

~~SECRET~~

INFORMATION RETRIEVAL ON URBAN AFFAIRS*

by
HERMAN G. BERKMAN, A.I.P., PROFESSOR OF PLANNING
GRADUATE SCHOOL OF PUBLIC ADMINISTRATION

The importance of information retrieval on urban affairs cannot be overstated. This is not only because in his most recent budget message to Congress the President called for the establishment of an Urban Information and Technical Assistance Program at a cost of \$6 million. More fundamentally, our concerns are important because, as Mayor Henry Maier put it at NYU's first conference on Applications of Electric Data Processing to local governments, we have within our grasp new management science techniques and technology which will help:

1. Promote greater economy and efficiency in present programs to aid our cities.
2. Develop the information needed to develop the new programs needed.
3. Provide, and this is our major concern here, more accessible information on which to base public management decisions.

An underlying concern, however, is related to a deeply important philosophical and methodological question. This is the extent to which research in Urban Affairs can or should be organized, structured, and coordinated in attempts to eliminate duplication, to optimize on the work of others by avoiding their pitfalls and building from others processings and findings.

*Prepared for "The Conference on Cooperation and Communication in Urban Research." Feb. 1, 1967, 10:00 A.M., Auditorium, Graduate Division City University of New York, 33 West 42 St. New York, New York.

This paper is in part supported under NASA Contract NGR 33-016-067.

N 68-28991

Cat. 34

13 Pages

Code 1

CH 85783

Cat. 34

HC MF

If this is the sole purpose of an information exchange, there is reason for some concern. Advances in knowledge cannot be achieved through excessive structuring. Research must be spontaneous and free to respond to the creative, intuitive processes which help differentiate man from computer processing. In this sense, research itself is a self-organizing process. Too much attempted direction can destroy creativity; spontaneity, diversification and even duplication have merit. Redundancy to a certain extent must be built into every successful system.

Pure research and operational research both have to be justified in terms of costs and benefits. However, there is a caveat to applying such analysis in research.

Efficiency and cost-effectiveness analysis in a given research project may be possible. However, if we enlarge the research system with respect to a specific programmatic subject such as urban affairs, the whole system may constitute a symbiotic, cooperative research effort where findings would not have occurred without research efforts interacting with each other. This should be the proper purpose of exchanging information on urban affairs.

Obviously cutting costs through sharing data bases, and information already generated is a sound objective of such systems.

However, here we must be careful not to impede research efforts either for programming, operational purposes or planning purposes or evaluation by introducing time-cost delays through excessive attempts at data standarization, or system capatibility. Nor should such impediments obstruct pure research efforts.

The development of several information systems, therefore, can be justified without waiting for some grand design. These are

[REDACTED]

justified on a priority of need and indigenous approach basis. Indeed, if such articulated subsystems for providing information exchange were developed, we might view the total of such systems as some form of time-sharing, on-line system which may tend to be integrated at some later time through executive programs and other integrating devices.

I am, therefore, pleased about distinctions in approach. Indeed, I hope the System Development Corporation in its work for the Department of Housing and Urban Development will recognize differences in local needs embedded in different localized environments so as to incorporate these within the general requirements for federal administration of housing and urban development program. The concept of "urban observatories" as proposed by Under-Secretary Wood for developing subsystems may be a usefull approach to urban research. Experience with local needs can then be integrated into a more meaningful total system. It is important to take advantage of diversified experience in developing information systems.

Obviously there are common problems which are difficult and expensive to solve.

There are at least three (3) common problem areas which are identifiable and about which sharing of knowledge is essential:

1. A communication problem
2. A pattern matching problem
3. An organizational problem

1

R.M. Hayes. "A Theory for File Organization" in On Line Computing. Edited by Walter J. Karplus. New York: McGraw-Hill Publishing Co., 1967, p.264,ff.

Because data to be sought in urban information exchange are exogenous to any given retrieval system, there are communication problems. The data to be sought must be described. It is not the data we are worried about here. Rather it is a description of the data which is the problem. Mechanisms must be, therefore, sought to aid in the process of selecting, including mechanisms for translating, to provide adequate descriptions. This is the communication problem.

This description must be matched with the selection of relevant data. In this matching, there is need to define relevancy and there is possible uncertainty and error in description of requests and data stored.

The problem of organization is essentially a technological and economic one. For it is hardly economic to scan every item in order to reach a decision point as to relevancy in accordance with the descriptors. The answer to this problem lies in so structuring a file as to provide indexing mechanisms and intermediate measures of match much less sophisticated than the ultimate measure of relevancy.

Size of file, response time, degree of selectivity and accuracy of response all interact in the problem of organization.

I recall the initial stages in the development of project Urbandoc. The project director asked for a meeting of urban planners. A group of us met in our studio at NYU where we sat around a table and threw out words which described the material in which planners might be interested. Here was a most natural way to begin our man-machine interaction. But I remember my dissatisfaction at the lack of controls in solving the problem

of communication.

While heuristically acceptable, lack of formal attack made me uncomfortable. In subsequent meetings of this group interested in this project, I reiterated my concern over the unstructured, intuitive attack made on this problem of communication. The project has made significant progress since those early attempts in uncovering the infra-structure.

We know that the communication problem can be broken into four (4) parts:

1. A vocabulary of descriptors must be defined.
2. A network of interrelationships adequate to the needs of translation must be defined.
3. Items must themselves be structured into formats and fields and then described in the vocabulary.
4. Descriptors must provide for many levels of pattern matching.

What we were doing in that studio was to provide the raw material which would have to be processed into terms we would have to reduce to descriptors recognized by and acceptable to the computer. On these bases a glossary of terms thus acceptable would be developed. This matching of descriptors from the profession with the terms the computer can accept marks the first significant potential leakage in the retrieval system. This leakage must become measurable if untoward loss is not to ensue. Techniques for measuring can be shared especially when applied to the same information base. Probability models are, therefore, useful and other lessons from information theory.

The terms of the glossary of machine acceptable terms must next somehow, in any relatively sophisticated system, be linked to each other. Generic and "degree" of association relations must be established. This is a mathematical processing. A hierarchical classification system is the key to a generic link. But usually a given term has more than one other term more general corresponding to it. Lattice theory is useful here.

The underlying lattice structure becomes a means for defining terms and for exploding them. This network of interrelationships between the rich vocabulary of the exogenous user and the more formalized, structured vocabulary of the computer must, therefore, be laid bare to communicate with maximum effectiveness. In urban information exchange, cooperative efforts in these areas and these directions should be possible.

The relationship between these items and their use is determined by the number of defined formats, the fields of data descriptors within each, the number of values acceptable to a given field, the degree to which a given field is repeatable and the number of roles which each field is assigned.

Finally, the level of description must be distinguished. The application of mathematical lattice theory is the first structured or formal processing of the data. These formal relations, as with the case of languages in data processing can be established with respect to a specific user or problem orientation. Hence, a sharing here would be desirable. The pattern-matching problems constitutes another area for mutual attack. For there must be a succession of screenings in accordance with rules for matching, each screen finer than the other.

In a field like urban affairs, the files of data are large and vocabulary is complex. Further, do to the richness of approach and interest, there is less formalization or specification to the descriptions.

With such extensive requirements imposed on retrieval, it becomes too expensive to apply sophisticated screening devices to every file item. However, a succession of screens is necessary.

Thus screening of words, phrases, and syntactic relations in that order becomes the basis for linking the outside descriptors to the system data base.

The characteristics of urban information systems thus described result in error, in the fact that all descriptive data can not be included, and in the fact that significant interrelations among descriptors may be missed.

Hence, weights may have to be introduced into the measures of degree of match. Difficulty of the pattern matching problems may than be summarized as follows:

1. A succession of measures of degree of match must be provided.
2. These must accommodate likelihood of error in description.
3. They must recognize the likelihood of rejecting items wanted.
4. They must also accommodate desired interrelationships among items described by the glossary.
5. They must be efficient insofar as processing time is concerned.

All user oriented urban affairs researchers face these problems in retrieval, whether systems are or are not computerized.

Another shared problem in the organization of a sharing of urban research in a retrieval system is file organization. In some respects, all previous considerations are embodied in the way the data files are organized. Required response time, degree of required reliability, value of information retrieval and cost are criteria for development of any file organization. In urban affairs information exchange, there are differences in the answers to these questions depending upon the particular system and its use and users. For each system, a specific answer to the following questions must be developed:

1. What is the relation between the value of the system and its response time to the user?
2. What is the relationship between the cost of building in reliability to the value of the information to the user?
3. What is the relationship between the value of an information system and the measure of information?

The value of an urban information system to a user is a function of time, cost, and measurement of the information that can be provided for a specific policy, operating or evaluational purpose. It may be assumed that the value of the decision is a base for judging the value of a system. But the value of the decision can be gotten at least in part from the amount of information upon which it is based. If we may assume that at least the value of a decision is a function of the amount of information on which it is based, we can begin getting some insights into how much we are willing to pay for such a system in any given case.

It is important to note that these problems exist as much in the establishment of such institutions as data banks for numeric

data to be stored, processed statistically into information and retrieved as they are in the establishments of urban information exchange systems where alpha-numeric data bases are inputs.

Hence, software and hardware systems developed for statistical information processing should be reviewed in developing sound alpha-numeric systems for urban information exchange.

Institutions and methods for expediting exchange of languages, system components, programs, and system design should, therefore, be developed.

This conference and its participating sponsors represent one such method. Others are the IBM SHARE program, the work of the Public Administration Service's "Public Automated System Service," the Bureau of Public Roads Computer Program Exchange Group, and the Conferences on ADP in Local Administration held under the auspices of the GPA and School of Continuing Education of NYU for example. The newly founded Association of Urban Information Systems promises also to be a sound supplement to other conferences held such as the more technically oriented sessions of the frequently held Joint Computer Conferences, the work of the Highway Research Board of the National Science Foundation, and the work of the American Documentation Institute among others.

I should like to close these remarks with another caveat to SIE and those trying to report on organizing research. Reporting on-going, uncompleted urban research in most areas in the social sciences may be at best a futile activity at this stage. At best high volatility occurs. For such research is likely to be:

1. Changed in direction
2. Dropped

The reasons for this are:

1. Lack of adequate long range, sustained funding.
2. Lack of adequately worked out program budgeting and estimation of time and cost.
3. Nature of the field-
 - a. It is richer, less structured or precise, than others at this point.
 - b. One encounters greater difficulty in obtaining relevant data.
 - c. The research processing generates changing content descriptors.

Recording of ongoing research in the natural and biological sciences appears to be more reasonable. Recording completed research in any area cuts down on the data base. I don't know whether anyone has made a study of the ratio of starts in research in the social sciences to completions but I would bet it is low. I submit recording of uncompleted research in Urban Affairs is costly, inefficient and may not be worth the effort. I reach this conclusion even though SIE may be dedicated to its objective as a referral center to those interested in Urban Affairs as well as providing a guide for urban research management, for federal agencies and other funding agencies. Urban Affairs may not be a field but rather an area of concern and program.

In "harder" sciences procedures for exchange of all types of research have been largely established. New exchange and abstracting systems are emerging almost daily. Reportage on ongoing research may occur where the findings are peculiarly important through special communication to editors of journals or through SIE

procedures. Even here directions change; content identification is being transformed by the research processing.

The long time sustained character of such research and the historical development of institutions able to cope with :
administering it result in greater effectiveness of information exchange. A simple transposition of these institutions to urban affairs is dangerous. Research in urban affairs is the baby of research efforts. The research process itself here should not be extensively burdened by continuous need for reporting and administration. Where reports to funding agencies must be developed to account for expenditures, evaluation and appropriation, such reports may be used as data sources. These essentially are built into current systems for information exchange where possible. But greater efforts must be developed to reduce administrative steps. Principal investigators who may change in the course of the research, usually report to central administrations which might be sources for information. Every University, for example, issues publications on research efforts.

There is a growing field of research administration which can consume valuable research time and talent unless its logarithmic growth is stopped.

Means for avoiding intrusion upon research time and manpower may lie in the development of a specialized field of research administration. Reports can be computerized. But such specialization can result in efficiency and economy only where the overhead costs involved can be sustained by a large scale program of research.

of research. This may mean that, unfortunately, the "business of research is a decreasing cost industry." Under this formula, research grants are concentrated in selected institutions.

This is unfortunate since the human system remains the "best" means to obtain answers, provided the intelligence system can be organized to combine the optimum man-machine components. The emerging science of Synnoesis may contribute to our handling of this system. Regardless of developments in that field, however, I feel that a minimum call on researchers time for reporting and administration should be an important criteria in the development of urban information exchange programs. In this light sources of input have to be carefully evaluated. Questionnaires, forms, and procedures for identifying and collecting and collating research may be convenient means for obtaining input. But minimizing and concentrating sources even if this means obtaining information indirectly, and shortening, clarifying and simplifying forms can reduce time intervals for obtaining such information. Computerizing information collection and dissemination through on-line systems must be developed. These might be organized as public utilities thus becoming external economies in research information processing.

It should be clear that I am speaking as a researcher who has been subjected to countless demands on my own time in the administration of research programs.

On the other hand, as a researcher in search of information, I obviously applaud and encourage those seeking to enhance, encourage and make more efficient and effective information exchange.

Information system comprehensiveness, integration, and unification are not the only criteria for effective exchange systems. The validity and applicability of input remain the chief aims.

"Garbage in-garbage out" remains a dicta to be heeded by those building urban information exchange systems.