Introduction

SUSAN LEIGH STAR AND GEOFFREY C. BOWKER

CLASSIFICATION IN THE WILD*

"CLASSIFICATION IS THE SLEEPING BEAUTY OF library and information science," said Hanne Albrechtsen (personal communication, November 1997). In some ways, at the most technical core of traditional library science, it also stands as a bridge builder between the past and the future of our field. Anthropologists have long seen classification as a tool for understanding culture. The distinctions that people make, as Lévi-Strauss (1969) argued in his famous report on *The Raw and the Cooked*, constitute cultural membership if not culture itself.

Classification involves the informal embedded in the formal and vice versa. If anthropological/cultural distinctions shape culture, the attempt formally to evaluate and improve those distinctions forms much of what we think of as information systems. We then take cultural cues from the systems so created.

In traditional library classification research, there have been two distinct challenges. The first is an ethnographic challenge: what distinctions does this specific group of clients use in forming their knowledge culture? How may we mirror in the thesauri, catalogs, and other search and retrieval tools we create for them? This challenge is one of *verstehen*—i.e., understanding the sense of vernacular terms. Here the information science researcher becomes an anthropologist—i.e, how to disambiguate

* These words are a play on Edwin Hutchin's excellent Cognition in the Wild (Cambridge, MA, 1995), which explores aspects of planning, coordination, and cognition from a social/organizational and material viewpoint.

Susan Leigh Star, Graduate School of Library and Information Science, 501 E. Daniel Street, University of Illinois, Champaign, IL 61820
Geoffrey C. Bowker, Graduate School of Library and Information Science, 501 E. Daniel Street, University of Illinois, Champaign, IL 61820

terms, decide what distinctions are really necessary, and which are reflections of ephemeral debates or fads?

The second challenge is a formal one in two parts concerning the structure of the information system or tool:

1. Occam's razor: how many terms can we afford given system capacity, user capability, and the means of distribution?
2. Structural soundness: are logical flows followed, are the branches of the tree carrying the right weight in proportion to the trunk, are the means of navigation easily grasped for accurate modeling?

There are no a priori solutions here: each scheme must be taken in its own context of use.

Classifications that work in the real world must meet both challenges simultaneously. For example, in studying the history of the International Classification of Diseases (Star & Bowker, 1994; Bowker & Star, in press), we noted that the designers of this global classification system must constantly make practical tradeoffs between the two challenges. In order to do justice to the range of subtle vernacular terms used by medical personnel around the world, a huge unwieldy list would have to be developed. In order for physicians and other users to actually employ the system, a much shorter key to filling out forms is the only possible alternative.

As the Internet, Web, and various digital libraries burst their boundaries and appear on desktops and in homes, the tension between these two challenges deepens. What do we understand about the interplay between vernacular classifications and the more formal structures underlying search engines, online catalogs, and other electronic guides? For groups of users that may be both global and unknown, what is the meaning of joining the two aspects of classification? What is usability in the context of *both* the Web and the intimate desktop?

The combination of the cultural and the formal in turn produces a third challenge—a moral and ethical one. For large-scale systems, whose voices will be heard and whose silenced? Whose culture will become the taken-for-granted and whose the exotic other? Where makers and users of classification systems do not address these questions, silent inequities prevail. The dominant voices may become the common sense of the designers or the loudest of the user voices (Forsythe, 1992).

The articles in this collection each address this set of issues from a variety of angles. Huber and Gillaspy's article tackles a core methodological issue surrounding the translation from vernacular systems of classification and vocabulary to the more formal controlled vocabulary systems such as those in LCSH, MESH, and ICD/DSM. Looking at AIDS and HIV vocabularies within the communities of gay men and IV drug users, they show how the political and historical situations of those affected help shape vocabularies and, in turn, the usability of more formal systems. They re-

port on their research-testing and refining vocabulary within a large-community AIDS provider organization via a series of face-to-face meetings with a range of users, from health care providers to patients and activists. They then move the findings to a large-scale vocabulary test sponsored by the National Library of Medicine. Grappling with the rapidly changing vocabulary, sensitivity to issues of stigma and race, and mapping the vernacular to the large-scale system, their research forms an exemplar of participant design and community-based research. As well, it points in some important directions for moral, political, and ethical aspects of applied classification work.

Jennifer Tobias's piece is similarly a manifesto for the examination of marginal, stigmatized, and experimental cultures under the rubric of "cyberspace." She continues in the tradition of Sanford Berman, challenging the worlds of cataloging and classification to keep up with cultural and social changes and to resist elitist or ethnocentric tendencies in large-scale systems. She argues that librarians need to become conversant in a range of specialized languages in order to provide the best possible services—and that this need is only heightened by the fluid nature of documents on the Web.

Another methodological piece, that by Star, provides a more abstract example of how "classification in the wild" could be joined with some foundational work in both library and information science on the one hand and sociology on the other. Star compares the faceted classification method developed for use in libraries by Ranganathan with the grounded theory method of qualitative analysis developed by Glaser and Strauss in sociology. Both systems struggle with the core dilemma posed above—i.e., how to braid the formal and the informal together in the study of classification. Future work in this area would, it is hoped, link some of the naturalistic findings of qualitative research with the developing technologies stemming from faceted classification and advanced networked information technologies for navigation (see Bradley & Sutton, 1993; McCombs & Maylone, 1998 for excellent overviews of research in the area).

Olson finds a theoretical mandate in the work of postmodern and feminist philosophy for some of the space mapped by the three discussions above. Perhaps instead of speaking in the "master voice" of the state-sanctioned list, she suggests that we might see the classification of marginalized domains as an exercise in cartography. Picking up on many of the same issues raised by Tobias and Huber and Gilaspy, she notes the implicit Western, and often sexist, constructions in the Dewey Decimal Classification. As an exercise in both imagination and method, she calls for a spatial imagery for neglected and stigmatized domains. Drawing on recent work in critical geography, she shows us both critique and a positive path for classification work and research.

Bowker's article adds a close reading of information-system as cultural artifact. He suggests that one can find traces of social and political debates in the classification of disease entities in the International Classification of Diseases (ICD). He argues that this encoding of social and political dimensions is a natural feature of such classification schemes, which thereby serve to encourage the development of some forms of knowledge while discouraging others. His article draws particular attention to the organization of time and space in the ICD—suggesting in turn that the scheme favors particular kinds of narrative of diseases, and that it best represents disease in the developed world.

Extending this theoretical contribution, Albrechtsen and Jacob conceptualize information in another inherently spatial fashion, that of information ecologies.[1] They note that classification systems play a key role as boundary objects (Star & Griesemer, 1989) in the organizations in which they are used—i.e., they serve simultaneously as lingua franca and as specialized tools in particular domains. Drawing on public library examples from the Book House Project and Database 2001, they emphasize especially the continually reconstructed nature of classifications as organization and knowledge tools.

Finally, Mark Spasser examines issues of agency and structure in psychiatrists' use of the *Diagnostic and Statistical Manual (DSM)* of mental disorders. Drawing on Gidden's concept of structuration, he suggests that the *DSM*, through its propagation of a particular biomedical reading of mental disorders, severely constrains the kinds of research that can be done by psychiatrists. He argues that Gidden's work provides an analytic purchase on this constitution of psychiatric discourse through classification and maintains that the concept of structuration will also provide a useful tool for understanding the development and change of library classification systems.

CONCLUSION

Library and information science stands at a historical crossroads. New information tools appear daily and are used in more kinds of settings than before. They are part of not only desktops or kiosks but, increasingly, of living rooms, gyms, cars, banks, and hospitals *inter alia*. The formal tools of classification construction and evaluation, and decades of experience and research in working with client populations, give us a unique suite of tools for understanding this phenomenon. As sleeping beauty wakes up in this new world, there is a unique opportunity for her to build bridges across its rivers and canyons.

ACKNOWLEDGMENTS

NOTE

[1]   Bonnie Nardi has also recently used the term "information ecologies" to describe her work with librarians. Susan Leigh Star uses a related concept in her edited volume, *Ecologies of Knowledge: Work and Politics in Science and Technology* (SUNY, 1995). See also Davenport, T. H. (1997). *Information ecology: Mastering the information and knowledge environment.* New York: Oxford University Press.

REFERENCES

Bowker, G. C., & Star, S. L. (In press). *Sorting things out: Classification and practice.* Cambridge, MA: MIT Press.

Bowker, G. C., & Star, S. L. (1994). Knowledge and infrastructure in international information management: Problems of classification and coding (pp. 187-213). In L. Bud (Ed.), *Information acumen: The understanding and use of knowledge in modern business.* London, England: Routledge.

Bradley, J., & Sutton, B. (1993). Symposium on qualitative research (Special Issue). *Library Quarterly, 63*(October).

Forsythe, D. (1992). Blaming the user in medical informatics: The cultural nature of scientific practice. In *Knowledge and society: The anthropology of science and technology* (vol. 9, pp. 95-111). Greenwich, CT: JAI Press.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Lévi-Strauss, C. (1969). *The raw and the cooked* (Translated from the French by J. Weightman & D. Weightman). New York: Harper & Row.

McCombs, G. & Maylone, T. (1998). Qualitative research (issue theme). *Library Trends, 46*(4).

Star, S. L. (Ed.). (1995). *Ecologies of knowledge: Work and politics in science and technology.* Albany, NY: SUNY.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111-134.