Knowl. Org. 31(2004)No.1
Birger Hjørland, Jeppe Nicolaisen: Scientific and Scholarly Classifications are not "Naïve"

55

Scientific and Scholarly Classifications are not "Naïve": A Comment to Begthol (2003)

Birger Hjørland and Jeppe Nicolaisen

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

Introduction:

Relationships between Knowledge Organization in LIS and Scientific & Scholarly Classifications

In her paper "Classification for Information Retrieval and Classification for Knowledge Discovery: Relationships between 'Professional' and 'Naïve' Classifications" (KO v.30, no.2, 2003), Beghtol outlines how scholarly activities and research lead to classification systems which subsequently are disseminated in publications which are classified in information retrieval systems, retrieved by the users and again used in scholarly activities and so on. We think this model is correct and that its point is important. What we are reacting to is the fact that Beghtol describes the classifications developed by scholars as "naïve" while she describes the classifications developed by librarians and information scientists as "professional." We fear that this unfortunate terminology is rooted in deeply anchored misjudgments about the relationships between scientific and scholarly classification on the one side and LIS classifications on the other. Only a correction of this misjudgment may give us in the field of knowledge organization a chance to do a job that is not totally disrespected and disregarded by the rest of the intellectual world.

The Nature of Scholarly and Scientific Classifications

The most respected and recognized of all scientific classifications is without much doubt *The Periodic Systems of the Element* in chemistry and physics. This classification is the result of research activities, which stands as a model for research, as defining the very nature of "real" science, of real progress in knowledge, as real pragmatic utility for mankind and of scientific consensus. To associate this classification with the adjective "naïve" is indeed misplaced. To add that they are "classifications in the wild" (p. 65) is an underlining that here is really something out of place.[1]

It should be said, however, that Beghtol does not consider the Periodic System or any other scientific systems for that matter. She writes: "the paper investigates a number of naïve knowledge discovery classifications as examples in order to compare and contrast them to information retrieval classifications, their purposes and methods. These naïve classification systems have been chosen from the humanities and social sciences because scholarly research and activities in those disciplines illustrates the distinction between artifacts and mentefacts and, further, are not constrained by the attributes of the natural world that constrain classificatory work in the physical sciences" (p. 65).

We cannot understand this argument. Scientists and scholars may *discover* certain attributes and relations in reality and may on that basis construe classifications that are both beautiful and widely accepted as strongly informative and of great practical utility. *They also form the basis of bibliographical classifications such as, for example, the UDC-classification.* Of course the natural world constrains classificatory work. If this were not the case classifications would be unjustified or abitrary constructions. It is exactly the reflection of objective attributes and relations that make classifications (or taxonomies) widely recognized as representing most valuable contributions.

In relation to Beghtol's argument, there is no reason to make a distinction between science on the one hand and social sciences and humanities on the other hand. Classifications are produced in both the sciences, the social sciences and the humanities and

56

Knowl. Org. 31(2004)No.1
Birger Hjørland, Jeppe Nicolaisen: Scientific and Scholarly Classifications are not "Naïve"

they are important for how "information retrieval classifications" should be designed. They are not "naïve" compared to library classifications, if anything it is the other way round.

Some examples on classifications produced in science and scholarship:

– In archaeology: Study of human artififacts[2]
– In biology: Taxonomies of plants and animals.[3]
– In linguistics: Classification of languages as well as their parts (e.g., classification of words into classes such as nouns, verbs, adverbs, etc.)
– In physical geography: Classifications of areas in zones such as tropical, subtropical and temperate.
– In cultural geography: Classifications of towns, countries, etc.
– In music: Classifications of music into genres and instruments into categories (such as wind instruments, string instruments, percussions, etc.)
– In psychology: Classification of abilities and other mental phenomena.[4]
– Classification of the social sciences into disciplines like economics and sociology (cf. Hjørland, 2000).

One of the domains that Beghtol refers to is the classification of religions. In *The Encyclopædia Britannica* there is a lengthy treatment of the classification of religions (Adams, 1994). The article discusses normative principles of classification, geographical criteria, ethnographic-linguistic principles, philosophical principles, morphological criteria, phenomenological principles, among others and it concludes:

– First, classifications should not be arbitrary, subjective, or provincial. A first principle of the scientific method is that objectivity should be pursued to the extent possible and that findings should be capable of confirmation by other observers.
– Second, an acceptable classification should deal with the essential and typical in the religious life, not with the accidental and the unimportant. The contribution to understanding that a classification may make is in direct proportion to the penetration of the bases of religious life exhibited in its principles of division. A good classification must concern itself with the fundamentals of religion and with the most typical elements of the units it is seeking to order.
– Third, a proper classification should be capable of presenting both that which is common to religious forms of a given type and that which is pe-

culiar or unique to each member of the type. Thus, no classification should ignore the concrete historical individuality of religious manifestations in favour of that which is common to them all, nor should it neglect to demonstrate the common factors that are the bases for the very distinction of types of religious experience, manifestations, and forms. Classification of religions involves both the systematic and the historical tasks of the general science of religion.
– Fourth, it is desirable in a classification that it demonstrate the dynamics of religious life both in the recognition that religions as living systems are constantly changing and in the effort to show, through the categories chosen, how it is possible for one religious form or manifestation to develop into another. Few errors have been more damaging to the understanding of religion than that of viewing religious systems as static and fixed, as, in effect, ahistorical. Adequate classifications should possess the flexibility to come to terms with the flexibility of religion itself.
– Fifth, a classification must define what exactly is to be classified. If the purpose is to develop types of religions as a whole, the questions of what constitutes a religion and what constitutes various individual religions must be asked. Since no historical manifestation of religion is known that has not exhibited an unvarying process of change, evolution, and development, these questions are far from easily solved.

With such criteria in mind it should be possible continuously to construct classification schemes that illuminate man's religious history" (Adams, 1994; bullets added). Such principles are, in our opinion, also important to consider in a LIS-context.

Library Classification is widely dependent on such scientific and scholarly classifications. Lack of subject knowledge in relation to such classifications may often lead to poor quality in information retrieval classifications. In such cases library classifications may be characterized as uninformed or "naïve." This fact was also recognized by the fathers of knowledge organization, who, for example, wrote:

I believe ... that the maker of a scheme for book arrangement is the most likely to produce a work of permanent value, if he keeps always before his mind a classification of knowledge. (Cutter, 1888)[5]

Knowl. Org. 31(2004)No.1
Birger Hjørland, Jeppe Nicolaisen: Scientific and Scholarly Classifications are not "Naïve"

57

Sayers expressed it in the following way:

> A book classification must hold the minuteness of the knowledge classification as an ideal to which it must approximate as nearly as possible" and further (p. 34): "It must be clearly borne in mind, however, that the classification of knowledge should be the basis of the classification of books; that the latter obeys in general the same laws, follows the same sequence. (Sayers, 1915, p. 31)

And Richardson said:

> In general the closer a classification can get to the true order of the sciences and the closer it can keep to it, the better the system will be and the longer it will last. (Richardson, 1964, p. 33)

The generalization of scientific classification principles and methods

We in LIS should obviously be concerned with generalized principles and methods of classification. Very often it seems however, as if we ignore the work done by scientists, philosophers and scholars.

Often scholars are painfully aware that, in spite of all efforts, their classifications are not satisfactory. In the social sciences Fenger writes:

> In the behavioral and social sciences, hundreds of classifications are published every year. Noteworthy examples are Bloom's taxonomy of educational objectives (Krathwohl et al. 1964[6]), as well as the *DSM* (*Diagnostic and Statistical Manual of Mental Disorders*[7]) and *ICD* (*International Classification of Diseases*[8]) classification systems used in psychology and psychiatry. None of these systems have been formally derived, however. Instead, they were generated based on 'experience.' The resulting classes are so heterogeneous that they acknowledge many exceptions. Also, a phenomenon called `comorbidity' shows that these classification systems are not optimal yet. It refers to the simultaneous existence of two or more disturbances in the same patient. If comorbidity is the rule rather than the exception, then the classification system loses plausibility and practicability. (Feger, 2001, p. 1968)

There is a close connection between the development of scientific concepts and classifications. When astronomy recognizes the different nature of stars and planets, for example, they reflected this in both their concepts and their classifications. This makes the study of the development of scientific concepts and conceptions highly relevant for LIS.

What research methods are being used to construe scientific classifications? The answer is, that there are many. One family of methods is statistical methods such as cluster analysis and factor analysis. These are directly "methods of classification." Often, however, classifications are arrived at using other kinds of methods and often more indirect methods. Frank C. Keil illuminates this:

> 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

58

Knowl. Org. 31(2004)No.1
Birger Hjørland, Jeppe Nicolaisen: Scientific and Scholarly Classifications are not "Naïve"

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)

The choice of scientific methods is related to epistemological views. In biology, for example, there are three major schools connected to classification: "There are two popular theories of taxonomy based on these evolutionary principles: *evolutionary taxonomy* and *cladism* (or phylogenetic systematics) and one based on statistical similarities between groups (*phenetics*):

> ...
>
> Phenetics is a classification based on the statistical similarities between organisms. All characters are given an equal weight and by measuring large number of characters, it was hoped that a stable classification based on overall similarities between organisms would be reached. This kind of taxonomy has received a great interest with the development of computers were later largely abandoned because phenetic classifications were arbitrary and unstable. However, as molecular techniques became popular and more refined, phenetics enjoyed a resurgence. The sequence of amino-acids in any protein, or the sequence of nucleic acids in the DNA provides a large numbers of equally weighted characters suitable for phenetic analysis. A similarity between organisms could be calculated on the bases of the changes or non changes in its proteins or DNA structure. (Anonymous, 2003)

Phenetics is a school that is closely related to classical empiricism compared to the other schools. There is another philosophical relation:

> Systematists have rediscovered a problem long familiar to philosophers. How can one know that a particular chunk of metal is gold unless one knows what gold is, and how can one know what gold is without inspecting some samples of gold? But if one does not know what gold is, one cannot decide what to inspect…. (Hull, 998)

Our answer (which is based on "pragmatic realism") to Hull's problem is that different methods may be used until we arrive at a theory that satisfies our demands and meets reasonable consensus among re-searchers. We define our concepts tentatively and revise our theories and conceptual systems when needed. As criteria we use the coherence of our theories, observations, and, in the end, pragmatic criteria.

A work such as Bryant's (2001) represents a serious effort to attack the general problems of scientific classification. Such a book should be considered in LIS. It should, for example, be reviewed in this journal.

The scientific investigation of "naïve" theories

While scientific and scholarly classifications are anything but naïve, laypeople may classify phenomena in "naïve" ways, based on naïve theories. This field of naïve cognition has in recent years been investigated by many researchers in cognitive sciences and artificial intelligence. Researchers speak of such things as "naïve physics," "naïve biology" and "theories of mind" as research topics in psychology.

> Current work in the field of cognitive development has seen an increasing number of researchers embracing the notion of domain-specificity as the key to understanding the organisation and acquisition of children's knowledge. It is accepted that young children have one domain concerned with naïve physics which encompasses knowledge of mechanics and physical matter. Another is concerned with intuitive psychology which pertains to understanding of the behaviour and actions of others (Theory of Mind). Contention among researchers arises in relation to the proposal of a third domain - that of naïve or intuitive biological knowledge. My research is especially concerned with exploring whether children have an intuitive domain of biology, how this develops from early childhood, and how we can promote development through intervention. (Williams, 2002)

Such research often represents phenomenological investigation of, for example, the naive-physical realm[9]. While it is of interest for "cognitive theories," we find it less relevant for knowledge organization in LIS. It is only mentioned here to illustrate a rather different conception of naïve classification compared to the way Beghtol uses that word.

Knowl. Org. 31(2004)No.1
Birger Hjørland, Jeppe Nicolaisen: Scientific and Scholarly Classifications are not "Naïve"

59

Conclusion

The Scientific Status of Classifications developed within LIS Library and Information Science claims to be a scientific (or scholarly) field, implying that classifications developed within this field should be based on scientific or scholarly methods and principles. Hjørland has many times formulated views on the status of classifications in our field. Clearly Beghtol, Hjørland, and all other professors in knowledge organization have a responsibility to contribute to the field, making classifications as sound as possible. Hjørland has, in different publications, tried to uncover the basic assumptions in different approaches. Such different approaches includes the facet-analytical approach developed by S. R. Ranganathan, the Classification Research Group, and others, bibliometric approaches, approaches based on information about developments in the social decision of labor in society, as well as other approaches. Like other fields, knowledge organization in LIS is influenced by different views or "paradigms" that should be compared. The relative strengths and weaknesses should be illuminated. Very little research is being done along such lines. Mostly, researchers in LIS have a favorite approach, which is applied without considering the possible contributions of other approaches.

The importance of subject knowledge is often sadly ignored and neglected in LIS. Often it is implicitly assumed that such knowledge is not necessary, that it can be obtained from users or from automated methods. Such professional ideologies may block for real progress. If our classifications should be "professional," we cannot ignore the contributions made by researchers in specific domains. Beghtol's paper is valuable in bringing scholarly and scientific classifications more in focus in LIS research and in exploring the relation between these different kinds of classifications. The way these classifications are named is our major complain. It is our hope that this paper may inspire more researchers in LIS to consider principles, methods and results of scientific and scholarly classification and their importance for LIS classification.

Notes

1 Beghtol (2003, p. 65) writes: "They are classifications 'in the wild' (Jacob, 2001, 78) in the sense that they have been created in a particular domain to enhance domain knowledge for the pursuit of scholarly activity and research."

2 See, for example Adams (2001).

3 Biology taxonomy is one of the most established fields of scientific classification. Consider works such as Anonymous [2003]; Dean (1979) & Schuh (1999).

4 Concerning the classification of psychology consider, for example: Braun & Baribeau (1984); Hjørland (1998); Krauth(1981, 1982) and Kuiken. Wild & Schopflocher (1992).

5 This quotation is cited from Grauballe; Kaae; Lykke Nielsen & Mai (1998, 18). We have been unable to identify or verify it. Concerning the relation between classification and philosophical realism see further in Hjørland (2004).

6 [Original taxonomy was Bloom & Krathwohl, 1956].

7 [See American Psychiatric Association Task Force on DSM-IV. (1994) and Kupfer; First & Regier (2002)].

8 [See World Health Organization (1978, 1979)].

9 Examples of research in naïve physics include: Boden,1990; Hager, 1985; Hayes, 1979, 1985a, 1985b; Hobbs & Moore, 1985; McCloskey, 1983, Smith & Casati, 1994 and Williams, 2002.

References

Adams, C. J. (1994). The study and classification of religions. *Encyclopedia Britannica*. Chicago, Illinois: Encyclopedia Britannica, Inc.

Adams, W. Y. (2001). Classification and typology. (Archaeological systematics). IN: *International Encyclopedia of the Social and Behavioral Sciences.* (Vol. 3, pp. 1962-1966).

American Psychiatric Association Task Force on DSM-IV. (1994). *Diagnostic and statistical manual of mental disorders 4th ed.; DSM-IV*. Washington, DC: American Psychiatric Association.

Anonymous [2003]. *Classification, taxonomy and phylogeny of animals.* http://web.archive.org/web/20020609074157/http://www.squ.edu.om/agr/OnlineCourses/Biol2020/Taxonomy/Taxonomy.html [accessed 10 July 2004 – *editor*]

Beghtol, C. (2003). Classification for Information Retrieval and Classification for Knowledge Discovery: Relationships between "Professional" and "Naïve" Classifications. *Knowledge Organization, 30*(2), 64-73.

Bloom, B. S. & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, by a committee of college and

university examiners.Handbook I: Cognitive domain. New York, Longmans, Green.

Boden, Margaret A. (ed.) 1990 *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press.

Braun, C. M. J. & Baribeau, J. M. C. (1984): The Classification of Psychology Among the Sciences from Francis Bacon to Boniface Kedrov. *Journal of Mind and Behavior, 5*(3), 245-259.

Bryant, R. (2001). *Discovery and Decision: Exploring the Metaphysics and Epistemology of Scientific Classification*. Madison, NJ : Fairleigh Dickinson University Press.

Dean, J. (1979) Controversy over classification: A case study from the history of botany. In: Barnes B, Shapin S (eds.) *Natural Order*: *Historical Studies of Scientific Culture*. Sage, Beverly Hills, CA, pp. 211-230.

Feger, H. (2001). Classification: Conceptions in the social sciences. IN: *International Encyclopedia of the Social and Behavioral Sciences.* (Vol. 3, pp. 1966-1973)

Grauballe, H.; Kaae, S.; Lykke Nielsen, M. & Mai, J.-E. (1998). *Klassifikationsteori. 2nd ed*. [Theory of Classification]. Copenhagen: Royal School of Library and Information Science. (Skriftserie I klassifikation; 1).

Hager, Greg 1985 "Naive Physics of Materials: A Recon Mission", in *Commonsense Summer: Final Report*, Report No. CSLI-85-35, Center for the Study of Language and Information, Stanford University, Stanford, California.

Hayes, Patrick, J. 1979 "The Naive Physics Manifest", in D. Michie, ed., *Expert Systems in the Micro-Electronic Age*, Edinburgh: Edinburgh University Press, 242-70, as repr. in Boden, ed., 171-205.

Hayes, Patrick J. 1985a "The Second Naive Physics Manifesto", in Hobbs and Moore, eds., 1-36.

Hayes, Patrick J. 1985b "Naive Physics I: Ontology for Liquids", in Hobbs and Moore, eds., 71-107.

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

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

Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization, 30*(2), 87-111.

Hjørland, B. (2004). Arguments for Philosophical Realism in Library and Information Science. *Library Trends*, issue on "Philosophy of Information", edited by Ken Herold, 52(3).

Hobbs, J. R. and Moore, R. C. eds. 1985 *Formal Theories of the Common-sense World*, Norwood: Ablex.

Hull, D.L. (1998). Taxonomy. IN*: Routledge Encyclopedia of Philosophy*, Version 1.0, London: Routledge.

Jacob, E. K. (2001). The everyday world of work: Two approaches to the investigation of classification in context. *Journal of Documentation, 57*(1), 76-99.

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

Krathwohl, D.; Bloom, B. S. & Masia, B. B. (1964). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Handbook II: Affective domain. New York: David McKay Co., Inc.

Krauth, J. (1981). Techniques of classification in psychology I: factor analysis, facet analyses, multidimensional scaling, latent structure analysis. *International Classification, 8*(3), 126-132.

Krauth, J. (1982). Techniques of classification in psychology 2: cluster analysis, typal analysis, configural frequency analysis, discriminant analysis, regression analysis. *International Classification, 9*(1), 1-10.

Kuiken, Don; T. Cameron Wild & Don Schopflocher (1992): Positivist Conceptions of Induction and the Rejection of Classificatory Methods in Psychological Research. (Pp. 47-56 In: Tolman, C. W. (ed.): *Positivism in Psychology. Historical and Contemporary Problems*. New York & Berlin: Springer-Verlag).

Kupfer, D. J.; First, M. B. & Regier, D. A. (Eds.). (2002). *A research agenda for DSM-V.* Washington, D.C.: American Psychiatric Association.

McCloskey, M. 1983 "Intuitive Physics", *Scientific American*, 248(4), 122-130.

Richardson, E. C. (1930/1964). *Classification: Theoretical and practical*. New York: The H. W. Wilson Co., 1930. (Reprinted unaltered 1964)

Sayers, W. C. (1915). *Canons of classification applied to "the subject" "the expansive", "the decimal" and "the Library of Congress" classifications : a study in bibliographical classification method.* London: Grafton & Co.

Schuh, R. T. (1999). *Biological Systematics: Principles and Applications*. Cornell University Press.

Smith, B. & Casati, R. (1994).Naive Physics: An Essay in Ontology. *Philosophical Psychology*, 7/2, 225-244. http://ontology.buffalo.edu/smith//articles/naivephysics.html

Williams, J. (2002). *Development of naïve concepts' and theories.* http://www.education.ed.ac.uk/research/projects/theme1/proj6.html (Visited April 30, 2004).

World Health Organization (1978). *Mental disorders; glossary and guide to their classification in accordance with the ninth revision of the international classification of diseases.* Geneva: World Health Organization (WHO).

World Health Organization (1979). *International Classification of Diseases. 9th edition.* Geneva: World Health Organization (WHO).