

## Chapter 8

# Benefits of a Comprehensive Phenomenon-Based Classification

With the broad outlines of a novel classification sketched in previous chapters, it is useful to note in this chapter that—while this approach to classification has its challenges—such a classification would have many advantages for classificationist, classifier, and user. We review first the advantages for scholarly users, and then for general users. We close the chapter with a brief discussion of the practical challenges of achieving adoption of the recommended approach to classification. It is argued that certain of the myriad advantages of the new approach should facilitate adoption.

### Advantages for KO and for Interdisciplinary Users

We have in Chap. 5 addressed the feasibility of a comprehensive phenomenon-based classification. That chapter and Chap. 7 necessarily noted certain challenges associated with the endeavor, but argued that they were surmountable. Now that we have sketched what such a classification might look like in practice it is possible to note that it has a variety of advantages as well.

Most centrally, the proposed system more closely captures the unique characteristics of (especially scholarly) works. If most works address how some phenomena affect others, then both classificationist and classifier struggle unnecessarily at present to capture the nature of works under a particular heading. Both will find it far easier if they can freely link concepts and especially phenomena and relationships. Recourse to a standard classification of relationships would spare the classificationist from having to develop a new class for works that engaged a novel relationship among existing things: they could simply employ compound notation of existing things and relationships. If a classifier confronts a book discussing the effect of attitudes toward punctuality on employment patterns in country X, both their task and user retrieval are best served by the use of linked notation rather than for both to try to imagine a unique class heading. The same argument can be made

with respect to works that explore the properties of a single phenomenon: compounding phenomena and properties will be both easier and better than inventing new classes for each possible combination.

To rephrase: we should move away from the idea inherited from the enumerative classifications developed in the nineteenth century that each work should be described by a notation that gives the impression but not the reality of a unique place in the classification. Many of the terms used within existing classifications are in fact compound terms (albeit expressed as if they were simple terms) that contain references to both things and relationships (Szostak 2011). Rather, classifications should begin from the idea that a work's uniqueness reflects the particular combination of phenomena and types of influence (and perhaps properties) that it addresses. An expressive notation, then, would strive to make these distinct elements clear, perhaps by using facet indicators, or spaces between elements, or a different notational base (letters versus numbers versus symbols) for the different elements.

As has been stressed in previous chapters, the classification will be a great boon to interdisciplinary researchers while also aiding disciplinary researchers. The same logic applies to any group: the proposed classification will encourage cross-group cooperation and understanding while also encouraging within-group conversation. The latter task will be enhanced if works are also classified in terms of authorial perspective (see Szostak 2014a).

In addition, such a classification solves many problems identified within existing approaches to classification. Coates (1988, 60), for example, stresses that we cannot, as Cutter had wished, rely exclusively on natural language for subject headings: these will often prove too ambiguous. But we have seen in Chap. 7 that various practices associated with our recommended approach serve to reduce the ambiguity associated with natural language terminology (Szostak 2015).<sup>1</sup> Later (1988, 174) when Coates reviews the shortcomings of existing classification schemes he stresses, 'makeshifts are resorted to in order to present an appearance of solving problems of subject interpolation.' In other words, logical hierarchies are deviated from (see also Mazzocchi et al 2007). Again, compound headings are the obvious solution (Cheti and Paradisi 2008 make a similar point).<sup>2</sup>

Coates on that same page also argues that ideally it should be much easier to add new entries to a classification. He notes that Ranganathan had hoped that at some point a classification would become self-sustaining: that new subclasses would be generated in a straightforward manner by new combinations of existing subclasses. New headings are regularly added to the Library of Congress Subject Headings (LCSH) (Leong 2010). Moreover, internet communities often create naïve

---

<sup>1</sup> Svenonius (2004) had argued that information scientists had to choose to either employ natural language ambiguously, or to provide formal precise terminology quite distinct from common usage. An approach that limits ambiguity potentially allows us to employ natural language precisely (Szostak 2015).

<sup>2</sup> The synthetic approach may also allow the out-performance of purely pre-coordinated or post-coordinated systems. See Szostak (2015).

classifications because they are working on new topics for which formal classifications do not exist. Only with the creation of a usable classification of relationships does it become possible to anticipate that complex new subjects can be readily rendered in terms of combinations of previously identified things and relationships.<sup>3</sup>

Notably, users will achieve better results whether they know precisely what they are looking for (for they can specify a precise causal relationship) or are performing an exploratory search (for they will find works across all disciplines that address a particular phenomenon or relationship in combination with any other thing or relationship). In the latter regard, a unified and comprehensive classification has great potential for enhancing literature based discovery: connections between existing but dispersed pieces of scholarly understanding that are of critical importance to the advance of scholarly understanding (Davies 1989). Coding works by relationship may prove especially valuable in this context.

The advantages noted above and some others discussed elsewhere in this book are summarized in Table 8.1.

## Coping with Information Overload

While the focus of this book is on interdisciplinarity, we have mentioned from time to time that the sort of classification envisaged here would be useful for general users as well. It is worth noting in this regard that the world we live in is itself interdisciplinary in nature. We each in our daily lives face problems that are complex in nature. In our further roles as members of society and citizens, complexity is ever-present. The users of public libraries and online databases of various sorts are thus often pursuing complex queries that span disciplines (Marshall et al. 2009). Certainly few general users limit their searches to only one discipline. And thus classification systems that presume disciplinary mastery disserve the general user.

Present systems of classification contribute to the sense that the world is simply too complex to be coped with. A classification system that clearly provided one and only one place for any bit of information would contribute to a healthier sense that humanity can cope—albeit imperfectly—with the world’s complexity. [As noted elsewhere such a classification would have to be transparent in structure so that users and/or computers could readily navigate it.] A classification system that reflected the efforts of scholars to understand the causal relationships among

---

<sup>3</sup> In recognition of the limitations of LCSH, especially in the digital age, a somewhat faceted application of LCSH, termed FAST (faceted application of subject terminology) has been developed which facilitates synthesis of common geographical, time-related, and other elements of LCSH. Notably the developers of FAST considered developing an entirely new system but decided instead to adapt LCSH in order to maintain its rich vocabulary (Dean 2003, Chan and O’Neill 2010).

**Table 8.1** Advantages of a comprehensive phenomenon-based classification

It more closely captures the unique characteristics of (especially scholarly) works.
The synthetic approach accommodates novel works without the necessity of creating a new class.
Communication across and within disciplines and social groups is encouraged.
Both precise and exploratory searches are enhanced, and thus literature based discovery.
Terminological ambiguity is reduced.
Logical hierarchy can be pursued.
We are able to achieve very precise classifications of a work with limited and expressive notation.
Users can judge the relevance of a particular work in terms of the relationships it addresses and the theories, methods, and perspectives employed.
By distinguishing different sorts of relationship, we enable searches by verb-like terms as well.
While LCC provides specific instructions in multiple places for coding by time or place or people, the proposed system would have a common coding for such elements. This renders both classification and searching easier.
The classification can alert users to closely related linkages.
The use of linked notation serves to place works within multiple hierarchies (and of relations as well as things). It thus facilitates a web-of-relations among works.
It should be possible to translate all search or entry terms employed in other classifications into the terms employed in the classification. Note that in addition we create the possibility of (fairly) automatically coding for new works or for existing works that are at present poorly classified.
It is possible that different databases could adopt this new system (since phenomena and relationships are of general interest), and thus cross-database searching would be facilitated. We expand on this idea below.
The classification aids both scholar and general user in coping with information overload (see below).
The classification appears well suited to the digital age (see below).
We focus above on the practical benefits. Szostak (2015) used a typology of ethical considerations developed by Fox and Reece (2012) to argue that the sort of approach recommended here is also ethically sound: it serves the needs of those facing the most difficulty in information acquisition, it encourages porous borders between disciplines and across social groups, it respects user rights, and it does no wrong.

phenomena would send an important signal that we are gradually advancing our understanding link by link. It would at the same time enhance the ability of scholars to do precisely that by enhancing their ability to find relevant information (see Szostak 2015). As noted in Chap. 2, such a classification would also decrease the chances that information is simply forgotten—either forever or to be unnecessarily ‘reinvented’ later—simply because it was not classified in a manner that made it accessible to those interested in it. In sum, present systems of classification decrease both our sense of coping and our ability to cope with complexity (Szostak 2014b).

In other words, the sort of classification recommended in this book quite simply better captures both the nature of the world and of scholarly understandings of this. The world, we all know, is not neatly divided into disciplinary compartments. Organizing our resources around phenomena and relationships captures the way the world actually works: a host of phenomena influence each other in diverse ways.

Organizing our understandings around disciplines instead supports a sense among users that we are not really grappling very well with the world we inhabit in its manifest complexity. Our collective understandings of that world are best understood in terms of a (large but finite) set of causal links. The recommended classification thus communicates correctly the idea that some set of scholars somewhere has and is studying (almost?) every possible causal relationship. Our collective understanding is imperfect, and more imperfect along some causal links than others. But a classification should communicate to users that there is almost certainly some insight out there into any causal relationship that they might wish to investigate.<sup>4</sup>

Scholars often disagree. It can indeed seem that scholars always disagree. This encourages skeptical attitudes about the possibility of us collectively understanding the world around us well enough to alleviate pressing problems. A better classification system would identify many instances where disagreement was only apparent: scholars were actually talking about different causal relationships or defining the same term in quite different ways. In cases where disagreement was real, the approach recommended in this book would serve to identify precisely what was being disagreed about. This alone is highly significant: the scholarly enterprise can be seen as not a congeries of discordant insights but a coherent exercise with some areas of consensus and some well-defined areas of disagreement. If works are classified in terms of perspectives applied, including theories and methods, then we will often identify the sources of these conflicts (Szostak 2014b). We then set the stage for interdisciplinary strategies for alleviating conflicts (see Repko 2012, Bergmann et al 2012).

Importantly, a classification that actually captures the nature of the world and our understandings of it can serve educative purposes. Students can be exposed to the broad structures of the classification. They will learn both about the nature of the world and about how they can explore our understandings of that world. A further indirect benefit would be an enhanced educational role for school (and other) librarians (Szostak 2015).

Lambe (2011) worries that KOSs need to become more complex as science does, but then they will exceed our human capacity to comprehend. This is an understandable problem within traditional approaches to classification, where scientific evolution generates a bewildering proliferation of new fields. But in a classification grounded instead in phenomena, new scientific fields generally just involve the more intense scrutiny of relationships among phenomena already classified (occasionally, such as on the frontiers of nuclear physics, new phenomena are posited, but the accretion of these is slow). The user, importantly, need not understand the

---

<sup>4</sup>Though beyond the scope of this book, Szostak (2012) urged the pursuit of comprehensive survey articles along each causal link as a further strategy for coping with information overload. These might be updated annually. They would need to be jargon-free and might thus narrow the gap between scholarly and popular understanding, and support better public discourse around policy issues. An informed public could in turn suggest questions, point to shortcomings in scholarly analysis, and suggest possible ways forward.

entire scholarly enterprise (though the proposed classification would help them gain an appreciation of its broad contours) but rather need only know where to find what they are looking for. The accretion of knowledge is not a threat but a benefit within a phenomena-based approach to classification.

Users will often know that they want to find something like ‘stop dogs from attacking mail delivery person,’ and can readily access the relevant works if these are coded in terms of the relevant things (dogs) and relationships (attacking). The more adventurous or scholarly user may instead search across all instances of attacking (a task not facilitated within existing classifications) and find some previously unappreciated similarity or difference across the attacking behavior of different animals. Farradane (1967, 297) noted that ‘The relations between concepts often appear to be absent, but if more than one word is used in indexing or in a search there is clearly an implicit relationship in the mind of the indexer or questioner, and other relations possible between the words would lead to false drops.’ That is, failure to be explicit about relationships in a classification will often lead users astray. Green (2008) notes that even when relationships are captured in a classification the type of relationship is usually not specified; failed searches are thus common. She urges the specification of particular relationships.

One emerging area of research in information science is ‘exploratory search’ and in particular how to use visual aids in guiding users who are exploring possibilities. [Lambe (2011) appreciates that visual design, as well as better classification, can reduce information overload.] A user that starts with some curiosity about ‘dogs’ might be presented with a visual representation of the causal (or other) relationships which connect dogs with other phenomena. Causal links that have received the most attention might be emphasized, but the curious could readily follow links that have received less attention. Causal links to dogs could be distinguished by color from causal links from dogs. Other relationships, including hierarchical, could be represented in other colors or styles. Users could be shown where their search terms fall within relevant hierarchies, and then decide whether to search broader or narrower terms.<sup>5</sup> Such visual aids can be employed to great effect in conjunction with a classificatory approach that stresses phenomena and relationships. One problem faced by researchers in this area is that documents are often not coded in terms of all of the relationships they wish to display. The proposed classification would alleviate this problem as well.<sup>6</sup>

---

<sup>5</sup> The approach recommended in this book usually leads to relatively compact hierarchies. But in areas such as classifying species, complex multi-level hierarchy is unavoidable. The techniques for compacting hierarchy identified by Julien et al. (2013) might then be useful.

<sup>6</sup> Zhang (2014) explored how users might utilize the Medline database in order to find health information. Zhang’s key recommendations for improvements are: a more natural interface to facilitate user questions, more effective hierarchies to facilitate query reformulation, focusing the search results page on accessing answers rather than documents, and providing multiple schemas to help users navigate. We have touched on all of these here: recall that the approach recommended in this book would classify documents (and thus facilitate search) in terms of the key insights that documents contain.

In sum, the proposed KOS would greatly facilitate the general user's search for information of various types. Importantly it would—both in reality and in appearance—support a sense that human understandings are coherent and that we are progressing in our understanding of the world. It thus enhances the educational role of libraries in general and school libraries in particular. Visual aids will work well with the proposed KOS and further enhance the ability of general users to comprehend their world.

## Seizing Digital Opportunities

We noted in Chap. 5 that digitization creates an opportunity for the development of a new approach to classification. But digitization has sometimes also seemed a threat. The increasing use of full-text searching has caused many users to eschew subject searching entirely.<sup>7</sup> Library administrators have wondered if traditional subject catalogues will become obsolete (LaBarre 2007).<sup>8</sup> But full-text searching is not perfect. The terminological ambiguity that we have had much cause to discuss in this book ensures that users find much that they do not want and miss much that they do. Scholars of information retrieval appreciate that more structured searching may be advantageous (Wallach 2006). The sort of classification recommended in this book, which is more easily mastered (and taught) would encourage a much greater use of controlled vocabulary in searching.

Similar arguments can be made with regard to machine indexing of web-based materials. The sheer volume of digital material renders manual classification a challenge. But machine indexing is problematic. As our discussion of the Semantic Web below indicates, the sort of classification urged in this book may both encourage and facilitate manual classification of digital material.

One important development is that there are now a host of accessible digital databases: libraries, archives, museums, government agencies, and a variety of private and non-profit organizations have each placed enormous amounts of information online. Users often want to search across diverse resources.<sup>9</sup> They can use internet search engines but these face the problems of all types of free-text searching. There are also concerns about biases in the algorithms employed in

---

<sup>7</sup> Even here, Sandgrind (2010) makes an interesting observation: “The explosion of information made available by the Internet makes the role of parliamentary libraries and research services more, not less important, as busy parliamentarians need people to filter information for them and to do so in a timely, accurate, and politically neutral way.” This point resonates far beyond parliamentary libraries. The Internet creates challenges as well as opportunities for users of all types. And the best antidote to information overload is information organization.

<sup>8</sup> LaBarre proposed user studies of scholars with an emphasis on how faceted classification might aid them.

<sup>9</sup> Lambe (2011) notes that it is increasingly important to provide access to ‘behind-the-scenes’ science. But this sort of science is generally not to be found in library databases but elsewhere.

internet search engines; among other things these sort results by the links to a website rather than the content of a website. Each resource inevitably organizes its material around some sort of classification system, but different resources employ different systems. It is thus exceedingly difficult to search across different resources.<sup>10</sup>

Indeed the challenge of searching across different resources and databases is quite analogous to the challenge of searching across different disciplines. We again face the difficulties inherent in different terminology and different perspectives. It thus makes sense to seek the solution in the same direction. Database managers recognize an advantage in facilitating search, and thus using a familiar classification. But we can hardly expect firms or NGOs, or even most archives and museums, to master the details of the Library of Congress Classification, or any other major classification used widely in the world.<sup>11</sup> A classification that was organized around phenomena and relationships might be much more attractive (Szostak 2016). A classification grounded in the nature of the world (ontology) rather than the nature of disciplines (epistemology) is much more supportive of the interoperability of databases (and also less subject to change over time) (Gnoli 2012).

Special attention should be paid here to the Semantic Web. The idea behind the Semantic Web is that computers should be able to explore and draw inferences across databases. But it is not hoped that they will do so through free-text searching. Rather each database needs to be purposelessly coded in a manner such that a computer would know what sort of information the database contained. Indeed the founders of the Semantic Web appreciated the advantages for search of having documents coded in terms of controlled vocabulary rather than relying on vague keywords. 'Instead of text, which cannot be processed by a computer without analysis by complex natural language processing algorithms, information is published on the Semantic Web in a structured format that provides a description of what that information is about' (Hart and Dolbear 2013, 29). Coding occurs in terms of RDF 'triples' of the form (phenomenon) (predicate or property) (phenomenon). Computers can only navigate seamlessly across databases if the terminology employed in RDF triples is identical across databases or can readily be translated.

Marcondes (2013) recognizes that the immediate purpose of the Semantic Web is to allow search engines to better identify what a website is communicating. This, though, is the means to a greater end: to 'assist the evolution of human knowledge.' He sees the Semantic Web as the purpose of formal ontology. He cites Gnoli on the need to classify in terms of phenomena rather than disciplines in order to facilitate

---

<sup>10</sup> Even an individual database may face controlled vocabulary challenges. Greenberg et al. (2011) discuss how they developed the HIVE vocabulary to allow translations across the multiple controlled vocabularies used by researchers wishing to deposit data into the Dryad data repository.

<sup>11</sup> Glushko (2013, 290) notes that the standard classification system for supermarkets has only 300 categories, whereas libraries may have over a million distinct records.



the Semantic Web. We would concur with Marcondes in a causal chain that extends from phenomenon-based classification through ontology and the Semantic Web to progress in human understanding.

The Semantic Web is developing slowly, but arguably surely. Some libraries have begun to experiment with RDF triples, and the new Resource Description and Access standard for descriptive cataloging encourages RDF coding (Glushko 2013, ch. 4, 224). Two inter-related sources of delay are network effects and ontologies. The value of the Semantic Web to any adopter depends on the number of other adopters *employing the same controlled vocabulary* (or at least controlled vocabularies linked by a translation device). O'Hara and Hall (2012) appreciate that the network is not yet big enough globally to make the benefits of participating exceed the costs. Not surprisingly, no consensus has been achieved on a particular controlled vocabulary to employ on the Semantic Web. The Semantic Web has come over time to rely on formal ontologies, but there are many of these, and they are each hard to master (Hart and Dolbear 2013 suggest that research has been moving away from this focus on ontology because of these and other problems).<sup>12</sup> The detailed assumptions at the heart of formal ontologies are not only difficult to appreciate but offend the open and democratic values on which the Semantic Web was based (O'Hara and Hall 2012). It would seem that an easier approach to providing controlled vocabulary and syntactic rules to the Semantic Web would be highly desirable, and perhaps essential to its success.

Scholars not only wish to navigate diverse databases but increasingly have web presences that they hope will communicate their insights to a broad audience. Interdisciplinary scholars in particular may find attractive the idea of opening their insights to the Semantic Web by coding these with RDF triples. But—given the challenges to interdisciplinary communication discussed elsewhere in the book—they will appreciate the value of a widely accepted controlled vocabulary so that their insights will be captured by the widest possible audience. Likewise interdisciplinarians will want to take advantage of the Semantic Web in order to search for relevant information and will again appreciate the value of thus searching as widely as possible.

While it would be folly at this moment in time to predict the precise form that the Semantic Web might take in future, there are several reasons to think that the sort of classification recommended in this book (supplemented by a comprehensive thesaurus) might be admirably suited to the needs of the Semantic Web. These are summarized in Table 8.2. These build upon and clarify the general advantages for digitization noted in Table 3.3.

Table 8.2 emphasized classification systems. If we wish to allow database managers some flexibility in terminology, but yet allow the potential for computer navigation inherent in the Semantic Web, then we will also need a detailed and comprehensive thesaurus. After all, one great value of thesauri is mapping

---

<sup>12</sup> A handy introduction to both RDF and the OWL ontologies is at [LinkedDataTools.com](http://LinkedDataTools.com).

**Table 8.2** Suitability of the recommended classification for the Semantic Web

---

Most obviously, what the Semantic Web needs is classifications of entities (things), relationships, and properties. RDF triples utilize these three elements exclusively. Computers need to be able to recognize how the entities, relationships, and properties coded for one database relate to those coded in another database. So there needs to be controlled vocabulary and structured relations for these three elements.

---

We have discussed in earlier chapters how the recommended approach to classification might serve as a stepping stone to an ontology. It thus makes sense to see what other sorts of logical inferences are necessary in order for the Semantic Web to function (Hart and Dolbear 2013, for example, speak of how it may be useful to specify that a river generally flows into a lake or a sea, so that computers might draw connections between references to rivers and to lakes and seas), and add these to this sort of classification. This sort of ‘build up’ strategy might avoid unnecessary complexity in ontology construction. Some of the roles for ontology identified in the literature, such as establishing hierarchy and the nature and boundaries of classes, can be accomplished within a well-designed classification (and RDF allows scope notes), the definitional role of ontology can be accomplished to a considerable degree by logical classification, and the inferences or reasoning facilitated by ontology can arguably be accomplished with simple and flexible syntactic rules, without introducing the unnecessary complexity, confusion, and errors associated with the sort of formal ontologies applied to date (Szostak 2014c).<sup>a</sup>

---

Our approach generates very clear terminology. Upper level ontologies at present are often characterized by impenetrable jargon. ‘There are, however, some drawbacks to using upper ontologies, not least because it can be very difficult for an expert in a particular domain such as GI to understand exactly which of the oddly termed classifications to assign to their concepts. Should a County be classed as a Physical Region or a Political Geographic Object? Is a flood an enduring or a perdurant? It depends on your point of view. These quandaries become even more apparent when confronted with terms like “Non-Agentive Social Object” or “Abstract.”’ (Hart and Dolbear 2013, 13–4). We can achieve exhaustive coverage without resorting to vagueness in terminology.

---

Though the Semantic Web literature appears to pay little heed to the undiscovered public knowledge literature, one could hope that the Semantic Web would indeed facilitate the latter. A further advantage of the ‘build up’ strategy is that we lessen the possibility of excluding some important relationships from consideration (what if some rivers just peter out in the desert or disappear underground?). One of the principles of the classification we have been recommending is that it should be possible to freely combine all terms. A similar sort of freedom should be pursued on the Semantic Web to the degree possible.

---

Critically, computers will face enormous difficulty if a particular phenomenon used in RDF triples might have several different meanings to accord with its place in several different disciplinary hierarchies. It will be much easier if each phenomenon has one unique place in a comprehensive classification.

---

Computer navigation is also aided by the use of strictly logical hierarchies. The sorts of abuse of hierarchy identified by Mazzocchi et al (2007) would require at the very least that each particular deviation from logical subdivision be redefined in a more logical way.

---

A classification of relationships that allowed simple verbs to be combined in order to generate more complex verbs (as in Szostak 2012) would allow computers to readily appreciate connections between these simple and complex verbs.

---

It is worth noting that the literature on the Semantic Web also engages with the distinction between comprehensive and domain analysis. It recognizes that the Semantic Web needs at the very least highly interoperable ontologies, but also worries about the uniqueness of each domain (Hart and Dolbear 2013). The complementary application of domain analysis and pursuit of comprehensive classification urged in this book (see Chap. 6) should thus be well suited to the Semantic Web.

---

(continued)

**Table 8.2** (continued)

---

We will discuss in the next chapter how our recommended KO could provide greater coherence (in both appearance and reality) to the scholarly enterprise. The multidimensional map of scholarship thus envisaged could support a unified Semantic Web.

---

<sup>a</sup>SKOS is a first solution to representing thesauri and classifications in RDF. Being originally developed for thesauri, however, it does not allow representation of notation and its components typical of synthetic classifications. In particular, SKOS representation of ILC has been found to be problematic (Gnoli et al. 2011)

synonyms onto preferred terms (Shiri 2012, 16). We have discussed in preceding chapters how a comprehensive classification both calls for and supports a comprehensive thesaurus. And we have recommended that such a thesaurus include independent verbs, adjectives, and adverbs along with nouns and noun phrases; since RDF triples always include predicates or (adverbial or adjectival) properties, this approach will be critical if a thesaurus is to serve the Semantic Web.

We have also discussed the advantages of expanding the set of terminology through which thesauri indicate the relationships between terms. Shiri (2012, 26) celebrates how thesauri, with their rich semantic relations, can aid exploratory search. Tudhope et al. (2001) recognize that interoperability of thesauri rests on the simple set of relationships long employed, but that there has been with digitization increased calls for augmented set of relationships. A thesaurus with an expanded and precise set of relationships between terms would allow computers to draw much better inferences. Recognizing the costs involved, Tudhope et al. (2001) propose a measured augmentation. It should be noted that a thesaurus which clarified relationships of all sorts would serve some of the purposes of formal ontology.

This may be a moment in time—much like the late nineteenth century—when approaches to KOSs are developed for a particular environment but have long-lasting impacts. Work on the Semantic Web has been dominated by IT professionals but there appears to be an important role for input from experts on classification *at precisely this point in time* to ensure that the Semantic Web evolves in a manner that reflects our understanding of how best to classify. But we can only provide input by focusing on a format amenable to RDF triples.<sup>13</sup> And this means a classification that treats entities (things), relationships, and properties in a manner that facilitates their combination.

---

<sup>13</sup> We should in particular want the Semantic Web to be open to diversity of all sorts. The web-of-relations approach advocated by Olson (2007) in the interests of social diversity involves fortuitously precisely the components identified above as necessary for the Semantic Web (Szostak 2014a).

## Overcoming Classificatory Inertia

'Libraries resist radical change, in part because the existing knowledge structures reinforce themselves' (Searing 1992, 24). We noted in Chaps. 4 and 5 that one important barrier to the development and (especially) use of a new classification system is the cost of switching from systems now in use. The Library of Congress (LCC) and Dewey Decimal Classifications (DDC) have not just over a century of development behind them, but also a paid staff to both oversee adjustments to the classification in order to capture new topics (or better reflect changes in social attitudes) and to classify new works in terms of the classification. Other classification systems generally struggle to maintain viability. How, then, can it be hoped that a new system can be accepted? A system that is not widely adopted cannot support the maintenance required to maintain viability, but adopters will be wary unless confident of viability. There are also network effects, especially online, such that the value of a classification increases as more databases adopt it.

The previous section has provided one powerful answer to this question: There are a host of databases whose managers are unwilling to master a complex system such as LCC or DDC or UDC. But there is immense pressure from users for the adoption of some sort of controlled vocabulary that can facilitate searches across databases. There is thus a potential market for a new classification, where it need not compete with LCC or DDC. As noted above, the cross-database communication challenge is similar in important ways to the cross-discipline communication challenge. Moreover, the sort of synthetic approach grounded in basic concepts urged in this book is at least potentially much easier for database managers to master.

The Semantic Web may be particularly important here. The desirability of a shared controlled vocabulary is widely appreciated but has proven an elusive goal. Any classification system found to serve the needs of the Semantic Web will then be adopted across a wide array of databases. But only a classification of entities, relationships, and properties can do so.

Special note should be made of archives, galleries, and museums. Researchers want to know what types of artifacts each possesses. Archives have tended to classify their documents primarily in terms of provenance. They would perhaps be encouraged to provide more information on the subjects addressed in these documents if this were facilitated by an easy-to-use classification. The same is true of museums and galleries. These increasingly have an online presence, but struggle to precisely identify the uniqueness of the objects they possess (Szostak 2016).

These markets outside of bibliographic classification may facilitate the adoption of a new classification. Its advantages for bibliographic classification itself may then be easier to realize. As noted in Chap. 5, one possibility here is that a new approach could be seen as complementary to existing systems. We should also note that a number of small public libraries in the United States have in recent years switched from DDC to BISAC, the classification employed in most bookstores. They have felt that their users are not comfortable with DDC (Martinez-Avila et al. 2014,

Martínez-Ávila and Kipp 2014). Such libraries might prove open to a classification that was easier for their users to understand. So also might journal publishers prove open to an easy-to-use system for subject classification of journal articles that would facilitate retrieval across disciplinary boundaries.

We live in a world of change, where changes in technology, politics, and economic performance constantly surprise (just decades ago, apartheid, the Soviet Union, and sluggish economic performance in China and India seemed solidly entrenched, to name but a few major transformations). We should thus not presume that—simply because they have been around for over a century—classification systems like LCC and DDC will continue to dominate the world of classification. Kodak dominated photography for a century but failed to adapt to digitization. LCC and DDC themselves struggle to adapt to changes in both technology and scholarly practice. It would be naïve to ignore the challenges in introducing a novel approach to classification, but likewise myopic to doubt that a classification with myriad advantages cannot achieve success.

It must also be emphasized that the hospitality associated with the synthetic approach—new research topics can generally be handled efficaciously through a novel combination of existing terminology—will significantly reduce the maintenance costs of the sort of classification advocated in this book. These lower costs, in conjunction with the many possible uses of the classification, should ensure that this novel approach can succeed in the real world.

## Key Points

The classification outlined in previous chapters has myriad advantages for classificationist, classifier, and user. This chapter began by reviewing the advantages for (especially) interdisciplinary scholars, and then proceeded to discuss advantages for general users.

A classification designed to facilitate interdisciplinarity will also serve the challenges and opportunities of the digital age. It will facilitate searches across databases. And it is suitable in a variety of ways to the needs of the Semantic Web.

It may prove that such a classification is more readily marketed to serve these last needs than to facilitate interdisciplinarity within libraries. But there are also opportunities within bibliographic classification itself. The challenge is to gain a large enough user base to become sustainable.

## References

- Bergmann M, Jahn T, Knobloch T, Krohn W, Pohl C, Schramm E (2012) *Methods for transdisciplinary research: a primer for practice*. Campus, Berlin
- Chan L, O'Neill E (2010) *FAST: faceted application of subject terminology: principles and application*. Libraries Unlimited, Englewood, CO

- Cheti A, Paradisi F (2008) Facet analysis in the development of a general controlled vocabulary. *Axiomathes* 18(2):223–241
- Coates EJ (1988) *Subject catalogues: headings and structure*, 2nd edn. Library Association, London
- Davies R (1989) The creation of new knowledge by information retrieval and classification. *J Doc* 45(4):273–301
- Dean RJ (2003) FAST: development of simplified headings for metadata. [http://www.oclc.org/research/projects/fast/international\\_auth200302.doc](http://www.oclc.org/research/projects/fast/international_auth200302.doc)
- Farradane J (1967) Concept organization for information retrieval. *Inform Storage Ret* 3:297–314
- Fox MJ, Reece A (2012) Which ethics? Whose morality?: An analysis of ethical standards for information organization. *Knowl Org* 39(5):377–383
- Glushko RJ (ed) (2013) *The discipline of organizing*. MIT Press, Cambridge, MA
- Gnoli C (2012) Metadata about what? Distinguishing between ontic, epistemic, and documental dimensions in Knowledge Organization. *Knowl Org* 39(4):268–275
- Gnoli C, Pullmann T, Cousson P, Merli G, Szostak R (2011) Representing the structural elements of a freely faceted classification. In: Slavic A, Civallero E (eds) *Classification and ontology: formal approaches and access to knowledge: proceedings of the international UDC seminar*, The Hague. Ergon Verlag, Würzburg, pp 193–206
- Green R (2008) Relationships in knowledge organization. *Knowl Org* 35(2/3):150–159
- Greenberg J, Losee R, Pérez Agüera JR, Scherle R, White H, Willis C (2011) HIVE: helping interdisciplinary vocabulary engineering. *Bull Am Soc Inform Sci Technol* 37(4):23–26
- Hart G, Dolbear C (2013) *Linked data: a geographic perspective*. CRC, Boca Raton, FL
- Julien C-A, Tirilly P, Dinneen J, Guastavino C (2013) Reducing subject tree browsing complexity. *J Am Soc Inform Sci Technol* 64:2201–2223
- LaBarre K (2007) Faceted navigation and browsing features in new OPACs: a more robust solution to problems of information seekers? *Knowl Org* 34(2):78–90
- Lambe P (2011) KOS as enablers to the conduct of science. Paper presented at the ISKO-UK conference. <http://www.iskouk.org/conf2011/papers/lambe.pdf>
- Leong JH (2010) The convergence of metadata and bibliographic control?: Trends and patterns in addressing the current issues and challenges of providing subject access. *Knowl Org* 37(1):29–42
- Marcondes CH (2013) Knowledge organization and representation in digital environments. *Knowl Org* 40(2):115–122
- Marshall JG, Solomon P, Rathbun-Grubb S (2009) Introduction: workforce issues in Library and Information Science. *Libr Trends* 58(2):121–125
- Martínez-Ávila D, Kipp MEI (2014) Implications of the adoption of BISAC for classifying library collections. *Knowl Org* 41(5):377–392
- Martínez-Ávila D, San Segundo R, Olson HA (2014) The use of BISAC in libraries as new cases of reader-interest classifications. *Catalog Classif Q* 52(2):137–155
- Mazzocchi F, Tiberi M, De Santis B, Plini P (2007) Relational semantics in thesauri: some remarks at theoretical and practical levels. *Knowl Org* 34(4):197–214
- O'Hara K, Hall W (2012) *Semantic Web*. In: Bates M (ed) *Understanding information retrieval systems*. CRC, Boca Raton, FL, pp 325–344
- Olson H (2007) How we construct subjects: a feminist analysis. *Libr Trends* 56(2):509–541
- Repko AF (2012) *Interdisciplinary research: process and theory*, 2nd edn. Sage, Thousand Oaks
- Sandgrind G (2010) Introduction: the purpose, present situation and future of the Parliamentary library. *Libr Trends* 58(4):413–417
- Searing SE (1992) How libraries cope with interdisciplinarity: the case of women's studies. *Issues Integr Stud* 10:7–25
- Shiri A (2012) *Powering search: the role of thesauri in new information environments*. ASIS&T Monograph series, Medford, NJ
- Svenonius E (2004) The epistemological foundations of knowledge representations. *Libr Trends* 52(3):571–587

- Szostak R (2011) Complex concepts into basic concepts. *J Am Soc Inform Soc Technol* 62 (11):2247–2265
- Szostak R (2012) Classifying relationships. *Knowl Org* 39(3):165–178
- Szostak R (2014a) Classifying for social diversity. *Knowl Org* 41(2):160–170
- Szostak R (2014b) Skepticism and knowledge organization. In: Babik W (ed) *Knowledge organization in the 21st century: between historical patterns and future prospects* Proceedings of the 13th ISKO conference Krakow. Ergon, Würzburg
- Szostak R (2014c) The basic concepts classification as a bottom-up strategy for the Semantic Web. *Int J Knowl Content Dev Technol* 4(1):39–51, [www.ijkcdt.net](http://www.ijkcdt.net)
- Szostak R (2015) A pluralistic approach to the philosophy of classification. *Libr Trends* 63 (3):591–614
- Szostak R (2016) Synthetic classification of museum artifacts using basic concepts. Paper presented at the *Museums and the Web* conference, Los Angeles, Apr 2016
- Tudhope D, Alani H, Jones C (2001) Augmenting thesaurus relationships: possibilities for retrieval. *J Digit Inform* 1(8), <http://journals.tdl.org/jodi/index.php/jodi/article/view/181/160>
- Wallach H (2006) Topic modelling: Beyond ‘bag of words.’ Proceedings of 23rd international conference on machine learning, Pittsburgh
- Zhang Y (2014) Searching for specific health-related information in MedlinePlus: behavioral patterns and user experience. *J Assoc Inform Sci Technol* 65(1):53–68