Report of the CSSDA Consulting Committee*

This letter constitutes the report of a three-man committee appointed by the Council of Social Science Data Archives[1] to consult with the officers and staff of the Human Relations Area Files concerning the automation and broader utilization of the Files. Our report is based on a visit to HRAF on March 20 and on our familiarization with the files in preparation for the meeting. We wish to thank the people at HRAF for their cooperation and hospitality extended to us during our visit, which we found to be most helpful and informative.

In preparing this report, we have addressed ourselves to three separate questions suggested by Dr. Ford in our meeting. He asked us what we would recommend concerning (1) the maintenance, improvement, and extension of the basic files, (2) the establishment of an archive for quantitative data in anthropological research, and (3) the involvement of HRAF in document retrieval services. Our report will be organized into sections discussing each of these topics.

I. The Basic Files

At present, the prospective user of the files has no adequate way to learn the nature and extent of information contained on topics within his interest. The computerized index to coded information in the files, being prepared at Northwestern University under the direction of Raoul Naroll and now nearing completion,[2] is an important step toward making information in the files more readily retrieved and used. We recommend that the indexing be extended to include all the material in the

files, including the "modern European" countries originally ex-
cluded from the indexing project. Material on these countries is
likely to be of great interest to political scientists and other social
scientists who now make little use of HRAF data, and their use of
the files would be facilitated if this information was included
in the index. We also recommend that HRAF consider publish-
ing or otherwise distributing the index to its members and asso-
ciate members. One way of accomplishing the distribution would
be to copy the computer tape upon request so that the index
might be printed on a computer at the requesting institution.
But before distributing the index in any way, the HRAF staff
should review the format presently planned for the index for
the purpose of requesting changes to improve its readability.
Finally, some provision should be made for acquiring user re-
sponse to the usefulness of the index in preparation for subse-
quent decisions on revision or updating.

In addition to facilitating the retrieval of information from
the files, the computerized index and the associated data tapes
provide HRAF with an excellent opportunity for *analyzing* the
nature and extent of information contained in the files. The data
on these tapes will be amenable to frequency counts by sources,
cultures, and regions, for example, to determine the coverage
accorded to various categories in the *Outline of Cultural Mate-
rials*.[3] As companion information for the index, tallies should at
least be made for the frequency of occurrence of each OCM
category within each culture. In general, having these data in
machine readable form will open the way to a wide variety of
substantive and methodological studies not heretofore possible.
Moreover, these analyses will serve to introduce the HRAF
staff to data processing techniques, developing competence for
subsequent computer applications to HRAF operations.

Concerning the automation of the textual information in the
files, the committee was firmly of the opinion that, given the
state of the art, it was premature to plan for putting the actual
texts in machine readable form. The thought of keypunching
this vast material is frightening, and the promise of effective
retrieval of the information once in the computer is not that
great. However, the committee does recommend that serious

thought be given to recording the files on 16 mm. microfilm for automated retrieval with Eastman Kodak's MIRACODE equipment and system,[4] which would permit instantaneous retrieval of pages in the files according to logical combinations of OCM codes. While considerable costs would be involved in recording the files in MIRACODE form and in acquiring the necessary equipment for reading the film and retrieving the pages, considerable savings as well could be realized in production of new files and in the costs of maintaining the files in the field. It is conceivable that conversion to the MIRACODE system will actually result in *savings* rather than costs, not to mention the tremendous advantages in using the files once in the MIRA-CODE system. This may also be a way to increase the membership of HRAF while eliminating the distinction between full and associate members, for the paper files would be expected to fall into disuse with the advent of a MIRACODE installation.

While the preparation and distribution of the index to HRAF holdings will improve the prospects for use of the files by other social scientists, more can be done to make the files more inviting to research by scholars outside of anthropology. In general, we advocate disseminating more information about HRAF methodology to potential users of the files. Specifically, we recommend that the *Analysts' Handbook*,[5] which we found to be an impressive document with few, if any, counterparts in the annals of social research, be made available for study by those who are interested in the methodology supporting the files. We believe that the *Handbook* will do much to dispel criticism and suspicion directed at the files by those who are ignorant of its methodological underpinnings. If in some cases distribution of the *Handbook* produces inquiries and requests for clarification, then this too represents a gain by opening lines of communications between HRAF and its clientele, perhaps resulting in improvements in HRAF procedures. As an additional attempt to establish the quality of HRAF material, more attention should be given to the issue of inter-analyst (or inter-coder) reliabilities. The 1957 study in the *American Anthropologist*[6] which reported on inter-analyst reliability is far out of date both in time and in methodology. Because of its basic importance to the utiliza-

tion of the files, the extent of coding reliability should be re-evaluated through a new study and the results of the study should be disseminated to the users.

II. *Archiving Quantitative Data in Anthropology*

The committee recognizes HRAF's role of leadership in the collection, preparation, and dissemination of qualitative ethnographic information. It recommends that HRAF enlarge the scope of its function to include the collection, preparation, and dissemination of *quantitative* information in anthropology. As employed in our recommendation, "collection" refers not so much to sponsoring field studies to acquire fresh data but to acting as an archive or repository for quantitative studies already available in machine-readable form. In addition to archiving raw data, HRAF should collect, prepare, and disseminate appropriate documentation of the studies in its repository. Such documentation should include, at least, information on the unit of analysis, data sources and research methodology, publications reporting the data, variables and coding categories used, and the column locations of data on the cards. For ease of storage, processing, and retrieval, it is recommended that this documentation be itself put in machine-readable form, which should be comparable to the format for machine-readable codebooks suggested by the Standards Committee of the Council of Social Science Data Archives.

Our recommendation for HRAF to extend its scope to include the archiving of quantitative information in a sense suggests that the Human Relations Area Files should perform within anthropology a function similar to that performed by the Inter-University Consortium for Political Research, an impressive organization of more than seventy universities in partnership with Michigan's Survey Research Center.[7] Because of characteristics of cross-national anthropological data and because of the smaller community of users, we believe that a much smaller organization can perform a similar function within anthropology. Unlike the sample survey data which form the backbone of the Consortium's archives, quantitative cross-cultural data in anthropology involve relatively few cases for analysis. Consequently, we believe that HRAF can embark upon the archiving

function with relatively little data processing equipment, storage space, and technical expertise. A good first step would be to announce in the professional journals HRAF's intentions to extend its scope in this manner while inviting researchers to deposit with HRAF duplicate sets of data cards and specified documentation. Although it would be desirable to have direct access to computing equipment in order to maintain the archive and service requests for data and documentation, all that would actually be required in the initial phases of operation would be some basic unit record equipment, consisting of a keypunch, counter-sorter, reproducer, and a tabulator.

. If HRAF follows our recommendation to function as an archive for quantitative data, then we suggest that they prepare for their new role by consulting with people engaged in large-scale data management activities. The discussions might begin right at Yale by talking with Charles Taylor, Bruce Russett, and Hayward Alker at the Yale Political Data Program; Richard Ruggles in the Economics Department (who has developed a system for retrieving information based on titles of tables, charts, graphs, etc.); and Zvi Namenwirth in Sociology (who has been working with the retrieval of documentary material coded into numbers). To obtain technical help in working with large scale social science data banks, visits should also be made to M.I.T. to talk with Ithiel de Sola Pool and his associates; the Roper Center and Phil Hastings; Oak Ridge National Laboratory and Davis Bobrow in the Social Science Division; the Michigan Research Center and Ralph Bisco; and the Berkeley Survey Research Center and Charles Glock. Of special relevance to the problems of collecting, processing, and disseminating anthropological data would be visits and discussions with Douglas White and Robert Textor, both of whom have been concerned with operation of data retrieval systems.

III. *Bibliographical Services*

HRAF has already had considerable experience in preparing traditional subject-classification and author-alphabetized bibliographies. With the development of Dr. Koh's Korean bibliography project,[8] HRAF will soon be gaining experience in a computerized document retrieval system. In order to round out

its qualifications as the major research center for cross-cultural anthropological material, we recommend that HRAF expand its services to include an automated bibliographic service. As a beginning, HRAF might consider extending Dr. Koh's adaptation of the MEDLARS methodology to the current titles in the files plus the titles published under the HRAF imprint. Appropriate OCM categories describing the content of the documents might be included for search and retrieval purposes in conjunction with the MEDLARS capabilities.

Again, considerable assistance in establishing an automated bibliographic system can be obtained from others across the country who are concerned with similar projects involving other social science material. The newly formed committee of the American Documentation Institute on Information Handling in the Behavioral Sciences should certainly be contacted for inspiration and coordination.[9] The current bibliographic activities of the American Council of Learned Societies would have special relevance for any system that HRAF might contemplate.[10] Through the ACL's Constituent Society Journal Editors' Project, machine-readable abstracts of articles appearing in relevant journals, including the *American Anthropologist,* can be made available for inclusion in HRAF's own bibliography on anthropology. Some of the material, therefore, might be obtained for HRAF's specialized bibliography at minimum cost. Additional material needed to round out the bibliography can be culled from relevant journals as they are published. HRAF's task in establishing an automated bibliographic service would be made far more manageable by aiming only to keep up with current literature beginning, say, in 1968 instead of trying to go back in time to include the endless backlog of material that has already been published.

Unlike the relatively modest equipment requirements for establishing an archive for quantitative data, the establishment of a bibliographic service requires ready access to sophisticated computing equipment. It should not be undertaken in a large-scale fashion without assurances of the availability of such equipment, preferably located at or near HRAF. In order to implement the full range of recommendations in this section, therefore, HRAF may have to seek funds for installing on the premises

the necessary data processing equipment or a suitable terminal connected to a computer at Yale University.

Summary and Conclusion

After reviewing the history, operations, and present position of the Human Relations Area Files, our committee felt that our recommendations should not be limited to improving HRAF's present operations, for we hold the view that HRAF should play a broader role in aiding anthropological research. Although the scope of HRAF's activities already extends beyond the preparation and dissemination of organized files of cultural data, we suggest that HRAF enlarge its scope even more to include and give emphasis to archiving quantitative data and retrieving information from bibliographic files. Given certain technological developments in processing, storing, and retrieving information and the current growth of research centers already applying these developments within the social sciences, the Human Relations Area Files stands in danger of atrophying as newer centers provide modern services to the researcher. By failing to retool itself to undertake activities required in modern research, HRAF not only runs the risk of being circumvented but also denies itself the maximization of its goal "to promote research on man and his ways of life by making widely available primary source materials and by instigating comparative and interdisciplinary research on human behavior, social life, and culture," as stated in the *Function and Scope* of the Human Relations Area Files. We feel that there is no reason why HRAF cannot function in this broader fashion and many reasons why it should. This belief in an expanded role for HRAF is reflected in our recommendations.

Committee Members:

Lutz Erbring
Survey Research Center
University of Michigan

Kenneth Janda, *Chairman*
CSSDA Consulting Committee
Northwestern University

Charles Yarbrough
Survey Research Center
University of California, Berkeley

NOTES

[1] The Council of Social Science Data Archives is supported by the National Science Foundation for the purpose of planning, coordinating, and publicizing the activities of a confederation of social science data archives in the United States. A brief description of CSSDA is contained in Ralph L. Bisco, "Social Science Data Archives: A Review of Developments," *American Political Science Review, 60* (March 1966), 93-109. A more thorough discussion is given in William A. Glaser and Ralph L. Bisco, "Plans of the Council of Social Science Data Archives," *Social Science Information, 5* (December 1966), 71-96. HRAF is a member of the Council.

[2] See Donald G. Morrison, "The Human Relations Area Files Index," *American Behavioral Scientist, 10* (February 1967), 27-30.

[3] George P. Murdock *et al.* New Haven: Human Relations Area Files, 1961.

[4] Kenneth Janda, "Political Research with MIRACODE: A 16 mm. Microfilm Information Retrieval System," *Social Science Information, 6* (April 1967), forthcoming. Reprinted in Kenneth Janda, *Information Retrieval: Applications in Political Science.* Indianapolis: Bobbs-Merrill, forthcoming.

[5] Human Relations Area Files. New Haven: Human Relations Area Files, 1965.

[6] Brian R. Kay, "The Reliability of HRAF Coding Procedures," *American Anthropologist, 59* (1957), 524-27.

[7] See Warren E. Miller and Philip E. Converse, "The Inter-University Consortium for Political Research," *International Social Science Journal, 16* (1964), 70-76.

[8] Hesung C. Koh, "A Social Science Bibliographic System: Computer Adaptations," *American Behavioral Scientist, 10* (January 1967), 2-5.

[9] See Frieda B. Libaw, "Information Handling in the Behavioral Sciences: Report of a First Convocation of a Conglomerate Clan," *American Behavioral Scientist, 10* (February 1967), 8-12.

[10] See Thomas J. Condon, "Computers, Traditional Research, and the American Council of Learned Societies," *American Behavioral Scientist, 10* (February 1967), 4-7.