I am neither a librarian nor an information specialist. I am, however, associated with an organization that has for some time had one foot in the social sciences and one in the development of an information system. And I can speak with some eloquence as an addicted user of the Yale Library—at one and the same time enamored of its seemingly inexhaustible collections, and frustrated by the anachronistic and time-consuming system of getting access to them.

I want to comment on what appear to me to be certain trends in the social sciences and to derive from these one or two generalizations of possible significance for the problems faced by librarians as arrangers, and researchers as users, of information. I would also like to comment briefly on one recent development at the Human Relations Area Files which I think has a particular bearing on these information problems. Much of what I am going to say has probably been said before, possibly in other ways and in other contexts. I therefore apologize to those of you to whom this may be "old hat."

The social sciences in this country, like so much else today, appear to be caught up in a process of escalation. This is certainly true of anthropology. The American Anthropological Association currently lists 102 institutions in the United States offering graduate degrees in anthropology. In 1950 this number stood at just over 40. In the year 1967-68 alone, 12 new institutions were added to the list. These departments are all turning out graduates, many of whom in turn are starting new departments and new research centers. This of course means more publication. And the number of publishers and publishing outlets has increased greatly. What is true for the United States is also true to a greater or lesser degree for the rest of the world. The International Bibliography of Social and Cultural Anthropology for 1955 listed some 550 periodicals
consulted. Ten years later, in 1965, this figure stood at well over 1,500. The problems posed for all of us—librarians and scholars alike—by a growing volume of information are something of which we are, however, well aware.

I want to focus here not on quantitative changes, but qualitative ones—changes in the nature of social science data, and in our attitudes toward these data, and the implications of this for traditional library functions, including the bibliographic function.

I can best speak to this point from the experience of the Human Relations Area Files (HRAF), which can be thought of as a kind of library or retrieval system for anthropological and related social science data. The HRAF, with 150 regular and associate memberships, is an interuniversity research organization founded in 1950. As of next year HRAF will have been in existence 20 years—something of a record in the annals of interuniversity cooperation. Its purpose is to facilitate research in the sciences devoted to the study of man—specifically by collecting, organizing, and distributing data for the comparative analysis of human behavior, social life, and custom. This is accomplished by the application of rather extensive topical and areal coding systems to the basic source materials on a selected sample of over 200 of the world’s cultures. This work is done at the research center in New Haven, from where copies of the Files are shipped to each institutional member either in a paper or microfilm format.

In the course of accomplishing these tasks, the HRAF organization has been in a rather good position, I think, to observe and assess trends in the production of social science data and in the uses to which these data are put. We are also, like librarians, deeply involved in problems of data retrieval (e.g., subject indexing) and problems of bibliographic control.

Three aspects of the qualitative nature of social science data—three areas of changing emphasis and attitudes if you will—seem to me to represent a particular type of problem with respect to bibliographic control, and thus for the traditional bibliographic and retrieval functions of libraries. These are: area provenience, ethnic specificity, and data quality control.

Area interests in American scholarship really got underway in the decade before World War II. The rapid postwar development of university-based area programs and interdisciplinary area research is too well known to need much comment here. There are at present, for example, some 20 universities in the United States with programs of sufficient scope to qualify as African studies
centers. And in 1967 there were in this country 15 university libraries with holdings of Far Eastern materials exceeding 50,000 volumes each. In the past decade in particular there has been a parallel development in the study of so-called underdeveloped areas. Partly in response to these area interests, which in many cases received their initial stimulus and funding from government sources, social scientists have in recent years shown an increasing scholarly (theoretical) interest in nonwestern cultures—supplementing to a considerable extent the long-standing anthropological concern with exotic, primitive, tribal societies beyond the confines of Western civilization. This has taken the form of a comparative, behavioral emphasis in such traditionally western-oriented disciplines as sociology, economics, and political science. For example, there are now at Yale an Economic Growth Center and a Political Data Center, both of which are amassing data on a world-wide basis for the formulation and testing of hypotheses. And the Yale Sociology Department has recently proposed a program in the comparative study of social systems—to include fieldwork by graduate students and faculty in a variety of nonwestern cultural settings.

These trends in research mean that more scholars will be wanting access to data by both subject and area, e.g., Far Eastern history, political systems of Sub-Saharan Africa, and the like. Moreover, with more scholars who regularly read the vernacular languages of their area specialities, librarians will increasingly be faced with problems of acquisitions and cataloging in these vernaculars. And the volume of literature produced in the so-called underdeveloped nations is increasing yearly. Thus, the Cornell Library Wason Collection’s latest monthly Southeast Asia Accession list includes over 50 titles in Indonesian and some 30 titles in Thai.

Allied to this area emphasis is the increasing awareness of the importance of the ethnic specificity of data. The growing sophistication in this regard is largely, I think, the result of increased awareness of anthropology and of anthropological theory—of the anthropologist’s insistence on viewing a culture as a whole; his attention to the functional interrelations among the parts of a culture; his dictum that any aspect of behavior or custom must be understood with reference to the social and cultural system in which it is embedded. Within the nation-state of Ethiopia, for example, there are at least eight major ethnic groups, of whom the Amhara are the most significant. It makes little sense, from
the sociocultural standpoint, to collect nutritional data or to sur-
vey political attitudes on an undifferentiated population labeled
"Ethiopians"—particularly if the researcher intends to correlate
these data with other aspects of culture, e.g., the relationship of
food and dietary habits to religious attitudes. Or if he intends
to compare his data with those from other nation-states, e.g., a
comparison of "Ethiopian" diet with "Senegalese" or "Indian"
diets.

This increasing awareness of ethnic specificity means that the
bibliographic screens through which we process information will
in the future be required to pay attention not only to subject and
area, but also, for many types of data, to ethnic provenience as well.

The third trend I wish to mention is more in the nature of a
change in attitude toward data. This is going on in anthropology,
particularly among those researchers doing comparative or cross-
cultural research—which is, above all, library research. I refer
here to the growing importance of what has been called "data
quality control" and an increasing sophistication in devising means
of assessing the validity and reliability of the source material.
What concerns us here is the aspect of this assessment that focuses
on the author—indirect measures of data quality based on such
factors as the author's professional training (e.g., degrees held),
nationality (is he a national, writing about his own country?),
major research interest (is he writing on a subject on which he
has published prior research?), field work experience (is research
based on firsthand observation or on library research?), length
of time in the field (is the author attempting an analysis of an
entire social system based on only four months in the field?),
knowledge of vernacular language (e.g., did the author work
through an interpreter?).

These attempts at quality control, while not wholly new to the
social sciences, will be receiving more emphasis as the machine
handling of data becomes increasingly practical for social science
research. In large-scale comparisons in anthropology, for example,
it is now possible to feed into a machine large amounts of coded
data from a great variety of sources. In such cases it is of the
utmost importance to attempt to control for known variables in
the reliability of data. Here again, one can predict a future de-
mand that such quality control information be built into the bib-
liographic process—to a much greater degree than its traditional
incorporation within the library card catalog.

The changes I have been mentioning are, I think, representative
of an emergent trend in the social sciences with respect to the kinds of services these researchers and scholars are going to be wanting from libraries. More and more, requests are going to be for pieces of information; it is going to be less a matter of the retrieval of documents by author and more the location of bits of information from documents and about documents. The specificity of information requested is going to be greater than heretofore. The implication of this, it seems to me, is that if libraries are going to continue to serve their function as the "middlemen" in research, they must move beyond being essentially document retrieval systems toward being able to function also as information retrieval systems.

As it is now, the user must pretty well adapt his research problem to the nature of the library system. And in this system the arrangement of information is primarily from the point of view of the producer, not the user. The card catalog is the traditional entrée to this system. The subject headings one finds there are geared to the document as a whole and reflect, by and large, the use to which the author felt his book would be put—as implied by title, chapter headings, and the like.

The reference librarian constitutes another entrée for the researcher attempting to extract from the system the information he wants. While this service can be helpful, it can only be so to the extent that the request can be phrased in terms compatible with the existing system.

What the researcher is asking, I think, is that the existing system somehow be loosened up—that provision somehow be made to increase the scope of entrée into the system; and that this be oriented to the user—who will be wanting access to information in and about documents.

Librarians and information specialists are of course aware of these problems, and there has been a good deal of research and development in the areas of classification, documentation, and information retrieval with respect to libraries and the problems posed by the information explosion. Much of this is geared to the use of computers and the possibility of computerizing various aspects of library functions.

One aspect of these methodological and technological developments lies in what is frequently called "automated bibliographic control." The possibilities inherent in this area are of particular relevance to the kinds of problems I have been discussing here. With automated bibliographic control it is possible to greatly ex-
pand the “screen” (the imposed “grid” or inventory of headings) through which bibliographic information is fed into a system and through which it is got out on demand. And to further convert this information into machine-readable form and magnetic tape storage such that catalog cards, accession lists, retrospective literature searches, special bibliographies, etc., can be printed out on a periodic or demand basis.

That this whole problem of “control” is a practical one in terms of money and university budgets, can perhaps be dramatized by considering the cost continuum involved in the chain of events between the time a research scholar determines that he needs information on a particular subject and the time when the documents containing (hopefully) that information are in his hands. In the typical case, the division of labor, and thus the prorated portion of cost in terms of man hours, in this chain of events is overwhelmingly on the side of the researcher rather than the library. If the researcher be regarded as part of the system, and his costs as part of the system’s costs, then it appears highly inefficient, economically speaking, to attempt to reduce library costs by, for example, cutting down the number of subject headings on catalog cards—thereby increasing enormously the time a researcher must spend looking for what he wants.

I can perhaps best illustrate the potentials of automated bibliographic control by reference to the development at HRAF of what we call HABS (the HRAF Automated Bibliographic System). This is the work primarily of Dr. Hesung Chun Koh, director at HRAF for the past three years of an NSF-funded Korean Social Science Bibliography Project. Dr. Koh, a sociologist, initially got into this bibliographic area because she wanted to write a handbook which would incorporate comparative and interdisciplinary insights into an understanding of Korean society. Her attempt to compile a working bibliography led to frustration—frustration allied in part to the kinds of changing emphases and expectations vis-à-vis the use of social science data already mentioned, and her attempts to extract information at this level from existing card catalogs and other bibliographic resources. Essentially, Dr. Koh wanted a bibliographic entrée into pieces of information about Korean society and culture. The traditional library, document-oriented bibliographic services simply did not provide entrée at this level. The methodological aspects of the Korean project have followed closely those of the Columbia-Harvard-Yale Medical Libraries Computerization project, initially directed by Frederick
Kilgour. This was a pilot attempt to establish a network of computerized bibliographic retrieval systems, and promote cooperative, shared cataloging among these three libraries. The Kon project, however, has oriented the Kilgour methodology to social science research needs and the handling of sociocultural data—most specifically materials in Chinese, Korean, and Japanese languages.

The project started by systematically compiling a comprehensive social science bibliography on Korea, consisting of Xerox copies of existing bibliographic entries in books, copies of library card catalogs (e.g., the LC Union Catalog), and annotative or descriptive references copied from existing bibliographies. The project next worked out methods for converting these various types of bibliographic information, including references to journal articles, unpublished manuscripts and chapters in books, into a standardized machine-readable language. Work sheets containing the converted information were keypunched on cards and an IBM 870 Document-Writer used to produce copy for proofreading. Punched cards were then converted to magnetic tape storage which in turn is used for high-speed computer retrieval and sorting or for further printout. Printout can include book-form bibliography (a preliminary volume of 3,000 selected entries was produced early this year), card-form bibliography, and various types and levels of bibliographic information retrieval on demand.

The analytic information portion of the work sheet provides for controlled codes, including subject codes adapted from HRAF’s Outline of Cultural Materials, time period codes, regional subculture codes, ethnic group codes, etc. It allows also for incorporation, if desired, of systems of uncontrolled classifiers such as KWIC (Key Word in Context). Allowance is made for retrieval by generic concepts such as used in cross-cultural comparative research, as well as specific concepts peculiar to a given sociocultural context. Finally, the system includes quality control codes aimed at answering general-type questions relevant to the assessment of an entire bibliographic corpus—such as who has written on what subject, why, how, and under what conditions? And with what results? The system also codes information on authors and their backgrounds relevant to the assessment of the reliability of a single document or piece of information.

HABS has been consciously designed as a comprehensive system, complete with all descriptive, analytical and annotative bibliographic information in a single frame of reference—with each ele-
ment treated in equal depth. A computer can be programmed to suppress any class of descriptors if not needed for a particular purpose; or the system can be applied to a given bibliographic universe, starting out at any level of comprehensiveness desired, and additional areas of information added later in a retrospective fashion. But at least the attempt has been made to allow for all possible elements in the initial designing of a single comprehensive system.

The HRAF Korea bibliographic project is one example among many that could be mentioned—attempts by researchers to somehow achieve a semblance of bibliographic control over bodies of literature that are expanding at an alarming rate. Most bibliographic projects currently underway assume some kind of machine handling of data; some, such as HABS, have been designed jointly with librarians or with library needs and problems in mind; others are being developed independently of library collaboration. One thing is certain—the next decade will see the large-scale development of automated bibliographic control in the social sciences and in area studies. Whether this will be on a piecemeal or coordinated basis remains to be seen. The need for cooperation, communication, and collaboration among researchers, librarians, and information specialists in this formative stage is evident. And there are already encouraging signs in the direction of coordinated effort. The Association of Asian Studies, at its March 1968 meeting, established a Committee on Automated Bibliographic Control for East Asian Studies. This committee of librarians, scholars, and computer experts will survey existing systems, including HABS, and make recommendations for the adoption of a cooperative, comprehensive bibliographic control system for East Asian studies. A similar movement is underway in the field of African studies; an International Conference on African Bibliography, held at Nairobi in 1967, recommended the establishment of an international, cooperative, African bibliography program. Committees appointed at this conference are presently surveying existing automated bibliographic control systems. A National Academy of Sciences Committee on Information Needs in the Social Sciences, in its final report dated 1968, recommended the establishment of a national bibliographic center or centers for the social sciences.

The key word for the future, it seems to me, is cooperation. The information explosion problem is a joint librarian-scholar problem. If nothing else, the sheer volume of information is requiring that both resort increasingly to selection. The scholar, faced with lengthy bibliographic searches through an ever-expanding literature, needs
more efficient selection to free his time for more productive research; the librarian, faced with budgetary limitations, is more than ever in need of guidelines with respect to acquisitions. Both are increasingly in need of quality-controlled selection based on comprehensive bibliographic knowledge of a given information universe.

Librarians can no longer hope to, nor should they be expected to, acquire across the board. Some kind of interlibrary sharing of acquisitions and cataloging has got to come about. And it is the obligation of the scholar to help here. Perhaps one should put this more strongly and say to “push” in this direction. Not so much by pushing librarians as by “agitating” in university library committees, professional associations, and foundation circles toward this end.

Automated bibliographic retrieval systems are but one aspect of the larger prospect of computerizing the actual information contained in libraries. At present, the prospect of achieving the former looks a great deal more practical than the latter. As things now stand, however, we have the very real possibility of at least one-half of such a total information system. The problems of converting bibliographic information to machine-readable form—and of thus being able, for example, to produce library cards by machine—have been largely solved. And the hardware needed, although it must be of a high power for the job contemplated, is available. Certainly half a loaf is better than none—particularly when we consider our present inability to control the literature and to gain access to materials we know exist.

Bibliographic control of the kind here contemplated will make it possible to supplement the card catalog, the traditional bibliographic entrée to information stored within the library. Up to a point it can increase the usability of the card catalogs themselves by greatly expanding the use of subject headings. But there is a limit here, it seems to me, and librarians have been understandably reluctant to use subject headings for the simple reason that if you use enough of them you will have little room left for books. But the trend has been altogether in the opposite direction—to the extent that now, I believe, the Library of Congress is only averaging something like 1.5 headings per card. This situation certainly has not helped the researcher to gain access to information through library card catalogs. Some combination, then, of limited expansion of card-form bibliographies together with automated printout of accession lists, book-form bibliographies and specialized subject
bibliographies on a periodic and demand basis seems the best solution at this point. Until we reach the utopia of the library as a completely automated information system, with both bibliographic and substantive data stored on tape or in otherwise retrievable form, our best hope would seem to be in automated bibliographic retrieval systems. Perhaps the time is not too far off when we will have interlibrary networks of such systems, sharing acquisitions and cataloging, and serviced by bibliographic processing centers. A long step in this direction has already been made with the setting up in 1967 of the Ohio College Library Center, designed by Frederick Kilgour. This is a bibliographic information exchange system, with each of the fifty-one participating libraries having access to a central computerized catalog. Future developments in this direction will necessarily be the product of cooperative effort on the part of librarians, information specialists, and scholars, working together to solve the very real problems of information needs in the social and behavioral sciences.