NOTES ON THE DISTRIBUTION OF LABOR IN A SOCIAL SCIENCES DATA INFORMATION
NETWORK

Paul Evan Peters

Social Sciences Informatin Utilization Laboratory
University Center for International Studies
University of Pittsburgh

FRAMEWORK

INTRODUCTION

The concern of these brief remarks is to contribute to the deliberations having to do with networks and Machine-Readable Data Files (MRDFs) by calling attention to the fact that one way to proceed is to distribute labor judiciously among network participants. To distribute labor is to distribute a very large cost, and, therefore, investment, associated with network operation. Distribution of investment often shows an added benefit of distributing commitment with as many people as possible having a perceived stake in network success or failure. The distribution of labor also invites specialization by network participants so that those in the best position to accomplish certain tasks are encouraged to do so without distraction from other matters. This modularity fosters incremental growth and is thus well attuned to snifting financial arrangements and a changing population of participants. Most of all, though, careful distribution of labor is an organizing style uniquely appropriate to grass-roots movements struggling to legitimize their interests. It is style that can be found in the history and operation of IASSIST itself. For all of these reasons I would like to take a moment to pursue this thought to see what it has to offer.

The point-of-view which I apply to MRDFs and networks can be expressed fairly simply. Within the last five to ten years we have witnessed a very important adjustment in the aspirations of those involved in the "data library movement," to use a single descriptor to refer to those activities havng to do with secondary uses of social science data. We used to hear rather frequent mention of national and even international data libraries. Indeed, with varying degrees of success, some nations have embarked on such a course. More and more, though, we are focusing our attention on national/international efforts directed at organizing information about data rather than data themselves. I agree with this change of "target" and am optimistic that this dream can be realized.

The sort of thing I have in mind when I think of a "data information handling system" (DIHS) can be sketched as follows:

1. The information flow begins with the primary researcher(s). Through the efforts of this agent a set of social science data is created. If we are lucky, a codebook is also prepared. The flow ends with a secondary researcher, who saves time and money by learning of another

researcher's data and using them.

2. Within "organizational reach" of these researchers things are considerably simplified if some data handling agent (DHA) is available. This agent could be a data archive or library, a statistical consulting service, a computer center, an academic/research library, etc. Assuming the availabilty of such an agent, either knowledge of the researcher's data or actual responsibility for them can follow. If such an agent does not exist then knowledge of the existence of research data is more problematic; both the dissemenation and acquisition of data and information thereabout is more naphazard, resting upon an uncertain set of professional incentives.

3. DHAs offer a data information handling system an economy in the area of scope of coverage. Rather than conceiving of all possible researchers as needing to be contionually surveyed, the system need only keep up with all possible DHAs. (This is by no means a simple problem, merely simpler.) Another advantage is that DHAs can justifiably be regarded as both the primary producers and consumers of whatever bibliographic unit drives the system; this establishes a rare and quite potent identity

of system design, operation, and user groups.

4. A minimal data information system would be established if the DHAs "shared" information, the present state of affairs. By word-of-mouth and periodic catalog updates one keeps abreast of new findings and acqusitions. The problems with such a system are all too well known to this audience. My favorites are: a) a waste of DHA effort in that all have redundant entries in mailing lists (the tip of the iceberg); b) a waste of secondary researcher effort in that all have to go from DHA to DHA asking the same question repeatedly (again, the tip of the iceberg); and, c) an unnecessary bias toward larger, financially "robust" DHAs, i.e., those with enough money to underwrite marketing/promotion in addition to operations.

5. We have begun, therefore, to think in terms of a national or international data information handling agent (DIHA). Descriptive records of some type by some means would flow to the DIHA. There they would be entered into a central information bank (computer-based, of course). The information bank would be used to issue products and to provide services. The major product would be a publication containing

the descriptive records and a set of suitable indexes. Such a publication could take the form of a directory, whereby each edition replaced the previous one, or a periodical, whereby each issue reported new records and cumulative indexes are printed once per year. Two services would be provided, both of them involving custom-designed searches of the information bank: a) Desired data could be described and the entire bank could be searched (the "retrospective" search mode); b) Desired data could be described and recent additions could be searched on a periodic basis (the "current awareness" search mode).

The DIHS schematic just outlined resembles the structure of literature information handling systems (LIHSs) as the combined talents of library, computer, and information scientists have brought them into being:

1. A primary researcher can be viewed as an author while a secondary researcher can be viewed as a reader.

2. A DHA can be viewed as an academic journal or a book publisher. Both collect materials from authors and make them availabe for use by readers.

3. The DIHA can be viewed as a secondary publisher, such as Sociological Abstracts, or a computer

search service vendor, such as Lockheed, SDC, BRS, etc., or both.

It may seem that the problems faced by LIHSs are simpler than those faced by DIHSs. I submit that this impression can be traced to the fact that LIHSs have become a familiar feature of academic life and that this fact can cloud our recollection of how things used to be in the 1950s and even, for many, the 1960s through the early 1970s. I believe that the similarities are strong enough to invite comparison so that the "light" OF LIHS experience can be shed on DIHS development. Perhaps even a co-development of these two types of systems is in the offing; this is certainly my hope and I don't believe I am unique in this respect.

ANALYSIS

My analysis of the distribution of labor in DIHSs begins with the observation that IHSs always have one or more "bibliographic units" (BUs) which are the life-blood of their operation. I believe that three types of BUs can be found in a DIHS:

1. The most complete type is that of the actual data and codebook themselves. These can be collectively referred to as a "study documentation." It may be that these have been transferred by a primary researcher to a DHA. The DIHA should not assume that this is the case and should provide for participation by individual researchers who desire to maintain discretionary control over their data

even though they are willing to release information about them.

2. The second type can be called a "study description" (SD). By this is meant a somewhat exhaustive characteriztion of the data: the research problem for which they were produced, notable aspects of the methodology, time-frame, population, etc., and comments about condition, accessibility and so forth. I propose that the preparation of this be regarded as the responsiblity of the individual primary researcher, DHA, or both. I wish that it was reasonable to think in terms of a "standard SD scheme" but I think that the most we can hope for in the near future is the promulgation of a set of guidelines, much as academic journals offer to their contributors.

3. I regard this concept of the SD and its role in a DIHS much as I regard that of the individual journal article in a LIHS. The SD would be the bibliographic unit sent to the DIHA. At the DIHA the SD would be analyzed and a "study citation" (SC) would be prepared. These SCs wold constitute the actual entry to the growing information bank. They would be very similar to the entries bound in LIHSs, viz, author name, article title, journal citation, abstract, and keywords. The DIHA would

have to be as flexible in its receipt of SDs as a LIHA is in its receipt of journal articles; some variance in format and content should be expected and accounted for. It would be possible for the DIHA to create a microfiche/film file of all the SDs it had received and to offer SD reproduction services. This could be done on a one-at-a-time basis, for individuals, an entire collection basis, for libraries, or on some other basis in between these two alternatives.

I have refered to these SDs and SCs without having said much about what in specific they would look like. For the purposes of this paper this concern is not a pressing one. I believe answers to this question are forthcoming from a number of efforts in process and IASSIST, for one, has Action Groups in both areas. My interest is in describing what I think should be done with these BUs.

This discussion of the BUs and their processing by a DIHS brings to mind another question with division of labor implications. As I have presented things, a SD would arrive at a DIHS where it would be filed after a SC had been composed and entered into the information bank. The reason why I propose that all aspects of the SC be under the direct control of a DIHA, rather than having DHAs submit their own SCs, is that I believe that a valuable if not necessary uniformity would be achieved by this means. (This is particularly the case insofar as the "coding" of the SCs consistent with the provi-

sions of a vocabulary control protocol are concerned. Such protocols are now regarded to be a necessary requisitie of reliable IHS operation.) I also believe, on the other hand, that the majority of the information provided by the SC should be imbedded in the SD. This would ease the SC preparation labor requirement at a DIHA. Even though this is the case it is important that a DIHA be responsible for the accuracy, completeness, and general quality of the SCs. Specific, identified accountability in this area is extremely useful.

I have two additional observations on this subject. Part of my position vis-a-vis the division of labor attendent with SD and SC preparation is based on my experience with journal article abstracts in the LIHS represented by United States Political Political Science Documents (USPSD). This experience shows that, in general, there is quite a difference between abstracts written by authors and those written by document analysts in a LIHS. My belief is that the abstracts of analysts are better than those of authors in the senses that they are more accurately descriptive of an article's actual contents and that they are more adequate predictors of source document relevance. My second observation is that even though I have described a situation in which SDs arrive at a DIHA by mail, I believe that the DIHS should encourage submission of SDs in machine-readable form. This may also have a bearing on the standardization of SDs issue. It is possible to conceive of DHAs presenting SD data to a DIHA through use of an interactive program housed in the DIHA computer. It is also possible to conceive of a DIHA making a program available to accomplish the same thing using the computers of the DHAs. Whichever, the

simplification of SD preparation and transmission would hasten development of SD standards. The most advanced possibility along this line would entail the SC information bank being available in much the same way that the Ohio College Library Center (OCLC) has made available Library of Congress (LC) cataloging data. These ideas are exciting and technologically feasible. Nevertheless, without a firm base of support I fear that setting our sights on these equipment configurations as a first step of development is ill-advised. I believe, to say it once again, that we are involved in grass-roots organizing and should use distribution of labor concepts appropriate to this.

There are also division of labor concerns to consider at the output end of the system. In my scheme a DIHA would broadcast the entries in its information bank by means of publication, directory or periodical style, and search services, retrospective and current awareness. To every extent possible, the preparation of the publication should be done at a DIHA using advanced information-base management concepts such as those we use for USPSD. A sophisticated programming system, for which many models are available, can produce a SC listing and all desired indexes from the same "linear" file as is used for searching purposes. This means that searching and publishing can conveniently interact so that the information bank can be packaged and repackaged in a very dynamic fashion with little or no additionak keyboarding labor. The most desirable output from such a system would be "camera-ready copy" for the publication in question; given this, the only "external" cost to be incurred by the publication would be that of printing and

binding. If the DIHA did not have internal control of some sort of computer-based typesetting equipment, then the best alternative would be to produce appropriately encoded magnetic tapes for someone who does; most metropolitan areas have printers who do. Under no conditions should the DIHA accept an arrangement whereby the text of the publication is rekeyboarded; such a thing would be inordinately costly and it is no longer technologically necessary.

The reasoning applying to provision of search services differs from that of publication. Here the co-operation of a DIHA with other agencies is not only possible, it is desireable. The minimal method of service delivery would entail the contacting of the DIHA by someone interested in having a search performed. A discussion of the characteristics of desired data would be held with a "retrieval specialist" who would compose "search strategies," review results, and forward what seemed to be the most relevant entries in the information bank to the inquirer. The experience of LIHSs leads me to believe that, depending upon volume, the "turn-around" time for such a system would run from 10 to 17 days and would occassionally take even longer than this. Evidently, a more time-responsive approach would be preferrable. The manner in which this could be accomplished would entail the provision of access to the SC file by way of an on-line, interactive search and retrieval system. The DIHA could engineer such a system itself or it could transmit its information bank to a vendor of search and retrieval services, someone like Lockheed, SDC, or BRS to name the current three major commercial actors. I favor the latter approach primarily because it woud reenforce what I consider

to be an extremely promising development in the literature searching sector of the information industry. It would also reduce the costs which would have to be borne by the DIHS in its efforts to establish this sort of search and retrieval operation. Regardless, the important thing would be to enable an individual or a DHA to perform their own searches and to receive results as soon as possible. Some form of on-line, interactive search and retrieval system would be the answer.

Two footnotes can be added to this discussion. First, I believe, consistent with my reported beliefs throughout this paper, that the place to start is with the provision of search services through a retrieval specialist attached to a DIHA. It is a tried and true technique, it concentrates investment in labor rather than equipment, it can be initiated quickly, and it provides a powerful evaluative framework. Nevertheless, the eye of the DIHS would always be on the on-line, interactive approach and as soon as adequate financing became available and broad popular interest and support became evident the DIHS would move in that direction. It is also important to note that such a movement would not render obsolete the retrieval specialist role at the DIHA. My experience with LIHSs, again, provokes this observation. Some individuals and DHAs, for various reasons, will never want or be able to assume responsibility for their own searching and retrieving activities. Others will perform searches so infrequently that they will desire to consult with someone with more experience. In short, retrieval specialists attached to any IHS always add value to computer-based search and retrieval procedures in a way analagous to

the value added by a carpenter to a hammer.

My second footnote is that the DIHA should have the capability to transmit the SC information bank into any requesting agent. This means that suitably equipped and motivated individuals and DHAs could acquire the entire information bank and updates thereto so that they could process their own searches locally rather than remotely. For agents with a high volume of searching activity this possibility will allow greater cost control than the other arrangements discussed so far. It will also, need I say it again, distribute labor in a useful way.

CONCLUSION

These "brief" remarks have wound up as rather "extended" remarks but

I am left with the impression that much more needs to be said. My intention has been to direct attention to the question of who will do what in the DIHSs we are all envisioning. This seemed to be worth doing because much of what I have been reading and hearing recently has been distionctly oriented toward equipment concerns. It occured to me that it was once again time for someone to raise the issue of the people dimension of the data information production, acquisition, organization, and dissemination process under discussion. This is what I have attempted to do. I believe that this concept of "distribution/division of labor" which has occupied my attention in this paper is extremely important and can provide a very concrete principle guiding how we choose among competing courses of action. I hope that its articulation and examination will lead to further discussion.