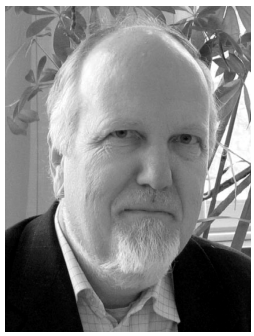
Fundamentals of Knowledge Organization[1]

Birger Hjørland

Royal School of Library and Information Science, Copenhagen Email: bh@db.dk



Birger Hjørland is Professor in information science (IS) at the Royal School of Library and Information Science in Copenhagen, Denmark. He was formerly professor in IS at the College University of Borås, Sweden. He has a Ph.D. in information science from the University of Goethenburg, Sweden and a Master of Arts in psychology from the University of Copenhagen, Denmark. For twelve years, Professor Hjørland worked as a research librarian at the Royal Library in Copenhagen, where he was a subject specialist in psychology and the coordinator of the library's computer based reference services. He has taught the history and philosophy of psychology at the University of Copenhagen and also taught information retrieval in the Institute of Applied and Mathematical Linguistics in Copenhagen. In spite of Professor Hjørland's experience in the field of psychology, his work is much influenced and oriented by philosophical and sociological perspectives. He is a member of ASIS&T and ISKO and is the author of Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science.

Hjørland, Birger. (2003). **Fundamentals of Knowledge Organization.** *Knowledge Organization, 30(2).* 87-111. 74 refs.

ABSTRACT: This article is organized in 10 sections: (1) Knowledge Organization (KO) is a wide interdisciplinary field, much broader than Library and Information Science (LIS). (2) Inside LIS there have been many different approaches and traditions of KO with little mutual influence. These traditions have to a large extent been defined by new technology, for which reason the theoretical integration and underpinning has not been well considered. The most important technology-driven traditions are: a) Manual indexing and classification in libraries and reference works, b) Documentation and scientific communication, c) Information storage and retrieval by computers, d) Citation based KO and e) Full text, hypertext and Internet based approaches. These traditions taken together define very much the special LIS focus on KO. For KO as a field of research it is important to establish a fruitful theoretical frame of reference for this overall field. This paper provides some suggestions. (3) One important theoretical distinction to consider is the one between social and intellectual forms of KO. Social forms of KO are related to professional training, disciplines and social groups while intellectual organization is related to concepts and theories in the fields to be organized. (4) The social perspective includes in addition the systems of genres and documents as well as the social system of knowledge producers, knowledge intermediaries and knowledge users. (5) This social system of documents, genres and agents makes available a very complicated structure of potential subject access points (SAPs), which may be used in information retrieval (IR). The basic aim of research in KO is to develop knowledge on how to optimise this system of SAPs and its utilization in IR. (6) SAPs may be seen as signs, and their production and use may be understood from a social semiotic point of view. (7) The concept of paradigms is also helpful because different groups and interests tend to be organized according to a paradigm and to develop different criteria of relevance, and thus different criteria of likeliness in KO. (8) The basic unit in KO is the semantic relation between two concepts, and such relations are embedded in theories. (9) In classification like things are grouped together, but what is considered similar is not a trivial question. (10) The paper concludes with the considering of methods for KO. Basically the methods of any field are connected with epistemological theories. This is also the case with KO. The existing methods as described in the literature of KO fit into a classification of basic epistemological views. The debate about the methods of KO at the deepest level therefore implies an epistemological discussion.

1. What is Knowledge Organization (KO)?

In the *Library and Information Science* community (LIS) Knowledge Organization (KO) means especially the organization of information in bibliographical records, including citation indexes, full text records and the Internet. Information Science (IS) is basically about the best way to construct such bibliographical

records (which is done in KO) as well as the optimal way to utilize given records (which is done in information retrieval IR).

KO is, however, a much broader concept. Knowledge is organized in, among other things:

- The social division of labour (e.g., in disciplines)
- Social institutions (e.g., in universities)
- Languages and symbolic systems
- Conceptual systems and theories
- Literatures and genres

Library and Information Science or just Information Science, (L)IS, has often ignored this broader meaning of KO and has thus failed to be based on such broader theories; or any theories at all. My central view is that such broader views cannot be ignored. Any attempt to develop fruitful principles for KO in LIS must be based on broader theories of KO.

The principal actors in IS are the knowledge producers (e.g., authors), the users and the intermediaries. It is their interaction with bibliographical records that is the focus of (L)IS. Each of those actors brings with him certain pre-understandings, views, concepts and languages mainly acquired during socialization in society. The success of the interaction depends on these pre-understandings, concepts and languages. This applies not only to the "match" of concepts but especially to their ability to support users' tasks. A theory of such broader forms of KO is therefore essential in order to construct efficient systems for KO in LIS.

LIS has claimed status as a field of study from about 1876 when Melvin Dewey first published his classification system and when schools of librarianship began to emerge. What are the accumulated results of more than one century of research in KO? Among the results are a number of standards and guidelines, some theoretical developments, for example, Cutter's rules (1904), advances in *facet analysis* and of course important changes in information technology (IT). It is, however, difficult to sketch the more theoretical and scientific progress in this field because it seems largely atheoretical and fragmented and because many different lines of thought seem to coexist. In my opinion KO lacks theories about its most fundamental concepts such as:

- Concepts
- Criteria for class inclusion
- Meaning
- Indexing
- Semantic relations
- Subjects
- Subject access points ("SAPs"), and so forth.

I also find that KO lacks serious explorations into the methods of KO and the methodological basis of KO. Developments in practice seem much more influenced by progress in IT than by progress in KO proper. Sometimes people seem even to expect that IT in itself can replace conceptual and theoretical concerns. In my opinion we may have the best overview of progress in KO if we consider five technology-driven stages in the development of KO. If this is true, it is an expression of a crisis in the field: that the field is not driven by developments in its own research. It is lacking accumulated knowledge, which can be transformed from one IT-platform to another.

2. Technology Driven Phases in the Development of Knowledge Organization (KO)

In library and information science (LIS) the concept of KO is connected to the development of classification and indexing systems in libraries, bibliographies, and electronic databases. Hjørland, (2000; Hjørland & Kyllesbech Nielsen, 2001) described five technology driven stages in the development of KO. Each stage is not replaced by subsequent stages, but continues to influence the field. Because of this fact the technology beyond those stages may also be seen as one among other causes for the development of more or less supplementary or competing approaches in KO.

The stages are:

a. Manual indexing and classification in libraries and reference works

Cataloguing in libraries and organizational principles in reference works go far back in time. One can trace the beginnings of the use of alphabetical order to antiquity, although it was not generally adopted until well into the Middle Ages, and it was not the only method of classification possible (Daly, 1967). However, in the Middle Ages and at the beginning of modern time catalogues were probably seen as inventories designed for controlling the collection and for information needed for new acquisitions, rather than as tools for subject searches.

A professionalisation of the classification and indexing of books in libraries appeared around 1876 with the publication of the *Dewey Decimal Classification*,

and with the foundations of library schools (and thus "library science"), and so forth. Charles A. Cutter (1837-1903), Melvin Dewey (1851-1931), Henry E. Bliss (1870-1955) and S. R. Ranganathan (1892-1972), among others, have been recognized as founding fathers.

It is generally recognized that this stage produced principles for KO that are still valid and important. Among them are "Cutter's rules" and Ranganathan's approach to the classification of subject fields. Other principles have of course become obsolete, and some are controversial. Frohmann (1994), for example, considered Dewey's approach to be very harmful because it ignored the semiotic nature of classification and just proposed an empty formalism: "Dewey's subjects were elements of a semiological system of standardized, technobureaucratic administrative software for the library in its corporate, rather than high culture, incarnation" (pp. 112-113). It must be added, however, that there is very little overall discussion of the importance of these (and other) principles. They have been considered standards or canons rather than research-based principles subject to investigation, modification and revision.

b. Documentation and Scientific Communication

The documentation movement developed from about 1895. The founders were Paul Otlet (1868-1944) and Henri Lafontaine (1854-1943), who established *The International Institute of Bibliography*[2] and in 1905-1907 published the first edition of the *Universal Decimal Classification*, UDC, which was an extension of the Dewey Decimal Classification designed for deeper subject analysis. The *American Documentation Institute* was founded in 1937 and by 1968, had changed its name to the *American Society for Information Science* (ASIS). It is difficult to describe the exact differences between librarianship and documentation. Documentalists were less interested in libraries and their collections; they were more interested in bibliographical control, in scientific communication and scientific documentation, in information services to industry and in the utilization of the knowledge in the documents. Implicitly they shared a view of KO as based on subject knowledge and subject literature. Documentalists often regarded themselves as more service-minded, more technology-oriented and more advanced than librarians. Where traditional librarians often had an orientation towards the humanities, documentalists were mostly affiliated with science, technology and business. They indexed single articles in journals and books and played a central role in the establishing of international abstracting journals[3]. They were less interested in keeping books for their own sake or for broad cultural purposes, and were highly interested in establishing services which could stimulate the application of knowledge to specific purposes. The foundation of user studies (Bernal, 1948) and bibliometrics (e.g., Bradford, 1948) is also part of this stage/tradition, which may first and foremost be characterized by a more specific subject approach, a deeper level of indexing, more emphasis on modern technology and a more scientific attitude towards goals and problems. This stage laid the foundation for the later development of online bibliographic databases.

c. Information Storage and Retrieval by Computers (mainly 1950-)

Computer science[4] has its own development, which should not be confused with the development of library, documentation and information science. As a theoretical discipline computer science was founded in the 1930s, well before the appearance of the modern computer. The founding fathers were the logicians Alonso Church (1903-1995), Kurt Gödel (1906-1978), Stephen C. Kleene (1909-1994), Emil Post (1897-1954), and Alan Turing (1912-1954). This early work has had a profound influence on both the theoretical and the applied development of computer science, but these persons are not directly connected to the history of LIS or KO, in my view. Computer science has different educational programs, scientific journals and conferences compared to LIS and it also has somewhat different goals compared to LIS. Arbib, Kfoury, and Moll (1981) write that "*Computer science seeks to provide a scientific basis for the study of information processing, the solution of problems by algorithms, and the design and programming of computers.*" In my opinion the goal of LIS is to optimize the utilization of documented knowledge. The primary purpose of libraries is to provide physical and intellectual access to information. The intellectual access is provided by the organization of either the physical documents themselves or by organizing document representations in catalogues, bibliographies and databases. LIS is not primarily focused on constructing algorithms, but on informing people about documents. LIS is supposed to be more open to different views, more reflective and meta-oriented and demonstrate gaps and uncertainties in knowledge to users. This is very different from making, for example, an expert system that performs optimally by reflecting some generalized cognitive models.

With the advent of computers in the 1950s, LIS and KO became influenced by this new technology, and many people felt that the future of LIS must be as a part of computer science. There is an indication that "information scientists" did not regard themselves as part of library science or as dealing with documents and their representations (cf., Hjørland, 2000a). They felt much more closely related to computer science. This may have confused the theoretical development of KO, because theories related to "information" versus theories related to "documents" are related to quite different kinds of outlooks. As a theoretical concept, "information" tends to move LIS and KO towards theories about control, feedback, coding and noise in transmitting messages, while "document" tends to move LIS towards theories about meaning, language, knowledge, epistemology and sociology. Therefore, in LIS there may be a whole paradigmatic conflict hidden in those words.

This third stage in the development of KO as a field has been influenced by experimental approaches in which recall and precision are well-known measures, by extensive use of statistical models of the properties of the document representations, by approaches that try to automate KO using Artificial Intelligence (AI) and expert systems, by applying natural language processing (NLP) techniques and the like. An offshoot of this stage is also the cognitive view that tries to model the users' cognitive functions in the computers.

The establishing of computer based abstract services such as Chemical Abstracts and MEDLINE in the 1960s is an important achievement at this stage. The development of descriptor based and free text retrieval (mainly based on titles and abstracts), Boolean logic, field specific subject access, measuring recall and precision and other innovations were extremely important developments in document retrieval. Information retrieval as a research tradition started with the ASTIA and Cranfield experiments in the 1950s, and today's TREC full text experiments may be seen as a continuation of this tradition. This third stage improved information services and research efforts in IS in a most important way. In my opinion these technological innovations and the goals they try to solve define the core area of what is termed information science.

Computer technology made it possible to use many kinds of Subject Access Points (SAPs), both of the traditional kinds produced by information specialists and of words from the documents themselves (e.g., titles and abstracts). This removed the librarians'/information specialists' monopoly on subject access, and established a direct competition between SAPs produced by different agencies.

An underlying philosophy at this stage has often been that the length of the searchable record in itself was the most important parameter in retrieval (Lancaster, 1998, pp.6-8). SAPs were often seen as merely "semantic condensations" of the texts represented (implying that the ultimate goal was full text representation and nothing more). Research was dominated by quantitative methodologies while little research concerning qualitative differences (semantics or meanings) between different kinds of SAPs was established at this stage. The underlying philosophy has been empiricist in more than one sense of this word. It was first and foremost empiricist in its attempt to measure the efficiency of subject retrieval points empirically, for example, by measuring "recall" and "precision." It was also empiricist in its avoidance of "metaphysically" based classifications, and in its favour of "atomist" SAPs such as the Uniterm system devised by Mortimer Taube in 1951 and similar systems depending on specific words from the document themselves.

One associated tendency at this stage or approach was the attempt to formalize and automate retrieval and to eliminate human interpretation and subject analysis. We must distinguish between the financial pressure to automate practical systems on the one side and the scientific evaluation of the performance of various aspects of human based and mechanized retrieval systems on the other side. It is a legitimate and highly desirable goal to reduce costs and improve efficiency in information systems. Basic research should, however, illuminate basic strengths and drawbacks in different approaches, and not be blinded by the pressure to use automated or cheap solutions. Because of such tendencies important approaches related to interpretation and qualitative aspects have been much neglected to this stage, and the research has not accumulated as satisfactory a body of knowledge as is needed.

d. Citation based retrieval and KO (1963-)

Eugene Garfield's introduction of the *Science Citation Index* in 1963 marks the fourth important stage in the development of SAPs. The possibility of retrieving documents according to the citations they receive represents a real innovation in IR, and this technique is able to supplement all forms of term-based retrieval in very important and qualitative new ways. This in-

novation has also contributed to research in motives for citing other documents, in sociological patterns in citing, in the relative role of terms and references as SAPs and in the semantic relations between citing and cited papers. Overall, this stage has brought the concept of *networks of documents* into the focus of KO.

In this way, citation based retrieval has changed our understanding not only of subject relatedness, but also of the concept of subject matter and of the fundamental aim of IR itself. Since it may be relevant to cite papers, which have no words in common with the citing papers (or no simple semantic relation such as narrower terms, broader terms, synonyms, etc.), naïve conceptions of subject relatedness or subject matter can no longer persist. Semantic relations may be implicit or latent. Semantic relations in science are determined by theoretical advances, which may change the verbal description of the research phenomena completely. This is the reason why statistical patterns in vocabulary may sometimes turn out to be a less efficient measure of subject relatedness than patterns in citations.

Citation behaviour is extremely important as a theoretical object because the goal of IR is to provide the references which are useful in solving a specific problem. A scientific article is a documentation of the solving of a specific research problem. The problem is formulated in the article, and this problem has determined what kind of information was needed by the author in order to contribute to this problem. Based on the information need, information has been sought and selected, and the documents actually used are finally cited in the article. Each of the thousands of articles produced weekly is in a way a case study in IR. Not only does every article pose a definite IR problem, but the list of references provided by the author is the key to how that particular person has solved the problem. Thus it is possible to check theories of IR against how they match the actual documents cited! In the traditional (positivist) philosophy of science, studies should be able to "predict" future behaviour or events. In our case this principle would imply that theories and models of IR should be able to predict what citations will turn up in the single papers. Most research on relevance and on IR seems to have overlooked this fact. From what we do know, it seems extremely unlikely that an algorithm should be able to select references from electronic databases and end up with just the set of references represented in a given article. From this point of view, theories of IR seem naïve and unrealistic and the goal of prediction seems to be wrong. A more detailed study of citation behaviour can illuminate the real problems of IR: That cited documents are not simply a set of documents sharing a fixed set of attributes which are not represented in the non-selected items. Documents which are similar from the point of view of retrieval algorithms do not need to be co-cited, whereas documents which are not similar are often co-cited. Ordinary retrieval algorithms and citation practices seem simply to reflect different theories about subject-relatedness.

Since authors may cite other papers in order to flatter or to impress, the prediction of which references a given author will select in the end for a given paper cannot be used as a valid criterion in IR. Criteria for IR should not be based on social-psychological motives, but on epistemological principles for the advancement of public knowledge! In this way, our insight from citation indexes has, in a very profound way, changed not only the methods of IR but also the concept of subject relatedness itself and the basic aim of retrieving information. We can no longer regard the prediction of individual use as the ideal criterion for IR, nor can we regard IR as a value free technique. Instead, we have to face the insight that the goals of IR are deeply rooted in epistemological norms for what should be regarded as good science and good citation behaviour.

Pioneers in the integration of bibliometrical methods with the more traditional methods of KO are Mike M. Kessler (1965), Miranda Lee Pao and Dennis B.Worthen (1989), Pao (1993) and Lorna K. Rees-Potter (1987, 1989,1991). I also see my own research as pioneering in this respect both theoretically and empirically (e.g., Hjørland, 1992, 1993, 1998a+b, 2002b). It has subsequently also been used by, among others, research students of Peter Ingwersen (e.g., Schneider & Borlund, 2002).

e. Full text, Hypertext and Internet (mainly 1990-)

Full text retrieval marks the fifth and final step in the development of SAPs. Until this point space limits were a major constraint in the development of subject access systems because length of record in itself is an important parameter in retrieval. At this stage, every single word or every possible combination of words in full text documents is a potential SAP, as is every thinkable kind of value-added information provided by authors, readers, or intermediaries. Given full-text representations the first important theoretical problem to arise is whether *any* kind of value-added information is necessary? Can such extra information as

provided by abstracting and indexing – at least in principle – increase recall and/or precision? If this is not the case, then it seems as if we have reached the end of the line, that no further contributions from research or practice in IS are needed any longer.

The answer to this question is closely linked to theoretical views on the concept of subject. In the view of Claus Poulsen (1994) a subject is something that is expressed in the literature in a transparent and self-evident way. By defining subjects this way it is impossible to even ask whether a given text always represents the optimal representation of itself. By defining subjects as informative or epistemological potentialities Hjørland (1992, 1997) establishes the possibility that documents may be implicit or even wrong about their own subject matter and hence that there still exists a need for information professionals. To take an extreme example: A document about Jews written by a Nazi author should not only be indexed as being about Jews, as it claims to be. It is important to make the Nazi view visible in the subject analysis (and, for example, index it as Nazi propaganda about Jews). Subjects are not objectively "given" but are influenced by broader views which are important for the information seeker to know and should therefore ideally be part of the subject analysis. Whether or not this is also a practical, economical and realistic solution is another question that must be illuminated by evaluating specific subject access systems.

The introduction of full text databases has emphasized the theoretical problems concerning subject matter and meaning and has also made composition studies highly relevant to IS. In this phase, qualitative aspects become more important than ever before. It is not just the problem of making efficient algorithms for IR, but also the problem of identifying the underlying values and goals that such algorithms are going to serve (see Hjørland, 2003b).

*

The above description of the five technology-driven developmental stages in KO provides, in my opinion, a clear picture of what KO means with LIS. In viewing those five stages together, we have a rather well-defined picture of the overall topic with which we are dealing. Behind those different traditions is a certain general goal that can be expressed as the optimisation of KO in libraries, databases, reference works and on the Internet. This goal is important to keep in mind because it is this goal that gives KO in LIS its importance for practice.

It is important to realize, however, that the development has been very much technology-driven. This is not a satisfactory circumstance for a discipline with the ambition of being a science and producing general theories and principles. Although those five stages define a topic and a common goal, they do not, however, define a cumulated fund of findings, theories and principles. On the contrary, there are often latent conflicting views between those stages or traditions. Those traditions have only to a very limited extent been in the form of a dialogue. This is why overall concepts for thinking about KO have been missing or at least strongly underdeveloped. In the next sections we shall examine some of those concepts, theories and principles that seem most important for this task.

The goal of LIS is to write a history of its theoretical development *abstracted* from the concrete technologies in which its principles have been studied. This is difficult because the general knowledge base is not well established. Also, the tendency has been that library and information science has passively used the technology without contributing much to its development. If LIS is to be able to contribute valuable knowledge, its focus must be abstracted from concrete technologies.

3. Information Processing in the Social Division of Labour: The Intellectual and the Social Organization of Knowledge.

Hjørland (2002d) analysed the official definition of information science given by the *American Society for Information Science and Technology* (ASIST), which states:

> Information science is concerned with the generation, collection, organization, interpretation, storage, retrieval, dissemination, transformation and use of information, with particular emphasis on the applications of modern technologies in these areas (Borko, 1968; Griffith, 1980).

This definition does not discuss how professions other than information scientists are concerned with "the generation, collection, organization, interpretation, storage, retrieval, dissemination, transformation and use of information". In fact, we know that many professions could claim to be equally concerned. By neglecting this issue, the definition fails to define information science since a definition of IS should specify the special role of information professionals in studying or handling information.

Astronomers, for example, can be seen as experts who identify, process and interpret information from the universe. Astronomers "read" both nature and books, but nature is mostly considered the key information in the sciences. As products of their activities they may publish their empirical and theoretical findings. The library, documentation and information profession, although interested in all kinds of documents, has a core interest and expertise connected to communication of published documents. Information scientists are *not* experts in interpreting information from the stars, but are, for the most part, experts in information documented by, for example, astronomers (e.g., indexing and retrieving astronomical documents). In this example information has been defined in a broader sense than usually implied in information science.

Just as astronomers can be said to handle information professionally, all other professions can be said to do so as well. Publishers, researchers, historians, lawyers, teachers and anybody else can be said to be professionals in handling information in some way or another. The role of information specialists may be relatively clear when the target group is, for example, astronomers: Information specialists are experts on forms of publications, databases, reference tools, and so forth. In the case of, for example, historians and lawyers, borderlines are much less clear because the key information that these professions seek, interpret and use is itself contained in publications and other documents. The historian, not the librarian or information specialist, is the expert in seeking, organizing, interpreting and utilizing the documents (mostly unpublished) needed in his or her professional work, but an information professional is more professional regarding some specific problems such as databases, cataloguing, and so forth. A basic difference in knowledge about information sources is typically that the subject specialist starts from a narrow point and works bottom-up to more general information sources, while the information specialist starts from a broad overview and works top-down to more specific information sources. In this way their competencies are supplementary.

The special focus of LIS is on documented knowledge produced by human beings in some kind of documents of potential use to other human beings. Light from the stars is not information for the LIS-community, but astronomical information as produced and used by astronomers is. This distinction may seem subtle, but is important in order to construct a firm theoretical basis for KO.

KO involves two kinds of organization: The intellectual organization of knowledge may also be termed the cognitive organization of knowledge. This is basically the organization of knowledge in concepts, conceptual systems and theories. If we use the periodical system of chemistry or the zoological taxonomy of biology as the basis of indexing systems, we are using intellectual or "cognitive" systems of KO. The social organization of knowledge, on the other hand, is basically the organizations in professions, trades and disciplines. If we refer to disciplines in our knowledge representations, we are using "social" systems of KO. It should be mentioned that this social organization principle has been the basic one in traditional library classification such as the Dewey Decimal Classification (DDC):

> *A work on water may be classed with many disciplines, such as metaphysics, religion, economics, commerce, physics, chemistry, geology, oceanography, meteorology, and history.*
>
> *No other feature of the DDC is more basic than this: that it scatters subjects by discipline* (Dewey, 1979, p. xxxi).

There is very little serious debate about this principle in the literature, supporting the view that many librarians and information specialists seem to be uncomfortable with it. Whether or not this principle is approved it should be clear, however, that the social organization of knowledge in disciplines is very important for KO and for optimising information seeking. When it has been established which discipline a given query belongs to, the hardest part of the retrieval task may well have been completed. The social organization of knowledge is indeed important to study, and apart from some research in bibliometrics (see Hargens, 2000), this study has hardly begun in LIS in spite of the significance expressed in the quotation given above. Today, the "empirical" basis for classifications such as DDC is mainly the literature classified by the system. One way to expand the social dimension could be to study the development of occupations in society, for example, as displayed in the *Standard Occupational Classification* System (U.S. Department of Labor, Bureau of Labor Statistics, 2002).

There are of course different theories or conceptual frameworks on both the intellectual and the social organization of knowledge (as well as about their interrelatedness). One view sees scientific concepts, theories and fields as reflecting a neutral and objective reality. This might be termed *science as a mirror* metaphor, and it is related to rationalism. Scientists may

organize themselves according to such pre-established fields. Basically, however, the intellectual organization is not connected to – let alone formed by – the social division of labour. Another view might see scientific concepts, theories and fields as useful tools constructed in order for human beings to accommodate the demands of life. This is the *science as a map metaphor*, which is related to pragmatism.[5] Even if a map is a reflection or representation of a reality, a map is still not a mirror. A map is first and foremost a tool for certain human activities. The kind of activities and interests the map is going to serve has a major influence on how the map is made. From a pragmatic perspective the intellectual organisation of knowledge is deeply rooted in and connected to the social organization of knowledge.

One influential view today is *social constructivism*. This view is related to the pragmatic view presented above. Often, however, social constructivism and pragmatism are opposed to kinds of realism such as scientific realism. One of the modern pragmatic philosophers is Richard M. Rorty. According to Rorty, scientific realism and pragmatism are two views that cannot be combined; pragmatism is seen as an anti-realist position. According to other philosophers (e.g., Dewey, 1929; Ellis, 1990), realism must indeed be based on pragmatism. This corresponds to my own view, which may thus be termed *pragmatic realism*. Scientific concepts and fields tend to represent parts of reality in a way which is functional for human activity. To the degree that social constructivism is opposed to realism, I disagree with that view. One should avoid the pitfall of sociological reductionism. However, I have found much research done under the banner of social constructivism deeply relevant for the understanding of the structure of many knowledge fields. For example, in order to illuminate the classification of mental illness it is relevant to do both experimental research and to do "discourse analysis." Experimental research may show that schizophrenia is not a fruitful aetiological concept because all major psychoses may have a common genetic basis (see Kringlen, 1994). "Discourse analysis" on the other hand may uncover the fact that most of our theories and concepts in this field are "social constructions" influenced by specific interests. In particular it may draw our attention to different cultural experiences and avoid tendencies to universalistic explanations. In this way it may open up for classifications that are better suited for a dialogue between different perspectives. In the end, our scientific concepts and classifications (e.g., "schizophrenia" versus "Einheitspsychose") are tools that should be evaluated according to the job they are supposed to do for us. As such they are parts of theories that have been developed in order to make us act in relation to specific problems.

The organization of knowledge in traditions, ideologies and paradigms may be seen as the combining concepts between the intellectual and the social organization. They are cognitive organizations based on social influences.

4. Documents and the Sociological Perspective of Knowledge Production, Knowledge Intermediating and Knowledge Consumption

It is important for information science to provide models of actors, institutions and information services in different discourse communities. The actors are knowledge producers, intermediaries, and users of knowledge. The institutions are research institutions, publishers, libraries, and so forth. Information services may be classified in primary services (e.g., publishing houses and journals), secondary services (e.g., bibliographical databases) and tertiary services (e.g., professional encyclopaedias and literature reviews). I have found the UNISIST (1971) useful. Figure 1 below is a modified version of the UNISIST-model from Trine Fjordback Søndergaard, Jack Andersen, & Birger Hjørland (2003, p. 303), in which the stippled ellipse symbolises a knowledge domain. Figure 2 shows the same model (not previously published) in which the Internet resources are highlighted as integrated with the printed sources.

At this point, I want to emphasize that such models provide one specific perspective on knowledge organization. By introducing the actors and the divisions of labour between different organizations and services, such models are sociological by nature. They help us to understand the different functions of kinds of literature, and they enable us to establish a meaningful typology of scholarly literature, including the well known differentiating differentiation between primary, secondary and tertiary literature.

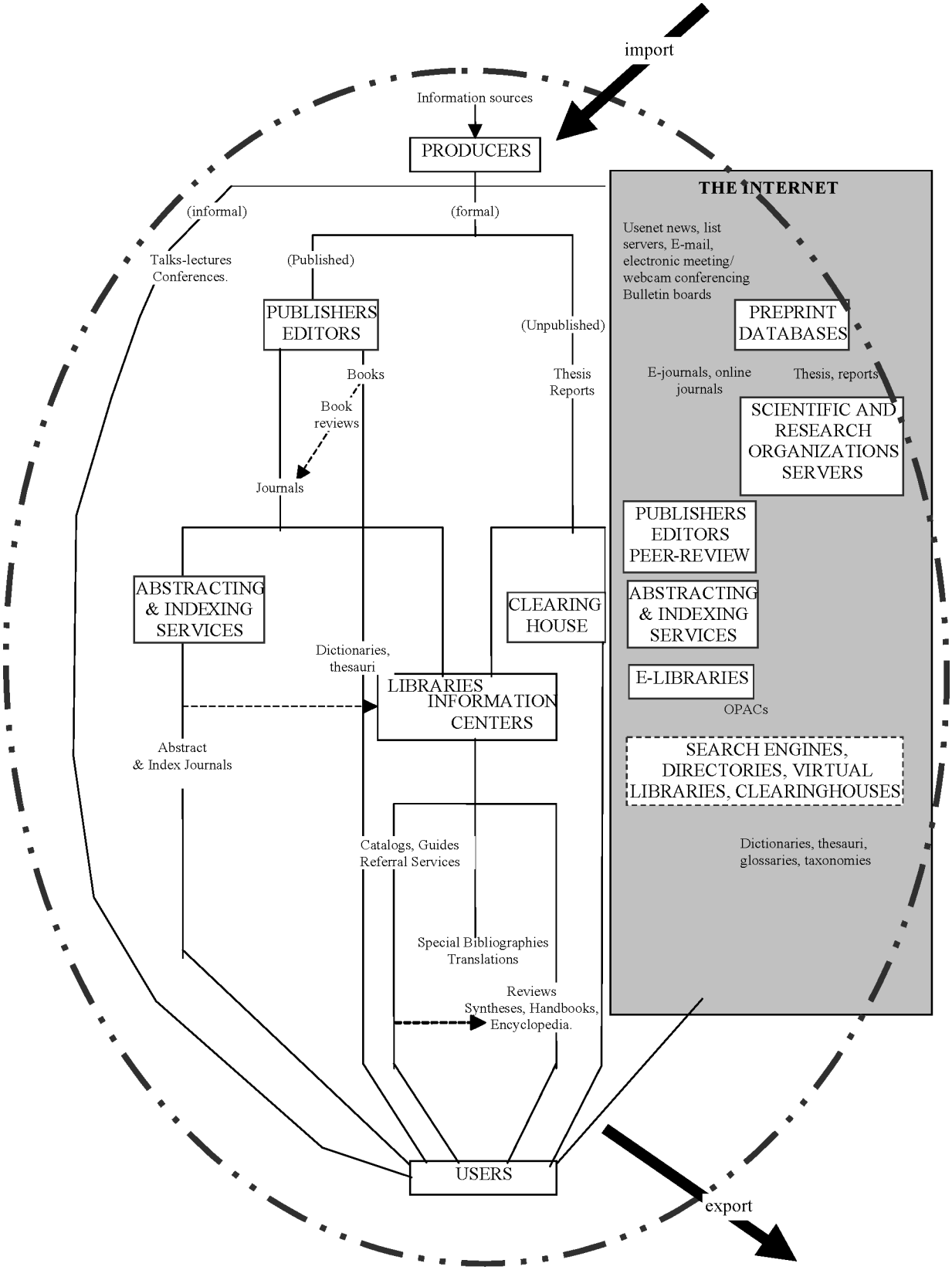


Figure 1: The revised UNISIST-model modified for the domain analytic approach

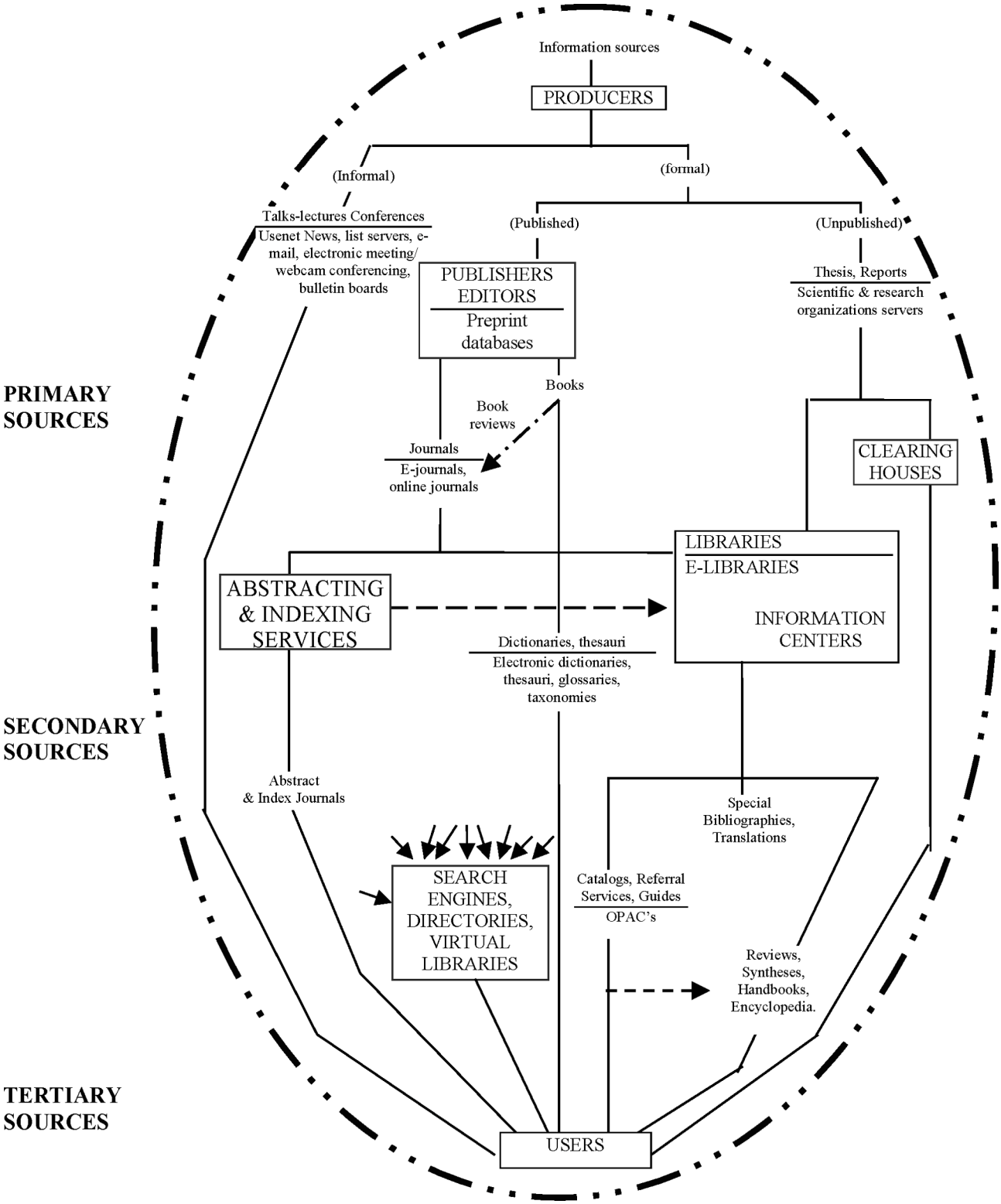


Figure 2: The revised UNISIST-model integrating printed and Internet resources and modified according to the domain analytical approach

5. Subject Access Points in Electronic Retrieval.

In electronic retrieval, words, phrases and texts are taken out of contexts and merged. Words and symbols from many different journals are merged; words from different sections and parts of texts are merged, and so forth. Words are taken out of their context, a context that contributes to their meaning. By using advanced technology it is possible to re-establish parts of this lost context (e.g., by means of proximity operators). Any string of characters and any symbol from the texts themselves or from so-called value added information become a possible subject access point. The art and science of information retrieval and KO is to utilize all of these possibilities in an optimal way and to add further value to the representation and retrieval of documents.

What we must emphasize in IS, is to study the ways in which words and symbols are given meaning by their specific contexts. Of importance are the way in which different disciplines construct their meanings, the way in which document composition provides meaning to words and symbols, and the way in which different controlled vocabularies construct meanings. If a word is employed, for example, in the title, in the abstracts, in the methodology section or in the conclusion, we might (sometimes) be able to attribute different priorities and meanings to the word, and this may help us to give different probabilitistic evaluations of whether documents should be retrieved or not. People, who know certain databases very well, including the subject literature and subject language they cover, probably employ this kind of knowledge in tacit ways. The job for information studies is to help to explicate the underlying principles. Just as ordinary people can speak a language and use grammar, linguistic experts have to explicate what competent users do.

In Birger Hjørland and Lykke Kyllesbech Nielsen (2001) we summarized research on the relative informational value of different kinds of subject access points (SAPs). It is important to realize that the relative value of, for example, terms from a text versus citation searching (based on the bibliographical references) is not constant. It varies according to disciplinary norms among other things. This implies that it is important to uncover such different disciplinary norms: In our above mentioned review we tried to integrate such knowledge from many different sources. Our approach implies that the rich flora of documents and domains should be investigated by information science. This is in contrast to the prevailing "universalistic," "reductionist" or "individualist" approaches. The purpose of studies is to explicate and represent the implicit meanings that are lost in merging.

Three major approaches to IR and KO are:

a) Traditional IR based on term frequencies in the texts themselves and in the total database.
b) Citation bases' organization and retrieval based on networks of citations between documents.
c) Traditional library classifications based on disciplinary divisions and disciplinary criteria. Those disciplinary criteria are by principle external to the documents themselves and are NOT derivable from either full text databases or hyper linked networks. You cannot, for example, construct a map of Spain from a collection of documents about Spain, whether you use co-citation analysis or any other techniques. Maps are produced by geographers, and their conceptual structures are subsequently applied in library classification.

All these approaches are often used concurrently and bibliographical records often contain a mixture of all of them. We should develop more (explicit) knowledge about their relative strengths and weaknesses for different kinds of queries.

KO involves two kinds of decisions: To make a classification system (or a thesaurus, etc.) and to decide where to classify a given document in that system. Both decisions are related and involve epistemological implications. Both decisions are made at the end, ultimately based on what is considered relevant: What classes are relevant? What kinds of connections between classes are relevant? What kinds of documents are relevant for given users or purposes to put together in a given class?

Different kinds of SAPs have different informational values in information retrieval. SAPs are more or less functional in revealing "the subject" of the documents that they represent. It is important to realize, however, that which documents are relevant to a given query depends, among other things, on the theoretical perspective in which the query is embedded. A description of "the subject" is not a neutral, objective activity, but is influenced by different theoretical views and interests. The consequence of this view is that "the subject itself" (the something that the SAPs are supposed to identify) cannot be regarded as being objective in the positivist sense. Based on such considerations, Hjørland (1992) defined the subject of a

document as the document's epistemological (or informative) *potentialities*. A document has given potentialities, and the job for indexers (and other SAPs) is to identify those potentialities. This is easily understood if we compare this with citation indexes. There may be many reasons to cite a given document, but some of these reasons may be rather widely varied. For the group of authors citing a given document for the same reason this document has the same function and the same "subject." Indexing is a process in the service of helping future authors to identify documents to be cited. This is done by identifying exactly those attributes that make it citable. The connections between cited and citing papers are relevance connections, and so also are (or should be) the connections between indexing terms and the documents indexed.

6. Social Semiotics and Activity Theory: The Teleological Nature of KO

The structuralist view of how our concepts are formed by our languages is shown in Figure 3 below. It shows that there is no one-to-one relation between meanings in different languages. When individuals learn a language, they learn the concepts of that language, and consequently they classify the world in the way that is given by that particular language. For example, Germans and Danes classify "tree" in different ways. By implication languages may affect the way we conceptualise the world. According to Louis Hjelmslev (1943) each language puts *arbitrary* borders on reality, while activity theory (AT) finds that our symbolic systems tend to capture functional aspects or affordances in the things we perceive (see Albrechtsen, Andersen, Bødker, & Pejtersen, 2001).

<table>
<tr><th>English</th><th>*German</th><th>*Danish</th><th>*French</th><th>Italian</th><th>Spanish</th></tr>
<tr><td>Tree</td><td>Baum</td><td rowspan="2">Træ</td><td>Arbre</td><td>albero</td><td>Árbol</td></tr>
<tr><td rowspan="2">Wood</td><td rowspan="2">Holz</td><td rowspan="3">Bois</td><td rowspan="2">legno</td><td>Leña</td></tr>
<tr><td rowspan="3">Skov</td><td>Madera</td></tr>
<tr><td>Woods</td><td rowspan="2">Wald</td><td>bosco</td><td>Bosque</td></tr>
<tr><td>Forest</td><td>Forêt</td><td>foresta</td><td>Selva</td></tr>
</table>

Figure 3: Cultural relativity in word meanings

* Originally presented by the Danish structural linguist Louis Hjelmslev (1943).
Extendet by information from Buckley (2001)

Hjelmslev's figure may be criticized; for example, it has been mentioned that he omits the Danish term "lund" (small forest) and that this omission may weaken this particular example. In this place we shall, however, take this basic structuralist claim for granted: that given languages affect our semantic structures and thus our classification of the world.

For human informative activities, the proper perspective of the meaning of "meaning" is very important. This is a difficult concept for which activity theory provides a fruitful understanding. The production of books, texts and other documents is a special development in literate cultures. Documents are tools having specific kinds of functional values in those cultures. The view of AT and social semiotic[6] theories is that meanings, signs and documents are developed to function in relation to standardized practices in communities. We use, for example, the Bible and the Hymn Book in our standardized religious practices, textbooks in our standardized teaching practices, law books in standardized legal practices, and so forth.

When we name something, we facilitate the use of that object. By naming churches (or books on churches) "tourist attractions," we facilitate a certain use of churches (or documents). We perform an act by means of language. Given names (or SAPs) will always facilitate some uses of a document and make other uses relatively more difficult. No terminology or technology or kind of SAP can ever be neutral.

Concepts and documents have more or less stable functional values in relation to such standardized practices. There are of course always different views of whether such standardized practices should be changed or remain unchanged, and there are always different kinds of possible changes of those practices. There are always dynamic developments in meaning. The most important cause of this is the purpose for which we use the concepts, which is connected to the practices we want to change.

Often scientific and technological development changes standardized practices in a rather uncontroversial way. In other cases, however, changes in practices are related to different political interests, to different theories or "paradigms." Different paradigms tend to influence given practices in different ways and by doing so they also tend to change our symbolic systems as well as our production of documents and the form, content, meaning and use of those documents. The proper study of symbols and documents is thus based on the study of the functions and interests those documents are serving.

Any given tool (including a sign, concept, theory, finding or document) has given potentialities in relation to new social practices and paradigms. Users of libraries may discover that previously neglected documents may be very useful to develop a new social

practice or to defend an existing practice. Or they may find that well-known documents may be re-interpreted and used in new ways. Especially the kinds of documents known as "classics" have such abilities to be reused and re-interpreted by each generation. But even the most humble documents may turn out to be of value. This does not, of course, imply relativism regarding the relevance of documents: tools also quickly become obsolete and "only of historical interest." There may of course also be a discrepancy between the potential value of some documents and the cost of storing, organizing and using them.

The meaning of any sign is its potential quality of for referring to some objects or states of affairs (Karpatschof, 2000, p.197). A tool is something that has a functional value for some human (sub)culture. Languages (and sublanguages) are also tools with functional values. In languages, there are terms for tools. The meaning of a word for a certain kind of tool is the functional value of referring to a certain functional value, defining the quality of the tool (Karpatschof, 2000, p.197). A hammer may, for example, be termed "hammer" (denotation) or "murder-weapon" (connotation). The word we use about a tool facilitates one or another use of it. Words may be more or less appropriately used for a specific object in relation to a given task, activity or discourse. In other words, to use words is a kind of act (verbal acts) often used to accomplish something extra-verbal. The meanings of the words we use may be more or less suitable for our purposes, and in those cases we try to develop new words or change the meaning of some old words. Such changes in meanings are visible in the study of different fields, traditions and paradigms.

It is very naive and reductionistic to disregard these kinds of cultural inter-mediating factors in people's relationship with information. The dominant traditions in both information science and in behavioural and cognitive sciences have, however, neglected these cultural aspects and just tried to study generalized human relations to something termed "information." This dominant approach may be broadly termed "behavioural" in spite of different attitudes to versions of behaviourism. In this tradition people are expected to react to something in a specific, mechanical way without considering the culturally-determined meanings and without considering the different goals and values in the meanings and in the documents. This has, in my opinion, brought about a situation in which we have inherited very little useful knowledge from these areas on which to advance our field and practice. That view is also expressed by Robert de Beaugrande (1997) in linguistics. "I would certainly be happier if my findings had turned out to be less sobering or disturbing; but I can only report what I have in fact found" (§89).

In the view of pragmatism (and Activity Theory) languages are tools that are adapted culturally to suit the needs of their users. General languages may be seen as adaptations that suit the needs of major parts of the populations, while languages for special purposes are adaptations that suit the needs of specific groups such as chemists, lawyers, musicians, and so forth. Both languages, (or more broadly symbolic systems) as also media, documents and information systems, are teleological or goal-directed. They are optimised to do certain functions, while relatively ignoring other purposes and goals. They should be seen as specializations in the social division of labour. This implies that no system of KO can be optimised to do all kind of tasks equally well. Although modern computer based retrieval systems are very flexible and seem efficient for almost all tasks, it is important to consider the limitations of each kind of system for different kinds of tasks and different domains. The way knowledge is organized in information systems must be relevant to the specific purpose of that particular system. Relevance must, in my opinion, always be regarded in relation to a goal. Birger Hjørland and Frank Sejer Christensen (2002) defined relevance this way: "*Something (A) is relevant to a task (T) if it increases the likelihood of accomplishing the goal (G), which is implied by T"* [p. 964]. Our conception of relevance is thus connected to meaning and to different views and traditions in society. It is a basic view in the pragmatic tradition that the nature of knowledge is to fulfil goals for the organism or system that possess the knowledge. All kinds of knowledge or information processing processes or institutions are basically telelogical by nature. They have to fulfil certain goals. Their relevance criteria are constructed backwards from those goals, and people learn those relevance criteria by being socialized and educated within a particular context and tradition.

7. Paradigms, Epistemologies and KO

The concept of "paradigm" became influential through Thomas Kuhn's (1962, 1970, 1996) book. Kuhn did not, however, recognize the social sciences as sciences or as paradigmatic. He considered the social sciences to be preparadigmatic. In his view a paradigm should be a set of *common* assumptions. As stated by Dogan (2001):

> Within a formal discipline, several major theories may cohabit, but there is a paradigm only when one testable theory alone dominates all other theories and is accepted by the entire scientific community. When Pasteur discovered the microbe, the theory of spontaneous generation collapsed: contagion became the new paradigm. In the social sciences, however, we see at best a confrontation between several non- testable theories. Most of the time there is not even a confrontation but careful mutual avoidance, superb disregard on all sides. This is relatively easy, owing to the size of scientific disciplines, and their division into schools; it is true for all countries, big or small. (p. 11024)

Nonetheless, his theory has been very influential in the social sciences, and the concept of paradigm is often used in regard to the different schools, approaches, systems and so on. Thus, the concept of paradigm is usually applied in a broader way. "Paradigm" may, in the philosophy of science, mean more or less implicit background assumptions concerning the object of research, concerning research methods, concerning the usefulness of research, and so forth. In an even broader meaning the concept may be used with more or less coherent views concerning a collective practice, such as teaching, divine service, administration of justice, politics, and so forth. Kuhn's 1970 postscript in the second edition defines among other things a paradigm as a constellation of beliefs, values, techniques and so on shared by members of a given community.

Håkan Tjörnebohm, in1974, (as cited in Andersen, 1999, p. 89) defined paradigms as systems of (explicit or implicit) basic assumptions and epistemic ideals in scientific disciplines. A paradigm is a super individual structure of meaning, which is formed and reproduced in disciplinary socialization, teaching and scientific communication. Components of paradigms include:

> 1) Ideals and beliefs about science (epistemic goals, methods and criteria in the production and evaluation of scientific results inside the discipline);
> 2) World-view hypotheses, basic ontological assumptions; and
> 3) Ideals concerning significance for society and culture, for practical use and for enlightenment.

"Paradigm" in the broadest meaning of the word is a very central concept for KO. Not only do scientific paradigms determine how we study KO in information science (see section 9 and 10 below); it is also the case that the knowledge that we are going to organize is already conceptualised and organized according to more or less invisible structures, determined by different paradigmatic influences. This is done in many "layers" in ordinary languages, in subject languages, in documentary forms, in networks of citing papers, in informal networks and so on. Different paradigms usually imply different values and goals. KO in information science must know those different values and be able to fulfil the different goals.

8. Concepts and Conceptual Relations as Units of KO

Concepts are normally seen as the units of thought and knowledge. This implies that the fundamental units of KO are the (semantic) relations between concepts. If knowledge is defined as justified true belief (as in the Platonic tradition) then real knowledge is hard or impossible to identify and to classify. It is more fruitful to speak of knowledge claims, rather than knowledge itself. To speak of *knowledge claims* as the things represented in the literature and the thing to be classified is a more careful way of speaking, and there is no real loss by this way of speaking.[7]

The basic function of concepts is to provide a basis for dealing with the world. Concepts provide borders and classes in a continuous world. "Blue" delimits some wavelengths, "castle" delimits some kind of buildings, and "music" delimits some sounds. Our theories, conceptualisations and paradigms tend to classify things and knowledge about things according to the same basic principles. Anders Ørom (2003, in press) has demonstrated how, for example, pictures in museums, the literature about art and the classifications used in libraries and bibliographies all tend to be influenced by the same paradigm (e.g. the stylistic paradigm, the cultural historical paradigm, the semiotic paradigm, the social historical paradigm, the feminist paradigm etc.).

A descriptor or a class represents a concept[8], and information retrieval is essentially concerned with semantic relations between queries, document representations and texts. The smallest unit is the relations between two concepts. The theory of concepts (meaning, semantics) is, however, probably one of the most difficult and muddled research fields today. It is important to understand the different views between traditional theories based on logical positivism and alternative views based on pragmatic theories. Temmerman (1997) describes some of these differences.

> Note the distinction between on the one hand logical and ontological *classification*, which in traditional [theory of] Terminology is supposed to be possible in the mind without considering or using language and before the naming of the concepts takes place, and on the other hand *categorisation* which is a result of the interaction between language and the mind (p. 55). [9]

The importance of concepts has been recognized in the KO community. Dahlberg (1991) is one prominent example. The journal *Knowledge Organization* has as subtitle "International Journal. Devoted to Concept Theory, Classification, Indexing, and Knowledge Representation". In spite of this, the contributions about concepts have hitherto been very limited in number, and the concept of concepts is much better explored in other fields like philosophy, cognitive and developmental psychology, artificial intelligence and linguistics.

Different theories of concepts are deeply linked with epistemological theories. Empiricism, rationalism, historicism and pragmatism have their different conceptions of concepts.

"Paradigms"	*In philosophy*	*In psychology*
Empiricism	Simple concepts correspond to simple sensations. There are no necessary relations between simple concepts. Simple concepts may be combined to complex concepts. *Nominalism*: General concepts are names, which we put on classes of things (empirical generalizations).	Classical associationism. Behaviourism. Connectionism (neo-associationism or parallel distributed processing, which works with inductive, self-organizing, "neural" networks for the processing of sensory input)
Rationalism	Simple are concepts, which cannot in fruitful ways be defined by other concepts. They are not experienced, but inborn or maturated. Complex concepts are defined from simple concepts. Fundamental concepts are concepts, which are necessary in order to describe a field. Simple, complex and fundamental concepts enter into certain necessary mutual relations. The differentiation between simple and complex concepts is absolute (independent of domain, interests, points of view etc).	Classical Artificial Intelligence. Cognitivism. Works with deductive, rule-governed algorithms for the processing of sensory input. "Classical" or "Aristotelian" concepts exhaustively defined by sets of necessary and sufficient attributes. Modern theory (Rosch, Lakoff) Prototypetheory. Concepts are more or less prototypical instances of things.
Criticism (Kant) (Empirico-rationalism)	Concepts represent knowledge of the world mediated by our forms of reason or categories like space, time, thing and cause.	Jean Piaget's "genetic epistemology". "Concepts mature in the individual. They grow out like the teeth"
Historicism and hermeneutics	Concepts are formed in a historical process on the basis of pre-understanding and holistic perception. There is circularity between the forming of simple and complex concepts. The relations between simple and complex concepts are relative in relation to interests. Traditions and social communities play important roles for the forming of concepts.	Psychoanalytical interpretation. A satisfactory analysis of the functions of cognition cannot just explore sensation, memory and thinking in isolation, but must involve the whole person, and his or her developmental history both individually and collectively. Concepts are thus formed by influence of personal characteristics such as sex and social class.
Pragmaticism and critical realism	Knowledge and concepts are formed by people's practical activity in relation to the objects of the activity.	Activity theory: Our concepts are not just determined by attributes in physical objects. They have also "historical depth". They are formed in the historical contexts of the objects, which have meaning.

Fig. 4: Basic Conceptions of "Concept"

Frank C. Keil outlines some very important developments in theories about concepts:

> ...one thing has increasingly emerged. Coherent belief systems, or "theories," are critical to understanding the nature of concepts and how they develop. Although philosophers and some psychologists of a more classical bent have often argued about the interdependence of theories and concepts, for the most part cognitive psychologists in the last 20 years have not. As a consequence of the work done in the 1960s on concept learning and the later "anticlassical" response to that paradigm, it became the norm to talk about concepts as consisting of combinations of features or as being the result of summarizing operations on exemplars and/or dimensions. Although the classical and newer probabilistic views clashed on the issue of whether features for natural concepts were necessary and sufficient, their debate masked a shared set of assumptions that the meanings of concepts could be fully described by lists of features and simple probabilistic or correlational operations on those features.
>
> As researchers began to look at the phenomena associated with concepts more closely, however, it became increasingly clear from several different lines of evidence that something was missing... . . Part of the meaning of a concept may be the apprehension of the theoretical relations that explain its internal structure... . such relations have always been fundamentally important in the philosophy of science. (Keil, 1989, p. 267)
>
> The history of all natural sciences documents the discovery that certain entities that share immediate properties nonetheless belong to different kinds. Biology offers a great many examples, such as the discoveries that dolphins and whales are not fish but mammals, that the bat is not a kind of bird, that the glass "snake" is in fact a kind of lizard with only vestigial limbs beneath its skin. In the plant kingdom it has been found, for example, that some "vegetables" are really fruits and that some "leaves" are not really leaves. From the realm of minerals and elements have come the discoveries, among others, that mercury is a metal and that water is a compound ...
>
> In almost all these cases the discoveries follow a similar course. Certain entities are initially classified as members of a kind because they share many salient properties with other bona fida members of that kind and because their membership is in accordance with current theories. This classification may be accepted for centuries until some new insight leads to a realization that the entities share other, more fundamentally important properties with a different kind not with their apparent kind.
>
> Sometimes it is discovered that although the fundamental properties of the entities are not those of their apparent kind, they do not seem to be those of any other familiar kind either. In such cases a new theoretical structure must develop that provides a meaningful system of classification.
>
> There are many profound questions about when a discovery will have a major impact on a scheme of classification, but certainly a major factor is whether that discovery is made in the context of a coherent causal theory in which the discovered properties are not only meaningful but central (Keil, 1989, p. 159).

This long citation is important, I believe, because it demonstrates the deeper reality of kinds and concepts that science discovers. It has important implications for the methodology of KO. First and foremost it challenges many user-oriented and empiristic approaches.

To the degree that this view is correct, the relations between two concepts are thus relative to the theoretical systems (or paradigms) in which they are embedded. We cannot say objectively and once and for all, for example, that "schizophrenia" and "manio-depressive disorders" are narrower terms (NT) in relation to "psychosis". It has been suggested by Kringlen (1994) that aetiologically schizophrenia might not be a useful concept, while one should probably alternatively operate with the old German concept "Einheitspsychose". It follows that schizophrenia is not NT in relation to psychosis. To the degree that this is correct the semantic relations between "schizophrenia," "manio-depressive disorders" and "psychosis" are synonym relations, not generic relations.

We may conclude that the basic units in KO, the semantic relations between two concepts, may be relative to the perspective and the theory from which they are considered. Because of this fact, KO cannot be done just from successive combinations of ele-

ments, but must reflect broader perspectives and theories.

9. Criteria for Likeliness in KO.

A fundamental principle in KO is that like things should be brought together, while different things should be separated. "Likeliness" is a concept that may also be expressed by other terms such as "similarity", "sameness" (used by James, 1890), "resemblance" or "equivalence". Many writers in LIS have defined classification and KO by using the concepts of likeliness. For example:

- Ernest Cushing Richardson (1964, p. 1) defined classification as the "putting together of like things, or more fully described, it is the arranging of things according to likeness and unlikeness. It may also be described as the sorting and grouping of things".
- Henry E. Bliss (1935, p. 3) wrote: "In dealing with the multiplicity of particular things, actualities, and specific kinds, we find that some are alike, in general characters and in specific characteristics; and we may consequently relate them in a class, or classes, that is classify them".
- Frederick Wilfrid Lancaster (born 1933) defined classification as "sorting items into 'conceptual classes'" and "forming classes of objects on the basis of their subject matter". (Lancaster, 1998, p. 17).
- Arlene G. Taylor: "The placing of subjects into categories; in organizing of information, classification is the process of determining where an information package fits into a given hierarchy and then assigning the notation associated with the appropriate level of the hierarchy to the information package and its surrogate." (Taylor, 1999, p. 237).

How do we decide what things are alike? How do we develop our criteria for likeness? The literature on KO in LIS seems to ignore this core problem. Is this because the problem is seen as obvious? Is this based on a kind of naïve realism: that things are what they look like, and that people's immediate sense of likeness is adequate as the basis for KO?

But when are two things alike? Naïve realism confuses seeming similarity with similarity in an objective sense. Some metals might look like gold, but might not be precious metal. Chemical analysis (not common sense) has to make a distinction between "what looks like gold" and "what is gold". Classification made for children may be based on more superficial attributes – e.g., "the big book of trains" which considers most aspects of the "railway system". Scientific classifications, on the other hand, reflect some deeper properties, such as the classification of chemical substances in organic and inorganic compounds, precious metals, etc., based on atomic theory. When scientific principles of classification are applied, seemingly related objects may be separated and seemingly different objects may be grouped together. For example, whales were once classified as fish, but are today – influenced by evolutionary theory - classified as mammals.[10] These examples demonstrate that naïve realism is not adequate as a method to classify documents.

If we consider different examples, we will see that conceptual relations have different kinds of motivations.

- "Institutions for Information Science" are generically subordinate to "institutions". This relationship seems to be motivated by purely logical relations (or by relations inherent in a given language).
- "Copenhagen" is part of "Denmark" (Whereas Malmö in Sweden is not. This last example is motivated by the fact that Denmark lost this part of territory to Sweden in 1658). The parts of a country are thus defined by social arrangements. This type of semantic relations is in other words based on human conventions.
- Whales are today classified as mammals. The explanation of this semantic (generic) relation is due to evolutionary theory.[11]
- Psychology (and also psychopathology) may be classified as part of neuroscience (natural science), as part of the social sciences or as part of the humanities (or otherwise). Such differences are, for example, visible in how such fields are placed in the organizational structures in universities. Such classifications (and semantic relations) often involve professional interests. It is, for example, partly a question of professional power whether a field (e.g. psychopharmacology) is monopolized by a profession. At the deepest level, the question of whether psychology is a human science or a natural science is a scientific question related to theoretical questions within psychology. It is well known that psychology is divided over this question. Different paradigms in psychology have different answers. For behaviorism psychology is clearly a part of the natural sciences (cf. Watson, 1913). For humanistic psychology it is clearly a part of the humanities. In classifying psychology as

belonging to science, or social science or humanities, one is actually involved in a theoretical battle between paradigms. (Which most people find rather uncomfortable).
- Classifications and semantic relations may be established by empirical generalizations. One example is Berlin & Kay (1969), who established the empirical generalization that human languages reflect the classification of colour sensations in essentially the same ways universally, regardless of historical and cultural differences. Such empirical generalizations avoid the uncomfortable theoretical issues addressed above. There are, however, other kinds of problems with this approach, which are closely connected with the basic assumptions in empiricism.
- Classifications and semantic relations may also be purely accidental or ad hoc. Such classifications may serve some purposes very well. In general, however, classifications that reflect essential characteristics in the objects are the most valuable. (This "essentialism" should not be confused with an objectivism that ignores human activity).

We may conclude that there exist many different kinds of criteria for likeliness. They may be conventional, logical, psychological and so on. Regarding natural kinds, however, they should especially be seen as domain-specific criteria which are discovered by science. They are not just something that can be extracted from users or from statistical investigations.

10. Methods of Knowledge Organization

By methods of KO we mean the methods of constructing systems of KO such as classifications and thesauri. Also the The processes of indexing and classification are also considered methods of KO.

What are the methods for classification and KO? We may first make a distinction between classification in the sciences (such as biology or archaeology) and in LIS. This paper is about KO in LIS. However, we shall argue that basically the methods of KO in LIS are connected to the same fundamental paradigms in epistemology as the methods of classification in sciences and other fields. First, however, let us shortly consider some methods:

- Standardization
- Computer based KO
- "Manual" or "intellectual" methods
- Quantitative methods
- Qualitative methods
- Text based methods
- People based methods
- Institution based methods (e.g. studies of university organisations)
- Bibliometrical methods
- Word frequency based methods
- Sociological methods
- Historical methods
- Pragmatic, epistemological and critical methods

It is important to realize that an important quality of a system of KO may be its function as a standard. Often classifications and other systems are made as compromises between experts in a field and declared a standard. Scientific arguments between researchers in the field are not published and discussed in the literature as a part of this process. This method is epistemologically a kind of "authoritanism": the quality of the system is claimed by reference to its authority as a standard. However, as researchers in KO we have to consider other qualities of KO too and should not accept "authoritanism" as a criterion of scientific knowledge.

One distinction, one that may be encountered in the literature, is between computer based methods and "manual" methods of KO. The term "manual" is a little peculiar, because what is meant is rather intellectual methods. This distinction is not in itself a fundamental distinction. If it is possible to develop some explicit rules for how to classify documents, then such rules can be used by humans or by computers. In other words, sometimes humans are not performing better than computers, which follow rather simple instructions. At other times, however, humans apply deep knowledge, which is not available to computers. And sometimes computers process huge amounts of data that are not possible for humans to process. We should thus not consider computer based versus manual methods as a basic methodological distinction. We should rather classify the methods of KO in other ways asking what kind of rules, what kind of knowledge and what kind of cognitive processes are involved in the process? In general our knowledge of how humans classify is limited.[12]

Another distinction is between quantitative versus qualitative methods. Quantitative methods may, for example, be based on word frequencies. Qualitative methods may, on the other hand, be based on interpretations of meaning. Again, I do not consider this a fundamental distinction. It is well known that any quantitative investigation presupposes an adequate

qualitative analysis. Qualitative methods, on the other hand, may result in a number of categories and thus approximate the quantitative approach. The sociologists Pierre Bourdieu uses many statistics (quantitative data) in his research; however, he explicitly denies doing this in a positivist way. There are thus important ways in which to use data that go deeper than the distinction between quantitative and qualitative approaches.

Systems of KO may use data from the literature, from people, from organizations, etc., in many different ways and in many combinations. Bibliometrical approaches are mainly based on literatures, and so are many conventional and "manual" methods (including the well known principle of identifying *literary warrant* for a given class or descriptor). The use of historical knowledge about the development of knowledge fields (as exemplified in Wallerstein, 1996 and Hjørland, 2000c) is a different way to use literature as data for KO. This last approach also uses the development of the organization of universities as a basis for constructing KO.

People may also be used as sources on which to base the development of KO. In cognitive science and artificial intelligence the methods of *knowledge elicitation* are well known (e.g. Cooke, 1994). This is, in my opinion, *an indirect method* because we have to know the methods by which the experts, from whom we obtain the knowledge, got this in the first place. In other words we must be able to argue which knowledge claims are best substantiated. If we do not address this issue, we are only indirectly qualified to construct and evaluate systems of KO. People may, however, also be used as sources in other ways. They may, for example, be seen as members of *discourse communities* and be studied as such. We may study the social division of labour and the dependency between different people and groups of people (compare Whitley, 1984).

When we are going to decide what kind of sources to build on (literature, people, organizations), we need to know something about how the relevant knowledge is distributed. Methods of knowledge elicitation seem to be built on the assumption that the needed knowledge is ready at hand within a group of experts or other people. This may be more or less the case. However, more scholarly approaches build on other assumptions. The meaning of words is traced in great detail and registered in historical dictionaries with quotations and citations from authoritative authors. In this case, the relevant information is not supposed to be ready at hand in the mind of experts, but something that might be constructed by working with the literature (not just as a kind of opinion poll). The question of cognitive authority is of course extremely important and difficult. Scholarly work is not just "subjective", choosing authorities according to personal preferences. The more qualified, the less "subjective" in this sense of the word. The "objectivity" of the decision of scholarly work is partly based on the knowledge of different traditions and metaperspectives and on the cumulated historical evidence. On could perhaps say that a fundamental difference between empiricism/rationalism and historicism/pragmatism is the emphasis on reading in the research process.

Basically, the methods of KO are related to fundamental theories of epistemology. All researchers in any field are always more or less influenced by certain ideals on how to obtain knowledge. This is also the case with regard to constructing systems of KO. If students of LIS have taken a course in the philosophy of science, they should be able to identify the dominant epistemic approach in any paper on KO (and also in any system of KO). Some papers and systems are mostly based on empirical generalizations. Systems based on word-frequency measures may be the best example. Other papers and systems are mostly based on rational rules and deductions (while often ignoring empirical issues). Facet-analytic systems in the tradition of Ranganathan may provide the best examples. A third kind of system is based on the study of the evolution of knowledge fields. To a certain degree the disciplinary based systems such as Dewey (DDC) can be said to conform to the ideal. On the other hand the updating of this system is more is and more influenced by the principles of the facet-analytic tradition (as discussed by Miksa, 1998). And in some sense the DDC is also empirically based on the basis of the literature, which is classified by the system. DDC is thus not as clear in its methodological principles as the former examples. The fourth basic ideal is the pragmatic epistemology. In this case the systems are developed on the basis of the analysis of goals, values and consequences. For example, female researchers may regard the goal of women's emancipation as the main goal of feminist scholarship. They may thus regard any relations that support this goal as important. "Similarity" is thus identified by classes of documents sharing similar functions in relation to this goal. Pragmatic classification may thus also be termed critical or political classification. It may at first seem strange and opposed to common scientific norms.

Pragmatic epistemology and pragmatic KO do not mean that a person (or a whole field) can just do things the way that suits his personal interests (or the interests of the researchers in the field). If this is done, if research just produces "social constructions" then reality will make those constructions incoherent. They will be opposed by empirical and theoretical arguments. The production of incoherent "knowledge" is not valuable and cannot be a serious goal. Therefore pragmatic philosophy is bound up with a form of realism. The pragmatic method is not opposed to aspects of empiricism, rationalism and historicism. It claims, however, that no isolated evidence is enough. The final criteria of truth are connected to human goals and activities. You cannot avoid considering such issues although they may seem uncomfortable.

The founder of pragmatism, Charles Sanders Peirce (1839-1914), wrote:

> The rational meaning of every proposition lies in the future. How so? The meaning of a proposition [its logical interpretant] is itself a proposition. Indeed, it is no other than the very proposition of which it is the meaning: It is a translation of it. But of the myriads of forms into which a proposition may be translated, what is that one which is to be called its very meaning? It is, according to the pragmaticist, that form in which the proposition becomes applicable to human conduct, ... that form which is most directly applicable to self-control under every situation and to every purpose. This is why he locates the meaning in future time; for future conduct is the only conduct that is subject to self-control. (Peirce, 1905)

In this way the pragmatic epistemology sees the meaning of words as connected to speech acts and to the goals that humans try to satisfy through their acts. In the same way it regards semantic relations and classifications as determined by their functions as tools for human goals. As the pragmatic philosopher John Dewey stated:

> So far nominalism and conceptualism - the theory that kinds exist only in words or in ideas - was on the right track. It emphasized the teleological character of systems and classifications, that they exist for the sake of economy and efficiency in reaching ends. But this truth was perverted into a false notion, because the active and doing side of experience was denied or ignored. Concrete things have *ways* [emphasis in original] of acting, as many ways of acting as they have points of interaction with other things. One thing is callous, unresponsive, inert in the presence of some other things; it is alert, eager, and on the aggressive with respect to other things; in a third case, it is receptive, docile. Now different ways of behaving, in spite of their endless diversity, may be classed together in view of common relationship to an end. No sensible person tries to do everything. He has certain main interests and leading aims by which he makes his behavior coherent and effective. To have an aim is to limit, select, concentrate, group. Thus a basis is furnished for selecting and organizing things according as their ways of acting are related to carrying forward pursuit. Cherry trees will be differently grouped by woodworkers, orchardists, artists, scientists and merry-makers. To the execution of different purposes different ways of acting and reacting on the part of trees are important. Each classification may be equally sound when the difference of ends is borne in mind.
>
> Nevertheless there is a genuine objective standard for the goodness of special classifications. One will further the cabinetmaker in reaching his end while another will hamper him. One classification will assist the botanist in carrying on fruitfully his work of inquiry, and another will retard and confuse him. The teleological theory of classification does not therefore commit us to the notion that classes are purely verbal or purely mental. Organization is no more merely nominal or mental in any art, including the art of inquiry, than it is in a department store or railway system. The necessity of execution supplies objective criteria. Things have to be sorted out and arranged so that their grouping will promote successful action for ends. Convenience, economy and efficiency are the bases of classification, but these things are not restricted to verbal communication with others nor to inner consciousness; they concern objective action. They must take effect in the world.
>
> At the same time, a classification is not a bare transcript or duplicate of some finished and done-for arrangement pre-existing in nature. It is rather a repertory of weapons for attack upon the future and the unknown. For success, the details of past knowledge must be reduced from bare facts to meanings, the fewer, simpler and more extensive the better... (Dewey, 1948, p. 151-154).

Figure 5 below summarizes the four fundamental methods of KO for which I have argued. It is important to realize that they are idealizations. The do not and they cannot exist in pure forms. Any empirical procedure must involve kinds of logic and non-empirical evidence. All rules must be applied to some empirical reality. All empirical and rational methods must be based on pre-understandings and meanings transmitted through language and cultural products. And any kind of pragmatism is limited by constrains set by the real world through empirical evidence.

By knowing the principal strengths and weaknesses of basic epistemological positions, the information specialist is equipped with some general understanding of the strong and weak sides of different methods of KO.

	"Scientific Classification"	*"Bibliographic Classification"*
Empiricism (Observations and inductions)	Classification provided by statistical generalizations (e.g. factor analysis) based in "similarity". Examples: Classifications of mental illness in psychiatry (DSMIV) kinds of intelligence in psychology based on statistical analysis of test scores.	Documents clustered on the basis of some kind of similarity, e.g. common terms in traditional IR or bibliographical coupling. Examples: "Atlas of science" and visualizations (White & McCain, 1998). "Research Fronts" I SCI and algorithms for information retrieval.
Rationalism (Principles of pure reason. Deductions)	Classification based on logical, universal divisions. Examples: Frame based systems in Artificial Intelligence. Chomsky's analysis of the deep structure in language-	Facet analysis built on logical divisions and "eternal and unchangeable categories" Examples: Ranganathan, Bliss II & Langridge. Semantic networks. According to Miksa (1996) the DDC has increasingly used this approach.
Historicism (Study of context and development- Explicating pre-understanding)	Classification based on historical or evolutionary development. Examples: Biological taxonomies based on evolutionary theory. Classification of the sciences on the basis of their history and organizational structures.	Systems based on the study of the development of knowledge and knowledge producing communities (the social division of (scientific) labour). Examples: Wallerstein (1996) (See also Hjørland, 2000c). The feature of the DDC that it distributes subjects by discipline.
Pragmatism (Analysis of goals, values and consequences in both subject and object)	Classifications based on specific values, policies and goals, e.g. feminist epistemology.	Systems based on "cultural warrant" or "critical classification". Examples: The French encyclopaedists, the Marxists, Classifications serving feminist collections.

Figure 5: Fundamental Methods of Classification

11. Conclusion

This paper has outlined the fundamental aspects of KO as a field of study. The basic assumptions and attitudes have been that the basic units of KO are semantic relations between concepts. Such semantic relations cannot primarily be established by universalistic assumptions, but much primarily be understood as domain specific, as uncovered by (and constructed by) scientific disciplines. KO in LIS cannot ignore concepts, theories and findings in specific disciplines, and the methods of KO in LIS are at the deepest level based on the same philosophical assumptions as the methods of science and scholarship. This implies that the fundamental discussion of the basis of KO in LIS is strongly connected to the discussion of different theories of epistemology. Epistemological studies have been rare in LIS, and it seems urgent for our field to upgrade our qualifications in this area.

Notes

1 This article is a slightly modified version of a paper presented in Salamanca, Spain (Hjørland, 2003a).
2 From 1937 named *Féderation Internationale de Documentation*, *FID,* and from 1986 *International Fédéra- tion for Information and Documentation, FID.*
3 The history of the abstract journal goes, however, back to 1665, cf. Manzer (1977).
4 Computer science deals with "information technology", IT, or with data processing. Computer science is in Denmark called "datalogi", the science of data.
5 I write this well aware that the pragmatic philosopher John Dewey used the mirror metaphor: "A classification of books to be effective on the practical side must correspond to the relationships of subject-matters, and this correspondence can be secured only as the intellectual, or conceptual, organization is based upon the order inherent in the fields of knowledge, which in turn mirrors the order of nature." (Dewey, 1929, p. viii)
6 Whereas traditional semiotics tends to be formalistic and to abstract signs from their contexts of use; social semiotics examines semiotic practices, specific to a culture and community, for the making of various kinds of texts and meanings in various situational contexts and contexts of culturally meaningful activity. For an introduction to social semiotics see, for example, Hodge & Kress (1988). The formulations in this paper have been inspired by Karpatschof (2000).
7 Sometimes a differentiation is made between the organization of knowledge and the organization of other "things" (including physical objects, processes, phenomena such as pain etc.). Other things than knowledge (claims) do not, however, classify themselves. They are also classified by our conceptualization of the world.
8 In the literature of cognitive science (e.g. Keil, 1989) a distinction is made between word meanings and concepts, or between concept structure and semantic structure. People are described as possessing two different systems: a semantic system of word meanings and a concept system. This distinction is not made in the present paper.
9 "Saussurian structuralist semantics believe that the best way to describe meaning is to describe the mutual deliminating of concepts (semantic relations). Traditional Terminology believes that the best way to describe concepts is to determine their position in a concept system which visualises logical and ontological relationships. On the basis of the position in the concept system a definition will be formulated" (Temmermann, 1997, p. 53).
10 Aristotle noticed that dolphins and porpoises had lungs. They did not lay eggs, and they nourished their unborn young via a placenta. He grouped them with the quadrupeds rather than fish. (Aristotle was a lone voice. For over 2,000 years afterwards, his idea was lost, and whales were fish. Only in relatively modern times were whales put back with mammals.)
11 Reveal (1999) describes different paradigms, methods and tools used in plant taxonomy and mentions among other things the use of (1) anatomy: 1800s, (2) chromosome numbers: 1920's, (3) comparative garden studies: 1940's, (4) palynology: 1950's, (5) phenetics: 1960's, (6) biochemical systematics: 1970's, (7) cladistics: 1980's and (8) chloroplast DNA and RNA: 1990's.
12 Klumpner (1993, p. 1) reports: "These problems led the Indexing Study Group of the American Psychoanalytic Association to an experiment. About a dozen seasoned analysts independently indexed a passage from the *Standard Edition* [the works of Sigmund Freud]. When they compared what they had done, all agreed that the failure to agree about which terms to index was humbling and impressive. The group did not even agree on which words required *see* or *see also* directives, or on the words that should follow those directives. A dead end?". Klumpner himself tried to solve the problem by a quantitative study of actual produced indexes. What I would like to emphaze is that in order to progress we need those kinds of disagreements recorded and discussed qualitatively in the literature. They are extremely rare, for which reason I claim that we know little about "manual" KO.

References

Albrechtsen, H.; Andersen, H.H.K.; Bødker, S.; Pejtersen, A.M.,(2001). *Affordances in Activity Theory And Cognitive Systems Engineering*. Denmark. Risø: Available at: http://www.risoe.dk/rispubl/SYS/syspdf/ris-r-1287.pdf

Andersen, H. (1999). Political Attitudes and Cognitive Convictions among Danish Social Science Researchers. *Scientometrics, 46*(1), 87-108.

Arbib, M. A.; Kfoury, A.J. & Moll, R. N. (1981). *Basis For Theoretical Computer Science*. Berlin: Springer-Verlag.

Beaugrande,R. de (1997). On history and historicity in modern linguistics. Formalism versus functionalism revisited. *Functions of Language, 4/2*, 169-213. http://www.beaugrande.com/

Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California Press.

Bernal, J. D. (1948). Preliminary Analysis of Pilot Questionnaire on the Use of Scientific Literature. In: Royal Society [Great Britain]. The Royal Society Scientific Information Conference: Report and Papers Submitted. 1948 21 June – 2 July; London, England. London, England: Royal Society. Pp. 589-637.

Bliss, H. E. (1935). *A System of Bibliographical Classification.* New York: The H. W. Wilson Company.

Borko, H. (1968). Information science: What is it? *American Documentation*, *19*(1), 3-5.

Bradford, S. C. (1950). *Documentation.* London: C. Lockwood.

Buckley, G. (2001). *Semantics.* http://www.ling.upenn.edu/courses/Spring_2001/ling001/semantics.html [visited 2003-01-08]

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies , 41*(6) , 801-849.

Dahlberg, I. (1991). Philosophical foundations of conceptual ordering systems. *Advances in Knowledge Organization, 3*, 102-119.

Daly, L. W. (1967). *Contributions to a History of Alphabetization in Antiquity and the Middle Ages*, Brussels: Latomus.

Dewey, J. (1929). Introduction. IN: H. E. Bliss: *The Organization Of Knowledge And The System Of The Sciences*. New York, Holt (pp. vii-ix).

Dewey, J. (1948). *Reconstruction in Philosophy*. Enlarged edition. New York: Beacon. (Original work published 1920)

Dewey, M. (1979): *Dewey Decimal Classification and Relative Index*. Edition 19. Albany, N.Y.: Forest Press, 1979. Vol. 1.

Dogan, M. (2001). Paradigms in the Social Sciences. IN: *International Encyclopedia of the Social & Behavioral Sciences*, (vol. *16*, pp. 11023-11027). Ed. By N. J. Smeltser & P. B. Baltes. Amsterdam: Elsevier.

Ellis, B. (1990). *Truth and Objectivity*. Oxford: Basil Blackwell.

Fjordback Søndergaard, T.; Andersen, J. & Hjørland, B. (2003). Documents and the Communication of Scientific and Scholarly Information. Revising and Updating the UNISIST Model. *Journal of Documentation, 59*(3), 278-320.

Frohmann, B. (1994). The Social Construction of Knowledge Organization: The Case of Melvin Dewey. *Advances in Knowledge Organization*. *4*, 109-117.

Griffith, B. C. (Ed.). (1980). *Key Papers in Information Science*. New York: Knowledge Industry Publications.

Hargens, L. L. (2000). Using the literature: reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65 (December), 846-865.

Hjelmslev, L. (1943) *Omkring sprogteoriens grundlæggelse*. København: B. Lunos bogtrykkeri a/s. (Many later editions and translations, e.g. *Prolegomena to a Theory Of Language*. translated by Francis J. Whitfield. Baltimore : Waverly Press, 1953).

Hjørland, B. (1988). Information Retrieval in Psychology. *Behavioral and Social Sciences Librarian*, Vol 6 (3/4), 1988, 39-64.

Hjørland, B. (1992). The concept of "subject" in Information Science. *Journal of Documentation*, vol. 48, no. 2, 172-200.

Hjørland, B. (1993). *Emnerepræsentation og informationssøgning. Bidrag til en teori på kundskabsteoretisk grundlag*. Göteborg: Valfrid,

Hjørland, B. (1997): *Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science*. Westport & London: Greenwood Press.

Hjørland, B. (1998a). Information retrieval, text composition, and semantics. *Knowledge Organization, 25*(1/2), 16-31.

Hjørland, B. (1998b). The Classification of Psychology: A Case Study in the Classification of a Knowledge Field. *Knowledge Organization*, 24(4), 162-201.

Hjørland, B. (2000a). Documents, Memory Institutions, and Information Science. *Journal of Documentation*, vol. *56*(1), 27-41

Hjørland, B. (2000b). Library and Information Science: Practice, theory, and philosophical basis. *Information Processing and Management, 36*(3), 501-531.

Hjørland, B. (2000c). Review of Wallerstein et al. (1996). Open the Social Sciences, report of the Gulbenkian Commission on the Restructuring of the Social Sciences. Stanford, Calif.: Stanford University Press. I: *Knowledge Organization, 27*(4), 238-241

Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology. 52*(9):774–778.

Hjørland, B. (2002a). Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of Documentation, 58*(4), 422-462

Hjørland, B. (2002b), Epistemology and the Socio-Cognitive Perspective in Information Science. *Journal of the American Society for Information Science and Technology, 53*(4), 257-270.

Hjørland, B. (2002c) The Methodology of Constructing Classification Schemes. A Discussion of the State-of-the-Art. IN: *Proceedings of the Seventh International ISKO Conference 10.-13. July 2002, Granada, Spain*. Ed. By María J. López-Huertas. Würzburg, Germany: Ergon Verlag, pp. 450-456. (Advances in Knowledge Organization, Vol. 8).

Hjørland, B. (2002d). Principia Informatica. Foundational Theory of Information and Principles of Information Services. IN*: Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)*. Ed. By Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, Colorado, USA: Libraries Unlimited. (Pp. 109-121).

Hjørland, B. (2003a). Fundamentals of knowledge organization. IN: *Tendencias de investigación en organización del conocimiento / Trends in Knowledge Organization Research.* (Pp. 83-116; +English abstract p. 51 + Spanish abstract p. 19). José Antonio Frias & Críspulo Travieso (Eds.). Salamanca, Spain: Ediciones Universidad de Salamanca.
Presentation available at: http://www.db.dk/bh/publikationer/Filer/Presentation.ppt

Hjørland, B. (2003b), Social And Cultural Awareness and Responsibility in Library, Information and Documentation Studies. IN: *Aware and Responsible:Papers of the Nordic-International Colloquium on Social And Cultural Awareness and Responsibility in Library, Information and Documentation Studies (SCARLID)*. Edited by W. Boyd Rayward, Joacim Hansson, Vesa Suominen. Lanham, MD: Scarecrow Press, (in press).

Hjørland, B. & Albrechtsen, H. (1995). Toward A New Horizon in Information Science: Domain Analysis. *Journal of the American Society for Information Science*, *46*(6), 400-425.

Hjørland, B. & Kyllesbech Nielsen, L. (2001). Subject Access Points in Electronic Retrieval. *Annual Review of Information Science and technology, vol. 35*, 249-298.

Hjørland, B. & Sejer Christensen, F. (2002). Work tasks and socio-cognitive relevance. A specific Example. *Journal of the American Society for Information Science and Technology, 53*(11), 960-965.

Hodge, R. & Kress, G. (1988). *Social Semiotics*. New York; Cornell University Press.

James, W. (1890). *The Principles of Psychology*. (2 vols.). New York: Henry Holt.

Karpatschof, B. (2000). *Human activity. Contributions to the Anthropological Sciences from a Perspective of Activity Theory*. Copenhagen: Dansk Psykologisk Forlag. ISBN: 87-7706-311-2

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, Massachusetts: The MIT Press.

Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation vol. 16*, no. 3, 223-233.

Klumpner, G. H. (1993). *A Guide To The Language Of Psychoanalysis. An Empirical Study Of The Relationships Among Psychoanalytic Terms And Concepts*. Madison, Connecticut: International University Press, Inc.

Kringlen, E. (1994). Is the concept of schizophrenia useful from an aetiological point of view? A selective review of findings and paradoxes. *Acta Psychiatrica Scandinavica, 90* (suppl. 384), 17-25.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press. 2rd. eition 1970. 3rd edition 1996.

Lancaster, F. W. (1998). *Indexing And Abstracting In Theory And Practice. 2nd* ed. London: Library Association.

Manzer, B. M. (1977). *The Abstract Journal, 1790-1920: Origin, Development and Diffusion*. Metuchen, N.J.: The Scarecrow Press.

Miksa, F. L. (1998). *The DDC, The Universe of Knowledge, and the Post-Modern Library*. Albany, NY: Forest Press.

Pao, M. L. (1993). Term and citation retrieval: A field study. *Information Processing & Management*, *29*(1), 95-112.

Pao, M. L., & Worthen, D. B. (1989). Retrieval effectiveness by semantic and pragmatic relevance. *Journal of the American Society for Information Science*, *40*(4), 226-235.

Peirce, C. S. (1905). What pragmaticism is. *The Monist, 15*, 161-181.

Poulsen, C. (1994). *Informationens skygge og foran: informationskvalitet, informationseksplosion og online kataloger*. Roskilde, Denmark: Institut for Datalogi, Kommunikation og Uddannelsesforskning, Roskilde Universitetscenter.

Rees-Potter, L. K. (1987). *A bibliometric analysis of terminological and conceptual change in sociology and economics with the application to the design of dynamic thesaural systems. Vol. 1-2*. Ph.D dissertation. *Dissertation Abstracts International, 48*(06), SecA, p.1344.

Rees-Potter, L. K. (1989). Dynamic thesaural systems: a bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural sys-

tems. *Information Processing & Management, 25*(6), 677-691.

Rees-Potter, L. (1991). Dynamic thesauri: the cognitive function. Tools for knowledge organisation and the human interface. *Proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990.* Part 2, 1991, 145-150.

Reveal, J. L. (1999). PBIO 250 Lecture Notes. Principles of Plant Taxonomy. http://www.inform.umd.edu/PBIO/pb250/prin.html (hentet 10-02-2003)

Richardson, E. C. (1964). *Classification. Theoretical and practical. Third edition.* Hamden, Connecticut: The Shoe String Press, Inc. (Reprinted unaltered from 1930 edition).

Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.

Schneider, J. W. & Borlund, P. (2002). Preliminary study of the potentiality of bibliometric methods for the construction of thesauri. IN*: Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)*. Ed. By Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, Colorado, USA: Libraries Unlimited. (Pp. 151-165).

Taylor, A. G. (1999). *The Organization of Information*. Englewood, Colerado: Libraries Unlimited, Inc.

Temmerman, R. (1997). Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes. Journal of Linguistics, 18*, 51-93. Available at: http://hermes.asb.dk/archive/FreeH/H18_04.pdf (Visited 2003-01-11).

Törnebohm, H. (1974). *Paradigm i vetenskapernas värld och i vetenskapsteorien*. Göteborg: University of Göteborg.

Unisist (1971), *Study Report on the feasibility of a World Science Information System*. By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Paris, Unesco.

U.S. Department of Labor. Bureau of Labor Statistics (2002). *Standard Occupational Classification* (SOC) System. Washington. D.C. http://www.bls.gov/soc/ (Visited December 30, 2002).

Wallerstein, I. (1996). *Open the Social Sciences, report of the Gulbenkian Commission on the Restructuring of the Social Sciences*. Stanford, Calif.: Stanford University Press.

Watson, J. B. (1913) Psychology as the behaviorist views it. *Psychological Review, 20*, 158-177.

Whitley, R.(1984). *The Intellectual and Social Organizations of the Sciences*. Oxford University Press.

Ørom, A. (2003) The organization of knowledge about art. Paper to appear in *Knowledge Organization* 2003.