DOCUMENT RESUME

ED 154 784 IR 005 840

AUTHOR Pogrow, Stanley
TITLE Implications of a Resource Approach to Data Management for Improving Intergovernmental Data Flows in Education.
PUB DATE Mar 78
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Canada, March 1978); Document may not reproduce clearly due to unevenness of type

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Computer Storage Devices; *Data Bases; Data Collection; *Educational Administration; Electronic Data Processing; Government (Administrative Body); *Government Role; Improvement; *Interagency Coordination; *Management Information Systems

ABSTRACT

Problems associated with uncoordinated proliferation of data collection procedures will not be alleviated until educational agencies and legislative bodies, at both state and federal levels, view data as a scarce resource which must be managed efficiently. Better linkages must be made between the data transfer requirements generated by the legislative process and effective administrative procedures. A centrally controlled, computerized data base system with decentralized data access could optimize efficiency by condensing and storing data, while maintaining the same levels of data and administrative report generating capability. (Author/CMV)

Implications of a Resource Approach to Data Management for Improving Intergovernmental Data Flows in Education

by

Dr. Stanley Pogrow
Assistant Professor
Department of Administration, Higher, and Continuing Education
University of Illinois at Urbana

The proliferation of new federal and state programs in the late 60's and early 70's promoted geometric increases in the amount of data flowing between, and stored at, federal, state and local educational agencies. The continued growth of reporting requirements combined with a leveling off of federal funds to help meet the administrative costs of providing such data has led to a condition in which the capability of agencies at the local and state level to provide required and requested data is becoming strained. Compounding the problem of growing reporting requirements is the constant change in the format of existing ones. Another seldom discussed contributor to the problem is the inadequacy of existing data management procedures at the state and federal levels. The rapid growth of the responsibilities placed on these agencies by the mushrooming categorical programs outstripped their administrative capabilities. The categorical nature of the programs encouraged a form of coping characterized by inter-level communication between harried administrators of a given program at the state and federal levels at the expense of intra-agency communication. The management systems to support the administrative effort developed in a similar fashion, on a program-by-program, application-by-application (within programs), basis in relative isolation from each other. Efforts to improve the degree of coordination at the state level led to the passage of E.S.E.A., Title V, which provided funds for State Educational Agencies (S.E.A.) to establish offices of evaluation and planning. As opposed to the impact of evaluation offices, planning offices have had little organizational impact in the area of program coordination in most states. This has been largely due to the weak technological underpinnings of, and the absence of external political constituencies for, the planning function. The amorphous concepts of long-range planning and coordination tended to be operationalized in the form of cliche-ridden master plans which had little effect on the degree of coordination between the increasingly politically powerful and entrenched program units. Some states attempted to improve the degree of coordination between programs by proposing various types of reorganization plans for their state education agency. For the most part, such plans were not implemented, or were implemented on paper only, due to the political strength of the program units.

The failure to improve the coordination between program units at, and between, governmental levels assured that data requests and collection procedures would be uncoordinated. This accelerated the rate of proliferation of data requests.

Despite universal agreement that a growing problem exists in the ability of educational agencies to provide required and requested data, there are major deficiencies in the solutions being proposed. The most commonly proposed solutions are the following:

a) National groups such as the Chief State School Officers (C.S.S.O.), through their Committee on Evaluation and Information Systems (C.E.I.S.), should monitor new federal requests for data and make recommendations as to the utility of the requested data.

b) Develop a more standardized set of terms and codes for reporting data.

c) Consolidate existing programs and data reporting requirements.

d) Refuse to accept federal funds.

e) Complain, pass the buck onto the next level, and hope the problem will go away.

It would appear that all the above proposals are limited. Proposal "D" is not a viable one for those many school districts in which federal funds constitute a significant portion of the budget. Proposal "e" is unrealistic since the problem will not go away. While proposal "a" might tend to limit further growth of the problem, it offers little solution to the existing problem. In addition, it depends on external oversight by representatives of agencies whose own data management techniques are highly uncoordinated. Proposal "b" appears to be of greater utility for facilitating comparative analysis of collected data as opposed to simplifying the data reporting and transfer process. Proposal "c" is the only one which offers the potential to alleviate the existing problem (assuming that political roadblocks to consolidation can be overcome). However, even this proposal doesn't deal with the fundamental problem of inadequate data management techniques which would continue to persist among the consolidated programs. In addition, this proposal does not provide any insight as to how to improve the degree of coordination. The proposal "hopes" that coordination will automatically improve as a result of the reduction in the number of programs.

In his book, Designing Complex Organizations, Jay Galbraith notes that the redesign of information systems is an alternative to changing organizational structure for improving task coordination and performance in organizations. The purpose of this paper is to suggest a new approach to operationalizing the concept of improved coordination which has potential for relieving the existing problems in interagency data transfers and intra-agency data utilization. Rather than using organizational units, the basic focus on this approach is the data itself, and the use of modern management techniques, integrated with the legislative process, to improve the ability of agencies to handle and process information. It is an approach which accepts the reality of educational organizations operating in an environment in which the need for and uses of data are expanding and changing. In addition, a research agenda for testing the feasibility of the recommended approach is included.

A Resource Approach to Data Management

As the number of administrative applications in an organization proliferates, the development of separate systems (whether computerized or manual) and data sets for each application inevitably produces increasing rates of data redundancy, as the same data elements are used in a number of data sets (applications). In addition to the wasted resources needed to store and collect the redundant data, it also represents an imposition on those from whom it is collected. Different update periods for the redundant items produce inconsistencies in the value of the same data item for different applications. The redundancy and proliferation problems inherent in the separate systems approach also hinder the organization's ability to coordinate applications or to respond to changes in applications. As applications change, organizations in this chaotic environment may find it easier to develop a new application and a new data set rather than try to revise the old system, a condition which exacerbates the proliferation and redundancy problems.

In modern data base management systems, data are viewed as an organizational resource which is pooled independent of applications or organizational units. The data base consists of those unique (non-redundant) data elements which the organization needs to support its administrative applications. Organizational units indicate which data elements they will need to process their applications. These requests are consolidated centrally, and a redundancy-free coordinated data collection process initiated. High level data base management software (computer programs) with simple-to-learn English-like commands automates the information storage, updating, and retrieval processes. In addition, the management software

performs the general processing functions that are common to most administrative applications such as report generation. This drastically reduces the need for application programs.

The capabilities of modern computers make it feasible to store massive amounts of data on-line (readily available) which can be accessed simultaneously by multiple users. The data base software makes it feasible to decentralize the data retrieval and processing functions. Clerks and administrative staff in the organizational units can directly access the central resource by themselves, (due to the English-like nature of the commands) to work with the specific data elements which they need for a given application (security provisions to restrict access can be built in where necessary). Changing applications which require working with different combinations of data elements are automatically accommodated by the existing system. Such a system offers the greatest flexibility for organizational adaptation to a changing environment. Multiple access to a resource pool of unique data elements reduces the need for structural coordination to accomplish information processing and dramatically reduces the problem of data redundancy.

The key to successful operation of such a system is prior anticipation of which data items will be needed over a period of time. A data base administrator is responsible for determining the composition of the data base as well as assigning responsibility for collecting and updating of the specific data items to be included. Anticipation must be accurate for responding to the data demands and requests made by other agencies as well as supporting its own internal data needs. An additional responsibility which could be envisioned for an S.E.A. data base would be to help L.E.A.'s meet state and federal requests for data. However, improved integration of agency data management practices with the legislative process is necessary if data base administrators are to have adequate lead time.[1] Such integration would require reforms in both branches of government.

Governmental processes in this country have a long tradition of recognizing the need for integrating agency requests for resources with legislative processes for allocating them. The basic objective is to allow for informed legislative decision-making about agency requests while providing sufficient lead time to the agencies to make plans to effectively utilize the allocated resources. However, such procedures are used primarily for fiscal resources. There are no equivalent legislative procedures for attempting to provide the same oversight and coordination of the use of what is an increasingly important resource in the governance process--data. The distinction in governmental procedures between fiscal and non-fiscal issues developed at times when money was the only significant resource in the delivery of governmental services. Applying basic governmental procedures to data management at the federal and state levels would encompass the following:

---

1 Data base software provides the capability to generate linkages between different data sets if unique identifies (e.g. Social Security number) exist. In an application area where there is a set of unique identifiers, the need to anticipate all needed data elements is less critical. Should it become necessary to collect some additional unanticipated data elements, these new elements can be collected and linked to the existing (anticipated) areas. Work on the Soundex numbering system in Maryland suggests that it may be possible to develop a unique identifier code for students which maintains confidentiality.

a) Agenda preparation for the coming fiscal year--Paralleling the process where departments prepare budget requests which move up the hierarchy and are consolidated at the top--each department would identify the data elements which they need to collect and would present justifications of the need for that data. These data requests would be consolidated, presented to a legislative committee and approved prior to the beginning of the fiscal year. This would provide lead time for the recipient agencies to organize their management systems to respond to a coordinated set of data requests.

b) Data accounting--As with fiscal resources, agencies would be expected to furnish periodic reports accounting for how requested data elements were used.

c) Data impact statements--When legislation has fiscal resource implications, legislative committees include cost recommendations (authorizations). This is to allow a process of deliberation whereby the potential benefits of the legislation are weighed against its cost. A data impact statement attached to each new piece of legislation would indicate how much new data would be required to support the legislation, who would be responsible for collecting it and whether this data collection task would be integrated with existing data collection tasks. Hopefully this would provide for a deliberative legislative style which asks whether the benefits of proposed legislation are commensurate with the data collection processes which it would initiate.

d) Standardization of data collection timeframes--At the present time there are few systematic procedures governing administrative discretion in the interpretation and implementation of legislative mandates. The most important procedure whose application would have the most potential for improving data flows would be standardization of the point(s) in time in which data status will be measured. At present, each program unit uses different points in time to measure program status. Additional simplification of the data transfer process would result if standardization of federal reporting dates coincided with the dates which L.E.A.'s were required to supply student and program data to meet state requirements.

It would appear that using data base management techniques, combined with improved governmental procedures for providing adequate lead time to coordinate data requests and for data base administrators to organize data collection procedures has potential for improving the existing data management capabilities of state and federal agencies.[2] The next section examines the organizational implications of implementing these concepts.

[2]Although the National Center for Educational Statistics will be developing an integrated data base management system for its operation, it is not clear whether this system, or approach, will be utilized by the Office of Education in its data collection and reporting responsibilities for program administration. The development of separate data bases for administrative operations and statistical reporting could possibly defeat the unification potential of the data base approach. Indeed, the maintenance of separate data bases for each governmental function could possibly generate redundancy and coordination problems as bad as under the present practice of separate systems by application.

Organizational Implications

Previous attempts to improve program coordination failed because of the political strength of the existing organizational units, and the tenuous technological underpinnings of the functions performed by the appended units which were responsible for coordination improvement. However, the concept of viewing data as a resource, and of using a centralized data base provides an alternative approach to improving the degree of intra-organizational coordination.

Organizational strategies for improving coordination which attempt to consolidate or reorganize existing units do not recognize differences in types of resources or types of control which units exercise over these different resources. Such strategies attempt to remove control over all types of resources (particularly fiscal) from the existing unit. Political retaliation is inevitable.

A resource approach to the problem of coordination recognizes that there are a variety of resources which organizational units possess and a variety of types of control which they exercise over each resource. Under the proposals outlined in this paper, existing organizational units would continue to exist and would continue to control their monies. What would have to be relinquished is their control over the collection and storage of data to a data base administrator; however, they would maintain total access to needed data elements and would retain control over how the data was used to make decisions relative to their programs. This conception of the reorganization needed to improve coordination differs from previous ones in that it partials out the key specific resource allocation procedures that must be reorganized rather than proposing that all resource control be removed from existing units, a goal which tends to be politically unfeasible and which is really not necessary or sufficient.

The major structural change that needs to be made in the organization is the establishment of a Data Base Administrator. The technology of the tasks to be undertaken by that individual is clear. The delegated responsibilities constitute the primary political resources of the Data Base Administrator. To be effective, such an individual will have to work closely with the heads of the organizational units.

Some S.E.A.'s have begun the process of coordinating the data collection process by establishing a central forms approval office. However, these agencies haven't yet addressed the issue of central storage of collected data.3 It is time that educational bureaucracies transcended the Weberian conception in which each office has its own files "preserved in their original or draught form."

3Some agencies have equated storage of L.E.A. provided data with establishing statewide data processing systems to provide data processing services to school districts. The proposals contained herein deal solely with a point in time coordinated data collection and storage process for the purpose of generating requested and required reports. They do not envision the state attempting to provide transactional on-going data processing services such as attendance reporting and financial accounting to L.E.A.'s since more flexible and cost efficient alternatives are available. The latter form of service also has the potential for constraining L.E.A. autonomy.

Conceptual Framework

There has been little research into how organizations use data and what their data needs are. Despite the widespread concern that has been expressed about the amount of data actually being requested and the degree of redundancy and coordination in the data being requested, stored, and processed at the different levels. In order to determine the efficiency of current data handling procedures and the potential for a data base approach to improve the situation, answers to the above questions must be determined. The problem can be operationalized as a production function whereby a given number of distinct elements (x) can satisfy a certain percentage of the organizations' applications (y). Assuming a large value of y, the smaller the value of x needed to support a given value of y (i.e., the greater the degree of redundancy) the more feasible the data base approach. For example, if 70 distinct elements can support 80% of the data reporting requirements, then the data base is more feasible than the situation in which 150 collected and stored distinct elements could only service 30% of the agencies' administrative applications.

Another consideration is the level and locus of aggregation, as well as the unity, of the collected data elements. The need for aggregation arises because most data at the point of origination describes individuals (students, teachers, etc.) but is used in aggregated form (school, district, region etc.) for most administrative applications at the state and federal levels. The locus of aggregation refers to the point in the data transmission process where the data are processed into aggregated form--whether at the point of collection (L.E.A.) or at the point of storage (S.E.A. or U.E.) where a file of unaggregated data could be used to produce other files with different levels (school, district, region, etc.) of aggregation. The degree of data unification refers to the extent to which elements are collected on a simultaneous common basis. In a disjoint system separate data collection processes are initiated for different applications relative to the same basic type of entity (e.g., students) so that data for Title I students are collected separately from data for bilingual students. In a unified system, common data elements are collected from all students.

There are clear flexibility and cost tradeoffs to the collecting and receiving agency as to where the locus of aggregation occurs and the degree of unification in the data collection process. Aggregation requires processing effort and it is more difficult to collect data in a unified manner. However, unaggregated unified data (aggregation responsibility at the point of storage at the state and federal levels and a unified data collection process on the part of L.E.A.'s) is the most flexible type of data to work with. With such data, aggregations can be produced, not only for the individual elements, but also for previously unanticipated combination of elements. For example, if data on Title I students are collected disjointly from data on bilingual students and aggregated, there is no way to determine which Title I students are also enrolled in bilingual programs without an additional data collection process should an unanticipated request for such data be made by a federal agency or individuals conducting a state level school finance study.4

4 If it would be possible to develop a system of unique student identifiers, an additional data collection process would not be required. The data base management software would generate a linkage between the Title I and Bilingual files and enable the needed data to be extracted from the existing files.

The feasibility of adopting the unified unaggregated approach to data collection would depend on the degree of overlap in the data needed for different applications associated with basic entities such as students, and the absolute number of data elements that must be collected. The greater the degree of overlap, and the smaller the number of elements, the more feasible the approach, particularly for small and medium sized districts (operationally, the best way to collect such data would probably be to use a mark sense form, one per individual). Such a data base would not only have maximum utility for the S.E.A. fulfilling its own administrative applications but for also generating the administrative reports for L.E.A.'s requested by state and federal agencies.5 The processing of the forms would be a logistical nightmare but a combination of techniques could be used to facilitate the process. Some districts could send the forms to the state system, while regional mobile equipment could transfer the forms to magnetic tape in other districts. Large districts could rent the needed processing equipment for the period of time (about one month) needed to generate magnetic tapes. All the tapes would then be sent to the state computer system where they would be transferred to an on-line storage medium such as magnetic disc packs.

Since the successful application of the data base approach is dependent on the ability of the data base administrator to anticipate which data elements will be needed, the timing of intergovernmental data flows becomes critical. Ideally, there would be a point in time by which most/all of the data requests from external agencies would be early enough in the school year, and a sufficient time lapse before the deadline for responding to those requests, to allow a data base structuring and data collection operation to be conducted within that time lapse. Clearly the ideal does not exist, but the question is how much of an imbalance exists between data requests and the ability of a data base to respond. If a sufficient imbalance exists, the alternatives are to reject the utility of a data base approach or to develop governmental and legislative procedures that would lead to a more coordinated approach to data management. Since it is unlikely that any other approach would fare any better, the latter choice would appear to be the more fruitful one. The extent of the imbalance could be determined by developing cumulative distributions of data requests, and requests for unique data elements, over time (school or fiscal year) to establish the relationship between time of year and the percentage of known required data elements relative to the total number of elements which will eventually be needed. For each of the requested elements, due dates would be plotted along with data on the points in time by which the data base could respond to the request for a variety of cut-off dates for initiating data collection procedures. Response time would be a function of needed lead time and whether knowledge of the need for that specific element existed by the cut-off date.6 Simulations could then be run for a number of cut-off points and statistics generated on the percentage of data requests filled. For the optimal cut-off date (the one which minimized the number of unfilled data requests), further analysis could be

---

5 Even with such a data base it would be difficult to use the stored files to generate longitudinal comparisons of similar unaggregated data elements (which would be particularly useful for evaluation purposes) would probably be difficult to accommodate due to the mobility of students. If a student-by-student progress report were needed, the pre-post data would probably have to be generated simultaneously at the point of data origination (L.E.A.).

6 There are, of course, ways to anticipate the need for a specific element prior to its actually being requested. Techniques such as automatically including all elements used the previous year could be built into the analysis.

conducted to determine whether the specific data requests creating the imbalance appear to result from lack of coordination within the requesting agency or whether it resulted from ill-timed legislative mandates.

While the previous sections outlined a possible approach to alleviating some of the existing problems of data management and some criteria for determining the feasibility of the approach, the next section describes the data collection task that would be needed to test the feasibility criteria.

Data Collection Procedures

In order to determine the degree of redundancy, aggregation needs and timing of the data flows, a series of studies could be conducted in several states in which all data requests received at the S.E.A. and several L.E.A.'s from external educational agencies would be logged. A smaller scale study could focus solely on the S.E.A. and include the additional step of logging all data requests made by the S.E.A. to external agencies. Each data request would be broken down into data elements and the following information compiled for each element:

a) Name of element.

b) Date requested.

c) Requesting agency and unit (within the agency).

d) Response date and data collection date.

e) Level of aggregation required.

f) Type of application used.

g) Reason for request (administrative initiative, legislative requirement, etc.).

h) Types of linkages needed to other elements (longitudinal or point in time).

Consolidation and analysis of the above data collected over a period of a year would provide the first quantitative glimpse at the extent of the data management problem in education and would provide answers to the feasibility questions posed in the previous section. In addition, it would provide insights as to the types of remedies which are needed.

Summary

The basic premise of this paper has been that the problems associated with the uncoordinated proliferation of data collection procedures will not be alleviated until such time as administrative agencies, and legislative bodies, at both the state and federal levels, begin to view data as a resource. Furthermore, it must be viewed as a scarce resource which, like our other resources, must increasingly be conserved, husbanded and used more efficiently. To accomplish such a goal, new procedures for managing data collection and storage operations must be implemented in these agencies. Better linkages must be forged between the data transfer requirements generated by the legislative process and these administrative procedures. Some preliminary ideas for implementing such a dual approach have been discussed and a research agenda for testing their feasibility proposed. If feasible, implementation of these and other related proposals have the potential to dramatically reduce the amount of data transferred (and the number of collection operations) between agencies while maintaining the same levels of administrative report gen-

erating capability. At the same time the data base could be utilized for policy analysis and evaluation studies. Additional opportunities would exist for S.E.A.'s to provide a wide range of needed services to help local school districts generate state and federal reports.

The approach proposed in this paper accepts the reality of change and recognizes some of the political realities that previous proposals for improving the degree of coordination in the administration of educational programs. The data base management system approach provides the potential for changing the couplings of data items in response to changes in federal and state forms. It also allows existing organizational units to maintain most types of resource control over all of their resources. The only change is that data storage is centrally controlled with decentralized data access and utilization maintained. Finally, the ideas contained herein provide the basis for reconceptualizing/elaborating the concept of "organizational coordination" and for generating new strategies for improving intra-agency coordination.