494

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

A Grammatical Approach to
Subject Classification in Museums

Rick Szostak

University of Alberta, Department of Economics, Tory Building 9-18, Edmonton, Alberta,
T6G 2H4, CANADA, <rszostak@ualberta.ca>

Rick Szostak is Professor and Chair of Economics at the University of Alberta. He is the author of fifteen books and fifty journal articles in economics, history, interdisciplinary studies, information science, and several other fields. He has studied the theory and practice of interdisciplinarity for two decades and has emphasized in the last decade the ways in which knowledge organization systems might better facilitate interdisciplinary research and teaching. He has long argued that a phenomenon-based synthetic approach to classification is both feasible and desirable; he has more recently stressed the use of grammatical construction in performing that synthesis. He recently co-authored *Interdisciplinary Knowledge Organization*. He has created the Basic Concepts Classification (BCC), and is engaged in efforts to evaluate the BCC and compare it to other classification systems.

Abstract: Several desiderata of a system of subject classification for museums are identified. The limitations of existing approaches are reviewed. It is argued that an approach which synthesizes basic concepts within a grammatical structure can achieve the goals of subject classification in museums while addressing diverse challenges. The same approach can also be applied in galleries, archives, and libraries. The approach is described in some detail and examples are provided of its application. The article closes with brief discussions of thesauri and linked open data.

1.0 Introduction

This paper seeks to address two key challenges in knowledge organization for museums. First, it is often recognized in the literature that it would be advantageous to utilize the same knowledge organization system (KOS) across the entire GLAM sector since users increasingly search across galleries, libraries, archives, and museums. McMarty (2014, 615) for example, argues that museums, libraries, and archives now face a very similar set of user expectations as a result of each developing an online presence; he cites a number of conferences and special issues of journals focused on how to respond. Yet, the KOSs developed for libraries are often thought to be ill-suited to the classification of objects. Second, museums often have limited resources to devote to classification, and often do not have staff trained in knowledge organization. Museum training has tended to focus on understanding the artifacts rather than knowledge organization—though of course there has always been some interest in and familiarity with knowledge organization in museums (Urban 2014). We would want, then, a KOS that is capable of addressing both objects and documents, and yet is easy for both cataloguer and user to navigate.

The solution that will be proposed involves a synthetic and grammatical approach to subject headings. This would allow cataloguers in museums, archives, and libraries to move fairly directly from a sentence in an existing object or document description to a synthetic subject string that orders terms grammatically. Though objects and documents differ in many ways, they are each commonly described—by publishers, authors, or curators—in a few sentences. Users in turn can move fairly directly from a query stated as a sentence to the most relevant subject string. Since object descriptions, document descriptions, and user queries are all formulated in sentences, there is obvious value in using sentence structure also in the subject headings that mediate among these.

The next section of this paper provides more detail on the challenges of museum classification. The succeeding section then addresses the present state of museum classification and foreshadows the potential advantages of a new approach. We are then able to expand on the nature of a synthetic and grammatical approach to classification that addresses the challenges identified. Examples are

provided of how this new approach could be imple-
mented. We then address the possibility of an associated
thesaurus and the benefits of the recommended ap-
proach for the world of linked open data.

2.0 Challenges in museum classification

Zoller and DeMarsh (2013) note that museums have tradi-
tionally focused on documentation—providing minimal in-
formation on what the museum possesses and where it
came from—rather than cataloguing, which provides users
with multiple access points. They cite a 1984 Report of the
American Alliance of Museums, which urged "Informa-
tion sharing among museums … similar to the library in-
formation system that can locate all the books anywhere in
the country on a certain subject is an ultimate goal." The
authors perform a survey and find a widespread belief that
cataloguing requires curatorial expertise in a field. The au-
thors worry that curators lack time, and that experts in
cataloguing may not wish to offend curators. They further
note that cataloguing gets little attention in museum train-
ing. Only a small minority of museums had a position for
cataloguer, and it was rare to want or require library-and-
information science (LIS) training. One oft-voiced concern
was that cataloguing would not respect the uniqueness of a
collection. This has proven to be a serious barrier to the
development of a common approach to classification. Mu-
seums differ a great deal in size, type of artifact, guiding
mission, and in a host of other ways. It might seem an im-
possible task to facilitate searching across institutions while
respecting institutional individuality. A common approach
to cataloguing that was easy to master and yet allowed
uniqueness to be represented would clearly be advanta-
geous.

McMarty (2014) worries about the different goals of
museums compared to those of libraries and archives.
The last two have always had a greater interest in guiding
users to relevant documents. Museums have instead fo-
cused on placing artifacts in a context that museum visi-
tors had found by browsing. The museum might only
need to tell visitors what sort of artifacts would be found
in particular rooms. Libraries could be satisfied with pro-
viding just enough information to guide users to a docu-
ment they could then read. Museums instead focused on
providing educational description for the objects they
displayed.

Museums around the world have launched websites
over the last decades. These websites have encouraged
museums to take a much greater interest in metadata
(Hider 2012, 68). Yet these websites are only rarely useful
in guiding users, and especially professional researchers,
to appreciate the detailed contents of a museum's collec-
tion (Menard et al. 2010). Indeed, these websites often do

not strive to do so. We will find at multiple points in this
paper that website redesigns often decrease subject ac-
cess. Menard et al. (2010) worry that when information is
provided on artifacts this is often jargon-ridden and ob-
scure. The authoritative "Categories for the Description
of Works of Art" (CDWA, Baca and Harpring 2016)
urges a more user-centered approach: "Information for
display is assumed to be in a format and with syntax that
is easily read and understood by users." The CDWA rec-
ognizes that jargon may be appropriate when detailed de-
scriptions are developed for internal use but must be es-
chewed when developing finding aids for users. And
Hider (2012, 49) notes that most object descriptions in
museums were developed for the use of curators rather
than users (and those for public display served educa-
tional rather than organizational purposes). This suggests
the value for subject headings of breaking the complex
concepts employed by professionals into basic concepts;
concepts that have broadly shared understandings across
different groups of experts as well as non-experts.
Menard et al. (2010) appreciate that some sort of syn-
thetic classification—which combines simple terms in
order to generate complex subject strings (that is subject
headings composed of a "string" of simple terms)—
could allow museums to provide much better informa-
tion regarding their possessions. They further note that
there are few tools available to museums wishing to pur-
sue such an approach.

Museums were in the past often organized chronologi-
cally, and their artifacts thus classified primarily by time pe-
riod. Museums (like art galleries) are increasingly organized
thematically. It is thus of increased importance that their
artifacts be classified by subject. This will among other
things increase the ability of curators to identify pieces to
borrow for thematic exhibits. We should note in this regard
that many museums have hundreds or thousands of arti-
facts for which they have lacked the staff to properly de-
scribe or classify. While our main concern in this paper will
be with allowing museums to move fairly quickly from an
existing description to a subject heading, we should appre-
ciate that some museums may be interested in providing a
minimalist description that can guide a subject heading,
with the intention of developing a fuller description of ar-
tifacts that prove particularly important to the museum's
mission (or that might be leant to another museum). This
huge backlog of undocumented artifacts also serves as a
reminder of the resource limitations that many museums
face.

Though some—especially national—museums have a
broad remit, most museums focus on materials generated
by a particular group or within a particular region. Yet, of
course, scholars (and many general users) are often inter-
ested in comparing artifacts produced by different people

or in different place—and thus would like to compare across the holdings of multiple museums that each have a special focus. There has, of course, been a huge debate within knowledge organization as to whether it is possible to identify a controlled vocabulary that can be understood in similar ways across disciplines or cultural groups. This debate has obvious implications for the museum community, which often faces pressure from particular social groups to respect the meaning which that group attaches to particular artifacts. Is it possible to provide descriptors for similar artifacts from quite different societies such that all potential users could understand what these descriptors mean? The solution I have proposed elsewhere focuses on "basic concepts" (Szostak 2011). Whereas complex terms such as "globalization" may be understood in diverse ways, basic concepts such as "trade flows" or "American movies" are subject to broadly similar understandings across individuals or groups. And thus a synthetic approach to subject classification in which basic concepts are combined can potentially allow users from different disciplines or cultures to find what they are looking for in both museums and libraries. "Adobe" is a familiar term, but many will not know that it is (clay)(for)(building), whereas "clay" and "building" are broadly understood concepts.

Synthetic classification will prove especially useful for artifacts whose nature may only be fully appreciated by experts in a field. We would want anyone searching for "tools for scraping" to find the "beamer," a bone implement used for scraping by indigenous peoples in the Americas. Another example is "abrader"—(tool)(for)(smoothing). To the unfamiliar, "Apache tear" will hardly signify (round) (nodules)(of)(obsidian). Synthetic classification is also particularly useful when terminology is ambiguous. Someone searching for awls used in working leather would like to readily distinguish these from the quite different awls used for working wood.

One challenge that many museums face is that different groups may attach different meanings to the same artifact. As an example, European explorers in nineteenth century Australia collected many Aboriginal artifacts but did not record the spiritual meaning that was attached to these. Aboriginal groups understandably find that these artifacts may be both mis-described and mis-classified in museums (Cameron & Robinson 2007). A synthetic approach may be helpful here in allowing a subject string that can combine both the practical use and spiritual meaning of an artifact (I argued in 2014 of the more general value of a synthetic approach grounded in basic concepts in addressing concerns around classification and social diversity). There may, of course, be artifacts for which meaning—or other elements of subject—is contested. The advice of CDWA is to openly discuss areas

of disagreement in the object description. The subject heading may then omit elements of subject that are controversial, though it would also be possible within a synthetic subject string to use "or" between differing interpretations if this element of the subject were judged of particular importance. As noted above, Szostak (2011) argued that ambiguity and controversy can be reduced by breaking complex terms into "basic concepts" for which broadly shared understandings exist across individuals and groups; an approach that relied on basic concepts might often be able to clarify areas of disagreement such that these could be captured in a subject heading.

Cameron and Robinson (2007) are deeply concerned with the contested nature of many museum artifacts. Indeed, they suspect that objects are inherently polysemic and that no curator can fully fathom their meaning. Yet, they recommend in the end a mix of "modernist" approaches to provide user access to objects and "postmodern" approaches that then guide the user to appreciate multiple interpretations of the object. That is, we must strive for some sort of "objectivity" in subject access but can then appreciate multiple interpretations in detailed descriptions. The user should be urged to read the detailed description once the subject heading has guided them to a particular object. Though this paper focuses on subject headings, it respects the desirability of detailed object descriptions and the importance of guiding users to these. Cameron and Robinson also appreciate the value of linking information resources, especially via linked open data (LOD), so that users can easily be guided to related information. This paper will address these issues also below, noting that a synthetic approach facilitates linking in general and LOD in particular.

Museums hold some artifacts because they are typical—say, the sort of sword used in a particular army—and other artifacts, because they are unusual: say, a golden sword used for ceremonial purposes. Note that a synthetic subject string employing basic concepts can serve to identify both. A user seeking to compare a particular kind of artifact across time or place can thus readily identify these artifacts in many museum holdings. But a user interested in unique items can enter a very precise search string. We will provide many examples of possible subject classifications of museum artifacts below.

Museum artifacts might be usefully classified in terms of their purpose (both practical and symbolic), material composition, and (for some items) methods of manufacture (provenance is beyond the scope of this paper, though it should be noted that a synthetic approach is also useful in designating the group that created an artifact or the place of creation: say, along a lake in southeastern Europe). A synthetic approach is useful for all three: (axe)(for)(war); (wooden)(shaft)(steel)(head); (mass)(produced). A synthetic

approach also allows these to be combined into one longer subject entry. This approach would allow the cataloguer to stress the most important characteristics of a particular artifact—but necessarily then allowing and perhaps encouraging one or two of the three elements to be ignored. Such an approach may also better identify artifacts whose specialness lies in unusual combinations of the three elements: (golden)(axe). Since objects can potentially be classified along at least three key dimensions (design might be considered as a fourth key dimension; see below), and users will often wish to search for particular combinations across these dimensions, a synthetic approach to classification will best allow users to find what they want.

Museums with different missions might in their object descriptions stress different aspects of the same object: a museum of technology might stress how it was made, a museum of daily life might stress how it was used, and a museum focused on a local industry might stress what it was made of (see Cameron & Robinson 2007). A synthetic approach that encourages cataloguers to reflect on these three key elements may facilitate access across quite different institutions. As long as the museum's object description refers to all three elements (which will not always be the case), a cataloguer can capture each in the subject string—though the cataloguer may also reflect institutional mission by emphasizing the element(s) that the museum values most. We may thus achieve the hoped-for balance between providing user access while respecting institutional differences.

It deserves to be emphasized here that a synthetic approach does leave scope for under-resourced museums to provide less detailed subject descriptions than others, while nevertheless employing the same controlled vocabulary. Some museums may also have less remarkable collections than others and, thus, require less extensive subject strings.

3.0 Existing approaches to museum classification

It has not proven easy to develop a common KOS that can address each of the challenges outlined above. The temptation of many museums has thus been to rely on some combination of full-text searching of object descriptions and crowd-sourcing whereby users can tag items. Yet, we in the knowledge organization community know that both strategies can be problematic. In particular, both lack precision. Given that museum staff and taggers may both use a variety of terms to describe the same thing, users may fail to find what they are looking for. Only controlled vocabulary can achieve precision in search. This might be important both when users search for a particular item that they have read or heard about somewhere or when they are wondering if a particular

combination of facets can be found anywhere: an axe fashioned from gold but actually employed in battle. Hider (2012, 54) reminds us that users with only a vague idea of what they are seeking can also suffer from reliance on full-text searching: they would potentially benefit from the structure that a hierarchically organized controlled vocabulary can provide.

The standard resource for *Cataloguing Cultural Objects* (*CCO*, VRA 2017), has been developed by the Getty Research Institute. It self-describes as follows: "CCO is designed for use by professionals in museum collections, visual resource collections, archives, and libraries that have a primary emphasis on art, architecture, and material culture." It devotes chapter six to subject classification. The *CCO* thinks subject classification important, and recommends controlled vocabulary but leaves museum cataloguers to choose which subject authorities they use and the degree of specificity in classification. The Getty's "Categories for the Description of Works of Art" provides some dozens of recommended terms but also encourages museums to choose among many sources of controlled vocabulary for additional terms. Such a recommendation naturally limits the ease of searches across museums. Yet, if museum cataloguers will be urged to consult subject authorities then we can potentially achieve ease of access by encouraging instead the consultation of one easy-to-master classification.

We might also mention CIDOC, the committee on documentation of the International Council of Museums. It has produced guidelines on documentation of museum artifacts. Though these are not explicitly focused on subject classification, I have shown elsewhere how each of CIDOC's metadata elements can be readily translated into basic concepts (CIDOC CRM Special Interest Group 2015; Szostak n.d.)

There are inevitably a host of subject classification schemes developed by different museums. These can potentially facilitate the development of shared controlled vocabulary (see below) but at the moment serve to encourage different museums to utilize differing controlled vocabularies. For example, the United States National Park Service (USNPS) has employed a list of artifacts several pages in length that it has borrowed in turn from the Arizona State Museum. The examples that we explored in the preceding section often came from this list. A scan of the list allows us to draw some important conclusions regarding artifact classification. Importantly we can note that many items in the classification will be needlessly obscure not only to non-archaeologists but also to archaeologists of non-American civilizations.

We can identify three broad types of artifact. The most important for our purposes are those that have obscure titles. These will not be discovered by anyone but

498

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

experts (and perhaps only experts in a particular people). It is desirable that these be coded in terms of more basic concepts. They can of course be classified as long as the term used at present automatically invokes the basic concept string. These strings are usually a handful of terms in length but sometimes longer. We provided examples of beamer, abrader, and Apache tear above. "Atlatl" is (weapon)(for)(throwing)(spear)

The "armor slat" will likely be appreciated only by experts in Native American warfare. It is a flat wooden piece, tied to others, employed as armor. Similar armor may well have been used in other parts of the world but be referred to by different terminology. It is thus useful to render this compound as (flat)(wood)(for)(armor), where armor is itself coded as (protection)(for)(war). Note that there were other kinds of Native American armor. Some used wooden rods tied together—this similarity with slat armor will be captured by a synthetic approach. Many used hides, sometimes hardened with other substances, sometimes padded with cotton—again a synthetic approach captures both similarity and difference.

A second type of artifact involves those with familiar titles. Here a translation into even more basic concepts is less critical but may still be useful in identifying similarities between objects (as, say, between arrows and spears). Examples are numerous:

– Arrow is (long)(round)(pointed)(projectile)(weapon) where (projectile) is (object)(propelled).
– Bag is (flexible)(container)(for)(carrying) [bags should then be distinguished by material].
– Basket is (container)(for)(carrying)(made of)(strips) (intertwined).

We can of course imagine a continuum of recognizability rather than the two distinct types of artifact identified above. One advantage of always defining terms with respect to more basic concepts is that the cataloguer need not constantly ask themselves which of the two types above an artifact belongs to. "Adze" is a good example here; many people will know that this is a tool for carving wood but others will not. Fewer will know that the main characteristic distinguishing an adze from an axe is that the blade is perpendicular to the shaft (such that one swings an adze sideways). Even fewer will appreciate that the mattock is similar to an adze but has a blunter blade.

A third type of artifact involves terms that are ambiguous. Those who recognize that an awl is a long pointed spike will still wonder if a particular awl was designed for making holes in wood, carving wood, making holes in leather, or other uses. Axes have an even wider range of uses. In particular, those designed for cutting or carving wood tend to be quite different from those in-

tended for battle. Of course, tools by their nature have multiple uses. But it is nevertheless often invaluable to identify the primary use for which a given tool was designed. It is often also critically important to distinguish the materials from which a tool was made.

Particular artifacts from any of these three categories may benefit from yet further clarification. They may, for example, be decorated, intended for display or symbolic use, associated with a particular person or event or group, and so on. Even if some of these elements are captured in other metadata elements, they should be recognized in a subject search as well.

It should be noted that many of the entries on the USNPS list are themselves compounds: adze blade, arrow shaft, bag handle. It is often useful to search by the component terms, but this will only be possible if these are explicitly compounds of simpler terms. A particular user might have an interest in blades or handles or even shafts.

We should close this section by appreciating that this is a historical moment in which significant change in cataloguing procedures might be possible. As noted above the online presence of museums has exposed museums to user demands for enhanced access (McMarty 2014, 618):

> They [users] want to be able to say, "I'm writing a paper about Hercules," or "I'm researching the evolution of glass-making technologies," and find all the relevant resources in one search, in person or online, regardless of the type of collections where the records they need may be stored. They do not want to learn that most information systems are not geared toward answering these kinds of questions, and they especially do not want to discover how difficult it still is today for cultural heritage organizations to share information about their collections and enable searching across multiple institutions.

Museums have experimented with software that allows searches across institutions employing different controlled vocabularies. Educational institutions have responded also. It is increasingly common for museum students to learn about knowledge organization and for information schools to compare GLAM sectors or even offer integrated programs addressing GLAM as a whole (McMarty 2014). As we have seen there are many challenges in providing subject access to museums. But with attention increasingly focused on this issue, an innovative solution that addresses these key challenges might prove attractive.

4.0 A novel approach to museum classification

We have in preceding sections identified several desiderata for a museum classification. We have suggested at many

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

499

points that a synthetic approach employing basic concepts might prove advantageous. Such an approach clearly has advantages in achieving clarity; it allows us to translate the quite different terminology that museums might use into basic concepts for which there will be broadly shared understandings across users and cataloguers. It should be stressed here that this approach simultaneously addresses several other desiderata. Though museums have not—and likely could not, given their emphases on different peoples and regions—identified a common controlled vocabulary of complex terminology, we can aspire to a common classification of basic concepts. I have argued elsewhere that this approach is also feasible for libraries (Szostak 2011) and galleries and archives (Szostak 2016a); it thus potentially fulfils the goal of facilitating search across the GLAM sector. A synthetic approach employing basic concepts allows us also to: deal with both typical and exceptional artifacts; translate professional jargon into subject strings that users can comprehend; clarify terms that are ambiguous and artifacts that might have multiple uses; identify purpose, mode of manufacture, material, and design; and may even facilitate and encourage documentation of undocumented artifacts.

We want an approach to classification that is both easy-to-use and respects the individuality of different museums. It should thus be stressed that the approach recommended here allows the cataloguer to move fairly directly from a sentence in an object description to a subject classification. The cataloguer need not master a complex knowledge organization system. Rather, they translate each term in the object description directly into controlled vocabulary. As noted above, complex terms such as atlatl or beamer can be used as long as they are also translated into basic concepts. Museums that have fewer resources or simpler artifacts will produce shorter synthetic subject strings than museums with more resources and complicated artifacts—but both can employ the same controlled vocabulary.

We do not, of course, want the cataloguer to translate an entire description of a paragraph or longer in length into a subject string. The idea is that the cataloguer selects—or concocts—one sentence that captures the essence of the artifact in question. McMarty (2014, 618) reports a common joke in museum informatics: "if museum curators ran libraries, no one would be able to check out a book until it was first explained to them what it was about and why they should read it." Museums will need to appreciate (as librarians long have with respect to documents) that the subject string allows users to find an object; they can then be directed to a longer description that places the object in context.

We have not yet in this paper discussed grammar at great length. But we have provided subject strings above that read like sentences or at least sentence fragments, and

we have now urged museums to move fairly directly from object descriptions to synthetic subject strings. We can also note that users concoct their queries in grammatical form as well: "bronze axe used in war." Our ability to link users to the precise objects they seek will be enhanced if we guide users to employ the same order of terminology that cataloguers employ—and provide users with a search algorithm that privileges the order in which search terms are entered. With regard to the latter, note that such a search algorithm circumvents the main criticism against a post-coordinated approach to classification (that is, one where the cataloguer combines simple terms synthetically rather than choosing from a set of complex headings): that users searching for "philosophy of history" will be guided to many useless (for them) works on history of philosophy. As I have noted elsewhere (Szostak 2015), this result holds only if we limit ourselves to search algorithms available in the 1960s. We should instead design our classifications to accord not with antiquated search techniques but to work with search algorithms designed in concert with the classification. A group of undergraduate computer science students has developed for me a search algorithm that prioritizes the order in which terms are entered in a query.

How can we best guide cataloguers and users to employ the same word order? We can encourage them to structure subject strings grammatically. That is, rather than requiring both to master some set of rules to govern the order in which they enter terms, we can encourage them both toward the standard grammatical approach that they employ in almost every other act of communication in which they engage. Is there enough regularity in grammar to achieve common word order? In Szostak (2017a), I surveyed the basic rules of English grammar and identified the following list of adjustments that a cataloguer (or computer) might make in moving from an object (or document) classification to a standard grammatical format:

- Translating interrogative, imperative, and exclamatory sentences or clauses into declarative format.
- Ignoring pronouns and most determiners.
- Using only the most specific form when nouns are repetitive.
- Translating verbs into the infinitive.
- Using combinations with auxiliary verbs to capture verb tenses.
- Translating phrasal verbs and idioms into synonyms (a task for a thesaurus).
- Placing simple adjectives before nouns but post-adjectival phrases after.
- Using compound adjectival forms to capture gradation.
- Translating adjectival phrases with "that" (or similar words) into adjectival phrases using prepositions or infinitives.

500

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

– Ignoring or translating the rare adverb that does not appear after a verb or before an adjective or adverb.
– Using an extra set of parentheses if necessary (or some other notational device) to clarify whether a modifier is an adjective or adverb.
– Distinguishing adverbs from prepositions when the same word can be used for each.
– Ignoring the first component of a correlative conjunction.
– Addressing inverse verbs, ideally by preferring one form over its inverse.

Most of the time, people employ very similar grammatical constructions. None of the adjustments listed above are then necessary. When people do not employ a standard word order it is quite straightforward to achieve this through a clear and manageable set of adjustments. Note, moreover, that if we fail on occasion and user and cataloguer use different word orders to express the same idea, we are no worse off than if we had no access to anything better than Boolean operators when performing subject search (and we could if desired provide users with the option of employing different search algorithms).

The experience of PRECIS (Preserved Context Indexing System), an indexing system developed by Derek Austin and colleagues for use in the British National Bibliography in the 1970s, is instructive. Though the purpose of PRECIS was quite different—it was designed to identify a number of different subject headings with different lead entries—the designers of PRECIS found it useful to employ grammatical construction within the key elements of their subject headings. This was not the original intent; they had experimented with other types of word order. But since they also were moving directly from document descriptions to subject headings, PRECIS cataloguers found over time that grammar worked best. "The fact that general rules of this kind can be deduced and applied in practical indexing would seem to indicate that natural language is endowed with a greater measure of underlying logic than many classificationists would allow" (Austin 1977, 82). It is also noteworthy that PRECIS was successfully translated into French. It would appear that the differences in grammar between the two languages were not overwhelming. And PRECIS was used across multiple media such as films (Dykstra 1989). PRECIS fell out of use for reasons that had nothing to do with its use of grammatical construction; the success that it had with a grammatical approach—and in moving directly from document descriptions to subject headings—over a period of decades thus suggests that the approach recommended in this paper is quite feasible. Indeed, this approach is far more feasible in the twenty-first century than in the 1970s for we no longer need to assign pre-

eminence to the first term in a search query as was the case when printed indices and card catalogs were the entry point to collections (Szostak 2017b).

The use of grammar has some further important advantages that deserve to be briefly mentioned here:

– Humans think in sentences. We spend our lives uttering, hearing, reading, and writing sentences. We may even have been genetically selected to appreciate basic grammatical constructions. We certainly master these at a very early age (even if we had trouble consciously mastering rules of grammar in elementary school). A subject heading formulated in accord with standard rules of grammar will thus be more readily and accurately comprehended.
– Linguists stress that sentences provide context for the terms within the sentence and thus serve to clarify the meaning of those terms. Scholars of knowledge organization worry a great deal about the ambiguity of individual terms. I argued in Szostak (2011), and again briefly above, that we can reduce ambiguity by focusing on basic concepts. We can now reduce ambiguity even further by placing these basic concepts within subject strings with a sentence-like structure.
– Grammar is a kind of facet analysis. A subject string that follows a grammatical format, and which draws its controlled vocabulary from logically organized schedules of things, relators, and adjective/adverbial properties will have one clear place for each facet appreciated in the literature on facets. (And since grammar itself captures some facets it is much easier to pursue logical subdivisions in our schedules.) A cataloguer—whether of an object or document—that first identifies a sentence that captures the essence of the object or document and then translates this into a grammatical subject string will have identified the key facets of the object or document without having to explicitly perform facet analysis (Szostak 2017a).

The fact that humans think in terms of sentences is particularly important, because information scientists might be tempted to look at the sort of subject strings generated in this paper and think "that's not what a subject heading looks like!" We have become accustomed over the last century and a half to subject headings that defy grammatical conventions. As noted with respect to PRECIS, such subject headings were necessitated in an age of card catalogues or printed indexes, because it was then absolutely essential to search by the first term in a subject heading. But we should not infer from the characteristics of a particular time and place that there is some natural shape to subject headings. It is reported, for example, that the famed Library of Ashurbanipal at Nineveh c.

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

501

1180 BCE had subheadings such as "sheep with arthritic hips" (McNeely and Wolverton 2008). It should thus be stressed here that there is absolutely no reason why a subject heading cannot take the format described in this paper. It is a historical convention that subject headings take the form that they do. Computers now allow searches for any term in a subject string and indeed for combinations of such terms.

5.0 Reprise

Note the advantage here is that each museum starts from a description that it has already fashioned, identifying a key sentence(s) in/from that description. We do not require that these descriptions be rewritten. Nor do we force museums to fit their object into some set of complex subject headings. We only insist on common grammatical phrasing and the use of a controlled vocabulary of basic concepts. Staff can move fairly directly from their own object descriptions to a subject classification. They can very quickly learn to ignore pronouns and follow a very small set of other rules that are required to transform a sentence into a subject description. They can also quickly learn how to employ a thesaurus or search the compact schedules for controlled vocabulary. The grammatical approach will be particularly valuable for museums that lack dedicated staff with knowledge organization training or responsibility.

Yet, we need not worry overmuch that we will end up with idiosyncratic subject classifications. As we have seen, the grammatical approach achieves the aim of facet analysis without requiring cataloguers to actually perform facet analysis. Each facet is represented by its place within a grammatical heading and/or within compact logical hierarchies. For example, in a sentence of form (noun)(action)(noun)(action)(noun), it will be clear that the action verbs fulfill the "operations" facet (in the *Bliss Bibliographic Classification*), the first noun is the "agent," the second the "patient," and the last the "product." The grammatical approach thus guides museum cataloguers to stress key facets. And these in turn will be captured by a shared controlled vocabulary.

The user then enters the terms that are of greatest import to their search in the form of a sentence. If the search interface guides them to controlled vocabulary and standard grammatical format (see below) then they can achieve great precision in their search without any need to comprehend the classification system in use. But if they wish to browse, they can be guided to flat and logical schedules within the classification—or shown how changing one term at a time in their search query guides them to related artifacts.

6.0 Some Examples

In Szostak (2017c) I showed how grammatical subject strings could be derived from document descriptions of a handful of recently published books. I showed that it was easy first to identify (or construct) a sentence that captured the essence of the document, and then to move to a grammatically constructed subject classification employing terminology from the BCC. In Szostak (2016b) I provided examples of synthetic strings for samples of museum, gallery, and archive holdings (though I was not stressing grammar at that point in time).

One item addressed briefly in Szostak (2016b) was a chalcedony cylinder seal. This used to be the first item listed under the Achaemenid dynasty of ancient Persia on the website of the British Museum (but is now harder to find as the museum has moved to the Google cultural heritage interface which stresses virtual reality tours). We can usefully imagine that the curator of this collection wished to provide subject headings for each item in the collection. The subject heading could capture the material, form, and purpose of the item with (chalcedony)(cylinder)(seal). Scope notes in the schedules could define the nature of a seal (to indicate official support for a particular document). Chalcedony is a compound crystal of quartz and moganite. We would want chalcedony to be in our schedule of materials but linked by a thesaurus to crystal, quartz, and moganite. Users searching for these other three terms should be advised of the connection to chalcedony. Users with a general interest in seals or especially cylinder seals, or in items made of chalcedony, and especially cylinder seals made of chalcedony, would be guided to this item—and to similar items held in this and other museums elsewhere in the world. Note that each term in the subject description is a basic concept (though admittedly chalcedony requires a very detailed classification of crystals, or must receive a compound rendering itself). It might also be indicated that such a seal was only (used)(by)(officials). Since seals were generally used by officials, this qualifying phrase would likely be judged redundant. But we can imagine that for other objects, a phrase clarifying who used an object might enable users to perform a more precise search. I discuss in Szostak (2016b) how in a phrase such as (chalcedony)(cylinder)(seal), which has the form (adjective/adjective/noun), there can be some ambiguity as to whether the first adjective qualifies the second adjective or the noun—but suggest that this will generally be clear in context. I noted in Szostak (2017a) that we can indicate the materials facet (one of thirteen facets identified in the *Bliss Bibliographic Classification*) by providing a unique notation to our schedules of materials within our logical classification of things. Since materials modify things, it is thus clear that "chalcedony" is indicating the material of which the seal is

502

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

made. The cataloguer might choose to indicate the high quality of this particular seal—as noted in the object description—by adding yet another qualifier (perhaps "beautiful" or "elaborate") to the subject string. Such an addition may be less useful, especially if there is a tendency for museums to each laud the quality of the items in their collection.

The most important way in which this (chalcedony)(cylinder)(seal) might be distinguished from others is in terms of the pictures inscribed on its surface. Here, the object description provides clear guidance: "The most common subject is a crowned figure wearing Persian dress, here shown fighting a lion alongside a hero in Babylonian dress shown fighting a bull." We could capture these key elements also: (crowned)(Persian)(male)(fighting)(lion) (beside)(Babylonian)(male)(fighting)(bull). These are all basic concepts, and diverse users should attach very similar meanings to each term in the subject string, especially since the string itself clarifies the meaning of each; we know what kind of fight it is because there is a man and a bull involved (The Flora and Fauna schedules of the BCC have not been fleshed out to the level of "lion" and "cow"— bull would be (male)(cow)—but will be; crowned would at present be a compound of (place)(crown)). So the cataloguer might well opt for (chalcedony)(cylinder)(seal) (associated with)((place)(crown))(Persian)(male)(fighting) (lion)(beside)(Babylonian)(male)(fighting)(bull). This is, admittedly, a fairly long subject string. It is quite likely that museum objects will often be described with longer subject strings than library documents (archival documents may also benefit from long strings that describe who produced the document and for what purpose and maybe where and when). Szostak (2016b) showed that notations of manageable length are generated even when several terms are combined in a subject string—because flat logical hierarchies allow short notations for each term. Such a string allows both the user seeking chalcedony seals and the user with a more specific interest in certain types of pictorial representation to find this particular artifact. And, of course, the user with a precise interest in that type of picture on that type of object will achieve great precision in their search. The importance of word order should be obvious: We would lose a great deal of precision in our search if the crowned male were linked as strongly to the bull as the lion in our search algorithm, and especially if he might be thought to be fighting a chalcedony seal. And imagine if the Babylonian were kissing the bull rather than fighting it; it would then be even more important that he be associated more strongly with "kissing bull" than "fighting lion."

Note that the cataloguer has a choice as to how much detail to provide in the subject string. Note further that adding detail does not interfere with the essence of the shorter original string. A user searching for chalcedony seals will still find this particular object whether the details of pictorial representation are provided in the subject string or not. Most importantly perhaps, we moved directly from sentences in the museum's own object description to develop our subject string. The cataloguer could potentially perform this task in about the time it took to read these paragraphs—especially if aided by a thesaurus (see below). Yet, that subject string is composed entirely of terms that are in a shared controlled vocabulary that can be used by all museums and understood similarly by all users. We have thus managed the seemingly impossible task of respecting the individuality of museums while facilitating search across all museums. And since the same exact procedure can be employed also in galleries, archives, and libraries, we can facilitate search across the entire GLAM sector.

The Smithsonian Institution used to have a "highlights" list, but this has disappeared sometime in the last year or so—signalling yet again that museums often redesign webpages and sometimes make it hard to find objects that one found there very recently. The third item on that list was the well-known photograph "Migrant Mother" by Dorothea Lange. It might be captured by (photograph)(of) (poor)(migrant)(worker)(mother)(in)(1930s). But this photograph is well-known, precisely because it captures the woman's desolation, presumably regarding her limited ability to provide for her family. This fact is signaled in the photograph's description. So, we could hope that a cataloguer would add at least (despair) and perhaps best (despair)(because)(poor) to the subject string (replacing just "poor"). Again, this addition does not interfere with those looking for photographs of poor people or poor migrants or poor mothers or poor migrant workers. But it allows precise search for those seeking photographs of "desolation because poor migrant worker mothers." These might be users with some vague familiarity with the photograph in question or simply users with a very clear idea of what they are looking for.

7.0 Thesaurus

The cataloguer in moving from a sentence in an object description to a subject string will need to locate controlled vocabulary. They will be aided in this task by the flat logical hierarchies of the BCC. Yet, the task could be further facilitated by the development of a thesaurus that could guide the cataloguer directly from terminology in the object description to controlled vocabulary. The approach recommended in this paper can facilitate the development of such a thesaurus. As Richmond (1976) noted with respect to the use of (what we would call) basic concepts within PRECIS: "The fact that PRECIS focused on terms

Knowl. Org. 44(2017)No.7

R. Szostak. A Grammatical Approach to Subject Classification in Museums

503

rather than complex headings means it is possible to have a 'true' thesaurus just as if it were a post-coordinated system." That is, rather than trying to link terms to complex and contested subject headings we can strive to identify synonyms and near-synonyms for basic concepts for which fairly precise and shared understandings are possible. If a close synonym cannot be achieved we might then guide cataloguer (and user) to the relevant schedule from which they can choose the most appropriate term. Fairly exhaustive lexicons such as WordNet already exist which can be harnessed to the task. There are also thesauri and object lists within the museum domain—most notably the *Art and Architecture Thesaurus* developed by the Getty institute (http://www.getty.edu/research/tools/vocabularies/aat/), the *British Museum Material Thesaurus* (http://www.vocabul aryserver.com/materials/index.php?letra=H) and *Object Names Thesaurus* (http://collectionstrust.org.uk/resource/ british-museum-object-names-thesaurus/), but also multiple editions of *Nomenclature for Museum Cataloguing* (Bourcier and Dunn 2015)—that can be harnessed to the task of connecting object descriptions to BCC terminology.

8.0 Linked open data

The museum community is intrigued these days by the possibilities of linked open data (LOD). Yet, efforts to employ LOD are often tangential to the purposes of subject classification; objects are linked to their creator or locale rather than readily compared and contrasted or connected to other objects. Linked open data are useful only if different databases employ the same (or interoperable) controlled vocabulary. This is possible at present with respect to names and places; we can all agree on a unique identifier for "William Shakespeare" or "Stratford on Avon." And, thus, we can fairly easily connect an object to a person or place. This may allow the user to move fairly easily from a paintbrush used by a particular painter in one museum to a painting by that painter in a gallery to a series of sketches by that painter in an archive to books about that painter in an archive.

But what if the user wants to identify which painters painted a particular subject or to connect differences in subject with differences in artistic style? The combination of grammatical structure and controlled vocabulary recommended above potentially allows cataloguers and users to draw such connections (especially if supported by appropriate visualization techniques), to identify objects that differ in one important way but are otherwise similar, or to identify objects that go together (say, because they were part of the same ceremony). A user might, for example, wish to compare seals made from chalcedony with seals fashioned from bronze. Such searches require a controlled vocabulary for subject headings which can then be

represented by unique identifiers for each term in that controlled vocabulary.

It is hoped that linked open data will allow computers to draw inferences across databases. If one website asserts that swans are birds, and another that birds have wings, the computer can conclude that swans have wings. Notably, the developers of the semantic web appreciate the limitations of keyword searching—if the different databases in the example above used different terminology for "bird," no deduction would be possible—and, thus, advocate coding with the use of controlled vocabulary. The second point to note is that the form of this coding centers on "RDF triples" of the form (subject)(predicate or property)(object). That is, databases are to be coded in terms of combinations of things, verbs, and adverbs/adjectives. This is exactly the sort of approach recommended above for "coding" the subjects of museum artifacts—though admittedly our subject strings will often be longer than three terms in length. It is possible, though, to translate longer strings into combinations of RDF triples. We can thus potentially provide subject classification of museum artifacts that facilitates not only searches by human users but also searches by computers. Computers, that is, can potentially link artifacts in one museum to artifacts in others, and to books, archival documents, and works of art. And a controlled vocabulary adopted in the museum community would likely be employed more widely, for the semantic web community has signally failed to achieve consensus on controlled vocabulary (Hart and Dolbear 2012).

9.0 Concluding remarks

An approach to subject classification for museums that synthesizes basic concepts according to grammatical rules achieves a diverse set of goals identified in the literature on museum classification: it respects the uniqueness of individual collections; it nevertheless facilitates search across museums—and also potentially libraries, archives, and galleries; importantly, it potentially facilitates search by computers as well as human users; the recommended approach is easy to use both by users and by cataloguers in museums who often have limited resources or training in information science; the approach is flexible such that different museums might provide different degrees of detail in their subject descriptions; the approach allows details of material, manufacture, and use of objects to be combined with details regarding design in one subject heading.

The recommended approach has been applied to small samples of objects or documents from across the GLAM sector. It seems to be entirely feasible—though the resultant subject headings may seem unusual to information scientists accustomed to the format more commonly pur-

504

Knowl. Org. 44(2017)No.7
R. Szostak. A Grammatical Approach to Subject Classification in Museums

sued. The recommended approach, which combines innovation in classification with innovation in search algorithm (and ideally innovation in thesaurus construction), allows greater precision than existing approaches to subject classification. This is because it allows cataloguers to move fairly directly from an object or document description to a subject string. The user likewise can search for precisely constructed combinations of search terms.

References

Austin, Derek. 1974. *PRECIS: A Manual of Concept Analysis and Subject Indexing.* London: Council of the British National Bibliography.

Bliss Classification Association. 2017. *Bliss Bibliographical Classification: Using the Scheme.* 2017. http://www.bliss classification.org.uk/bcclass.shtml

Baca, Murtha and Patricia Harpring, eds. 2016. *Categories for the Description of Works of Art (CDWA),* rev. Patricia Harpring. J. Paul Getty Trust; College Art Association. http://www.getty.edu/research/publications/electronic _publications/cdwa/

Bourcier, Paul and Heather Dunn, eds. 2015. *Nomenclature 4.0 for Museum Cataloging: Fourth Edition of Robert G. Chenhall's System for Classifying Man-Made Objects,* 4th ed. American Association for State and Local History Book Series. Lanham, Md. : Rowman & Littlefiedl.

Cameron, Fiona & Robinson, Helena. 2007. "Digital Knowledgescapes: Cultural, Theoretical, Practical, and Usage Issues Facing Museum Collection Databases in a Digital Epoch." In *Theorizing Digital Cultural Heritage: A Critical Discourse,* ed. Fiona Cameron and Sarah Kenderdine,. Cambridge, MA: MIT Press 165-91.

CIDOC-CRM Special Interest Group. 2015. "Definition of the CIDOC-CRM Conceptual Reference Model." Last modified May 2015. http://www.cidoc-crm.org/ Version/version-6.2

Dykstra, Mary. 1989. "PRECIS in the online catalog." *Cataloging & Classification Quarterly,* 10, nos. 1-2:81-94.

Hart, Glen and Catherine Dolbear. 2013. *Linked Data: A Geographic Perspective.* Boca Raton: CRC Press.

Hider, Philip. 2012. *Information Resource Description: Creating and Managing Metadata. Chicago, IL:* American Library Association.

Marty Paul F. 2014. "Digital Convergence and the Information Profession in Cultural Heritage Organizations: Reconciling Internal and External Demands." *Library Trends* 62, no. 3:613-27.

McNeely, Ian F., and Lisa Wolverton. 2008. *Reinventing Knowledge: From Alexandria to the Internet.* New York: W.W. Norton.

Menard, Elaine, Sabine Mas, and Inge Alberts. 2010. "Faceted Classification for Museum Artefacts: A Methodol-

ogy to Support Web Site Development of Large Cultural Organizations." *Aslib Proceedings* 62:523-32.

Richmond, Phyllis A. 1976. "Classification from PRECIS: Some Possibilities." *Journal of the American Society for Information Science* 27:240-7.

Szostak, Rick. 2011. "Complex Concepts into Basic Concepts." *Journal of the American Society for Information Science and Technology* 62:2247-65.

Szostak, Rick. 2013. *Basic Concepts Classification.* https:// sites.google.com/a/ualberta.ca/rick-szostak/research/ basic-concepts-classification-web-version-2013

Szostak, Rick. 2014. "Classifying for Social Diversity" *Knowledge Organization* 41:160-70.

Szostak, Rick. 2015. "A Pluralistic Approach to the Philosophy of Classification." *Library Trends* 63, no. 3: 591-614.

Szostak, Rick. 2016a. "Employing a Synthetic Approach to Subject Classification across Galleries, Libraries, Archives, and Museums," *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society; Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil,* ed. José Augusto Chaves Guimarães, Suellen Oliveira Milani, and Vera Dodebei. Advances in Knowledge Organization 15. Wurzburg: Ergon, 359-67.

Szostak, Rick. 2016b. "Synthetic Classification of Museum Artifacts Using Basic Concepts." Paper presented at *MW2016: Museums and the Web 2016: The Annual Conference of Museums and the Web, April 6-9, 2016, Los Angeles, CA, USA.* http://mw2016.museumsandtheweb.com/ proposal/synthetic-classification-of-museum-artifacts-using-basic-concepts/

Szostak, Rick. 2017a. "Facet Analysis without Facet Indicators" in *Dimensions of Knowledge: Facets for Knowledge Organization,* ed. Richard Smiraglia and Hur-li Lee. Wurzburg: Ergon, 69-86.

Szostak, Rick. 2017b. "Theory versus Practice in Facet Analysis," In *Faceted Classification Today: Theory, Technology and End Users; Proceedings of the International UDC Seminar 2017, London (UK), 14-15 Sept. 2017,* ed. Aida Slavic, and Claudio Gnoli,. Würzburg: Ergon 259-69.

Szostak, Rick. 2017c. "Facet Analysis Using Grammar." In *Proceedings of the NASKO Conference, Champaign, IL, 14-15 June, 2017,* ed. R.P. Smiraglia and L. Ridenour. https://iskocus.org/publications.php

Szostak, Rick, trans. n.d. *Translation of CIDOC Reference Model.* https://sites.google.com/a/ualberta.ca/rick-szostak/ publications/appendix-to-synthetic-classification-of-museum-artifacts/national-gallery-us-highlights/ translation-of-cidoc-reference-model

United States National Park Service. 2016. *Archaeological Object Name List used by the US National Parks Service,*

Knowl. Org. 44(2017)No.7

R. Szostak. A Grammatical Approach to Subject Classification in Museums

505

annotated by Rick Szostak. https://sites.google.com/a/ualberta.ca/rick-szostak/publications/appendix-to-synthetic-classification-of-museum-artifacts/archaeology-object-name-list-used-by-the-us-national-parks-service

Urban, Richard J. 2014. "Library Influence on Museum Information Work." *Library Trends* 62, no. 3: 596-612.

Visual Resources Association. 2017. "The CCO Commons: *Cataloging Cultural Objects.*" http://cco.vrafoundation.org/index.php.

Zoller, Gabriela and Katie DeMarsh. 2013. "Museum Cataloging from a Library and Information Science Perspective." *Art Documentation* 32: 54-70.