in two pieces, I have in my cabinet. These bottles are called lachrymatorij, or tear-bottles, because the friends and relations of the defunct were in ancient time accustomed at the funeral to carry each of them a lachrnyatorio in his hand, to save his tears that he shed for his deceased friend, and then leave those bottles behind them with the immuralld corps. David seemeth to have allusion to this ancient custom when he saith, Psalm 56, 8, ‘Thou hast put my tears into thy bottle.’
(Canterbury Cathedral Library and Archives 1676, ff. 19–22, emphasis added)

For reference and analysis, it is helpful to be able to locate quickly and conveniently the core reference to the object in question by means of a string of text that can serve as a handle (analogous to a name), such as “small ancient lamp,” or “small bottle with a long neck.”

Following the principal of adopting (and if necessary, adapting) what is already available in the TEI, we might consider first the possibility of using <name> with a @type="object" as we do for the mentions of people and places by name. There are a few limitations here. Our needs for “naming” most certainly stretch the semantic intent of the <name> element: these are not names. Very often there is nothing even analogous to a name, so it will be necessary to use an arbitrary string for identification.

There are a couple of options for identifying a string of text to serve as a convenient “handle.” One option, which semantically suits the present case, is <objectType> because most of these objects are not named so much as categorized. They really are “types”: a lamp; a girdle; gaters; a loadstone; a bird’s skull. However, <objectType> is not allowed a @type attribute (probably because it is itself a type), but in the case of collections of curiosities, this attribute would be convenient. For example, objects in seventeenth-century collections were commonly classed into three types: natural, artificial, and objects that fall between these two categories (natural-artificial hybrids). It would be useful to have the option of making such distinctions. Moreover, in some cases, the text string might not connote a type of object: for example, “a coin” or “the horn of a rhinoceros.” A more general and flexible alternative is the <rs> (referencing string) element, which is more “general purpose” and perhaps less semantically specific than the <name> element. It is simply a string of text that references a real-world object, and this element can bear @type and @subtype attributes, enabling two levels of categorization for an object:

<rs type="object" subtype="artificial">a very ancient Æsculapius, in brass</rs>

But again we run into complications. How do we determine the extent of a handle—that is, what constitutes the minimal core that is the referencing string proper? In many of these cases, it will be difficult to determine what should be included: “a very fair small ancient lamp,” “small ancient lamp,” “ancient lamp,” or just “lamp?” The question is crucial because everything else that
comprises a mention or treatment of an object (minus the naming) is context. In other cases, it is hard to identify which string of text constitutes the core of the reference, and thus an appropriate handle. For example:

   Leg with hoof of an animal of the deer family. Perhaps the Elk or Elend of the Germans
   
   (The Book of the Principal of Brasenose, in MacGregor et al. 2006, 97)

The handle should probably be `<rs type="object">Leg with hoof of an animal of the deer family</rs>`, which seems to be the core of the reference: the rest is added in apposition to this handle, even though it provides a more precise (if tentative) identification of “Elk or Elend.” And then, how does one create a handle when the class of object is assumed (as in, for example, a list of coins) and where the contents of an entry are simple reports of what the object contains?

   King James I on the obverse and his son on the reverse.
   
   (The Book of the Vice Chancellor, in MacGregor et al. 2006, 1)

Can one create a handle that is meaningful in cases where there is no meaningful noun to serve as the core?

5. Relation between a Mentioned Object and its Context

For curating in TEI, then, we must be able to define what constitutes a mention of an object and be able to identify a handle for ease of reference and analysis, but we must also be able to relate each mention to its relevant contexts, most crucially to object groupings. Sometimes a mention pertains simply to a single object, but we have already seen that mentions of objects often come in the context of sets and groups. In the example above, we have parts (“Leg with hoof”) forming a single referenced object, but understood in relation to the whole (the “deer”). These contexts are essential to understanding the individual objects.

To return to our first example, the whole mention—everything that is (misleadingly) identified as the “Item 22”—comprises in fact a group of objects that includes the cravat, shass or girdle, and gaiters, as well as the chains of the redeemed merchant. The first three can also be properly considered a set. Collectively, this set, together with the chains, constitutes a group of objects. The set of objects belongs together by design or by some essential relationship. The group of objects belongs together by a looser association based on some common element—in this case, their connection with a common event—but also by virtue of Bagnell’s document structure, which
identifies the group as constituting an *item* in his catalogue. The semantic distinction is obvious enough here, but it can be complicated in other instances where one has, for example, a “box of shells.” Adding further complication, there is a second mention of the set of objects (“cravat, etc.”) along with a vague reference to another group of objects: “divers other things.” Thus we have:

```xml
<seg type="objectGroup">
  <seg type="objectSet">
    <seg type="object">a</rs type="object">cravat</rs><seg>,
  <seg type="object">a</rs type="object">shass or girdle</rs><seg>, and
  <seg type="object">a small</rs type="object">pair of gaiters</rs><seg>...
  </seg>
</seg>
```

Here the reference strings are limited to the smallest unit of reference, analogous to `<name>`, but the vague group (“divers other things”) and chains require a longer string of text. The larger context—the entire context of a mention of either a group (as in this case), or a set or individual object—is delimited as a `<seg>`. Another option for indicating relationships between objects is, again, to use some form of stand-off solution, either in the header of the TEI document, or in RDF or in a database, which would enable explicit identification of relationships, rather than relying on parsing of containing structures to articulate relationships.

Once objects are identified with handles, one might select only the mentions of individual objects and present them in a browsable list: cravat; shass or girdle; pair of gaiters; chains of the redeemed. Or one might elect to see the individual objects in their set and group contexts, such as the cravat in the context of its set: “cravat, shass or girdle, pair of gaiters.” Here again questions arise regarding how to limit the identification of an object:
• In considering the “pair of gaiters,” we sidestep the issue of what constitutes a discrete object. Strictly speaking, there are two objects mentioned as one. In one sense the pair is considered a whole, but in an important sense they are two: it is impossible to mark them as two using an `<rs>` element. A compromise solution would be to add an indication of quantity, but the `@quantity` attribute is not allowed here. The only available solution is `<num value="2">pair</num>.

• There are other objects here, but I elect not to identify them because they are not collected objects. These objects are: butcher’s shops (mentioned twice, the second time in more general terms as “shops”); porcupines, both as animals and as a type of meat; money (10,000 pounds sterling); and Bargrave’s grave. If one were to identify every mention of an object in these sorts of documents, they would be considerably more plentiful than either names of people or names of places. It would not be practical, nor to the purpose of identifying and selecting mentions of collected objects within and outside of their contexts. For the purposes of my project, these other elements are part of the context and need not be identified as objects in their own right.

• Finally, should the “&c.” following “cravat” somehow be marked and identified as constituting an object, a group of objects, a subset, or should it not be marked at all?

More complicated are cases where there is intervening content between a discrete mention and other content directly relevant to the object. To what degree should we attempt to distinguish and relate particular mentions in context? In most cases, it is enough simply to define the context that contains the mentions:

(20). Item, two large loadstones, one armed with steel, in a black velvet case, which I have hanging in my study upon a piece of silk, in a perpendicular thread, when it standeth still, the north point hangeth still due north; by which I found that our cathedral of Christ Church, Canterbury, doth not stand due east and west, but the east end is at the least 2 poyns of the compass too much to the southward. Now, where it is generally received that the loadstone draweth iron to it, by this perpendicular posture of the stone upon a thread, and putting a key or any other piece of iron to it, the iron draweth the loadstone quite round, as far off as you please, so that it seemeth there is no compulsion on either side, but a mutual reciprocal compliance between them both, which we are fain to call sympathy. Now, on the other side,
have in my cabinet another triangular, unequilateral, bumped-up, large loadstone that weigheth almost half a pound, which is a rude thing to look on, but of good value. This is unarmed, but it is strange to see how great an antipathy there is between the north point of this stone and the other that hangeth perpendicular in the velvet bagg, this making that (at a great distance) fly from it with violence as often round as you please; and, on the other side, there is a great sympathy between the south point of the one stone and the north point of the other. For this seemeth strange to me, that every loadstone, be it in pieces bigger or less, have still their north and south point, according to the two poles axill of the world. With the hidden qualities of these 2 stones I used sometimes to make sport with young gentry in telling them their fortunes, &c. as if there had been an intelligence between them and me — ‘If so and so, then do so and so.’ And truly it is wonderful to me to think that it was the loadstone that found out America and the Straights of Megallan, and by virtue of which several nations, especial[y] England, have almost found out the north-west passage of the West Indies, and so to go a much shorter cut from England by the West Indies to the East. And if the Terra Incognita, or the fifth part of the unknown world, be ever found out, it must be done by virtue of the loadstone.

(Canterbury Cathedral Library and Archives 1676, ff. 13–6, emphasis mine)

In this instance we have a collective mention of the two objects as a group or class—“two large loadstones”—and then commentary on the “one armed with steel, in a black velvet case” before turning to “another triangular, unequilateral, bumped-up, large loadstone” on the other side of the cabinet, followed by commentary pertaining to it. But then Bargrave turns to talking about the two stones together and their “antipathy” for each other, and then how he used to have fun with young gentlemen by manipulating the “hidden qualities of these 2 stones.” Marking out mentions of these individual objects and their grouping is not exactly straightforward. The whole paragraph could constitute a mention of a group, and “two large loadstones” can serve as a handle for this group. But then this corporate handle must serve as an implied referent for the first mentioned loadstone: “one armed with steel ....” The second yields a more complete handle (“another triangular, unequilateral, bumped-up, large loadstone”), although without the corporate referent, the pronoun “another” makes the reference incomplete: “another triangular, unequilateral, bumped-up, large loadstone.” As difficult as it is to mark out the mentions of the
individual objects in relation to the mention of the grouping of the two, it becomes even more difficult to sort out the rest of the paragraph, where the discussion of the two objects blends into a discussion of the two together.

Closely related to the set, but more theoretically and practically complicated, is the case of the composite object: an object that is one object comprised of two or more objects. This object is from a seventeenth-century inventory of John Paston’s (1631–1693) collection at Oxnead Hall, Norfolk:

A shell cup engraven, with ivory handles, with a tortois-shell foot and cover.

(Repton 1884, 150)

In this case we have an ambiguous relationship between constituent objects, parts of an object, and materials. The principal object is the whole cup, but the main part of the cup is also an object in its own right—a shell—and not simply “material.” By this analogy, the two tortoise shells (one for the foot, and presumably another for the cover) are also objects and not simply materials. Should the “ivory” then also be considered an “object” even though it is stated adjectivally? If the tortoise shell is an object in its own right, should the “foot” and “cover” also be considered objects, or should they be considered parts of an object?

In addition to sets and immediate groupings, documents also sometimes provide categorical groupings of objects. The easiest case is when an object is situated within a defined structure in a document, as is often the case with early modern catalogues of curiosities, such as Nehemiah Grew’s catalogue of the Royal Society’s Repository, which provides a hierarchical structure beginning at the top level with “Animals,” “Plants,” “Minerals,” and “Artificial Matters” (a fairly standard taxonomy for collections of the time):
Curating Object-Oriented Collections Using the TEI

(\textit{Musaeum Regalis, 1681}) Every object, then, can be selected and identified in a complex categorical grouping. Similar, but somewhat more complicated, are cases where objects are arranged in a document according to spatial groupings, such as Robert Paston’s inventory:

\begin{verbatim}
<div type="catalogue">
  <div type="animals">
    <div type="quadrupeds">
      <div type="viviparous">
        <div type="object">
          <!-- ... -->
        </div>
        <!-- ... -->
      </div>
      <!-- ... -->
    </div>
    <!-- ... -->
  </div>
  <!-- ... -->
</div>
\end{verbatim}

(Repton 1884, 150) Representations of objects in space are usually not as regular and structured as Grew’s taxonomic structure for his catalogue. Again, for these complications, stand-off markup is probably the best solution.

6. Conclusion
As Øyvind Eide argues, “it is hard to understand [a modeling standard] at a deeper level without using it in practical work” (Eide 2014–15, para. 12). The preceding examples are as much about elaborating the challenges as providing solutions, deriving from my experience in building an archive of primary materials related to what was arguably the first major cultural movement to pay close and critical attention to the world of material objects. In order to do things with this virtual museum, to be able to curate it in the way one might curate a material museum—that is, to package materials for public use—we must first be able to select objects for manipulation. I conclude with some initial requirements to facilitate this core function of selection in order to treat object entities using the TEI with a curatorial sensibility:

1. We must be able to determine and define the limits of what constitutes a mention of an object. This is crucial so that we retain all information that is immediately relevant to that object, but (ideally) no more than is pertinent.

2. We must be able to identify a “handle” for each mention of object in the same way that we can identify a mention of a person or place.

3. Each mention must be addressable. That is, each object must have a unique identifier that enables references to it and correlations between it and other entities and units of information.

4. We must be able to define the relevant and relative contexts of each mention of an object:

   4.1 To enable an articulation of an object’s place in a hierarchy;
   4.2 To enable identification of relevant and related information; and
   4.3 To enable the articulation of events involving other entities (e.g., people, places, and other objects).

BIBLIOGRAPHY

Balzer, David. 2014. Curationism: How Curating Took Over the Art World and Everything Else. Toronto: Coach House Books.

Beagrie, Neil. 2006. “Digital Curation for Science, Digital Libraries, and Individuals.” International Journal of Digital Curation 1(1) (February 12): 3–16. doi:10.2218/ijdc.v1i1.2.
Brown, Bill. 2001. “Thing Theory.” *Critical Inquiry* 28(1): 1–22.

Cameron, Fiona. 2003. “Digital Futures I: Museum Collections, Digital Technologies, and the Cultural Construction of Knowledge.” *Curator: The Museum Journal* 46(3): 325–40. doi:10.1111/j.2151-6952.2003.tb00098.x.

———. 2005. “Digital Futures II: Museum Collections, Documentation and Shifting Knowledge Paradigms.” *Collections: A Journal for Museum and Archives Professionals* 1(3): 243–59.

Canterbury Cathedral Library and Archives. 1676. Lit. Mss. E16a. “Rara, Antiqua, et Numismata Bargraviana.”

———. 1658. Miscellaneous paper, “Memorandum April 28, 1658.”

Ciula, Arianna, and Øyvind Eide. 2014. “Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory.” In DATeCH ’14: Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, 35–41. New York: ACM. doi:10.1145/2595188.2595207.

Deeran, Stefan. 2013. “What Is Social Media Content Curation?” (Blog post.) *The Huffington Post*. Accessed July 14, 2015. http://www.huffingtonpost.com/stefan-deeran/what-is-social-media-cont_b_3383706.html.

Eide, Øyvind. 2014–15. “Ontologies, Data Modeling, and TEI.” *Journal of the Text Encoding Initiative* 8. http://jtei.revues.org/1191. doi:10.4000/jtei.1191.

Eide, Øyvind, and Jon Holmen. 2006. “Reading Gray Literature as Texts: Semantic Mark-up of Museum Acquisition Catalogues.” Paper presented at CIDOC Annual Conference, Gothenburg, Sweden, September 10–14. Accessed 4 July 2015. http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2006/Eide_Holmen.pdf.

Findlen, Paula. 1994. *Possessing Nature: Museums, Collecting, and Scientific Culture in Early Modern Italy*. Berkeley: University of California Press.

Grew, Nehemiah. 1681. *Musæum Regalis Societatis, Or, a Catalogue & Description of the Natural and Artificial Rarities Belonging to the Royal Society and Preserved at Gresham Colledge*. London: Printed by W. Rawlins for the author.

Impey, O. R., and Arthur MacGregor. 1985. *The Origins of Museums: The Cabinet of Curiosities in Sixteenth- and Seventeenth-Century Europe*. Oxford: Clarendon.

MacGregor, Arthur, ed. 1983. *Tradescant’s Rarities: Essays on the Foundation of the Ashmolean Museum, 1683, with a Catalogue of the Surviving Early Collections*. Oxford: Oxford University Press.

MacGregor, Arthur, and Moira Hook, with John Davies, Stephen Harris, Chris Howgego, Malgosia Nowak-Kemp, Philip Powell, and Donald Sykes. 2006. *Manuscript Catalogues of the Early Museum Collections, 1683–1886. Part II: The Vice-Chancellor’s Consolidated Catalogue 1695*. BAR International Series 1569. Oxford: Archaeopress.

Ore, Christian-Emil, and Øyvind Eide. 2009. “TEI and Cultural Heritage Ontologies: Exchange of Information?” *Literary and Linguistic Computing* 24(2): 161–72. doi:10.1093/llc/fqp010.
Pomian, Krzysztof. 1990. Collectors and Curiosities: Paris and Venice 1500–1800. Translated by Elizabeth Wiles-Portier. Cambridge, U.K.: Polity Press; Cambridge, MA: Basil Blackwell.

Repton, John Adey. 1844. “Oxnead Hall, Norfolk.” *Gentleman’s Magazine* n.s. 21: 21–24, 150–53.

Sturdy, David, and Martin Henig. 1984. The Gentle Traveller: John Bargrave Canon of Canterbury, and His Collection. [Canterbury, U.K.]: [Canterbury Cathedral].

TEI Consortium. 2013. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.5.0. Last updated July 26. N.p.: TEI Consortium. http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/.

NOTES

1 The model elaborated in this paper is based on the TEI Consortium’s *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (2013).

2 The research represented in this project has been funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

For more information on the project, see the Digital Ark Blog, http://digitalarkproject.blogspot.ca/p/documents.html. Collections of curiosities became a defined field of study in the late 1970s and early 1980s with the historical and theoretical groundwork laid by Krzysztof Pomian (1990) and O. R. Impey and Arthur MacGregor (1985).

3 See, for examples, MacGregor et al., *Tradescant’s Rarities* (1983), and David Sturdy and Martin Henig, *The Gentle Traveller* (1984).

4 CIDOC (International Committee for Documentation of the International Council of Museums), accessed April 4, 2017, http://network.icom.museum/cidoc/.

5 The documents referred to here can be found at http://digitalarkproject.blogspot.ca/p/documents.html

6 The Internet is brimming with blogs about curating social media. From a popular perspective, see Deeran (2013); from the GLAM perspective, Beagrie (2006); and taking a popular-historical perspective, Balzer (2014).

7 See, e.g., the various “Best of the Web” features at Museums and the Web, http://www.museumsandtheweb.com/best-of-the-web/, and any number of articles, such as Cameron (2003 and 2005).

8 http://www.tei-c.org/Vault/P5/2.5.0/doc/tei-p5-doc/en/html/ref-material.html
ABSTRACT

This article considers the possibilities and challenges in using TEI-based XML markup for curation of objects mentioned in historical documents such as catalogues and inventories, but also in unstructured forms such as diaries and personal correspondence. It takes as a case study documents related to early modern collections of curiosities. It first considers how far the current guidelines for manuscript description can be generalized for encoding other kinds of material objects and their contexts. It then examines what more is required for treating mentions and descriptions of objects in historical documents. It argues that the core affordance of curation for such materials is the ability to identify and select what constitutes a mention of an object and to relate that mention to its immediate context, including its relationships to object groupings.

INDEX

Keywords: curation, collections of curiosities, material culture, seventeenth century

AUTHOR

BRENT NELSON

Brent Nelson is Professor of English at the University of Saskatchewan. Most of his research occurs at the intersection of digital humanities and early modern literature (the John Donne Society’s Digital Prose Project) and material culture (The Digital Ark: Cabinets of Curiosities in England and Scotland 1580–1700).