

Ten Long-Term Research Questions in Knowledge Organization

Claudio Gnoli

University of Pavia, Department of Mathematics Library,
Strada Nuova 65, Pavia, Italy 27100, <gnoli@aib.it>

Claudio Gnoli has been working as a librarian since 1994, and is currently at the Mathematics department of the University of Pavia, Italy. His main interest is in theory of classification and its digital applications. He is chair of ISKO Italy chapter, and member of the Executive Committee of ISKO.



Gnoli, Claudio. **Ten Long-Term Research Questions in Knowledge Organization.** *Knowledge Organization*, 35(3/2), 137-149. 67 references.

Abstract: Research can benefit by periodical consideration of its status in a long-term perspective. In knowledge organization (KO), a number of basic questions remain to be addressed in the 21st century. Ten of them are identified and synthetically discussed: (1) Can KO principles be extended to a broader scope, including hypertexts, multimedia, museum objects, and monuments? (2) Can the two basic approaches, ontological and epistemological, be reconciled? (3) Can any ontological foundation of KO be identified? (4) Should disciplines continue to be the structural base of KO? (5) How can viewpoint warrant be respected? (6) How can KO be adapted to local collection needs? (7) How can KO deal with changes in knowledge? (8) How can KO systems represent all the dimensions listed above? (9) How can software and formats be improved to better serve these needs? (10) Who should do KO: information professionals, authors or readers?

1. Introduction

At the International Congress of Mathematicians held in Paris in 1900, the famous German scholar David Hilbert proposed twenty-three major mathematical problems that were unsolved at the time. This list was stimulating and influential to subsequent research in mathematics: one century after, 9 of those problems have been fully solved, 8 more have been solved partially, 4 are considered too loose to be ever declared solved, and only 2 remain completely unsolved.

This story sounds interesting in the context of this special issue devoted to our own field. Knowledge organization is a smaller field than mathematics (though potentially not less important), and it is obviously not our aim to emulate Hilbert's prestige and influence. Nevertheless, his case shows that it is sometimes productive to stop and consider a research field from a more general and long-term perspective. We will try to do this by identifying, on the basis of both currently available literature and reflection,

some relevant questions that look to the present author of more general and far-reaching interest.

2. Ten long-term research questions in KO

2.1 Can KO principles be extended to a broader scope?

Maybe the first thing that one can notice in a KO conference is a wild variety of topics and approaches. There are librarians and information scientists, but also philosophers, sociologists, linguists, information architects and Web designers. Everyone focuses on apparently different problems, though the trained eye can see many connections and similarities. People using different approaches unfortunately also use different terminologies, sometimes failing to realize that they are speaking about inherently analogous problems: a paradoxical situation, as the aim of KO itself is to provide access to information through standardized languages.

But this variety is also a richness, as the field is evolving from its documentary origins, to embrace a

much broader range of disciplines, and to take advantage of interdisciplinary confrontation and discussion. This is one important implication of speaking about knowledge organization instead of just bibliographic classification and subject headings. It also means that knowledge organization can hope to have a greater impact on other knowledge domains and on everyday life and society. Bliss (1929) had already realized that "this is not merely an intellectual interest but has social and economic value ... It is not merely a bibliothecal problem, nor on a higher plane is it a problem solely scientific or philosophic. It concerns all these and also the educational interests and those of social organization."

If we agree that the scope of knowledge organization is broader than traditional indexing of library documents, we need to consider whether the traditional methods and rules still apply to the broadened object of application. Towards the end of the 20th century, there was a great debate on the application of descriptive indexing to new carriers of information, like multimedial and digital documents: some have claimed that we just have to apply to the new materials our well-tested technical tools, like ISBD and MARC, while others believe that the new carriers and forms of knowledge also require fundamentally new ways of treating them, of which the first realizations are metadata element sets and XML; or maybe, the two approaches can be complementary levels of a larger structure, as is suggested by projects using both, such as OCLC's CORC. We are now facing a similar debate concerning subject indexing. Are the conceptual structures developed for indexing knowledge as contained in books and papers also applicable to other information carriers?

A first answer may be that they are, as content is something different from carriers, and can exist independently from it: the Yugoslav wars can be documented in photographs, videos, or websites as well as in books. However, we have also realized that the "content" of an image is something more subtle and complex than the explicit text of a book, as it includes several levels of meaning: primitive features such as colour and shape; objects represented; and inductive interpretations (Panofsky 1955, Greisdorf & O'Connor 2002) of which the exact formalization is problematic (Svenonius 1994; Rafferty & Hidderley 2004). Multimedial and hypertextual organization of contents, as is now common in CD-ROMs and websites, also poses complex problems of cataloguing.

Apart from the technological innovations in carriers, even more radical questions rise, depending on what we exactly mean by "knowledge". If we want to look beyond conventional library and information science, what should we consider to be a document containing knowledge? For example, are not manuscripts kept in archives, objects collected in museums (or better, the whole organized and illustrated expositions of those objects), and monuments visible in buildings and streets, also forms of knowledge, that we are not used to collect in libraries simply because of their format? Indeed, surveyors of the cultural heritage are also busy with making inventories, catalogues, and search interfaces for them (Angeli & Cuna 2006). The vast experience accumulated by libraries up to now, having produced sophisticated tools like cataloguing rules, data models, and online search interfaces, could have very useful results in those fields, where the work on standards and information sharing has begun more recently. On the other hand, a museum object or a monument clearly requires additional information elements not considered in the standard library tools. Thus, it seems that some unification of these methods and formats would be desirable in the future.

At this point the orthodox expert in subject indexing will stand up, to warn enthusiast knowledge organizers that current refined subject tools are not designed for non-bibliographical objects. Those are just objects, while documents are something more complex, as they involve both a physical manifestation and a subject dealt with (Mai 2004a). Indeed, the Classification Research Group (1978) pointed out that classifying a book on Chinese plates is essentially different than classifying Chinese plates themselves. The bibliographical treatment adds some further dimensions that are reflected in classification schemes to the phenomena treated. Apart from the characteristics of the paper medium, like page size, style of print, or illustrations layout, frequently noted components of documents are the disciplinary and theoretical approaches adopted: we will discuss these in sections 4 and 5. Another difference relevant to subject indexing was noted by D.J. Foskett (1958): documents "do not consist merely of descriptions of objects; they contain descriptions of objects in relation to one another. To give you one simple example. The classification of objects in a natural history museum enables us to detect identity between several objects.... But consider 'the corrosion of tinplate by acid fruit products', or 'the Direct method of teaching French in secondary modern

schools'. What genus can be said to contain each of these? What characteristics of division distinguish them? Where would you put them in a museum?"

Nevertheless, obvious connections exist between objects and documents dealing with them. Hjørland & Nicolaisen (2004) noticed that classifications based on the properties of real phenomena, like the periodic system of elements, "form the basis of bibliographical classifications such as, for example, the *UDC* classification. Of course the natural world constrains classificatory work". Thus, KO schemes are at least partially based on the classifications of objects that the appropriate science has developed. In a previous paper (Gnoli 2006b) I discussed two basic principles, common origin and similarity, used in scientific classifications of climates, of organisms, and of musical instruments, as well as in bibliographic classification by authors like Brown, Richardson, Bliss, Ranganathan, Austin, and Dahlberg.

Musical instruments are an especially interesting case of exchanges between bibliographic and object classifications. Indeed, the notation for the main classification of instruments, Hornbostel-Sachs, influenced editions of the *UDC*; in turn, there are music libraries finding it profitable to shelve bibliographic material on musical instruments according to the Hornbostel-Sachs classification. Issues essentially similar to bibliographic ones, like concept combination, facet analysis, and their representation in notation, can be found in several classifications of instruments (Ghirardini & Gnoli 2005). Another notable example of hybrid applications is the use of a well-known knowledge organization system (KOS), the *Getty Art and Architecture Thesaurus*, to index exhibits kept in art museums (Will 1992).

Another possible broadening in scope concerns the organization of knowledge in the form of the research, teaching, and administrative institutions dealing with it. These aspects of knowledge consist more of living processes than of written documents. Nevertheless, they are organized according to schemes, like hierarchies of university departments and schools, or lists of government ministries, sharing many features with bibliographic classifications (usually in more rudimentary and less consistent forms); for example, Granata (2004) suggested adapting the scheme of Scientific Disciplinary Sectors recently produced by the Italian government to organize books in university libraries. If KO researchers want their field to be better linked to social and cultural issues, instead of being limited to the technicalities of book indexing, they should consider

these broadened meanings of "knowledge organization".

2.2 *Can ontological and epistemological approaches be reconciled?*

Poli (1996, 1997) made clear the distinction between the ontological and the epistemological approach to knowledge. Ontology, in its philosophical meaning (not to be confused with the homonym schemes for machine treatment of semantic information), concerns the nature of the known things, especially in terms of the general categories to which they may belong. Issues like the subdivision of a class into kinds and parts, or the acknowledgment that a given concept consists of a process or a static entity, are ontological. Epistemology, instead, is about how humans know the world through their sense organs, and how they process knowledge according to categories both innate and culturally biased.

Knowledge is both epistemological and ontological, as it passes through human perception by its very nature, but also refers to real objects of the world having some intrinsic structure. However, authors in KO often emphasize either one or the other approach. Dewey's main classes follow an epistemological sequence, going back to Francis Bacon, as they are listed according to basic forms of the human intellect producing them, like reason, imagination, and memory; *UDC* main classes are also epistemological, as they are derived from Dewey. Other systems, like the *Bliss Classification*, the *Broad System of Ordering*, and the *Information Coding Classification*, base the sequence of their main classes on a supposedly natural sequence of increasing specificity and complexity of the known objects, hence they are primarily ontological.

In recent decades, Dahlberg (1974, 1978) has worked deeply on the theoretical and conceptual foundations of KO from a philosophical and ontological perspective. Both Dahlberg and Poli have made reference to the philosopher Nicolai Hartmann, who gave new life to ontology in the 20th century. A completely different epistemological approach has been that of domain analysis, recently spreading in information science (Hjørland & Albrechtsen 1995), which starts KO work by studying how domain-specific communities of scholars use terms to denote concepts. Epistemological knowledge organizers often take examples from language and its cultural relativity, and quote philosophers like Wittgenstein and the American pragmatists.

Thus it seems that in the philosophy of KO foundations there are two big “schools,” which flow parallel in quite independent streams. As we said that knowledge is both ontological and epistemological, some reconciliation between them should probably be sought. For example, Szostak (2007a) thinks that one can partially agree with Hjørland’s assumptions and still not give up searching for more objective and effective ways of representing scholarly knowledge. Hjørland & Hartel (2003) themselves acknowledge that taxonomies of naturally occurring phenomena, like living organisms or chemical elements, are an obvious base for bibliographic KOSs: “human knowledge is thus a product of both the world itself and of human interests and capacities”. Although the opposition of ontology to epistemology is somewhat perennial in philosophy, we may hope to see in the future of KO some more complementary integration of the two approaches.

2.3 Can any ontological foundation be identified?

Starting with the ontological approach, we may observe that it needs to provide some good foundation in order to justify itself. It is especially difficult for an ontological system to escape criticism from the contemporary perspective of multiculturalism, which provides a sound source of opposition to idealistic systematizations. Clearly, ontological foundations should be as culturally neutral and as generally agreed as possible, in the same way as the items of an international encyclopedia try to be, without falling into sterile relativism.

The “new ways of ontology” (Hartmann 1942) appear suitable to KO purposes in being pluralistic, in the sense that they acknowledge the richness and variety of the real world, and try to model it according to all its intrinsic categories (like time and space, but also value and function) without privileging any special concept (e.g. materialism or spiritualism). Notions developed in philosophy of science, like integrative levels, general systems, or complexity, can result in useful methods for the arrangement of known phenomena into a coherent system (Dahlberg 1974, D.J. Foskett 1978). Though inevitably connected with philosophical speculations, due to their need for generality, KO foundations should be independent on any specific philosophical system. Ideas like integrative levels or complexity can fit different philosophical views (e.g. reductionism or holism), and should be used in KO just as a general structuring principle (Beghtol 1994, 121-122, notes 4-5).

An aspect recently emphasized in ontology is the dynamism of the world (Seibt 2004). We no longer see entities as necessarily eternal and stable, but as the product of processes and interactions with other entities. As has already happened in many domains, an evolutionary approach to KO can help to represent the diachronical relationships between objects (Gnoli 2006b).

Supporters of cultural relativism insist that it is better to focus on domain-specific schemes, so as to be aware of their epistemological premises, as any scheme will be biased by the cultural environment in which it has been conceived (Hjørland 2004). However, reference to a general scheme is needed even while indexing special literature (D.J. Foskett 1991). If we really want to enable interoperability between different schemes and interdisciplinary research, we will always need some general scheme, at least as a switching device between systems based on different epistemologies. Thus, the need for ontological research in KO is far from being obsolete.

2.4 Should disciplines continue to be the structural base of KO?

Most KOSs are structured according to a list of canonical disciplines, trying to reflect how knowledge is organized by the community of scholars in each domain – what Bliss called the “academic consensus”. This, however, produces problems for cross-disciplinary knowledge retrieval, especially when (a) new interdisciplinary domains, not provided for in the existing schemes, arise at the boundaries between older disciplines, as in the case of biotechnologies, environmental sciences, and ethnomusicology, or as interests crossing disciplines, like women’s studies and Oriental studies (López-Huertas 2006); (b) the corpus of a discipline includes knowledge relevant to scholars of other disciplines, who are not familiar with its terminology and canonical organization. These situations are becoming more frequent in our age of globally shared information. “It is thus probably no longer possible to specify one clearly defined user group for an information resource. For this reason, classification research needs to curtail local emphases and to augment culturally neutral internationalization” (Beghtol 1998b).

To solve such problems, several authors over time have suggested that schemes should be defined more in terms of single phenomena than of disciplines. This idea appeared already in 1906 in James Duff Brown’s *Subject Classification*, and later informed the

Classification Research Group's attempt at building a new general scheme based on phenomena, which partially evolved into Derek Austin's *PRECIS*. Still, the most widespread classification schemes like *LCC*, *DDC*, and *UDC* are disciplinary, and inherit their top-level structure from the segmentation of knowledge as it was conceived in 19th century. Some of them offer ways to represent phenomena treated in an interdisciplinary way, and their editors think that these devices can be an adequate response to the needs of interdisciplinarity (McIlwaine 2000).

Others, however, wonder if this implies a need for more radical innovations, eventually leading to completely new, non-disciplinary schemes (Beghtol 1998a, 1998b, Williamson 1998, Gnoli 2006a, Szostak 2007b, ISKO Italia 2007). This view considers that the function of KO is not only to represent the status quo of how knowledge has been organized in documents until now, but is also to suggest new paths of research by connecting concepts previously studied in specific contexts. That is, to exploit the mass of the "interdisciplinary undiscovered public knowledge" hidden in published works of which the relations and implications have not yet been noticed (Davies 1989, Beghtol 1995, Szostak 2007b).

Thus, both existing and new systems should be equipped with ways to retrieve information on a given phenomenon independently of the disciplinary context in which it appears, as well as ways to specify the disciplinary perspective adopted in studying a phenomenon in a given document. It seems that this implies a separate representation of the two subject components of phenomena and disciplines, instead of merging both in a single concept listed in schedules. It should be made clear how phenomena are to be treated in a discipline-based KOS, as well as how disciplines are to be treated in a phenomenon-based KOS (Gnoli 2005).

2.5 How can viewpoint warrant be respected?

One outstanding claim of the epistemological approach is that knowledge organization can be different to different communities. Disciplinary main classes are taken as useful in that at least they represent the most widespread research approach of contemporary western scholars. On the other hand, they cannot be so helpful to users exploring innovative interdisciplinary fields, to those not adopting scholarship perspectives, like spare time readers (Hartel 2003) or children, and to those interested in perspectives other than the contemporary western dominant culture.

It has been shown how the terminology used in KOSs can be biased by culturally dominant groups, like middle-class white males (Olson 2002). This can produce problems in using them in different contexts, like women's studies and feminism (Kublik et al. 2004) or gay and lesbian communities (Campbell 2004). A cultural bias can even be observed at the level of the segmentation of the semantic space in different languages (e.g. Mai 2004a). This especially affects verbal KOSs, like keywords or subject headings. Thesauri add to the vocabulary component the specification of relationships holding between terms, thus producing a more abstract conceptual structure, though still focused on terminology. Severino (2005) has discussed the capitalist-biased use of the term "development" in five thesauri of international organizations, showing that they treat this concept only in economical terms, while failing to account for the human, social, and cultural sides of development. In classification schemes, a notational symbol can stand for a concept represented by one or more words or phrases in different languages. The possibility of crossing language boundaries is a factor explaining the wide use of *UDC* in countries speaking languages of limited diffusion, like those of Eastern Europe.

Of course, the use of symbols does not make classification completely independent from cultural biases, as its semantic structure will still be based on cognitive categories which are not necessarily universal, e.g. those dependent on the deep structure of Indo-European languages and culture. An excellent example of this is the experience with the Korean translation of *DDC*, where the necessity has emerged that calligraphy be a main subclass of the arts, with many specific subclasses for styles and types of writings, instead of being expressed only by a very specific number and poorly developed, as in western *DDC* editions (Kwaśnik & Chun 2004). Similar problems arose while translating kinship structures represented in *LCC* and *DDC* into fourteen different languages (Kwaśnik & Rubin 2004). An alternative approach to indexing materials of different cultures is to design a completely new scheme reflecting their categories better, as was recently done with the *Brian Deer Classification*, used in a library devoted to Canadian indigenous peoples (MacDonell et al. 2003).

To face such problems, Beghtol (1998b) has proposed the notion of viewpoint warrant, which "would presumably have the advantage of providing infinite hospitality for adding any viewpoint—cultural, multidisciplinary, disciplinary, or sub-disciplinary—that

might arise in future," "to be able to support multiple perspectives in a looser structure."

As noticed in section 3, one would always need a general system, that should be as neutral and colourless as possible, to act as a switching device between the different specific viewpoint subsystems. On the other hand, each of these could be used as the preferred KOS by implementers and users adopting a particular viewpoint. The concept of, say, the magic attributes of a given plant in a traditional culture could be accounted for in a place in the scheme reflecting the classification of the world from that viewpoint, and at the same time be linked to the definition of that plant in the basic neutral scheme. This would also serve the important ethical requirement of preserving cultural diversity by representing it appropriately in KOSs (Beghtol 2002).

Users of a system should be allowed to switch between different viewpoints, both to choose their preferred one, and to explore how related knowledge is expressed from different perspectives. Beghtol's suggestion has been echoed by Preuss (2004):

Integrating all these different viewpoints or layers of local knowledge into the universal tree of the classification would stimulate a more experimental and transdisciplinary approach to knowledge discovery, providing a tool for cross-fertilization of what seemed to be incommensurable approaches to knowledge organisation ... the underlying universal classification acting merely as a black box, an universal 'engine' for local 'mods', as you would say in the language of first-person-shooter [video-games].

2.6. How can KO be adapted to local collection needs?

A similar but distinct problem is to serve the preferences of local user communities. This should not be done at the level of the scheme itself, as international interoperability requires that the same document be always indexed by the same classmark, reflecting objectively its content, including the perspective adopted in it. Local adaptations should instead be applied to the arrangement of specific collections intended for a specific target, that is, they should have a standard classmark but a local shelfmark.

General KOSs often pose problems to local users. A Swiss library devoted to Alpine local documentation would hardly make use of the main classes U

and V in the *Library of Congress Classification*, meaning respectively military sciences and naval sciences, as Switzerland has been militarily neutral for a great many years and it is land-locked; on the other hand, it would require that typically local concepts often treated in its collection, such as Rhaeto-Romance languages, be treated with shorter symbols than the long ones provided in the general scheme (Zuccolo 2006). A nice example was offered by Langridge (1992), describing the arrangement of books in the Avalon Library specialized in occultism and New Age: every subject, like astronomy or health, is there viewed in the unusual perspective of the collection specialization.

This problem in itself is not new. Indeed, Ranganathan (1967, sections DG 34-35) solved it by providing his *Colon Classification* with a notational device to express the "favoured host class", that which is a priority for the library although not being a main class in the general scheme. By representing it with a symbol (0) having lesser ordinal value than the other classes, the documents dealing with it will be filed at the beginning of the shelves, or of the browsing interface.

What is more and more topical is the increasing worldwide availability of subject information, through online catalogues, Web directories, metadata formats and links between knowledge bases hosted in servers throughout the whole globe. The idea of a Semantic Web using all taxonomies provided by local knowledge publishers shows the current relevance of this trend. International knowledge exchange obviously requires some standard format for data, and this pushes us towards global KOSs. Organizing documents for local users is no longer the only purpose of cataloguing: the Internet makes remote uses, like interlibrary loan or direct online access, much more frequent, and conversely a single user needs to find and integrate materials from a great number of information sources scattered throughout the world. How can systems interoperate effectively without missing the richness and specificity of local knowledge? Again, it is a problem of mapping schemes designed for different purposes and allowing users to shift between them through appropriate relations.

2.7 How can KO deal with changes in knowledge?

A classic problem in KO is the developing nature of knowledge, which makes schemes obsolete as time passes. This requires that KOSs used for many decades be updated, producing new editions that have

the advantage of including more modern treatment of subjects, and the disadvantage of requiring changes of practice and expensive re-indexing of previous material. Successful KOSs like the *DDC* seem to have found some balance between the needs for stability and innovation, as their new editions now appear at regular intervals.

Apart from the practical issues, more fundamental problems arise from the observation that KOSs change in time. Is KO time-dependent? This appears to be the case, if we look at old systems like the 17th-century Wilkins's *Philosophical Language* (Vickery 1953). Classes that make us smile today were quite obvious and serious at the time they were conceived. This suggests that the same destiny is awaiting current systems: they could become nothing more than witnesses of the state of knowledge in our time, but be useless for their original purpose.

It's true that knowledge changes in time, and KOSs change with it. However, this change is not random, but rather oriented towards an increasing understanding of the many aspects of our complex world. At least, after some 150 years of published research in bibliographical subject indexing, we can learn from experience and foresee general trends in the future of our systems. According to Tennis (2006):

A vital part of classification theory inheres in its self-reflection on its place in the history of document use, information agencies, and human communication practices.... By understanding the similarities and differences of classification work throughout time and place, knowledge organization research gains another view into the nature of classification."

We know, for example, that the continued use and spread of a system has an additional value in itself, making it worth maintaining and adopting even in new projects despite its theoretical limitations. Classifying psychology as a subclass of philosophy is clearly not an optimal choice in the light of contemporary knowledge, still we can be willing to do it if the return is using a system, like *DDC* or *UDC*, shared with many others and equipped with good tutorial and distribution supports. A.C. Foskett (1996) emphasizes how the managing organization is one key factor in the success of a KOS.

The existence of successive editions of schemes, and of schemes conceived in different epochs, implies the need of ways to treat relations between

them. As economic limitations prevent continuously re-indexing documents according to the most updated system, we have to cope with an increasing mass of knowledge indexed according to older schemes. Some have even suggested that old documents should be indexed according to the scheme of knowledge which was valid at the time they were published, as it best matches the conceptual organization of their contents. Thus, this approach could be both practically useful and theoretically correct. We could move from the issue of re-indexing old documents to the provision for links between old and new schemes. In most systems users cannot see any link with how their subject was treated in previous editions, although they can be interested in it, both better to exploit the system, and to have a representation of the changes that have occurred in human knowledge itself (Tennis 2002).

These links appear to be similar to those between systems originating from different cultures and perspectives, the only difference being that the distance is chronological rather than geographical or ethnic. Thus we are again wondering whether different knowledge systems are incommensurate, or can be mapped in some way. "Traduttore traditore" is an Italian saying, meaning that every translation implies a loss in shades exclusive to the original language, that cannot be reproduced faithfully in the target one. Still, translators are respected and increasingly sought professionals, as they are the only means to provide a large number of people with access to foreign information. In the same way, the unattainability of a perfect and eternal KOS does not mean that all KOSs are equally good, nor that KOSs are useless.

2.8 How can KOSs represent all these dimensions?

We have so far reviewed several dimensions which appear to be relevant in the organization of documents: their material forms (1), the ontological status of the phenomena they treat (3), their disciplinary and theoretical approaches (4), the viewpoints they express (5), the local preferences of users accessing them (6), the historical changes in knowledge systems (7). Dimensionality is a desirable value in KOSs (Tennis 2002). Thus, if these dimensions are found to be relevant, they should also be represented in KOSs, in order to be accessed by those users who want to retrieve information on them.

This is not the case in most current KOSs: too often, some media cannot be indexed adequately, or ontological relationships are poorly represented, the

theoretical approach or the viewpoint is not expressed, or local needs cannot be served effectively, or historical changes are more a limitation than an additional access point, and so on. Knowledge is a complex thing comprising many layers, while most KOSs have a flat structure that forces contents into the Procrustean bed of only one or two of these layers. Thus the eighth question is how to develop systems more efficient in representing all the relevant dimensions of the content of documents.

One answer seems to be available already. Facet analysis is acknowledged to be a fundamental improvement in KO. More generally, facets, together with phases, common auxiliaries, and other elements, can be part of an analytico-synthetic system, that is, a system allowing for a free combination of concepts to build specific compounds reflecting carefully the indexed contents. Thus one could define analytic elements for all the dimensions needed to be represented, not only classic facets within disciplines, but also theoretical approaches, viewpoints, historical context, or degree of fictionality (Beghtol 1994, 1998a). Analogously, phenomena and disciplines could be connected by an accordion-like device (Gnoli 2005), and the theories and methods adopted could also be expressed separately (Szostak 2007b) and thus made searchable.

Going down this route, one can obviously analyze anything, and express any component separately. However, this will end by producing extremely long and complex headings. Experience with classification systems has shown that notation has to be reasonably simple and brief, if it has to be copied and remembered by users. Some have observed that in machine processing this is not a problem, as notation can be handled by machine, and users will focus on the verbal captions, both in searching and looking at displayed results. In any event, the problem is not really limited to notation, but is a more general cognitive issue.

Anything can be decomposed into semantic factors, and it is doubtful whether we will ever arrive at "primitive" elements, like those imagined in Leibniz's *ars combinatoria*. Water can be described as a combination of hydrogen and oxygen; but hydrogen can in turn be described as a combination of one proton and one electron, and a proton as a combination of quarks. Clearly this does not mean that a user interested in water concerns in desert countries will take advantage of getting information on quarks. He/she will rather appreciate those emergent properties which are relevant at the integrative level of

water, though not existing at the level of quarks. Therefore, the right degree of analysis must be identified, in order to provide our indexes with it, but not more than it. We have to find where the optimal boundary lies between analysis and synthesis.

2.9 *How can software and formats be improved to better serve these needs?*

One essential part of a KOS is its implementation in search interfaces. Much information is now available in databases searchable online, including library catalogues (OPACs), bibliographical services, factual knowledge bases and websites of institutions and projects. However, this does not guarantee that their knowledge is well organized and exploitable. Unfortunately, often the opposite happens: while KO experts are busy with developing and improving sophisticated systems, the bulk of actual information sources do not use them.

This is the case even with the most classic and consolidated KOS application, that is, library catalogues (Svenonius 1983). A recent survey has listed the wide bibliography concerning subject access in OPACs (Casson et al. 2004), and by checking a sample of 152 catalogues has confirmed that most of them still offer poor tools for subject searches, despite librarians being skilled in creating subject headings and *DDC* numbers. This situation is probably due more to organizational and policy matters than to technical limitations: database managers need to acknowledge the value of KOSs and their specific search needs, such as managing tables of equivalence between notation, captions and synonyms, or appropriately recording and displaying cross-references. Therefore, one first need is simply a greater integration and communication between indexers, reference librarians, computer scientists and information architects.

One second level of the problem resides probably in software and data formats. KOSs are complex objects, and they need to be appropriately represented in databases if they are to be fully exploited in search interfaces. This requires not just a flat table of terms, but a relational system able to manage all hierarchical, associative and equivalence relationships present in the system. In analytico-synthetic classifications, many relations can be represented by expressive notation, and can thus be searched in clever ways (Gödert 1991, Slavić & Cordeiro 2004, Broughton & Slavić 2007), as is the case of *UDC* but not of *BC2*. Additional fields, though, can also help to manage rela-

tionships not expressed in notation. Users should be given hints of links existing from the term they have searched for and other terms more or less strictly related with it (Bates 1998), including those lying at other integrative levels (*mountains* vs. *alpinism*), used in other disciplines (*water vapour* vs. *steam*), or by other discourse communities (*categorization* vs. *classification*). Alternative citation orders of facets and phases can be specified by machine-readable rules, in order to serve local preferences (Broughton & Slavić 2007). Projects like Hibrowse-VBS, Devadason's online classaurus, FACET, FATKS, and ILC have began to use faceted classification for online information retrieval, joining the database layer with search interfaces programmed with script languages. Their techniques could be extended to analytico-synthetic treatments of other dimensions of documents.

Standard formats should also represent these structures adequately. UNIMARC provides some fields (661-668) for the combination of concepts in *LCC*, *DDC* and *UDC*, but these are hardly used, and do not cover facets and phase relationships anyway (Cordeiro & Slavić 2002, Slavić & Cordeiro 2004). Thus, the semantic richness of KOSs gets lost when data are exchanged through different catalogues, or used in union catalogues and meta-catalogues which gather data from archives using different formats.

A similar problem is now pressing towards standardization of XML data and the publication of ontologies to be shared through the World Wide Web, in order to improve global interoperability (Zeng & Chan 2004). A key issue in this process is the representation of KOS structures in XML/RDF syntax (Slavić 2005, Schmitz-Esser & Sigel 2006), of which the SKOS language is a first important achievement.

The NKOS (Networking Knowledge Organization Systems) initiative is also focusing on this important but still underdeveloped terrain of integration between machine formats and the conceptual structures developed in decades of library and information science (Tudhope & Koch 2004). It seems that, at the present stage, we have all the pieces needed, but are still waiting to see the resultant building, that is widespread semantically rich information search and display.

2.10 Who should do KO?

Traditionally, the agents of KO are information professionals, who are trained in using KOSs and applying them to indexing documents. However, with the

explosive increase in the availability of digital documents, it is not possible for the information professional to keep pace with their publication. This means that a large portion of new documents, especially those published locally, or in digital form only, are not indexed by standard cataloguing agencies like national bibliographies. This has generated the need for providing at least some rough information about their content.

One solution is that authors themselves provide metadata for their documents. The Dublin Core Metadata Initiative has offered a standard format to record the main metadata, including semantic items such as "subject/keywords" and "description". These can contain uncontrolled terms, or even terms taken from well established KOSs. This practice has raised the obvious question that metadata provided by authors is far from being standardized and controlled: some authors say that in any event they are better than no indexing at all, while others observe that the selection of documents of relevance and quality making them worthy of being acquired and catalogued is part of the tasks of information management.

On the other hand, this practice is not entirely new, as for decades authors of specialized papers have been asked to provide their own documents with keywords; and other authors in the same specialized field review and index the documents of their colleagues, thus replacing professional indexers who are too few or too little specialized to cover all the literature. The bibliographic resources thus produced may not implement KO theory perfectly, but are widely used.

A more original concern is ethical: in many cases, especially where commercial activities are involved, authors can have a personal interest in indexing documents in a biased way. A simple example is the webmaster of a firm producing X who types in the metadata of the firm homepage a false statement, like "the only X producer in the region Y" while others actually exist. These cases show the valuable role of professional indexers, acting as a disinterested third party between authors and readers, in the same way as judges act as neutral third parties between prosecution and defence (Ridi 1999).

Only very recently has the third logical possibility become popular: that is, KO is done neither by information professionals nor by authors, but by readers. Indeed, network technologies make it possible that a mass of readers have access to a collection of documents, like blog posts or photographs shared by

other users, and add their own “tags” describing them. KOSs emerging by the accumulation of terms used by readers are called folksonomies (Quintarelli 2005) and are being increasingly discussed in KO conferences. Supporters of them emphasize their democratic aspect, as anyone can use his/her preferred terms, thus overcoming some of the problems of question 5, without being forced into the rigidities of a pre-produced scheme. Critics, on their side, emphasize the obvious lack of vocabulary control. To face the latter, the recent trend of folksonomies seem to go in the direction of some mediation by a central group of experts, trying to improve the uniformity of the system while still starting from bottom-up generated terms. One suggested way of organizing terms is, once more, facet analysis. A similar path has been followed by Wikipedia, maybe the most wonderful product of network information sharing: after the spontaneous production of any kind of information items, tools for top-down coordination and classified indexes are developing.

While these new forms of KO do not add much to the theory of our field, they are relevant for the socio-cultural use of information. The future of KO has to face not only technical, but also pragmatic questions. If the most sophisticated and developed KOSs, like general faceted classifications, are published and updated slowly by small organizations with important economic limitations, will they be able to survive the concurrence of no-cost indexing by non experts, and that of automatic classification? Or will they join forces with them to produce even better systems? Will we see any full implementation of them in easily available and numerically relevant information interfaces?

3 Conclusion

Hilbert's list, though acting as an important point of reference, did not foresee several outstanding developments in 20th century mathematics, and surely we are doing the same. Anyway, it is always useful to have a list of problems, even to discuss it, to change its terms, or to add more items. It is hoped that the present discussion will serve this purpose.

References

Angeli, Gabriele and Cuna, Andrea. 2006. *La catalogazione dei beni culturali: oggetti e opere d'arte*. Roma: Aracne.

Bates, Marcia. 1998. Indexing and access for digital libraries and the Internet: human, database, and domain factors. *Journal of the American Society for Information Science* 49: 1185-1205.

Beghtol, Clare. 1994. *The classification of fiction: the development of a system based on theoretical principles*. Metuchen and London: Scarecrow.

Beghtol, Clare. 1995. 'Facets' as interdisciplinary undiscovered public knowledge: S.R. Ranganathan in India and L. Guttman in Israel. *Journal of documentation* 51: 194-224.

Beghtol, Clare. 1998a. Knowledge domains: multidisciplinarity and bibliographic classification systems. *Knowledge organization* 25: 1-12.

Beghtol, Clare. 1998b. General classification systems: structural principles for multidisciplinary specification. In Mustafa el Hadi, Widad, Maniez, Jacques and Pollitt, A. Steven, eds., *Structures and relations in knowledge organization: Proceedings of the Fifth International ISKO Conference 25-29 August 1998 Lille*. Würzburg: Ergon, 89-96.

Beghtol, Clare. 2002. A proposed ethical warrant for global knowledge representation and organization systems. *Journal of documentation* 58: 507-32.

Bliss, Henry Evelyn. 1929. *The organization of knowledge and the system of the sciences*. New York: Holt.

Broughton, Vanda & Slavić, Aida. 2007. Building a faceted classification for the humanities: principles and procedures. *Journal of documentation* 63: 727-54.

Campbell, Donald Grant. 2004. A queer eye for the faceted guy: how a universal classification principle can be applied to a distinct subculture. In McIlwaine, Ia C., ed., *Knowledge organization and the global information society: Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK*. Würzburg: Ergon, 109-13.

Casson, Emanuela, Gnoli, Claudio, Slavić, Aida and Visintin, Giulia. 2004. *Opac semanticici. Bibliografia*. University of Pavia. Department of Mathematics. Available <http://www-dimat.unipv.it/biblio/sem/bibl.htm>.

Classification Research Group. 1978. CRG bulletin 11. *Journal of documentation* 34: 23.

Cordeiro, Maria Inês & Slavić, Aida. 2002. Data models for knowledge organization tools: evolution and perspectives. In López-Huertas, María J., ed., *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the*

Seventh International ISKO Conference 10-13 July 2002 Granada. Würzburg: Ergon, 127-34.

Dahlberg, Ingetraut. 1974. *Grundlagen universaler Wissensordnung.* Pullach bei München: Verlag Dokumentation.

Dahlberg, Ingetraut. 1978. *Ontical structures and universal classification.* Bangalore: SRELS.

Davies, Roy. 1989. The creation of new knowledge by information retrieval and classification. *Journal of documentation* 45: 273-301.

Foskett, Anthony Charles. 1996. *The subject approach to information.* 5th ed. London: Library Association.

Foskett, Douglas John. 1958. *Library classification and the field of knowledge.* London: Chaucer House.

Foskett, Douglas John. 1978. The theory of integrative levels and its relevance to the design of information systems. *Aslib proceedings* 30 no. 6: 202-208.

Foskett, Douglas John. 1991. Concerning general and special classifications. *International classification* 18: 87-91.

Ghirardini, Cristina and Gnoli, Claudio. 2005. Zampogne e libri sulle zampogne: Classificazioni diverse? *Bibliotime*, n.s., 8 no. 3. Available at <http://www2.spbo.unibo.it/bibliotime/num-viii-3/gnoli.htm>.

Gnoli, Claudio. 2005. BC2 classes for phenomena: an application of the theory of integrative levels. *Bliss classification bulletin* 47: 17-21. Also available on *DLIST*, <http://dlist.sir.arizona.edu/920/>

Gnoli, Claudio. 2006a. The meaning of facets in non-disciplinary classifications. In Budin, Gerhard, Swertz, Christian, and Mitgutsch, Konstantin, eds., *Knowledge organization for a global learning society: Proceedings of the Ninth International ISKO Conference 4-7 July 2006 Vienna.* Würzburg: Ergon, 11-18.

Gnoli, Claudio. 2006b. Phylogenetic classification. *Knowledge organization* 33: 138-52.

Gödert, Winfried. 1991. Facet classification in online retrieval. *International classification* 2: 98-109.

Granata, Giovanna. 2004. Classificazione e ordinamento delle raccolte nelle biblioteche universitarie: una proposta "user-oriented". In: Biagetti, Maria Teresa, ed., *L'organizzazione del sapere: Studi in onore di Alfredo Serrai.* Milano: Bonnard, 147-65.

Greisdorf, Howard and O'Connor, Brian. 2002. Modelling what users see when they look at im-

ages: a cognitive viewpoint. *Journal of documentation* 58: 6-29.

Hartel, Jenna. 2003. The serious leisure frontier in library and information science: hobby domains. *Knowledge organization* 30: 228-38.

Hartmann, Nicolai. 1942. *Neue Wege der Ontologie.* Berlin, 1942. Engl. transl.: *New ways of ontology.* Westport: Greenwood.

Hjørland, Birger. 2004. Theory of knowledge organization and the feasibility of universal solutions. Paper presented at the Eighth International ISKO Conference 13-16 July 2004 London, UK. Available on *DLIST*, <http://dlist.sir.arizona.edu/389/>.

Hjørland, Birger and Albrechtsen, Hanne. 1995. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science* 46: 400-25.

Hjørland, Birger and Hartel, Jenna. 2003. Ontological, epistemological and sociological dimensions of domains. *Knowledge organization* 30: 239-45.

Hjørland, Birger and Nicolaisen, Jeppe. 2004. Scientific and scholarly classifications are not "naïve": a comment to Beghtol (2003). *Knowledge organization* 31: 55-61.

ISKO Italia. 2007. The León manifesto. Available <http://www.iskoi.org/ilc/leon.htm>. Republished in *Knowledge organization* 34: 6-8.

Kublik, Angela, Clevette, Virginia, Ward, Dennis, and Olson, Hope A. 2004. Adapting dominant classifications to particular contexts. In Williamson, Nancy J. and Beghtol, Clare, eds., *Knowledge organization and classification in international information retrieval.* Binghamton: Haworth, 13-32.

Kwaśnik, Barbara Hanna and Chun, You-Lee. 2004. Translation of classifications: issues and solutions as exemplified in the Korean Decimal Classification. In McIlwaine, Ia C., ed., *Knowledge organization and the global information society: Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK.* Würzburg: Ergon, 193-98.

Kwaśnik, Barbara Hanna and Rubin, Victoria L. 2004. Stretching conceptual structures in classifications across languages and cultures. In Williamson, Nancy J. and Beghtol, Clare, eds., *Knowledge organization and classification in international information retrieval.* Binghamton: Haworth, 33-48.

Langridge, D. 1992. Bliss, the disciplines, and the New Age. *Bliss classification bulletin* 34: 8-13.

López-Huertas, María J. 2006. Thematic map of interdisciplinary domains based on their termino-

logical representation. The gender studies. In Budin, Gerhard, Swertz, Christian, and Mitgutsch, Konstantin, eds., *Knowledge organization for a global learning society: Proceedings of the Ninth International ISKO Conference 4-7 July 2006 Vienna*. Würzburg: Ergon, 331-38.

MacDonell, Paul, Tagami, Reiko & Washington, Paul. 2003. *Brian Deer Classification System*. University of British Columbia. School of Library, Archival and Information Studies. Available at http://www.slais.ubc.ca/PEOPLE/students/studnt-projects/R_Tagami/517/index.htm.

McIlwaine, Ia C. 2000. Interdisciplinarity: a new retrieval problem? In Beghtol, Clare, Howarth, Lynne and Williamson, Nancy J., eds., *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference 10-13 July 2000 Toronto*. Würzburg: Ergon, 261-67.

Mai, Jens-Erik. 2004a. Classification in context: relativity, reality, and representation. *Knowledge organization* 31: 39-48.

Mai, Jens-Erik. 2004b. The future of general classification. In Williamson, Nancy J. and Beghtol, Clare, eds., *Knowledge organization and classification in international information retrieval*. Binghamton: Haworth, 3-10.

Olson, Hope A. 2002. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer.

Panofsky, Erwin. 1955. *Meaning in the visual arts: papers in and on art history*. New York: Anchor.

Poli, Roberto. 1996. Ontology for knowledge organization. In Green, Rebecca, ed., *Knowledge organization and change: Proceedings of the Fourth International ISKO Conference, 15-18 July 1996 Washington, DC*. Würzburg: Ergon, 313-19.

Poli, Roberto. 1997. Framing ontology. In Corazon, Raul, ed., *Ontology: A resource guide for philosophers*. Available at <http://www.formalontology.it/essays/framing.pdf>.

Preuss, Michael. 2004. Locality and information. Unpublished paper presented at the Eighth International ISKO Conference 13-16 July 2004, London, UK.

Quintarelli, Emanuele. 2005. Folksonomies: power to the people. Paper presented at the First ISKO Italy-UniMIB Meeting, Milan, June 24, 2005. *ISKO Italia. Documenti*. Available at <http://www.iskoi.org/doc/folksonomies.htm>.

Rafferty, Pauline and Hidderley, Rob. 2004. A survey of image retrieval tools. In McIlwaine, Ia C., ed., *Knowledge organization and the global information society: Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK*. Würzburg: Ergon, 303-8.

Ranganathan, Shiyali Ramamrita. 1967. *Prolegomena to library classification*. Bangalore: SRELS.

Ridi, Riccardo. 1999. Metadata e metatag: l'indicizzatore a metà strada fra l'autore e il lettore. In *Proceedings conference The digital library: challenges and solutions for the new millennium, Bologna, 17-18 June 1999*. Available at <http://www.aib.it/aib/commiss/cnur/dltridi.htm>.

Schmitz-Esser, Winfried & Sigel, Alexander. 2006. Introducing terminology-based ontologies. Papers and materials presented at a workshop at the Ninth International ISKO Conference 4-7 July 2006 Vienna. Available at <http://eprints.rclis.org/archive/00006612/>.

Seibt, Johanna, ed. 2004. *Process theories: cross-disciplinary studies in dynamic categories*. Dordrecht: Kluwer.

Severino, Francesca. 2005. What thesaurus to define EU/ACP relations? Analysis of the term *development* in the thesauri of the EU and other international organizations. Paper presented at the First ISKO Italy-UniMIB Meeting, Milan, June 24, 2005. *ISKO Italia. Documenti*. Available at <http://www.iskoi.org/doc/development.htm>.

Slavić, Aida. 2005. Knowledge organization systems, network standards and Semantic Web. Available on *DLIST*, <http://dlist.sir.arizona.edu/1326/>

Slavić, Aida and Cordeiro, Maria Inês. 2004. Core requirements for automation of analytico-synthetic classifications. In McIlwaine, Ia C., ed., *Knowledge organization and the global information society: Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK*. Würzburg: Ergon, 187-92. Also available on *DLIST*, <http://dlist.sir.arizona.edu/1303/>.

Svenonius, Elaine. 1983. Use of classification in online retrieval. *Library resources and technical services* 27: 76-80.

Svenonius, Elaine. 1994. Access to nonbook materials: the limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science* 45: 600-606.

Szostak, Rick. 2007a. Classification, interdisciplinarity, and the study of science. *Journal of documentation* Forthcoming.

Szostak, Rick. 2007b. Interdisciplinarity and the classification of scholarly documents by phenomena, theories, and methods. In Rodríguez Bravo, Blanca and Alvite Díez, María, eds., *Interdiscipli-*

narity and transdisciplinarity in the organization of scientific knowledge: Proceedings of the Eighth ISKO Spain Congress 18-20 April 2007 León. University of León, 469-77.

Tennis, Joseph. 2002. Subject ontogeny: subject access through time and the dimensionality of classification. In López-Huertas, María J., ed., *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada.* Würzburg: Ergon, 54-59.

Tennis, Joseph. 2006. *Overview of the research agenda.* University of British Columbia. School of Library, Archival and Information Studies. Available at <http://www.slais.ubc.ca/PEOPLE/faculty/tennis-p/research-agenda.html>.

Tudhope, Douglas and Koch, Traugott. 2004. New applications of knowledge organization systems: introduction to a special issue. *Journal of digital information* 4 no. 4. Article 286. Available at <http://jodi.tamu.edu/Articles/v04/i04/editorial/>.

Vickery, Brian Campbell. 1953. The significance of John Wilkins in the history of bibliographic classification. *Libri* 2: 326-43.

Will, Leonard. 1992. Thesaurus principles and practices. In *Thesauri for museum documentation: Proceedings of a workshop held at the Science Museum, London, 24 February 1992.* Cambridge: Museum Documentation Association. Also available in *Willpower information*, <http://www.willpowerinfo.co.uk/thesprin.htm>.

Zuccolo, Lorena. 2006. CDD e minoranze linguistiche: il caso del Friuli. In *Classificare la documentazione locale: giornata di studio, San Giorgio di Nogaro, 17 dicembre 2005. ISKO Italia. Documenti.* Available at <http://www.iskoi.org/doc/locale1.htm>.