4

Data Administration Workshop Proceedings

Frankie E. Spielman, Editor

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Center for Programming Science and Technology
Institute for Computer Sciences and Technology
Gaithersburg, MD 20899

Feburary 1986

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

NBSIR 86-3324

DATA ADMINISTRATION WORKSHOP PROCEEDINGS

Frankie E. Spielman, Editor

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Center for Programming Science and Technology
Institute for Computer Sciences and Technology
Gaithersburg, MD 20899

February 1986

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, ***Secretary***
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, ***Director***

PREFACE

This report constitutes the proceedings of a two-day workshop on Data Administration, held at the National Bureau of Standards, Gaithersburg, Maryland on March 27-28, 1985. The workshop was sponsored by the National Bureau of Standards under the auspices of the Federal Data Management Users' Group (FEDMUG).

The purpose of the workshop was to provide a forum for Federal, State, and local government Program Managers, Information Resource Managers, Data Processing Managers, and Data Administrators to hear nationally prominent speakers and to discuss and share data administration ideas and experiences.

The Data Administration Workshop Steering Committee consisted of the following members:

Ted Albert, Department of Interior (USGS)
Jane Benoit, Department of Agriculture
John Coyle, Department of Interior
Carl Fritzges, Department of Defense (DIA)
Daniel Schneider, Department of Justice
Margaret Skovira, Department of the Treasury
Frankie E. Spielman (Chair), National Bureau of Standards

The following individuals also provided significant guidance and help to the Steering Committee:

Vincent DeSanti, General Accounting Office
Ronald Shelby, Department of Interior
Roxanne Williams, Department of Agriculture

Because the participants in the workshop drew on their personal experiences, they sometimes expressed their own opinions or views which do not necessarily reflect those of the National Bureau of Standards. Additionally, they sometimes cited specific vendors and commercial products. The inclusion or omission of a particular company or product does not imply either endorsement or criticism by the National Bureau of Standards.

We gratefully acknowledge the support and assistance of all those who made the workshop possible. The Steering Committee diligently worked for nine months to shape the program and organize the sessions. We wish to express our appreciation to the committee members, authors, discussants, recorders, session chairs and the organizations, both in the private sector and in Government, who supported the participants.

Frankie E. Spielman, Editor

TABLE OF CONTENTS

WELCOMING ADDRESS
Data Administration Workshop

James H. Burrows
Director, Institute for Computer
Sciences and Technology
National Bureau of Standards
Gaithersburg, Maryland

The Institute for Computer Sciences and Technology is pleased to host this Data Administration Workshop. Our goal is to discuss some current issues facing data administrators and to map out future opportunities for improving the management and sharing of organizational data.

In the early 1960s when I got involved in the development of large electronic data collections, we talked about databases, file handling, and file access. The major point, though, is that these were largely centrally managed technical activities. Today, data collection and use are decentralized activities that are no longer exclusively within an organization's technical data processing operations. Almost every organization is experiencing an explosion in the amount of data that is collected and stored, and a tremendous increase in the number of computer end-users. Many people within the organization are collecting data that is potentially useful to others within the organization. However, we have not always kept pace in managing the data environment to serve the needs of these users effectively.

Data administration is emerging as the organizational function that brings together the end user and the data that they need. We are beginning to understand that data administration is really a key element of the overall management of our organizations, and not an isolated technical function. Data administrators have an important role to play in helping users find needed data and in educating users about the concepts learned in the centralized environment. Data security and integrity are critical. Data that is shared must be accurate because other people depend upon it being accurate.

Meetings such as this help us identify common problems and share the solutions that we have discovered. We at the Institute for Computer Sciences and Technology hope to learn a great deal about what you are doing, and we hope that you will take back to your organizations what you learn here. I believe that the goal of computing is not just to use technology, but rather to make it possible to use data. That's what data administrators are helping us do. Thank you for coming.

DATA ADMINISTRATION IN THE INFORMATION AGE

Keynote Speaker

Robert H. Holland, Ph.D.
Holland Systems Corporation
Ann Arbor, Michigan

The purpose of this Data Administration Workshop is to describe the role of Data Administration in achieving organization goals. It is a significantly important role. In the past, the role of the Data Administration function in organizations around the world has been too narrow and needs to be broadened. What has happened is that it has drifted towards solving today's problems with the current technology and has been done without looking far enough ahead into the future business needs of the organization. As a breakthrough in technology occurred, a new methodology was added to solve the current data processing problems. The result has been that organizations have not done the long-range planning which is so drastically needed to properly steer the organization in the direction that should be taken to solve the data problems. In other words, the solutions to problems have been done in piecemeal fashion. The Data Administration's role is to change the culture of the organization so that productivity can occur using the budget that is available to provide more and better services. The Grace Commission report concluded that there is an increasing gap between the available revenue for government expenditure and the real needs of government (figure 1). So, there are a lot of people in government saying let's put data in the hands of the appropriate people so that we can leverage our work environment and make us more productive. Data Administration is responsible for seeing that this happens.

In the industry, Information Resource Management (IRM) is a term associated with the process of managing an organization's information resources. IRM is the set of management approaches and mathematical principles that have been evolving over the years. It is the capstone of technology management that facilitates the integration of the many different functions of data technology, such as those performed by programmers, analysts, operators, project managers, database managers, and teleprocessing managers. The IRM concepts and principles related to these functions enable an agency to operate as an integrated set of functions. Without IRM, there is no integration and the deficit gap between government revenues and need for outlays will widen. With IRM, this gap will narrow.

IRM is a very broad and encompassing process within the organization. Underneath IRM, there are elements of business, technology, direction setting, and auditing. The role of Data Administration involves all of these key elements. It is very important for the Data Administrator to understand both the business needs and technology in order to help the organization meet its goals. The Data Administrator must be the communications bridge between the business and the technological elements. There are basically three groups having IRM responsibility within an organization. The first group is the direction setters or executive team who set the overall direction of the organization. The second group is the middle management team which is more concerned about the daily problems and meeting the shorter term objectives. Finally, there are the builders of technology, the IRM implementation team. It takes all three groups to make IRM work. Data Administration is one of the key elements in moving the organization in the direction set by the executive team.

Given the overall IRM direction that is established, there are five architectures of information technology related to the IRM activities. The first cornerstone is data which Data Administration is responsible for. This includes Subject Databases and data that can be derived from them. The other four are: application systems, hardware, networking, and systems software. Other groups are responsible for these cornerstones, however, they do get input from the Data Administration staff. The IRM staff is responsible for integrating all of these architectures so that the objectives of the organization can be met.

In looking at the broad IRM data model environment, historically the emphasis has been on the detail technical level without looking at the strategic direction. Many Data Administrators know the details of technical implementation from a decomposed view but don't include the strategic business directions. There is a gap between this technical view and the strategic view which must be closed. It is Data Administration's responsibility to change the direction or to close this gap through top-down strategic data planning, bottom-up implementation, and auditing. Data Administration needs good access to the overall business and IRM direction.

It is extremely important that we move toward narrowing the gap between the technical and strategic views of data because the demand for information is on the increase. An estimate by IDC Corporation shows that at least 15% of an average organization's revenues are spent on information handling (figure 2). Of this, 84% is labor intensive which involves collecting, synthesizing, storing, and summarizing data into a usable form called information. Most organizations that spend 15% of their budget on a specific resource have someone at the top managing these resources, but they haven't been doing it for information resources. This is an area where IRM can be beneficial and make the difference in the way information is handled within an organization.

There are other dynamic factors affecting the scope of effort. As the information demand increases and the technology costs decrease, the cost benefits from applying electronic technology to problems will reach an optimal point of return (figure 3). In other words, the changes in technology and information demands are so dynamic that it is impossible to determine a stable operating point on the optimal curve. Instead we should operate within an acceptable bandwidth around the operating curve. The first step in this effort is education - this is a must. The people must be educated on the technology that is available to solve their problems. When their knowledge matches the level of available technology, then there is an optimal trade off between the two. The assessment between the information demands and technology costs must be done as an integrated whole and not done in a piecemeal fashion. The Data Administration role is to set the proper direction for the data environment. However, because the information demand and technology are changing at a rapid pace, this is extremely difficult to evaluate, like shooting at a moving target. Therefore, the data environment should be viewed more as a data utility, like a power utility, where a person needing data can just plug into the system containing the data resources or a network of "subject" databases. The trend is to put computing directly in the hands of the people who need the data.

In looking at the hardware trends (logic circuits per chip, microcomputers and desk-tops installed, and networks), we see that organizations are providing the technology. A lot of data is getting processed. If we also look at the trend in the number of keyboards per white collar worker (figure 4), it reflects that by 1986, four out of five workers will have a keyboard or electronic device. Later predictions even suggest five out of five, by 1990. Even with all of this technology we still have problems locating the trouble in satisfying the users. Technology doesn't solve everything, it takes a blend of many things. Data Administration's role is to help improve the management approaches that are being taken to manage data in an organization. An effective Data Administration staff can help reduce the credibility gap between the MIS department and the end-users. This gap has been highlighted many times by different business executives with comments such as those identified in the chart on Symptoms of the Information Crisis (figure 5). Given all of this, a key role of Data Administration is to superimpose the technology with the data requirements in order to aid the users in getting the data they really need to perform their job.

The more sophisticated organizations today are organized similar to the Data Resource Management Structure (figure 6) where the Data Administration, Database Management, Data Processing Management, and Telecommunication Management are separate groups in the organization. Their roles are further identified in figures 7 and 8. Historically, the organizational structures have evolved from attempts to solve technical problems. However, as organizations have realized the importance of viewing data from a business viewpoint, this typical structure seems to be emerging. The Data Administration function must look at data from both a top-down business and a bottom-up logical view. It is responsible for stabilizing, organizing, and synthesizing the data. Managing data through its life cycle is a role of Data Administration from inception until it is archived. On the other hand, Database Management is responsible for the details of creating, controlling, and monitoring the physical databases. Figure 9 depicts an organization with distributed functional area Data Administration, which works very closely with the centralized Data Administration function. Sample job descriptions of Data Resource Management team members are also included in Appendix A.

The organization must have a strategy for information resource development (figure 10) which should be done by Data Administration. Data Administration must develop or aid in the development of a top-down direction through a Strategic Systems Planning process. The implementation is done from a bottom-up design strategy and audited to the top-down results through the use of a Logical Database Design process. The two strategies must merge in the middle. Some organizations make attempts at this approach but fail to meet in the middle. They fail to carry the top-down strategy far enough down or the bottom-up strategy to high enough of a level.

The Strategic Systems Planning process results in the formalized direction or business model for the organization. This may also be called the Data Architecture or Subject Database Architecture for the organization. The architecture will be organized around subject databases (see attached Fruit Salad Analogy, figure 11). Everyone wants his fruit salad containing a mixture of fruit to suit his taste; data users want reports containing a mixture of information. In the past, databases (files) were built to support a specific application. These application databases contained all information for the application and thus contained a mixture of data that really should not be mixed in the database when needed for other applications. It would be like having to eat from a mixed fruit bowl or many such bowls when you're allergic to certain fruit; it is hard picking out the fruit you want. Likewise, in an application, it is often times hard to get the data out that you want because of this mixed effect. The goal is to create subject-oriented databases, each containing similar data, and letting the application pull the data it needs from the different databases. It would be like going to a fruit supermarket, picking out the fruit you want for your salad, and ignoring the fruit that you are allergic to. Each person could do likewise selecting the data for the job at hand.

In developing the top-down Data Architecture, there are three stratifications to consider. Some organizations may only consider two. The first stratum is the set of subject databases needed to support base-line or operational functions. The second stratum is the databases needed to support upper level management, the decision support requirements. The data requirements generally need to be more flexible to provide for management experimentation. Most of the time, the work may be done off-line outside the database environment. However, when the experimentation is done, the final results must be captured in subject databases within the first stratum. The third stratum may be required to support the middle level management needs. In all cases, the data architecture must identify all the entities required for each subject database. With this architecture, there is an integrated systems structure from which to start building the systems needed to support the organization's data needs. In other words, there is an integrated structure instead of a patchwork of systems (figure 12). The information systems and the data architectures must fit together. From this architecture, projects can be identified and described which show the relationship of information systems and subject databases needed to support the business. The priority of projects for implementing information systems and subject databases can be determined based on the time that they are needed and their precedence or dependency on other projects (figure 13). The time and precedence of data for each project must be reviewed. A project that implements an information system which creates or updates a database obviously would have to be implemented before other systems that reference the database. Who needs the data and when it is needed are important factors affecting the schedule of projects. From all of this, Data Administration and management can see the full scope of effort required by the organization.

With the Data Architecture defined through a Strategic Systems Planning process, the implementation of the systems to support the end-users is done in a bottom-up strategy through the Logical Database Design process. This bottom-up process is also the responsibility of Data Administration. The primary goals of Logical Database Design are to:

Provide shareable and available information which can be used by new systems as they are built; in other words, provide data that is standardized, accurate and consistent, and can be used for multiple purposes and functions.

Create a stable and an expandable environment which will increase productivity, allow for growth, and reduce the cost of handling data.

Maximize data independence which decouples the applications as much as possible from the data. The data must be called or referenced by its name rather that by its structure or how it is stored. This will minimize the changes that need to be made as new systems and subject databases are implemented or as business requirements change.

One test for an organization to see if IRM is working is to check for an inventory of entities in that organization. The inventory should include a list of entities by name showing unique identifiers and definitions, and the entities should be consistent across the entire organization. If an inventory doesn't exist, then the organization is not doing adequate planning and should do something about getting there. The Data Administration role is to ensure that the inventory of entities is created. This is a function performed during the Logical Database Design process. More advance systems that we will be seeing in the future, such as those using artificial intelligence and expert systems, will require this inventory in order to function.

There are four steps or phases required to perform a Logical Database Design (figure 14) and to build up the Data Dictionary containing the inventory of data entities. The first step is to identify and define data elements from existing applications and databases, end-user interviews, and required outputs. The second step is to develop and review user views, essentially identifying the data elements required for specific business and transactional views. The third step is to generate a logical data model which groups data elements required for the entities. The fourth and final step is to reconcile the differences between the logical data model and the subject databases. This last step is where the two strategies, top-down and bottom-up, come together at the entity level.

When the Strategic Systems Planning (figure 15) and the Logical Database Design (figure 16) processes are done, then the Management and Data Administration will have a clear picture of what the business does, the data that it needs, and the data that it currently has. They can then identify the projects needed to implement the strategies (figure 17), and finally develop a hierarchy of users (figure 18) within the organization for approving and implementing application systems. The whole process is a migration process under the direction of management but supervised by Data Administration. So the total data environment (figure 19) is a very important cornerstone of the IRM environment.

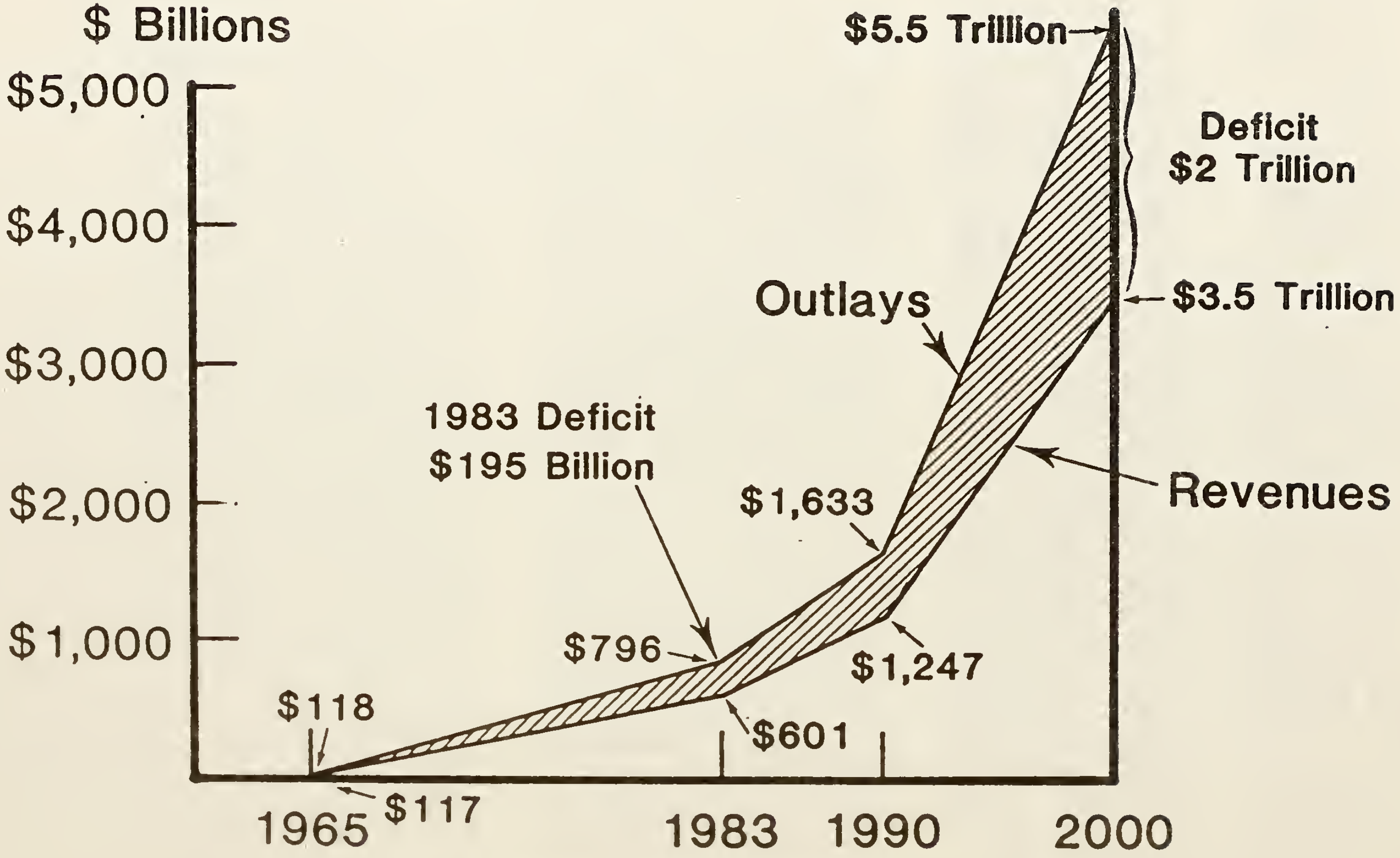
BIOGRAPHICAL SKETCH

Dr. Holland is best known for his pragmatic and academically sound systems design methodologies. Through his work with Fortune 500 companies, he has established proven approaches to data resource management, database design, distributed processing, and systems and data planning.

As president and chairman of Holland Systems Corporation, he directs all project resources and development methods. His other responsibilities include participating in the research, design, and development of Holland Systems' family of proprietary software products for automating application designs.

Dr. Holland is a widely published author, an international lecturer, and a regular contributor to industry periodicals.

Federal Revenues, Outlays and Deficits



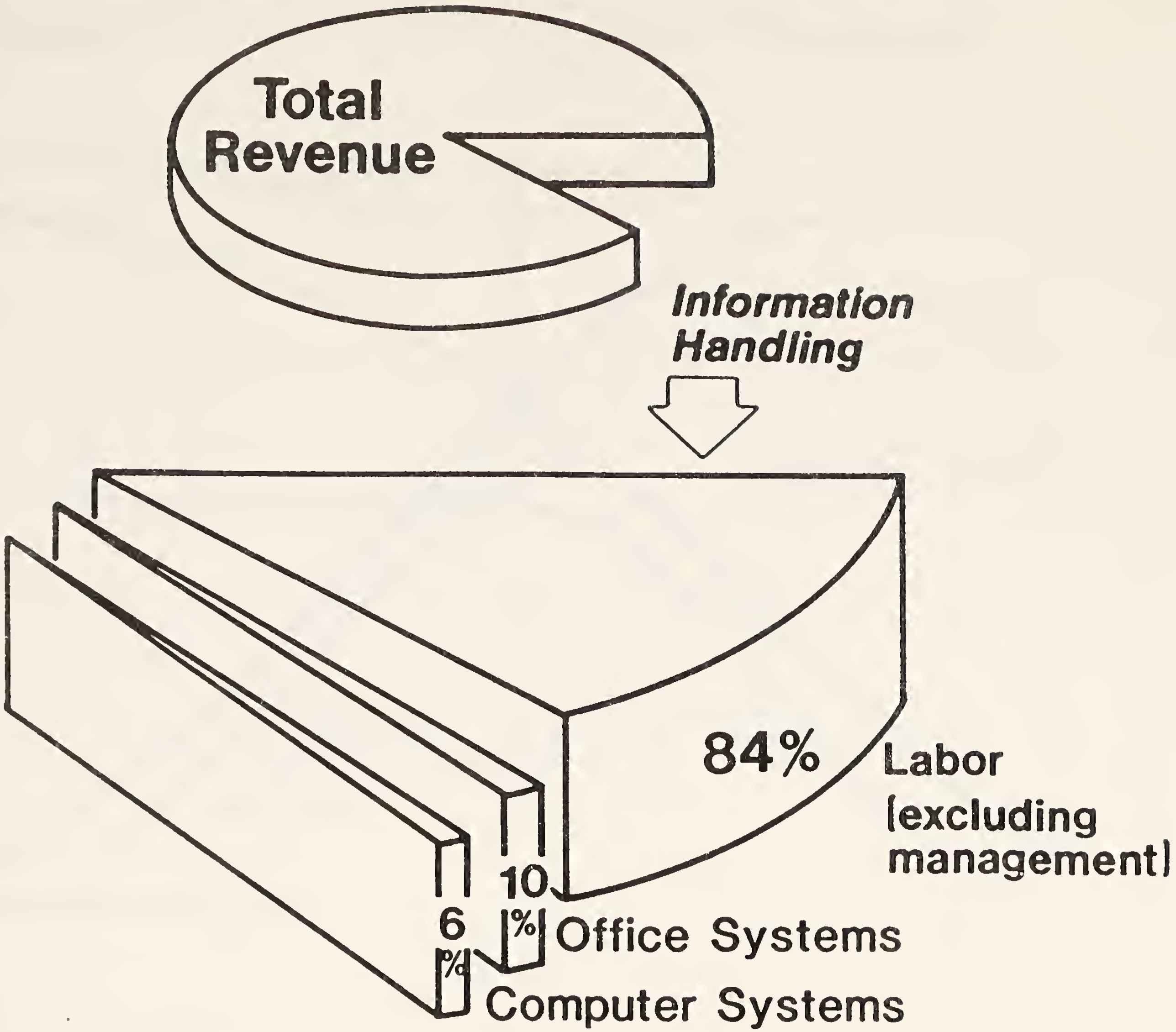
Source: Grace Commission

Figure 1

T1544

Reprinted with permission of Holland Systems Corporation

Information Handling Cost Breakdown



T0416 R1

Figure 2

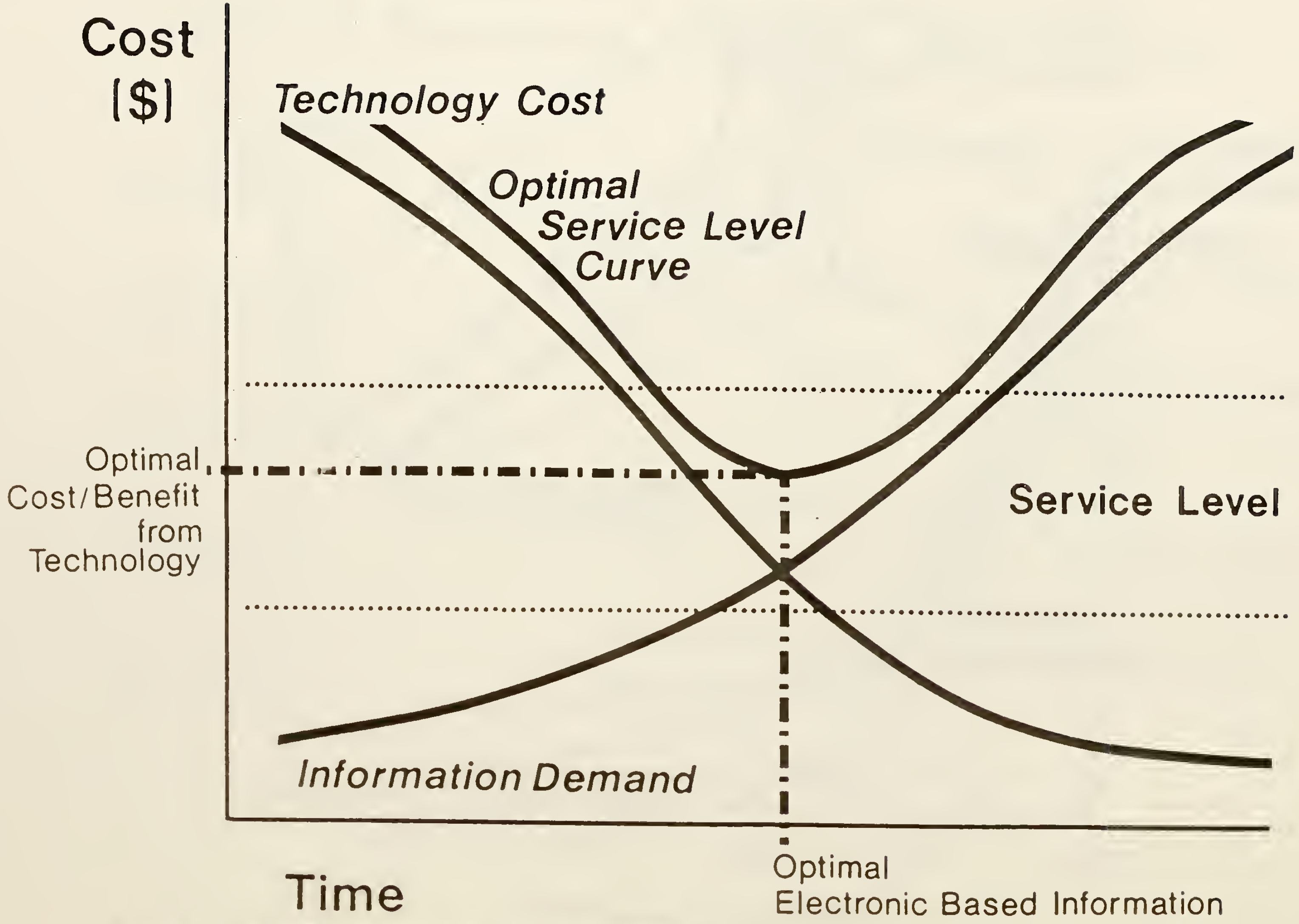
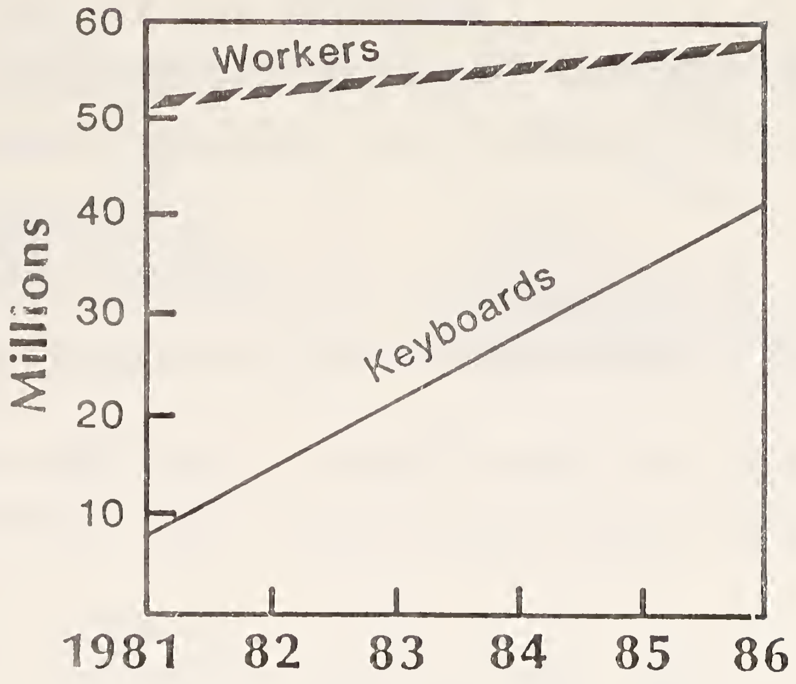

Cost
[$]
Technology Cost
Optimal
Service Level
Curve
Optimal
Cost/Benefit
from
Technology
Service Level
Information Demand
Time
Optimal
Electronic Based Information
T0416

Figure 3

Keyboards per White Collar Workers

1981:
1 in 5
Workers



1986:
4 in 5
Workers

T1007 R1 ©1984 Holland Systems Corporation

Figure 4

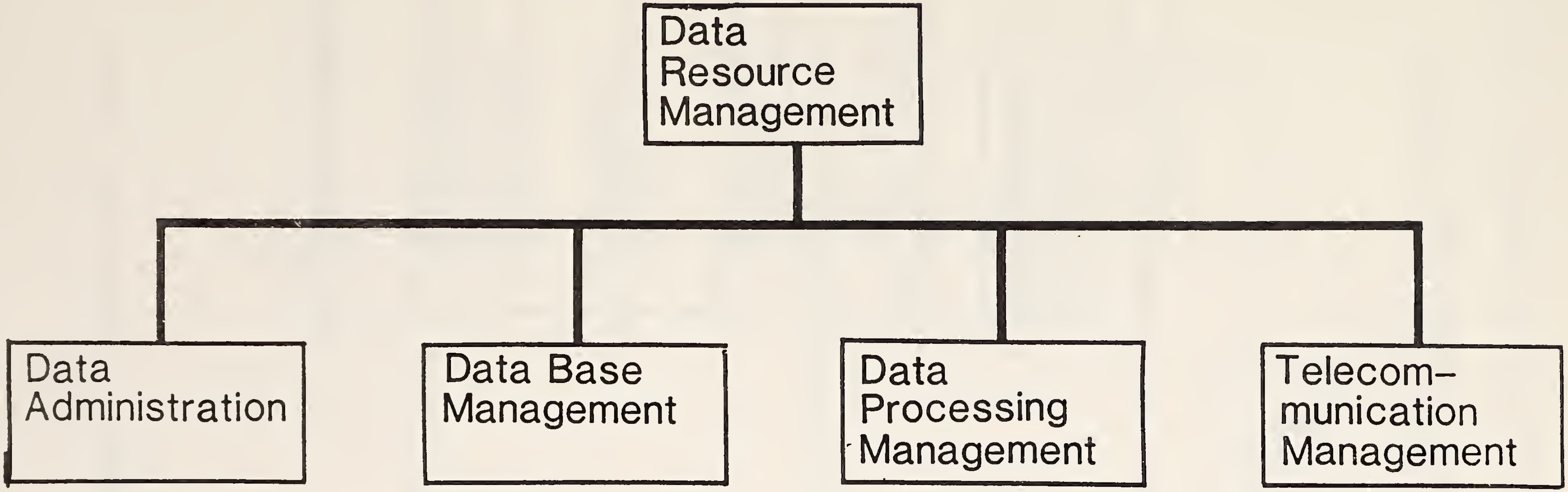
Symptoms of the Information Crisis

- *"Reports that cross my desk from different areas of the organization conflict in their data values when they should be consistent."*
- *"I need only a small portion of the data I now receive-the rest is useless."*
- *" In my company, the same data gets re-keyed several times, wasting time, money, and personnel resources."*
- *"I've just learned that the hardware and software we've been buying is totally incompatible across the organization."*
- *"The automated systems my people want and need take too long to put into operation, and the backlog of requests just keeps growing."*
- *"The dollars going into my computing budget are not paying off as they should."*

T1029 R2

Figure 5

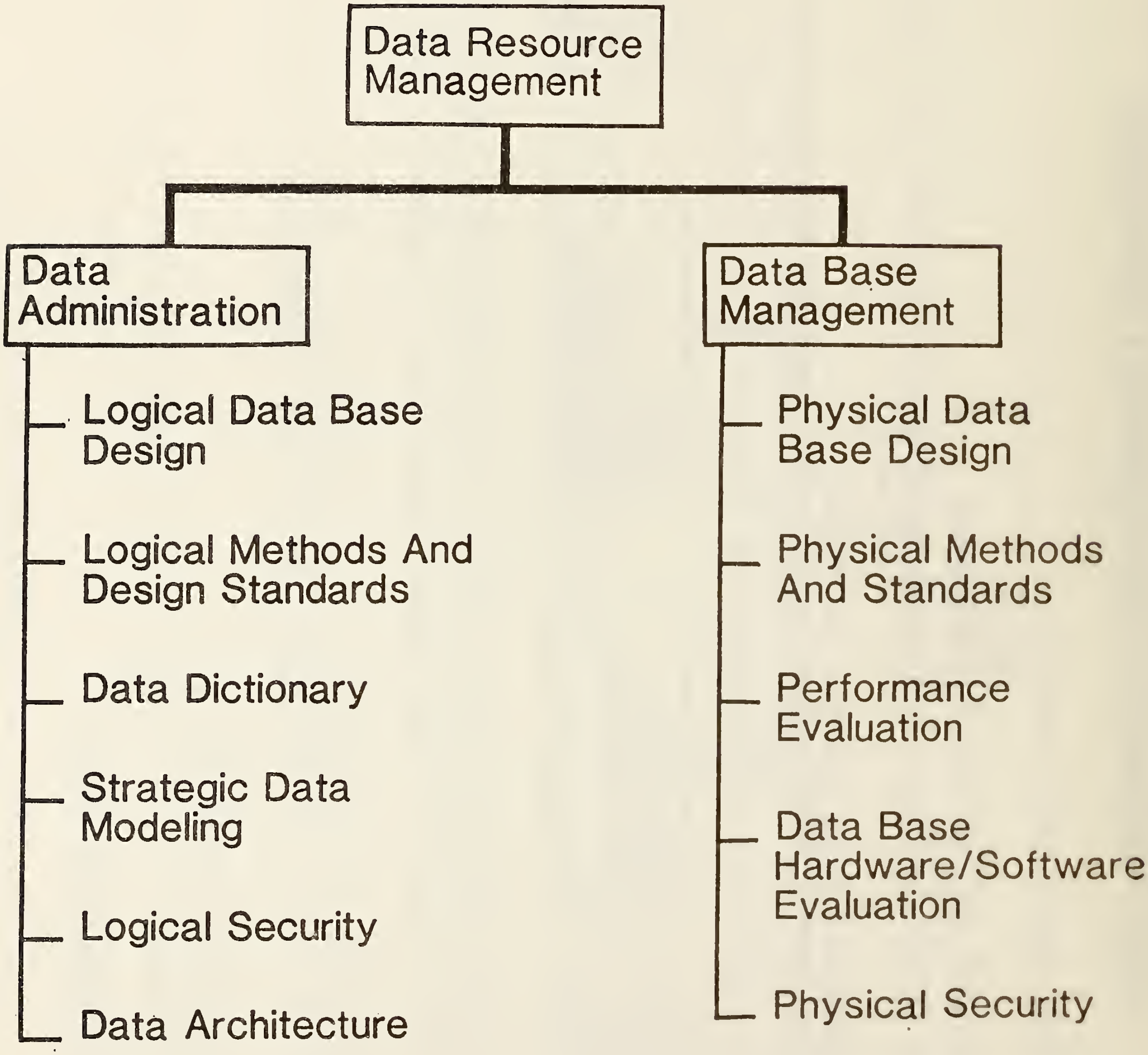
Data Resource Management Structure



T0162

Figure 6

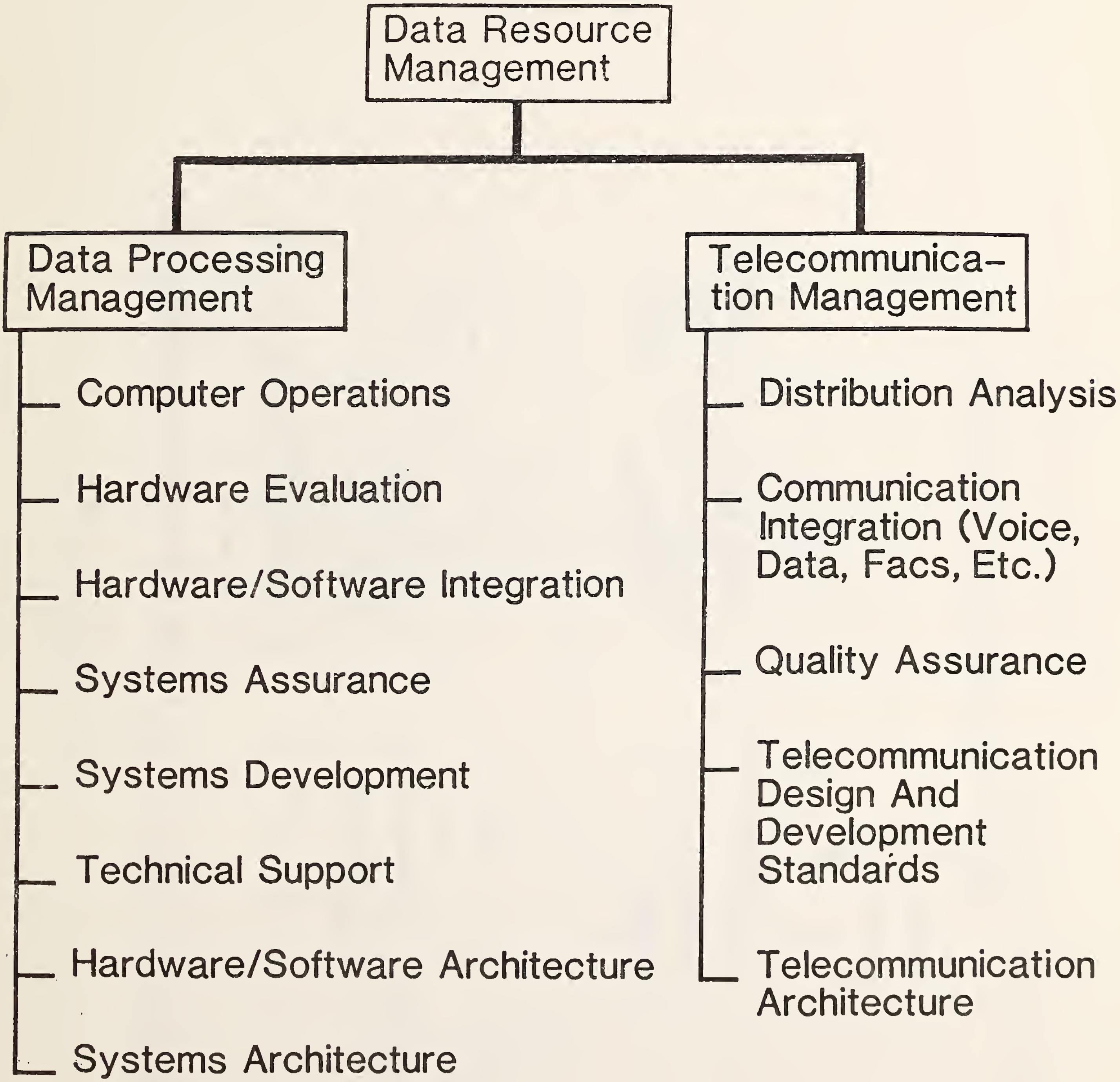
Data Resource Management Structure



T0163

Figure 7

Data Resource Management Structure



T0164

Figure 8

Data Management Structure

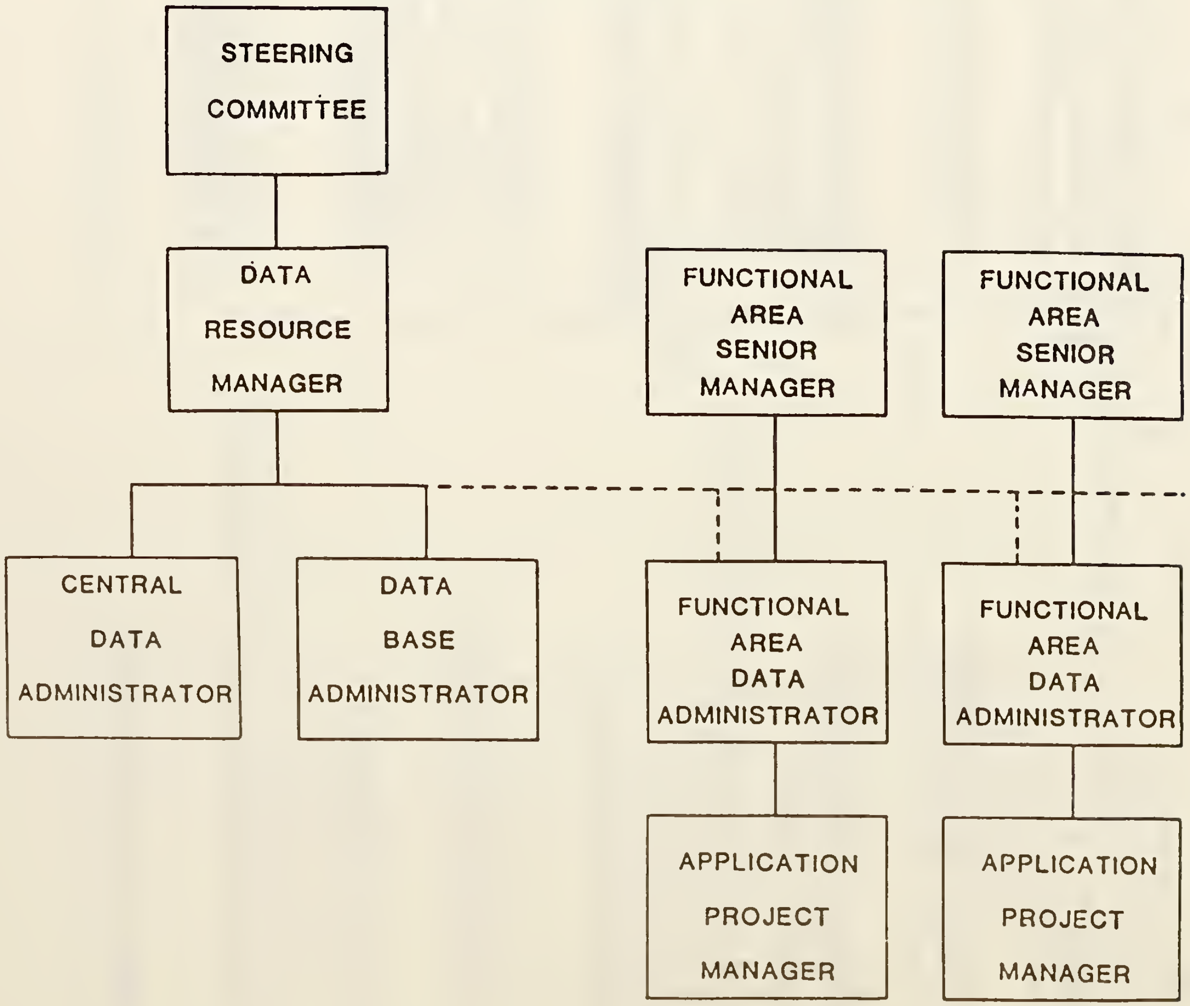
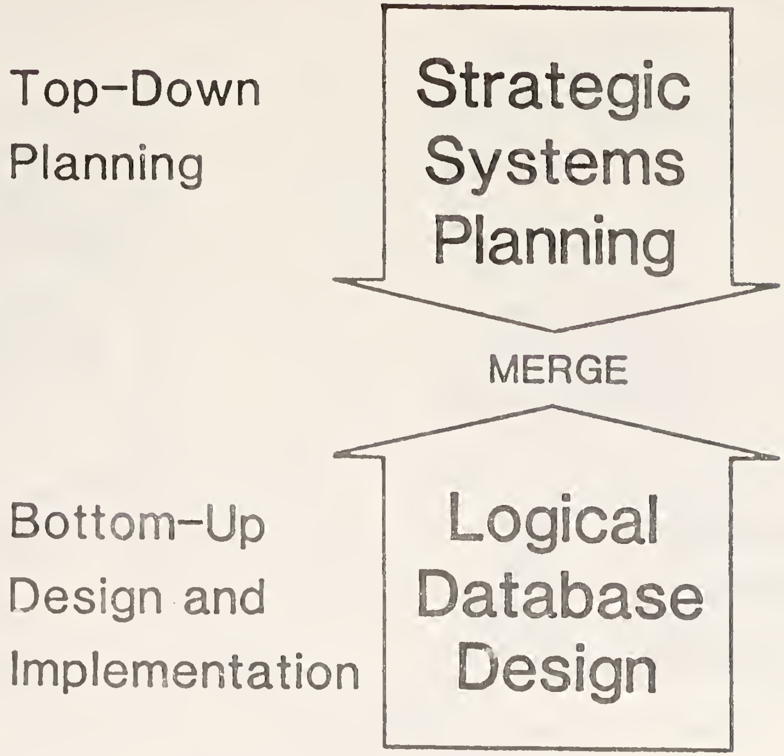


Figure 9

T0499

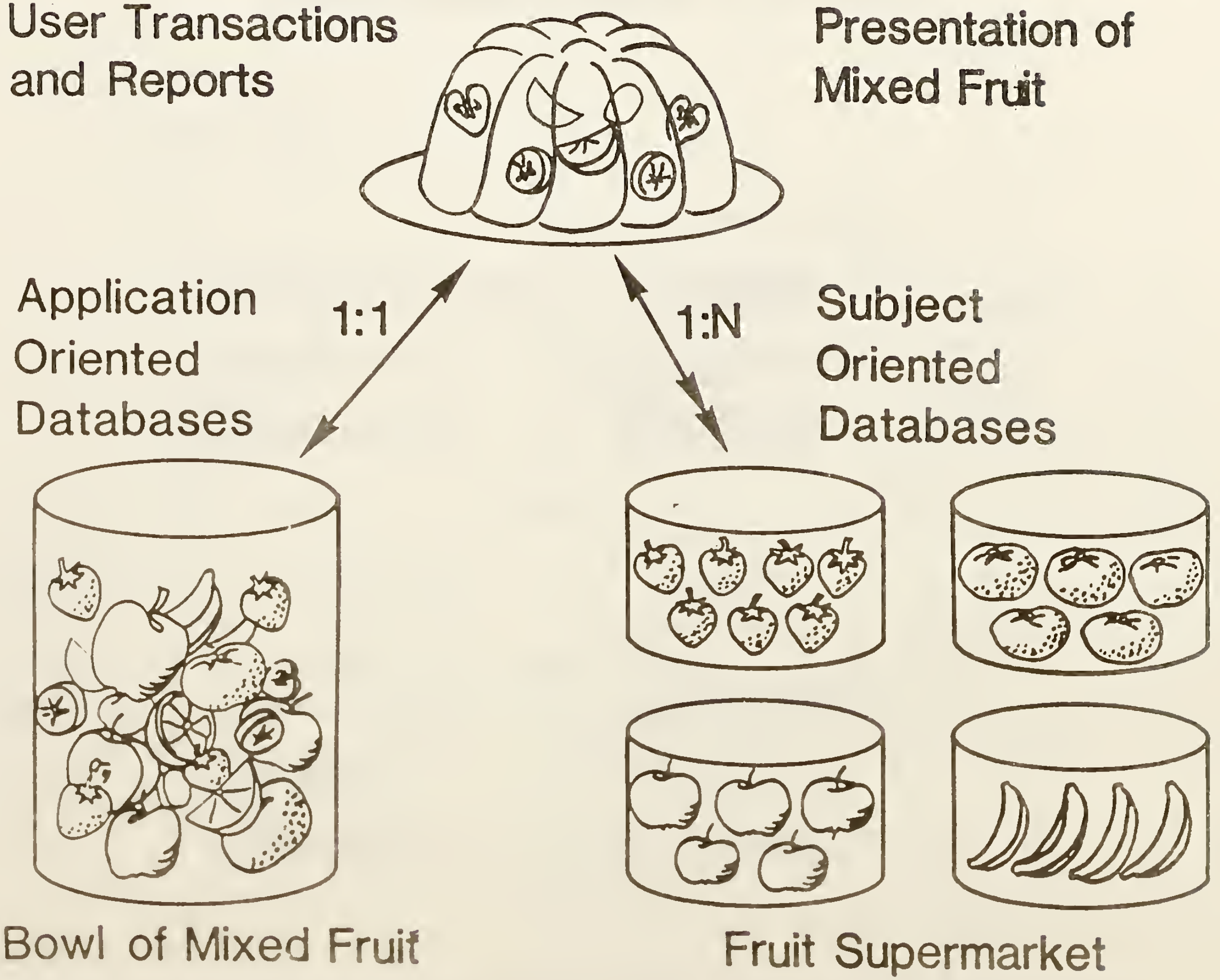
Strategy for
Information Resource Development



T0703

Figure 10

Subject Databases: The Fruit Salad Analogy

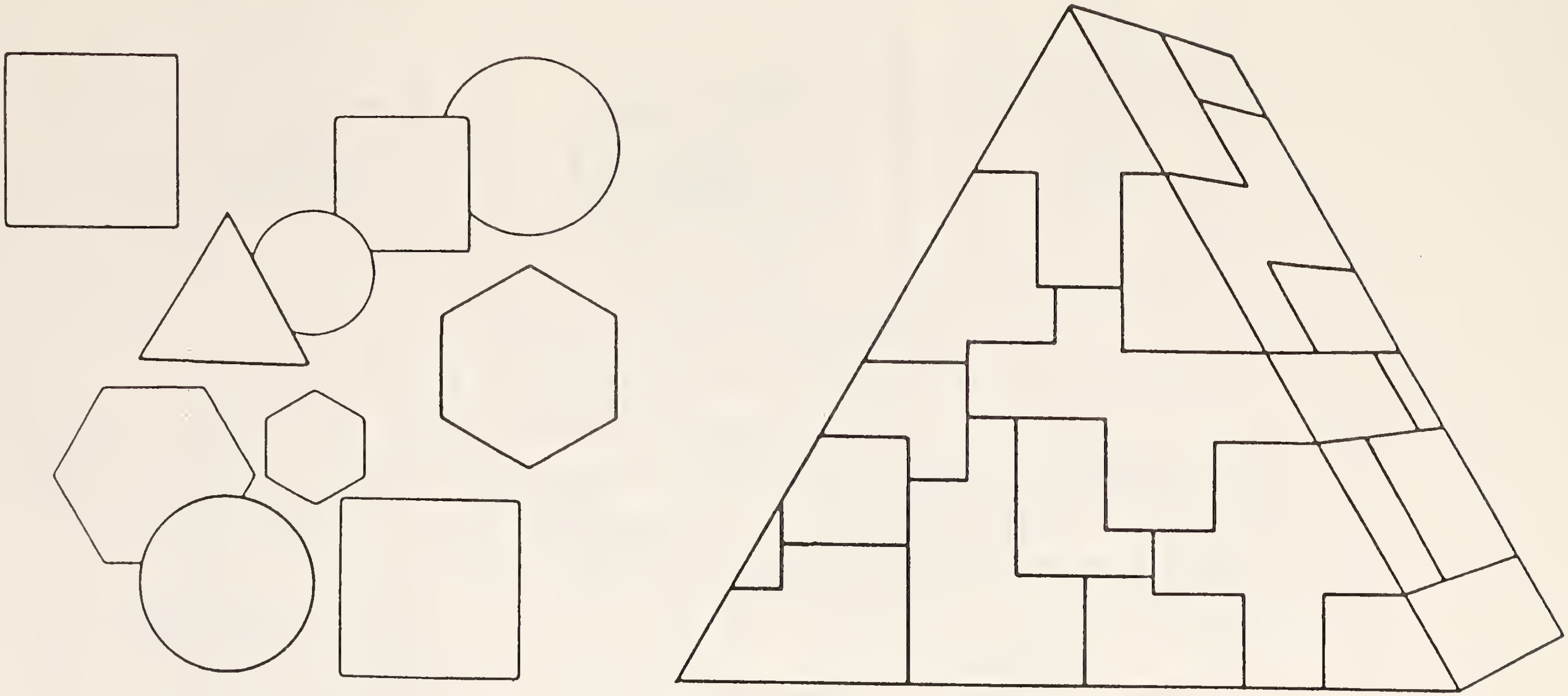


T0953

Figure 11

SYSTEM STRUCTURE

Bottom-Up | **Top-Down**



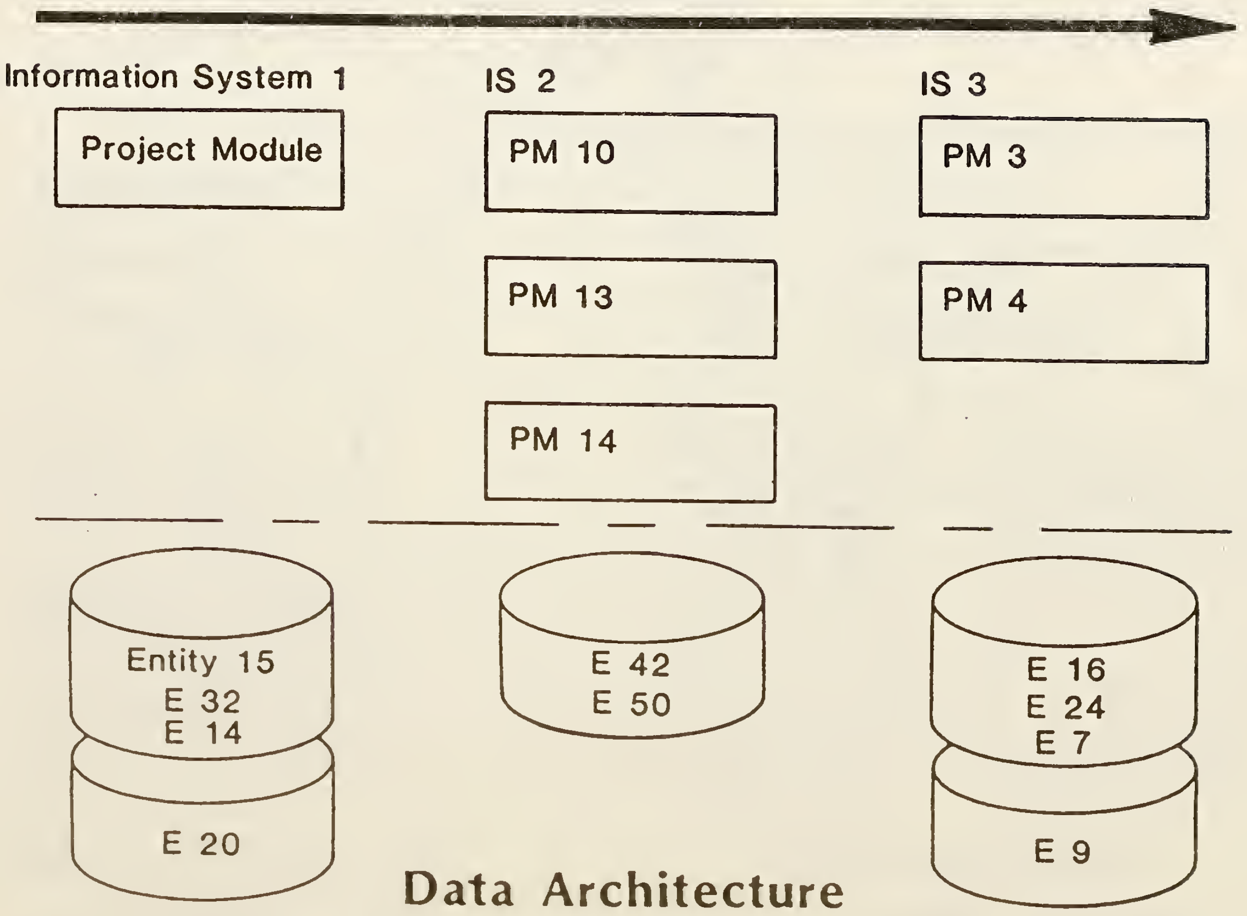
Patchwork | Integrated Systems

T1359 ©1984 Holland Systems Corporation

Figure 12

Time and Precedence Sequence

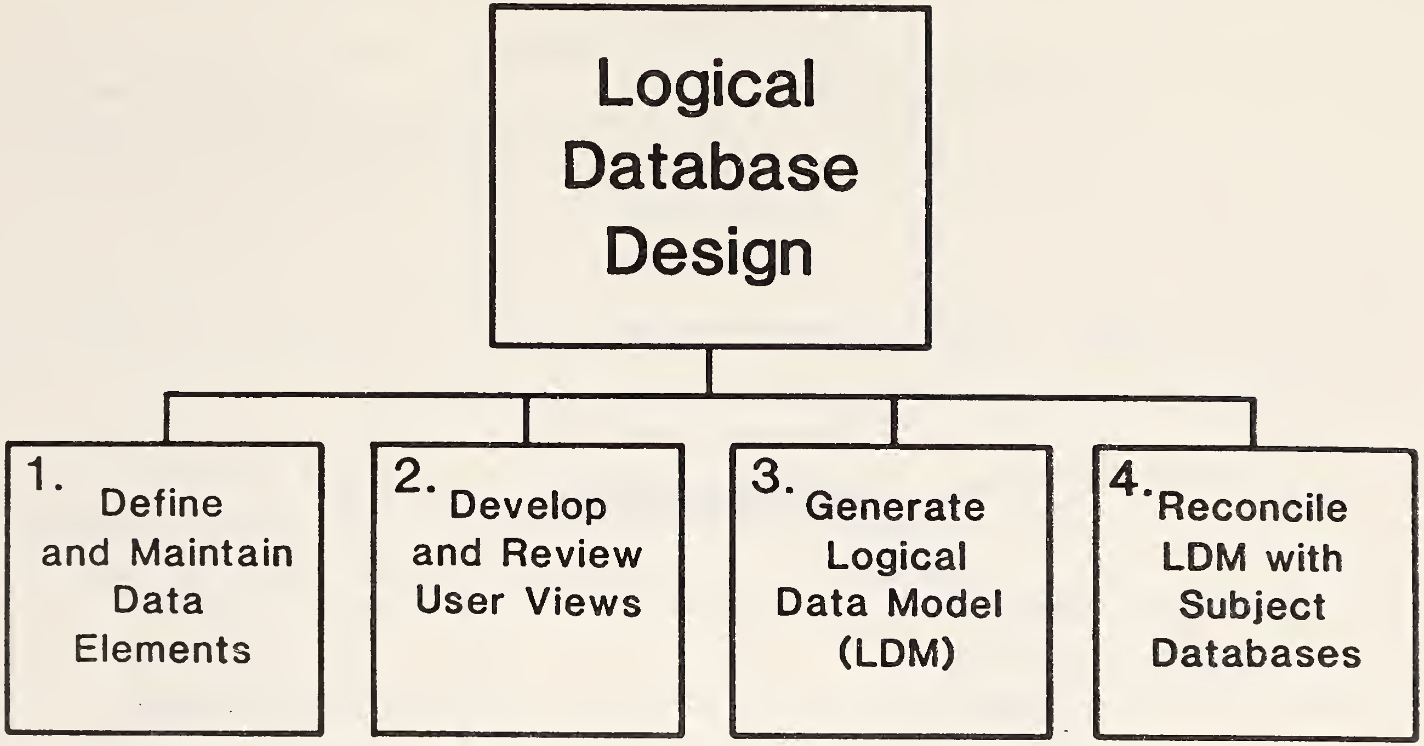
Information Systems Architecture



T0048

Figure 13

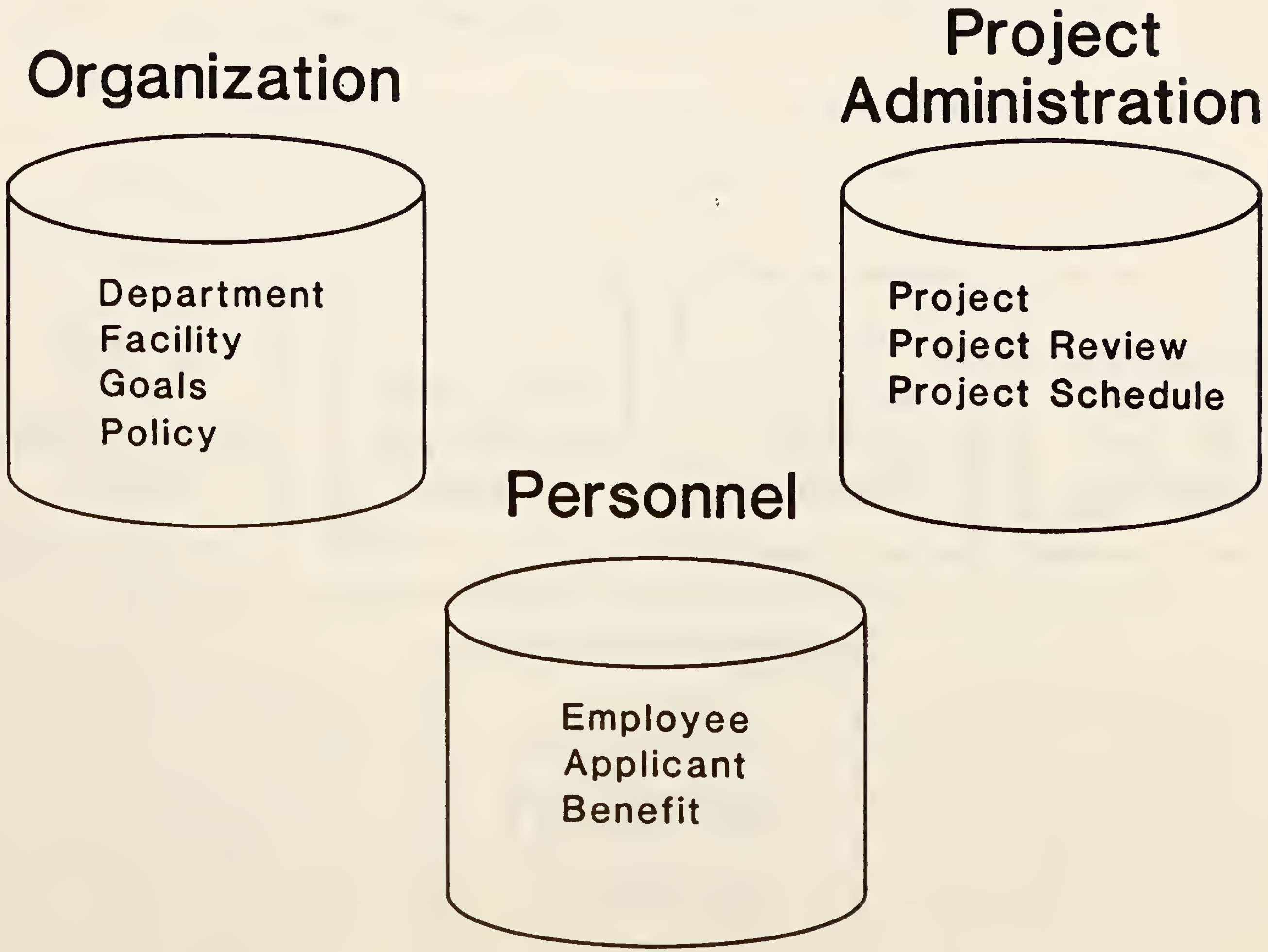
Logical Database Design Phases



T1123

Figure 14

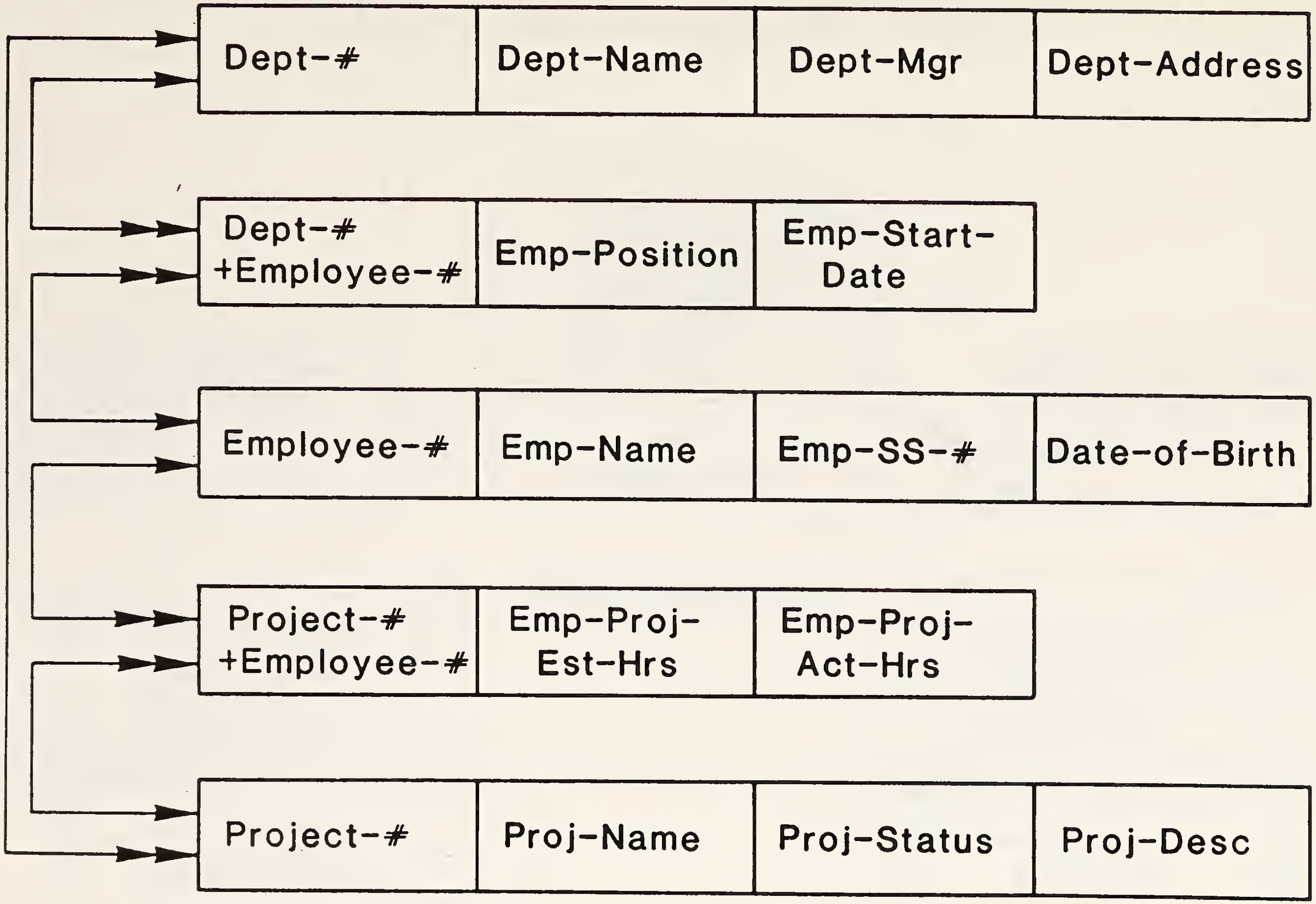
Top-Down Architecture



T1273

Figure 15

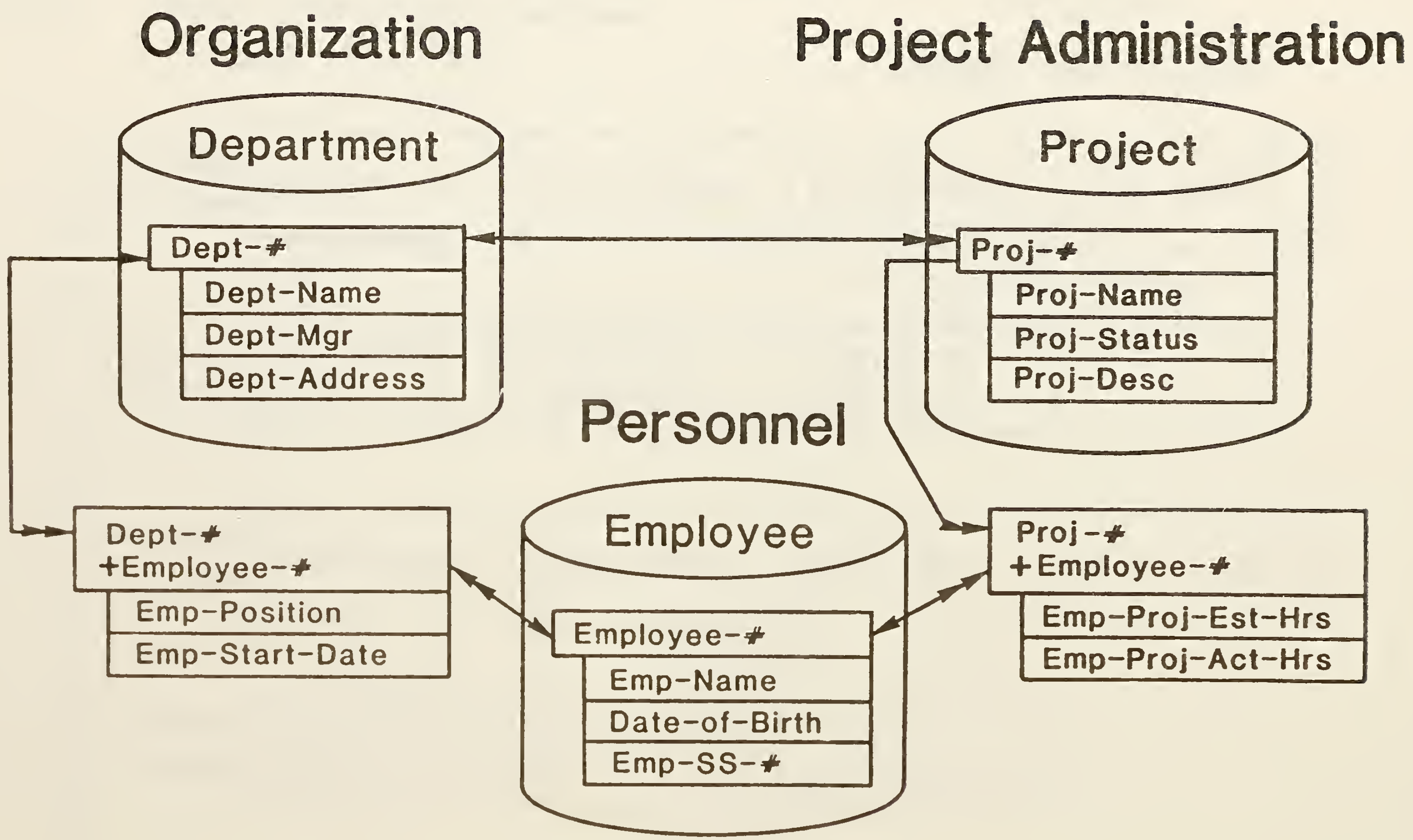
Logical Data Model



T1267 R1

Figure 16

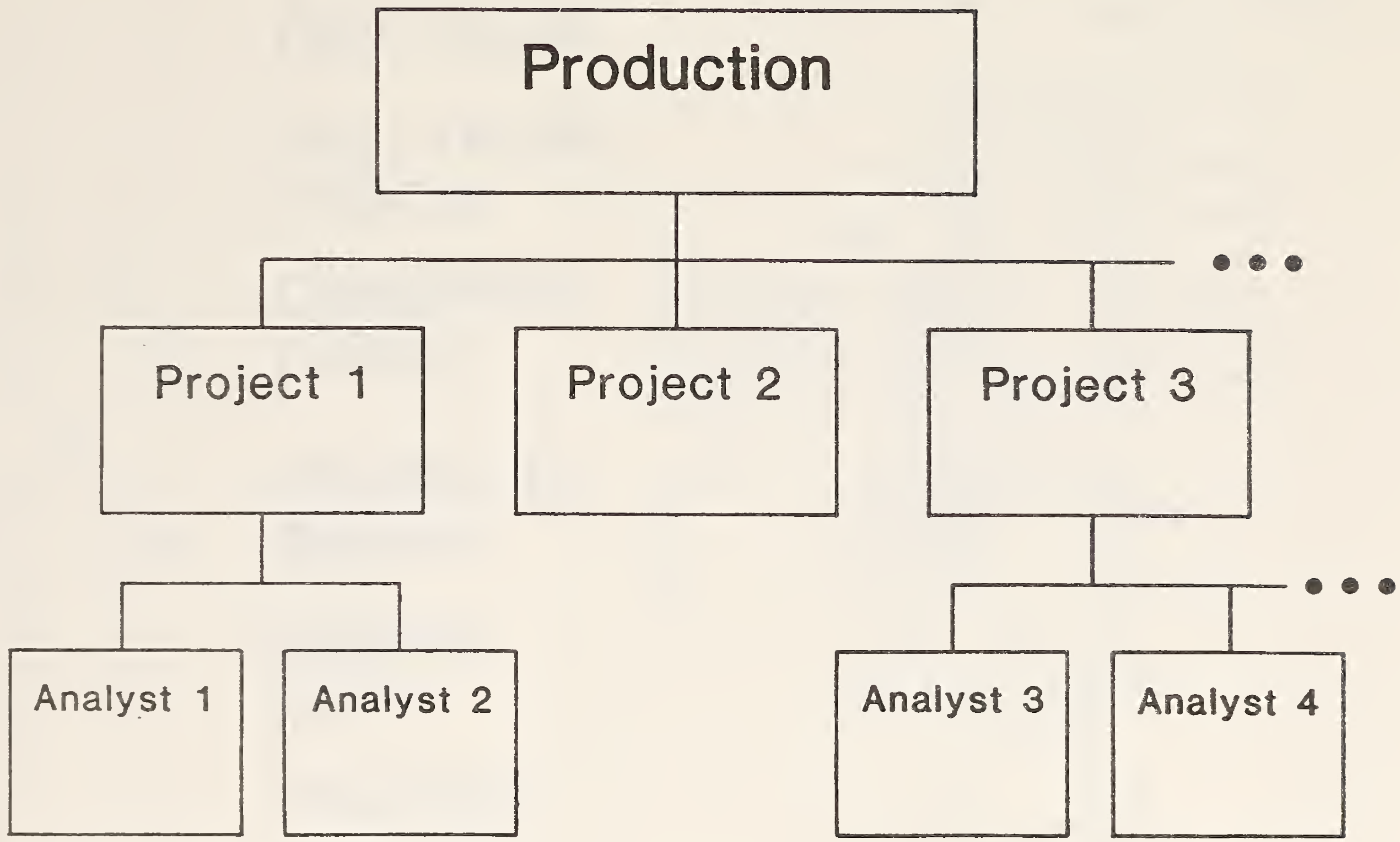
Data Groups Related to Entities



T1274 R2 ©1984 Holland Systems Corporation

Figure 17

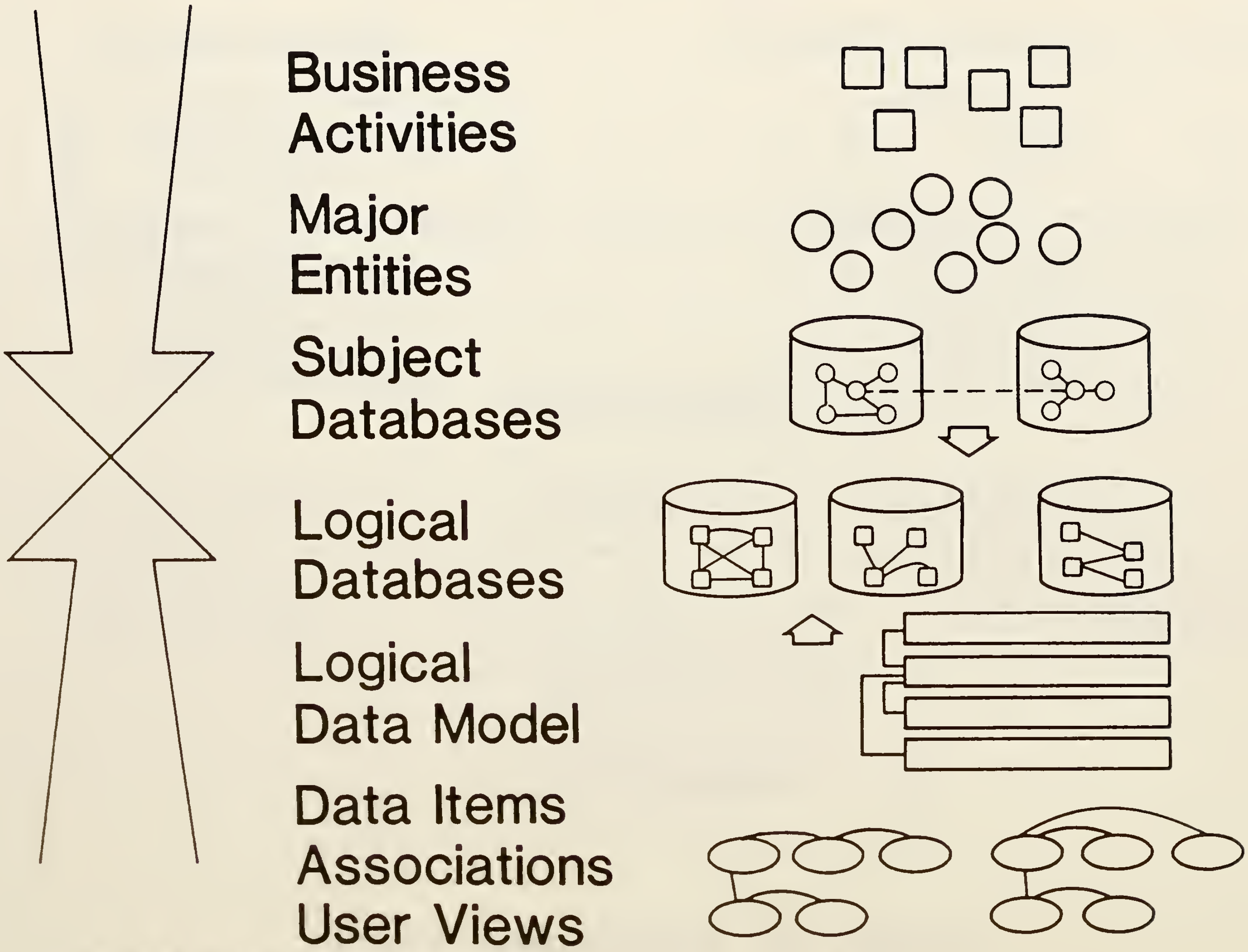
LDD Hierarchy of Users



T1217 R1

Figure 18

Top-Down, Bottom-Up Design Approaches



T0153

Figure 19

MANAGEMENT PERSPECTIVES OF DATA

Moderator

Daniel Schneider
Department of Justice
Washington, D.C.

Recorder

Josephine L. Walkowicz

EXTERNAL DATA AS A MANAGEMENT ASSET

Speaker

James P. McGinty
The Dun & Bradstreet Corporation
Washington, D.C.

ABSTRACT

A discussion of the importance to management of external data, and the opportunities, both personal and organizational, offered by this management phenomenon. External data is defined as those data sources that exist outside an agency, which when properly defined, structured, and transmitted, become a management asset by enhancing decision-making within the agency. The importance of external data is discussed, together with the personal and organizational opportunities external data affords in supplementing, complementing, and enhancing data internal to an organization. The management of external data is described as a five-step process within an organization. Functions associated with each step are identified, and placement of these functions within the organization is suggested.

My objective today is twofold: to look at your role in data administration from a different viewpoint, a viewpoint that I will call external data, and to look at opportunities, both personal and organizational, that external data may provide.

When I talk about external data, I refer to those data sources that exist outside of your agency, which when properly defined, structured, and transmitted, become a management asset by enhancing decision-making within your agency. Much of the decision-making processes by the senior managements in your agencies have to do with information that is outside of your agency. Today I want to talk about your responsibilities and opportunities in addressing that peculiar management phenomenon.

Some examples of external information in use by senior agency officials include: industry data, business data, economic data, demographic data, environmental data, financial data, and legal data. External data drives the decision processes of your agencies. There is hardly an agency here that does not have piped into it -- sometimes on-line, sometimes just in manual format -- some type of external information that drives the decision process in your agency. As a matter of fact, this is a wonderful footnote to Dr. Holland's talk because, if you look into the architecture of your information systems, rarely is there any node that talks about external data. Yet, external data is some of the most valuable data that an organization can have. Let's take a look at a few issues where external data is crucial. You are all familiar with the government competition and contracting processes in the procurement of ADP equipment. You have to know what EDP capabilities exist outside your own agency. The defense industrial base issue, for example, is one that the Pentagon has to understand before a new project is started. In order to understand the defense industrial base, the Pentagon has to have information on the U.S. industrial base, and this information exists outside the Pentagon.

The impact on tax policy is another issue that is debated constantly on Capitol Hill; again, this represents a requirement for external data. The impact on regulatory policy -- and many of you are from regulatory agencies -- is significant. In the banking regulatory process (FDIC, the Federal Reserve, and the Comptroller of the Currency), there is another tremendous appetite for external data as there is in the Securities and Exchange Commission, the Small Business Administration, and the Department of Transportation. It goes on and on and on. Essentially, every agency has a need for external data.

Now, why is external data so important? When you look at it from a data administration standpoint, there are three essential reasons for the importance of external data.

First, external data can provide a universe against which an agency can measure the coverage of its own data. The Securities and Exchange Commission (SEC) regulates 12,000-13,000 companies. There are five million companies in the United States which provide the environment for those 12,000 companies. The SEC provides the classic example of a need for somebody to do their job in terms of a regulatory environment. They need data not only on the agencies they regulate -- this is easy for the SEC because they can force the regulated companies to supply data. However, the SEC needs information on the external environments of the companies they regulate. This information comes from external sources.

The second reason is that external data can be used as a substitute for the development and maintenance of an internal database. The Small Business Administration (SBA), for example, wisely made this decision when it needed a database of all small businesses in the United States. This is a universe that comprises between 5 to 7 million business establishments, depending on the definition. To satisfy this requirement, SBA leased two data bases from private sector companies. This was extremely cost effective. Another agency, which I shall not mention, tried to create a similar data base and it cost five to seven times the amount of money spent by the SBA to lease their databases.

Finally, an external database can be used to enhance internal data. Why get into the coding business yourself when a database exists outside your organization? Later on in this workshop we will hear a presentation on computer matching of names. This is a perfect application of the use of external data. It does not involve matching names simply for the purpose of matching names. Names are matched so that data elements may be transferred. Frequently, external data elements exist outside your agencies, and it is cheaper to tap them rather than develop and maintain them yourself.

The above are three good reasons why, from a data administration standpoint, you want to pay attention to external data. However, it is not all good news. This is because of certain characteristics of external data. Generally external data is

- difficult to define;
- difficult to understand, principally because it is not always within your world;
- usually unstructured;
- very costly to create;
- and, like any database, can be costly to maintain.

Much of this, however, is changing right now. The world of external data has changed; something has been happening in the last ten years. Unstructured data and the associated cost of creation and maintenance are really being taken on. Structure is being added to external data by private sector organizations and within government, by NTIS, as an example. People are trying to get their hands around this external data world. They are providing systems whereby you can access and search external data, and, in general, they are operating on a cost curve which decreases the cost of external data. And that's happening in both the public and private sectors. When I joined the Information Industries Association fifteen years ago, there were eleven companies. Today there are 380 companies principally involved in creating information and disseminating information for use by others. It is a multi-billion dollar industry.

Here is what really happened in the environment of external data. First, it was hard to copy books and files. That's all senior management had to look at when they wanted to address a problem that required external information. Then we went to an era of machine-readable files and databases in the 1960s. In the 1970s, we got syndicated databases where one company, like Dun & Bradstreet, for example, would create a database and then syndicate or lease it out to various other organizations. Then in the late 1970s, on-line syndicated database services began to come about. An example is Lockheed's Dialog -- many of you are users of that tremendous service. BRS is another one -- there are several on-line databases. What is going to happen in the late 1980s? My guess is that it will be the compact disk -- the CDROM environment will really play a big role. We will be delivering databases to you on 5-1/4" disks with 550 megabytes of storage capacity. You will read and process this data on your own microprocessor.

That's what has been happening in the world of external data. Millions of published sources are still out there, thousands of machine-readable files. But look at this, 2000 on-line databases from almost 1100 different database publishers. A phenomenal occurrence in the last few years.

What does all of this mean to you in terms of a data administration function, a MIS function, an IRM function, or whatever function you are into? Somebody had better be addressing questions such as: Who is managing the use of external data? Who has the responsibility for identifying external data requirements? Who is responsible for integration of internal external data requirements? These are tough questions and in an IRM environment they are really telling questions. As a matter of fact, they are questions that the chief executive of the agency really has to ask and have answered. If the chief executive is not, then he or she is simply not doing his/her job in the 1980s.

The management of external data is a five-step process: (1) someone has to define the areas of interest; (2) someone has to identify available external databases; (3) someone has to formulate reporting policy; (4) someone has to execute some form of integration plan; and (5) someone has to manage vendor/source relationships. Don't forget, this data is coming in from outside your agency. It may have to come from a private source. Wherever it comes from you have to maintain a relationship with that source.

There are individual problems associated with each step in the above process. In defining the areas of the interest, again the chief executive must really articulate the mission and objectives, in other words, the business side of the agency. In the case of the Department of Defense acquisition process, for example, somebody at the top level must say, "Hey, MIS guys, I want to know something about the defense industrial base." That's how it happens. And then a whole bunch of people have to go scurrying around. The guidance is there. In my company, the boss says "I want to know what our market share is," and everyone jumps around to figure out what that market share is. Guidance and direction must come from the top; the responsibility is clearly that of the senior executive.

Functional users, people under them -- business people with whom the data administrator interfaces -- need to be involved. I like the term MIS here. It's probably a data adminstration function or a business function. The problem is getting people to do the work; it is as simple as that.

In the area of identifying available external information, once again, if you know the direction of your agency and you know what they are doing, you know that agency direction really does not change much from administration to administration. If you know that, someone has to identify all of the published sources. That's generally done for you. But the databases out there, both numeric and bibliographic, are proliferating at an unbelievable rate. Someone in data administration, in IRM, or in MIS has to have a handle on them. Again, if that is not happening in your agency, there is a real problem. The responsibility is that of the functional representative. If I am in the marketing department in my company, I have responsibility for external databases related to marketing. So, right down at the agency level, various functional as well as MIS people ought to be involved in this. This is really strategic top-down planning. Here is a very good use for a consultant because it is difficult for an organization to know what's out there externally. I know of one big accounting firm, for example, that has created an entire practice around external information and its use by organizations in an information system environment. That tells you something about the trend. Again, the problem here is getting someone to tell you what the data means. The data file may look good, the tape description may look good, but what does the data really mean?

Formulation of reporting policy is a traditional concept. This involves identifying things which should be done and specifying how they should be done. We have to determine how the data will be used, by whom, and the decisions it will support. Furthermore, the editing, processing, and all functions that are within the province of data administration must be specified. This responsibility is of an MIS or IRM type; the functional representative or the user also has to be involved. As you all know, the problem is, of course, finding people to do this work and maintaining user interest because users tend to disappear or lose interest when they find out the difficulty and intensity of the kind of effort.

My personal philosophy regarding the integration plan and its execution is prototype, prototype, and prototype. Or model, if we can use these two terms synonymously. When you model the environment, you have to involve the data source or sources, the user, the functional representative, and the MIS or IRM person. The latter is really the key. Finding the people, finding the money, and freezing the specifications are usually the problems in this effort.

Finally, maintaining vendor/source relationships is crucial for external data. You cannot go out to buy, lease, or use another agency's data and then walk away from it. You must maintain some form of relationship. I say that's the province of the functional user. It is not necessarily the province of the MIS or data administration functions, although it may be shared. I feel quite strongly about this, that it's the users of the data - the business application, if you will -- they are the people that should maintain the vendor/source relationship. Maintaining the interest of the user is sometimes tough, but that's part of the ball game. If the "user" is the one making decisions based on external data then the "user" must be involved with the data source.

That's a way of looking at the process and doing something with external data. Where do you fit in? Again, the audience is a very broad one and you could take a couple of positions at opposite ends of the pole. You could be someone who just sits back and waits until someone comes and requests data or you could assume total responsibility for leading the process of using external data in your organization. That's why I believe that there is a real career opportunity for people in the data administration area.

Final point, food for thought -- The President does a nice job, he appoints a lot of people to manage agencies; but I like to think that it's you folks who provide the information who make those appointees managing executives. If you keep that in mind you will understand why your work is so worthwhile and something which is vital to the management of our government.

Best of luck in your effort and thank you for your service to our country.

BIOGRAPHICAL SKETCH

As Vice President, Group Government Marketing, Mr. McGinty has corporate-wide responsibility for directing Dun & Bradstreet's marketing efforts in the Federal sector. He is also responsible for research and development of computer-based systems to serve the Federal sector. Prior to his current assignment, Mr. McGinty was Director of Corporate Government Services.

Before his assignment to Washington, Mr. McGinty was Vice President of the Marketing Services Division of Dun & Bradstreet. This Division provides computer-based systems and information services to the sales, marketing, and planning functions. During his eight years with the Marketing Services Division, Mr. McGinty held the following positions: Vice President, Market Management; Vice President, National Accounts; Assistant National Sales Manager; Product Manager of Computer Services; Divisional Sales Manager; and District Sales Manager.

Throughout his 15-year career with Dun & Bradstreet, Mr. McGinty has been active in the Information Industry Association (IIA). His service to the IIA includes membership in the Future Technology Innovation Council, as well as several subcommittee assignments relating to the industry's relationship with government. In November 1984, Mr. McGinty was named to the Board of Directors of the IIA.

ACCESSING NATURAL RESOURCES DATA

Speaker

Ralph J. McCracken
U.S. Department of Agriculture
Washington, D.C.

ABSTRACT

The Soil Conservation Service (SCS) of the U.S. Department of Agriculture is required by law to make periodic appraisals of the status, condition, and trends in soil, water, and related resources for use at local, state, and national levels in setting conservation policies and priorities. Each such appraisal is designated as a National Resource Inventory (NRI). The presentation provides detailed descriptions of two of the largest of several natural resource databases that are maintained by SCS. These two include the most recent (1982) in the series of National Resource Inventories and the SOILS-5 database. The complexity and size the SCS databases, coupled with the multiplicity of various data providers as well as users and the need for data sharing, are only a few of the factors that challenge management with a variety of problems and opportunities.

The Soil Conservation Service (SCS) was established in the U.S. Department of Agriculture in 1935 because of mounting concerns about wind and water erosion and perceived needs for conservation of soil and water resources. The Soil Conservation Act of 1935 (Public Law 74-46) authorized the Secretary of Agriculture, among other conservation activities, "to conduct surveys, investigations, and research relating to the character of soil erosion." This mandate has been used as the basis for collecting data on soil and water resources, for conducting soil surveys, and for maintaining databases on the properties and uses of soils. These activities and the continuing conservation concerns have resulted in the development of two large national databases: one on soil, water, and related natural resources; and the other on the nature and properties of the nation's soils.

These responsibilities were reinforced by the Soil and Water Resources Conservation Act (commonly known as RCA) of 1977 (Public Law 95-192). This legislation calls for periodic appraisals of the status, condition, and trends in soil, water, and related resources for use in setting conservation policies and priorities at the local, state, and national levels. Such an appraisal is identified as a National Resource Inventory (NRI). The database concerned with United States soils is updated on a continuing basis and is commonly designated as the SCS SOILS-5 database after the number of the form that is used for data collection. The formal name of this database is "Soil Interpretation Record," and interpretations of uses of soils are included in the database. The NRI and SOILS-5 constitute the largest of several national natural resource databases which SCS maintains. For this reason, they are the source of the examples used in this presentation to illustrate problems and opportunities that exist in providing and maintaining access to national resource data.

The 1982 NRI, the most recent in the series of National Resource Inventories, is now becoming available in final, fully summarized formats. It contains 22 natural resource data elements for approximately one million selected sample/sites covering the nonfederal lands of the country. Thus, the NRI database for 1982 contains approximately 22 million items of natural resource information. This information describes not only erosion status, but also vegetative cover, land use, and related resource conditions that are currently of high interest and importance. This database is maintained on the mainframe of the Washington Computer Center and of the Statistical Laboratory of Iowa State University which designed the sampling under a cooperative agreement with the SCS. The data were collected manually by visits of SCS personnel to each site. Estimated resources that were required to accomplish this task included staff time of over 300 person-years at a cost of approximately $15 million.

The Soil Interpretation Record of SOIL-5 database contains 15 to 20 data elements for approximately 13,500 kinds of soils now recognized in the United States. The data are stored on the mainframe computer of the Iowa State University which designed and maintains this database under cooperative agreement.

Because of the detailed and comprehensive nature of the information contained in these two databases, there is a high demand for access from a number of sources. Requests for access fall into three user categories: other Federal agencies, both within and outside the Department of Agriculture; state and local government agencies; and university and other non-government researchers, analysts, and interested persons.

Internally, ready access must be provided to local district conservationists and their cooperators who are located in most of the approximately 3,000 counties and parishes of the country. Also, the Soil Conservation Service frequently needs access to natural resource data acquired by other agencies in order to complement or provide for comprehensive coverage in their own data.

The problems associated with internal access to this data are:

1. the definition of selection criteria for storing and archiving natural resource data;

2. efficient relational databases for cross-referencing different types of natural resource data;

3. determination of requirements for downloading to field office microcomputers; and

4. definition of most efficient configuration of minicomputers, microcomputers (with and without hard disks), and departmental mainframes. (A pilot project is now underway at SCS on the uses and roles of minicomputers.)

The problems associated with providing access to the SCS data are:

1. Sensitivity to premature release of natural resource data precludes access by nonfederal personnel to the USDA mainframes.

2. The high cost and time requirements associated with responding to requests for data that is continually changing. Also the problem of unfamiliarity of university personnel with interpretations and procedural questions associated with the collection and analysis of the data.

3. Lack of adequate staff to respond to requests for copies of data tapes and to prepare appropriate documentation for the tape files.

4. Policy and procedural questions associated with charging of user fees either by the SCS or by public or private information brokers. Potentially, these fees could be very high because of the detailed and highly specialized nature of the databases. (The SCS operates on a non-reimbursable basis in providing information and technical assistance with conservation problems.)

5. Need for detailed explanations necessary for applications of NRI data and avoidance of misuse of the data. The related need for documentation on the sampling design, as well as sampling and measurement errors in statistical analysis. This problem is compounded further by recent improvements in design methodology which resulted in a lack of comparability of current data with data from previous inventories.

6. Problems in downloading the data from mainframes to microcomputers now being acquired by SCS field personnel.

7. Need for further development and applications of relational databases.

8. Problems associated with networking and data sharing with other agencies include:

   a. differences in definition of data elements and in the methodology of collecting the data;

   b. incompatibility of hardware and software in many cases; and

   c. concern by the other agencies, especially state and local agencies, about uses that might be made of their data.

The opportunities perceived by the Soil Conservation Service in providing access to their data include:

1. The possibility of a central repository for the natural resource databases, a repository that could also provide a technical information service. For the SCS, the National Agricultural Library could provide this service, and is moving in this direction.

2. Workshops and symposia which would provide a forum for potential users for the exchange of information on the SCS data, data collection methods, and similar matters. One such workshop on the 1982 NRI data was held in cooperation with the Board on Agriculture of the National Research Council.

3. Cooperation among agencies in solving definitional problems and in maximizing compatibility of hardware and software.

4. Improvement in the technology for linking of main-frames with microcomputers and data sharing by telecommunications.

5. Development and use of geographic information systems which incorporate several data layers in order to provide more highly integrated databases, reduce the likelihood of misunderstanding and misuse of particular data files, and promote compatibility to the fullest extent possible.

6. Interagency exchanges of personnel by temporary assignments of personnel from other Federal agencies, or university personnel on sabbatical or other types of leave.

7. Interagency agreements on definitions and compatibility of hardware and software.

8. Increased communication and cooperation with state and local agencies to reduce mistrust and misunderstanding regarding ultimate uses of natural resource data.

9. Workshops, like this one, and follow-up by agencies like the National Bureau of Standards with mandated responsibilities and capabilities for standardization.

BIOGRAPHICAL SKETCH

Dr. McCracken has had a long and distinguished career as a soil scientist and has spent many years in the academic environment. He was a professor and Head of the Department of Soil Sciences at North Carolina State University. He also served as Associate Director of the North Carolina Agricultural Experiment Station. His service with the Department of Agriculture began in 1973 as Associate Administrator of the Agricultural Research Service. Later he served as Associate Director of Science and Education at the Department. Since 1981, he has served as Deputy Chief of Assessment and Planning in the Soil Conservation Service. Dr. McCracken is a Fellow of the American Society of Agronomy and of the Soil Science Society of America. He won a Presidential Meritorious Executive Award in 1980.

INFORMATION VALUE/COMPUTER MATCHING OF DATA

Speaker

Morey J. Chick
General Accounting Office
Washington, D.C.

ABSTRACT

Computer matching is defined here as a comparison of data that exists in different files, for the purpose of creating new information. The new information that is created by a computer match is a factor that is measurable and that represents a value which may be added to intrinsic value of the information contained in the files that were matched. In an information resources management context, information value must be maximized and information costs must be minimized. In management, these factors, i.e., value versus cost, are often confused. Nonetheless, they must be measured; the question arises as to whether the value of information can be measured in terms of dollars. Results of some examples of computer matches cited in this presentation appear to indicate that this question can, in some cases, be answered in the affirmative. Several concerns about computer matching are also discussed.

The presentation is a distillation of the views of the General Accounting Office (GAO), the author, and other sources, on computer matching as a tool for the management of information. The views of the General Accounting Office are documented in their report HRD-85-22 entitled, "Eligibility Verification and Privacy in Federal Benefit Programs: A Delicate Balance." The author's views are partially reported in his article, "Information Value and Cost Measures for Use as Management Tools," published in Information Executive, Volume 1, Number 2, 1984. A copy of this article is part of this record of the presentation, Appendix B.

Computer matching is defined here as a comparison of data that exists in different files, for the purpose of creating new information. The files may belong to a single agency, to several agencies at various Federal, State, or local government levels, and/or the files may belong to non-government organizations. The new information that is created by a computer match is a factor that is measurable and that represents a value which may be added to the intrinsic value of the information contained in the files that were matched (figure 1).

Computer matching is really a type of data analysis. In the "old" technology, the process involves a simple match of files from database B against the files from database A on data elements that are common to both files. A match on these data elements generates new information which adds value to the value intrinsic in databases A and B (figure 2). The purpose of the new information is to detect errors, fraud, and/or internal control problems associated with the management of benefit programs in the Federal Government. Dollar values, here, can be measured by the savings resulting from the new information created by the match.

Figure 3 illustrates current technology as moving towards direct linkages of files via telecommunications lines. Location C on this figure represents non-government organizations, such as a credit bureau, a bank, or a school. What we have basically is a de facto centralization of data. Figure 4 represents a hypothetical link comprised of real providers of data. At present, there is no central information on all current linkages.

The concept of computer matching is not a new phenomenon; it has been in existence since approximately 1976. In the time that has elapsed since then, some 126 matches have been performed at the Federal level and some 1200 more at the state level. These matches were made on files that store information on a minimum of 136 Federal programs which benefit three out of ten Americans. The Federal share of total expenditures represented by these programs amounts to approximately $400 billion a year or 45 percent of the national budget. It is estimated that several billion dollars are overpaid annually because of abuse, fraud, error, and inadequate verification of applications for benefits. GAO historically supports matching when the benefits exceed costs and the rights of individuals are protected.

Figure 5 presents three examples of major Federal matches of data on income tested programs. The agencies involved were the Veterans Administration (VA) and the Social Security Administration (SSA). The VA pension program files were matched against earnings reports of state unemployment security agency files on at least four data elements: Wages, Social Security Number (SSN), Name and Employer. This match resulted in the detection of overpayments totaling an estimated $100 to $300 million. Benefits realized from two matches of Social Security files are reported in the form of reduction in payments of approximately $110 million per year in one case, and expected recoveries of $100 million in the other. Some $20 million of the latter figure have been recovered to date.

Figure 6 presents examples of three state matches. In the first example, New York City identified companies paying business taxes, but not rent taxes. The City matched the files from several of its own departments and collected $24.8 million in additional commercial rent payments.

What are the concerns related to computer matching? Some of them are:

- cost versus benefit ("added value");
- technology and centralization;
- privacy;
- security; and
- other concerns.

Cost/benefit analysis presents a very difficult problem, one which is under study by GAO at the present time. Measurable benefits are being identified and are continuing to be reported; recoveries represent real savings. Reductions in future payments present an added difficulty in that there is a lack of information on how long the benefit payments would have been made or even if they would have been made, in any given case.

Intangible benefits identified include the potential inherent in the use of computer matching as an internal control mechanism, as a means of testing of internal controls and as a deterrent factor. Benefits of such intangibles are very difficult to measure in dollar terms.

GAO is just now beginning to study the different kinds of costs involved in computer matching. Some of these are:

- cost of match (software, computer time, etc.);
- manual verification (e.g., employers, manual computations, etc.);
- file acquisition costs (from third parties, e.g., credit bureaus);
- costs of poor data quality;
- cost of reducing or deleting payments;
- cost of denying payments (e.g., litigation and related administrative procedures); and
- collection costs for recoveries.

The first of the above costs is the traditional one. The second, manual verification is now required by law for certain major programs. There are hidden costs associated with matching in cases where there is a need for employers to verify information. Poor quality of data is partially a result of the lack of data standards. Further costs are those stemming from data sensitivity and privacy issues, such as litigation and related administrative procedures. Currently, GAO is studying the situation, particularly from the standpoint of much-needed methodology for measuring value versus costs associated with computer matching.

In information management, the terms value and cost are often confused. The cost of information can be equated almost to the cost of producing a commodity from raw materials. Many accounting functions can be applied here, and information value can be described in terms of worth, merit, importance, etc. However, the question remains, "Can we measure value in terms of dollars?" In his journal article cited above, the author presents ways to measure, in some cases, the information value in dollars (figure 7). It should be done, where possible, for "effective management."

Computer matching does represent a de facto centralization of data, as figures 4 and 8 indicate. The figures also identify the many and various sources of information for matching purposes. This de facto centralization is not unconstitutional but does raise increased concern about privacy and security. The privacy issue is a very sensitive issue these days, one that is being hotly debated. The GAO report cited above addresses some of these issues. GAO's conclusion is that there is a delicate balance involved between detection of fraud on the one hand aimed at protection of the U.S. taxpayer and the privacy of the individual on the other hand aimed at protection of the U.S. citizen. In many cases, these are the same people.

The sources of citizens' rights to privacy are basically the Constitution, the Fourth, Fifth, Fourteenth and perhaps other Amendments, and Common Law. These are the real sources. The Privacy Act of 1974 (P.L. 93-579) is the legal source for Federal data only. The Privacy Commission provided opinion and clarified the principles. Section 552a of Title V of the Privacy Act defines routine use as, "The use of such record for a purpose which is compatible with the purpose for which it is collected" (figure 9). This is the part of the Act that provides for no disclosure without written consent of the individual citizen. However, there are 11 exceptions to that, and the routine-use clause of the Privacy Act is one of the exceptions. Executive interpretation is usually related to this clause and has basically increased and facilitated extensive Federal matching. State matches are not covered by this Act.

At this point, the author separated himself from the GAO and presented the views of some of the opponents of computer matching. Some of these views include the following:

- the real possibility of excessively broad interpretation of the routine-use clause;
- matching presumes crime, therefore it does not constitute reasonable search;
- the category of people is of interest to the government;
- fear of misuse of information (big brother);
- matching involves everyone in the file, including the innocent, and even people not receiving benefits, as in the case of credit bureaus, for example;
- purpose of match is to generate evidence of wrongdoing;

- not every program requires a direct notification of a match;
- notification via the Federal Register as required by the Privacy Act is inadequate notification;
- technology linkages increase security vulnerabilities; and
- there is no requirement for central approval of matching.

The Internal Revenue Service (IRS) has a concern about the confidentiality of tax information, as provided for in the Tax Reform Act (figure 10). Though opening of actual taxpayer information files (Forms 1040 and related schedules) is not in sight at the moment, the IRS is concerned about the impact of opening tax records. The potential losses in voluntary tax collection may be more than what may be saved through computer matching.

The last major item of concern in this area has to do with computer security. GAO is currently studying this area, and the author is involved in the study. Figure 11 lists the concerns associated with computer security. One of the items on this list is the personal data and privacy issue. The Privacy Act requires adequate technical, administrative, and physical safeguards for the protection of personal data. The last item concerns human safety considerations. Factors such as speed, error, system design problems, human response to speed, and automated decision making are major personal concerns.

Finally, some other major concerns in computer matching include:

- data quality in automated decision making and the associated practice of direct notification and elimination of beneficiaries without manual verification;
- the question of when to match;
- the SSN as the national identifier; and
- alternative verification techniques, such as telephone contacts.

The above concerns comprise basically the GAO report now being circulated. In conclusion, matching does represent a delicate balance.

BIOGRAPHICAL SKETCH

Mr. Chick is a manager in the Government-wide Information Technology Studies Group of the Information Management and Technology Division, General Accounting Office (GAO). He is a graduate of Pennsylvania State University, a Certified Public Accountant, and has 24 years of experience with the GAO. Mr. Chick specializes in studies of the problems associated with managing information, automatic data processing, individual information systems, and information policy issues in the Federal government. He has worked on a GAO task force to define and develop methodologies for studying Information Resources Management (IRM) in Federal agencies. He has also instructed GAO professional staff in IRM. Among other things, Mr. Chick has identified and reported to Congress significant problems in automated decision making, computer security, data standardization, and the implications of automation on employment. Mr. Chick has also prepared Congressional testimony on computer matching as used to detect fraud and error in Federal benefits. He has written extensively on the subject of the economics of information.

DEFINITION OF A COMPUTER MATCH

- DATA FROM DIFFERENT FILES
- MAY BE DIFFERENT AGENCIES
- FEDERAL, STATE, LOCAL
- OTHER FILES (NON-GOVERNMENT)
- CREATES NEW INFORMATION
- DONE BY COMPUTER

Figure 1

CONCEPT
"ADDED VALUE"

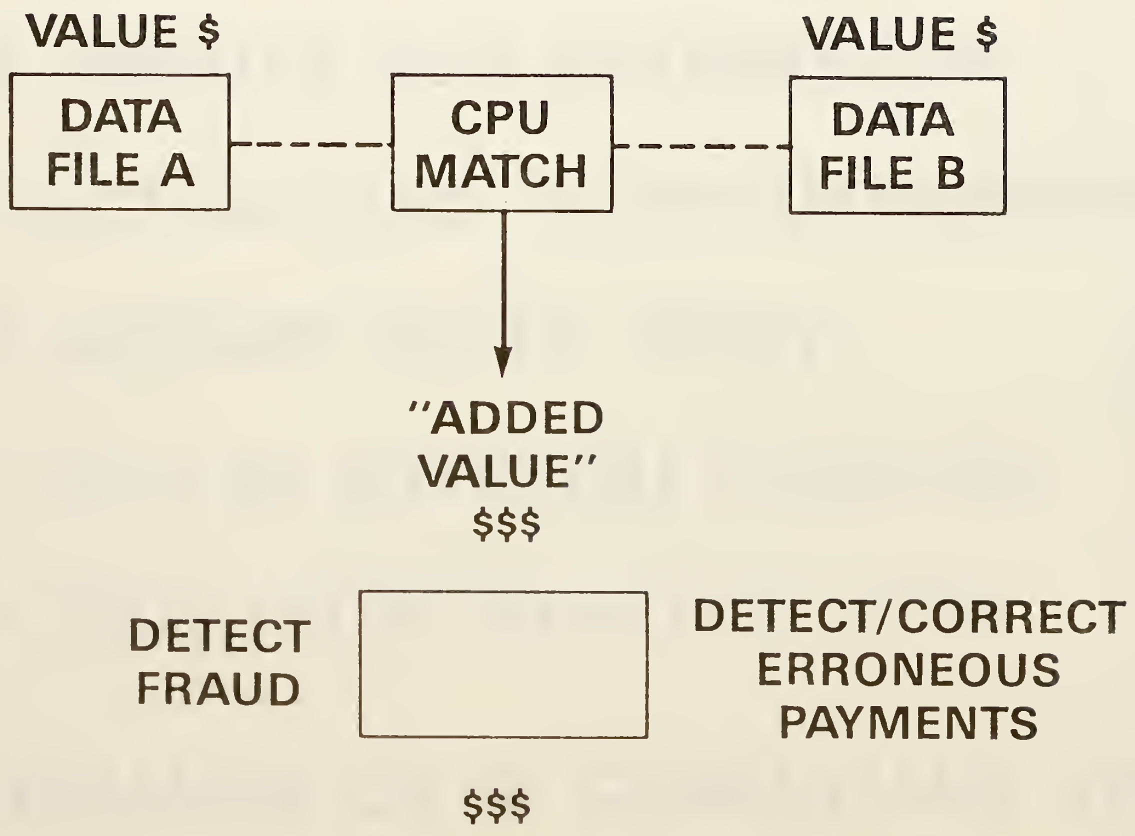
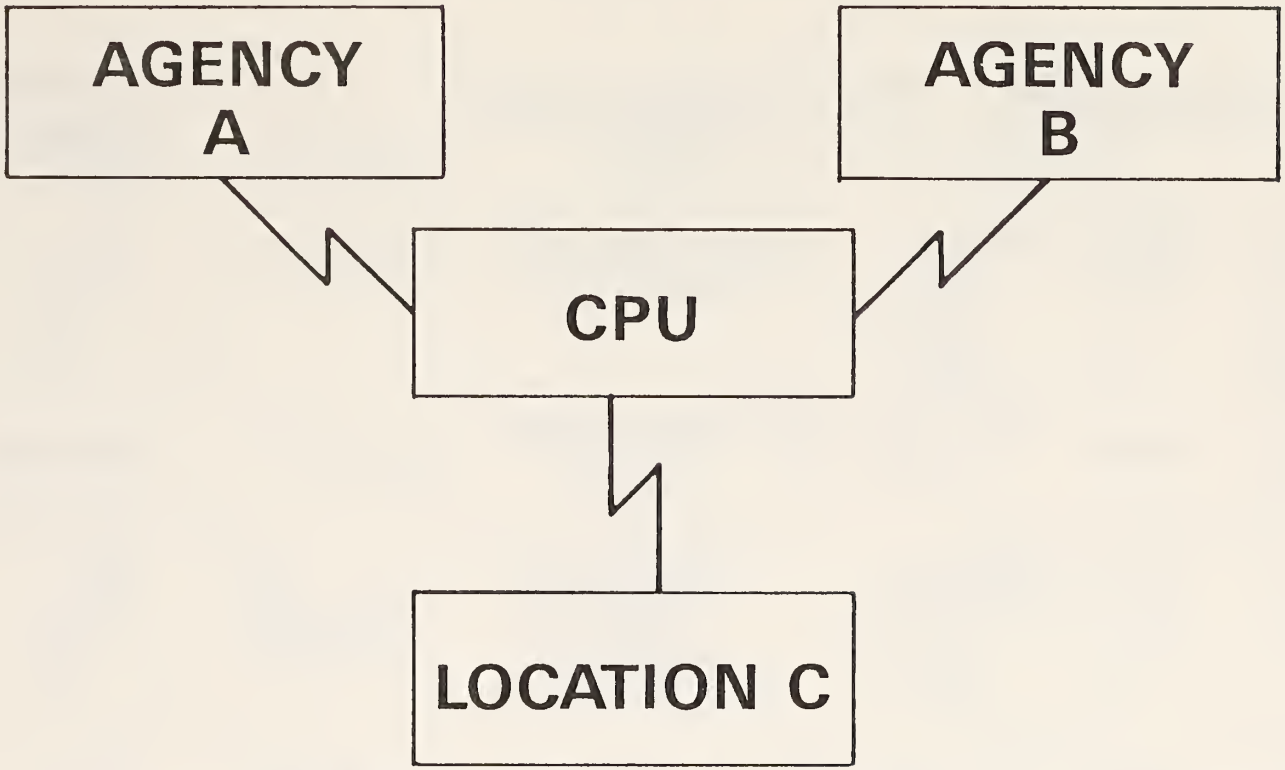


Figure 2

CURRENT MATCHING TECHNOLOGY

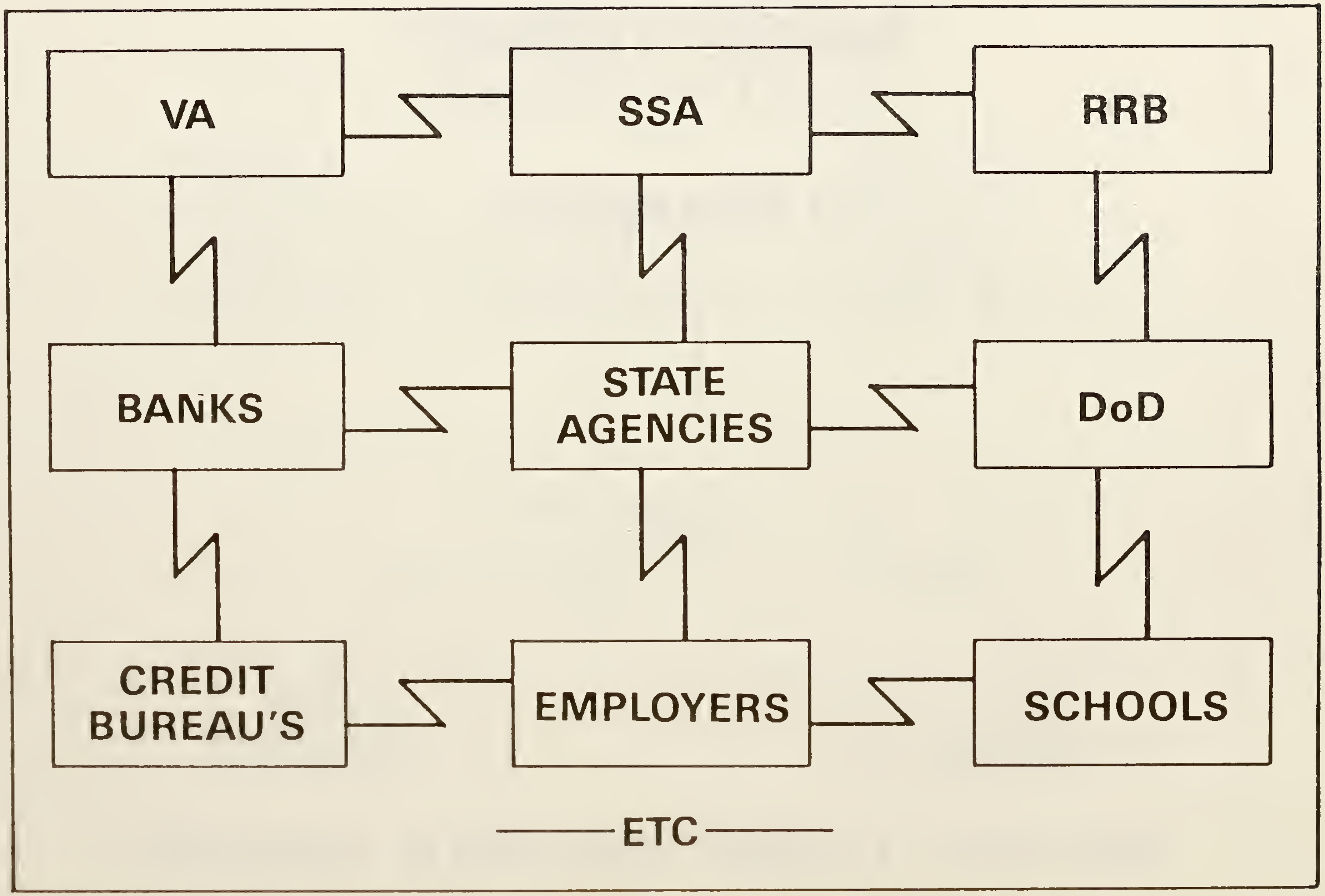
MOVING TOWARDS DIRECT LINKAGE



CREDIT BUREAU
BANK RECORDS
SCHOOL RECORDS
ETC

Figure 3

DE FACTO CENTRALIZATION OF DATA



NO CENTRAL INFORMATION ON ALL CURRENT LINKAGES

Figure 4

EXAMPLES OF MAJOR FEDERAL MATCHES (INCOME TESTED PROGRAMS)

YEAR INITIATED (APPROXIMATE)	PRIMARY AGENCY	PROGRAMS	MATCHED AGENCIES	DATA ELEMENTS MATCHED (EXAMPLES)	BENEFITS
1983	VA	VA PENSIONS	STATE UNEMPLOYMENT SECURITY AGENCIES	(EARNINGS REPORTS) WAGES, SSN, NAME, EMPLOYER.	FROM \$100 MILLION TO \$300 MILLION (TOTAL OVERPAYMENTS ESTIMATED)
1976	SSA	SSI (AGED, BLIND DISABLED)	VA (COMP./PEN) RRB (PENSION) OPM (PENSION) DoD (RETIREMENT)	UNEARNED INCOME (PENSIONS) SSN, NAME.	\$110 MILLION A YEAR (REDUCTIONS)
1979	SSA	OLD AGE SURVIVORS AND DISABILITY	STATE DEATH DATA (30 STATES SO FAR IN CENTRAL FILE IN HHS SSA ACT AMENDMENT OF 1983)	DATE DECEASED, SSN, NAME, DATE BORN.	EXPECT \$100 MILLION (RECOVERIES) \$20 MILLION TO DATE

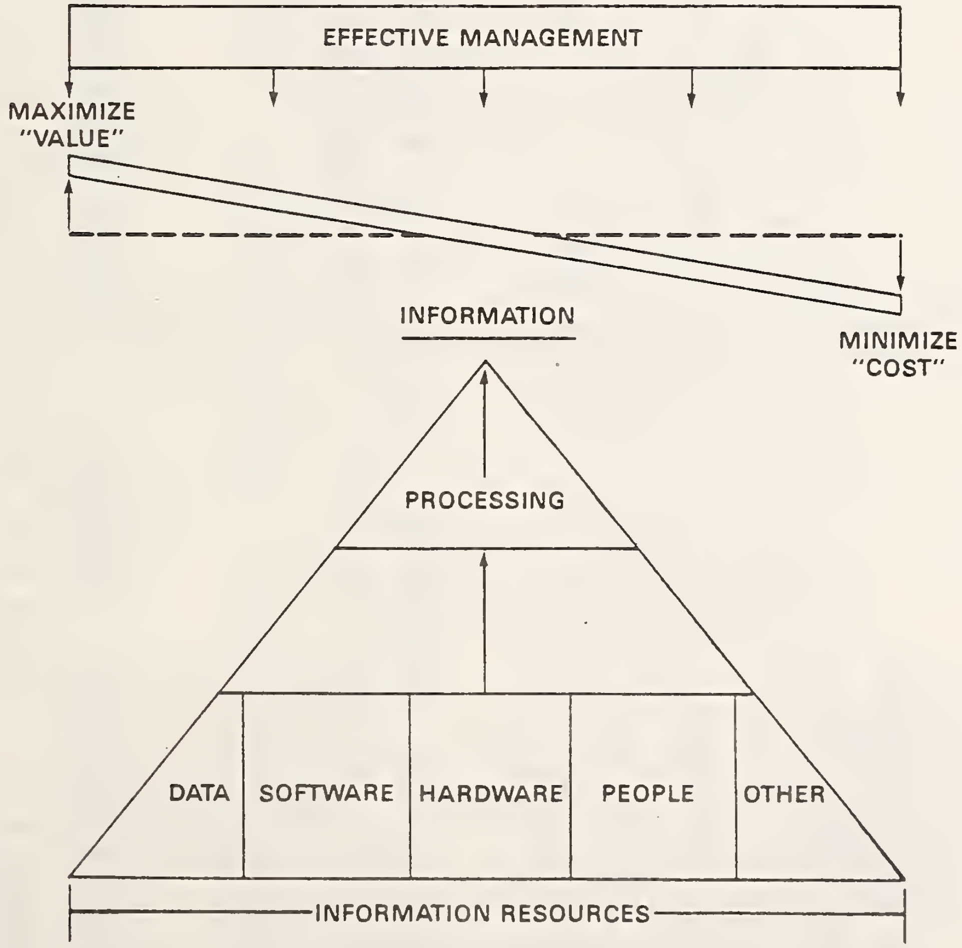
Figure 5

EXAMPLES OF MAJOR STATE MATCHES

YEAR INITIATED	PRIMARY AGENCY	PROGRAM	MATCHED AGENCIES	DATA ELEMENTS MATCHED (EXAMPLES)	BENEFITS
1981	NEW YORK CITY DEPARTMENT OF FINANCE (COMMERCIAL RENT FILES)	COMMERCIAL RENT TAX (OCCUPANCY TAX ON COMMERCIAL TENANTS)	SAME (BUSINESS TAX FILES)	CONTROL NUMBERS BUSINESS NAME ETC	$24.8 MILLION IN ADDITIONAL COMMERCIAL RENT TAX PAID
1981	NEW YORK CITY DEPARTMENT OF FINANCE	BUSINESS TAX FILERS IN 1976 (NOT FILING IN LATER YEARS)	SAME (SAME FILE, PREVIOUS YEARS)	CONTROL NUBMERS BUSINESS NAME	$20.2 MILLION IN ADDITIONAL BUSINESS TAX PAYMENTS
1984	MICH. DEPT OF SOCIAL SERVICES (WELFARE FILES)	STATE WELFARE STATE: AFDC FOOD STAMPS MEDICAID	SSA BENEFICIARY EARNINGS INDEX (BENDEX)	EARNINGS, BENEFITS, RETIREMENT AND DISABILITY SSN	550 CASES OF FRAUD IN ONE COUNTY ALONE (WAYNE) ESTIMATE: $6.3 MILLION IN FRAUD

Figure 6

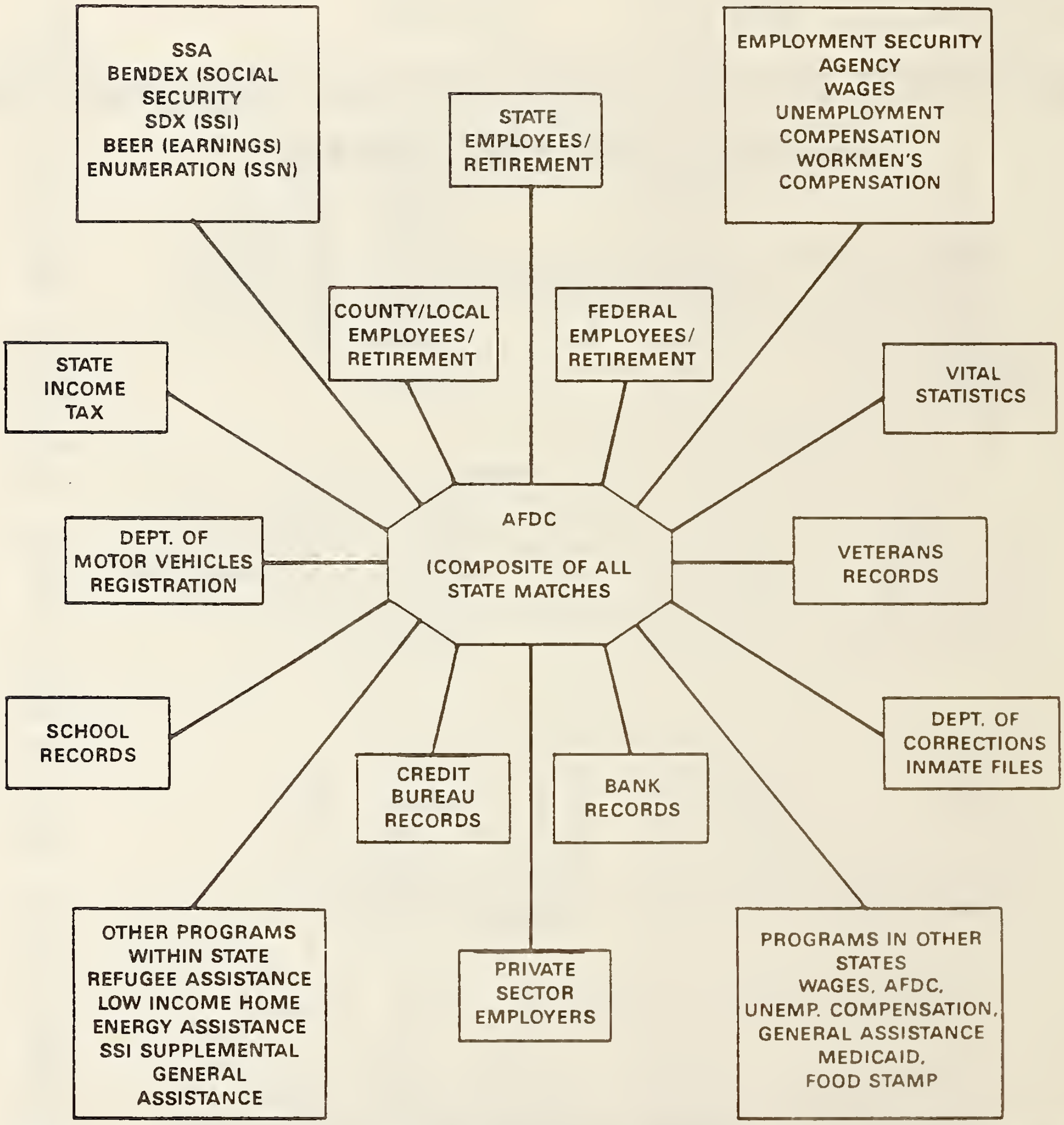
OBJECTIVES OF INFORMATION RESOURCES MANAGEMENT (IRM)



SOURCE: "INFORMATION EXECUTIVE" VOLUME 1/NUMBER 2/1984 PAGE 48

Figure 7

COMPOSITE OF DATA LINKAGES THROUGH COMPUTER MATCHES BY AFDC PROGRAMS IN VARIOUS STATES



NOTE: NO SINGLE STATE HAS ALL OF THESE LINKS, BUT EACH LINK OCCURS IN AT LEAST ONE STATE. WITH A FEW EXCEPTIONS, HOWEVER, THESE TYPES OF SOURCES COULD BE AVAILABLE IN EVERY STATE.

SOURCE: DEPARTMENT OF HEALTH AND HUMAN SERVICES, OFFICE OF INSPECTOR GENERAL, INVENTORY OF STATE COMPUTER MATCHING TECHNOLOGY; AND GAO OBSERVATION (HRD 85-22)

Figure 8

MATCHING AND DISCLOSURE PRINCIPLE (PRIVACY ACT) - SECTION 552a

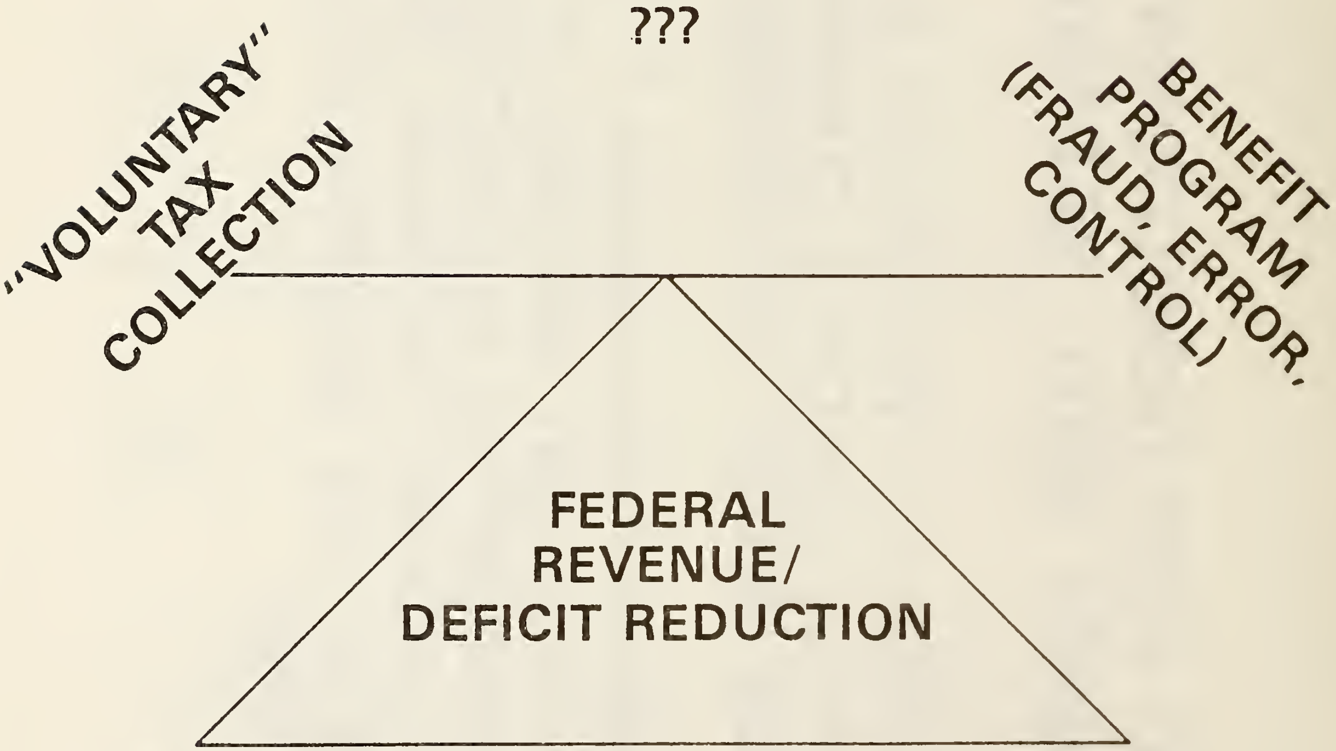
- "ROUTINE USE" *** THE USE OF SUCH RECORD FOR A PURPOSE WHICH IS ***COMPATIBLE WITH THE PURPOSE FOR WHICH IT IS COLLECTED.***"

- "CONDITIONS OF DISCLOSURE *** NO AGENCY SHALL DISCLOSE ANY RECORD *** (WITHOUT) ***WRITTEN CONSENT*** OF THE INDIVIDUAL TO WHOM THE RECORD PERTAINS, ***UNLESS DISCLOSURE*** OF THE RECORD WOULD BE *** ***FOR A ROUTINE USE***"

Figure 9

COMPUTER MATCHING

IRS VIEW ABOUT TAXPAYER RETURNS



WHAT IS THE PROPER BALANCE ???

WHAT IS THE POSSIBLE NET $ EFFECT ???

Figure 10

REASONS FOR CONTINUING CONCERN FOR ADP/TELECOM. SECURITY

- DEPENDENCY
- HIGH DOLLARS
- SENSITIVE DATA
- PERSONAL DATA ("PRIVACY")
- ADP/TELECOMMUNICATIONS INTEGRATION
- RAPID ADVANCES
- OLDER SYSTEMS
- LEGISLATION
- PAST/CURRENT PROBLEMS
- AGGRESSIVE HACKERS, CRIMINALS, AND OTHER UNFRIENDLY SOURCES
- HUMAN SAFETY CONSIDERATIONS

Figure 11

MANAGING DATA IN FINANCING HEALTH CARE

Speaker

John Parmigiani
Department of Health and Human Services
Baltimore, Maryland

ABSTRACT

The Health Care Financing Administration (HCFA) is responsible for a $100 billion-a-year program that provides health insurance protection to more than 50 million Americans. While all HCFA efforts can be linked to supporting five major agency missions, further analysis has yielded 19 critical functions, derived from these missions, prioritized, and categorized into four principal groups,that must be efficiently executed for the agency to meet its responsibilities. The management of data is the key to the successful performance of these functions. HCFA is currently developing an information resources management structure that will enable it to plan, manipulate, secure, exchange, and integrate its data. Management perspectives relative to the administration of HCFA's data universe are stressed in the following discussion.

This presentation includes a description of some of the management problems, with respect to data, that Health Care Finance Administration (HCFA) is faced with and what is being attempted to solve them. The HCFA is the Federal agency responsible for spending approximately 10 percent of the Federal budget! HCFA spends approximately $100 billion a year in carrying out its responsibilities for funding the Medicare and Medicaid programs. Over 50 million of the nation's poor, elderly, and disabled people will have their health care needs met through these programs. By 1986, these programs are expected to aid nearly one in every five Americans.

IMPORTANCE OF DATA TO HCFA

The present information systems environment suffers from a number of system-specific and agency-wide problems which hamper its effectiveness in meeting the Agency's programmatic, policy-making, and decision support information needs. A review of data flow in HCFA is shown in (figure 1). Approximately 120 application areas in four major system areas - health insurance (I), statistical (II), program management (III), and administrative (IV) - exist in HCFA. The flows in this chart can be described as follows:

Area I

The Social Security Administration (SSA) and Railroad Retirement Board (RRB) provide entitlement and status information to the Hospital Insurance/Supplementary Medical Insurance (HI/SMI) System. Providers and contractors transmit queries and bills to the HI/SMI System. HCFA transmits replies concerning entitlement, eligibility, deductibles, and remaining benefit days to the contractors and providers. HCFA pays the providers through the contractors and reimburses the contractors for ADP and administrative costs.

Area II

Information about the beneficiary is extracted from the Health Insurance Master (HIM) file, and information about utilization (medical procedures, costs) is extracted from the Medicare bills. These two sources provide most of the information in the Medical Statistical System (MSS); hence, HI/SMI and MSS are closely related.

Area III

Information from the MSS is passed on to, or used by, the Program Management Systems, but there is not a strong connection; hence, it is shown with dotted lines. The data from MSS is used to project long-range trends. The data to operate Medicare and Medicaid in such areas as cash flow, budget, and administrative costs, is obtained from the contractors on a current basis.

Area IV

The connection between the Program Management Systems and the HCFA Administrative Systems is also shown as a dotted line to denote the interchange between the overall budget managed by HCFA, OMB, and the program agent (Medicare Contractor, Medicaid State Agency, Peer Review Organization) budgets managed by components of the Associate Administrator, Operations.

In general, these systems exhibit certain shortcomings, some of which are:

- The amount of time required to provide Medicare contractors with information on beneficiary entitlement, eligibility, and deductibles from the HI/SMI System. Many HCFA intermediaries provide on-line query/reply to major providers (hospitals); but HCFA can give, at best, overnight service because of its tape-oriented, batch system.

- Inadequate support to HCFA's ten regional offices in the areas of Medicare/Medicaid, program management, and contractor monitoring operations.

- Delayed or inadequate support to top-level HCFA management in automated support for decision-making in such areas as projection of trends, monitoring cash flow, and estimating the impact of new legislation.

- HCFA has a number of ADP systems which have been developed in a bottom-up fashion to meet the needs of operating divisions for functions such a claims processing, statistical analysis, contractor management, and personnel management. Because of the way these systems were developed, the same data has been collected and stored in different ways, creating overlapping files containing redundant and frequently incompatible data. In using such systems, managers have been unable to obtain, in an automated fashion, summary information that cuts across division or departmental lines to support decision making. The data available from the bottom-up systems is frequently found to be inconsistent, lacking significant details, and not sufficiently up-to-date to support decision making.

Additionally, there are various general ADP problems attributable to an ADP support system that HCFA inherited when it was created from several agencies in March 1977 and which subsequently grew in piecemeal fashion in response to priority needs. These are:

- Reliance on obsolete tape-oriented systems and batch processing techniques.

- Software which is difficult and costly to maintain because it has been repeatedly patched, represents obsolete design concepts, and uses a variety of documentation standards.

- A lack of flexibility and the ability to adapt to changing needs in the design of application systems.

- A collection of overlapping systems and redundant tape files in the statistical and program management areas.

- Fragmented ADP operations in the statistical, program management, and decision support systems (DSS) areas.

Another glaring shortcoming of the current HCFA application system environment is lack of flexibility. Recent health care legislation which made fundamental changes in medicare operations highlighted the inflexibility of HCFA ADP systems. At any given time, the Agency is generally considering approximately 40 legislative proposals affecting Medicare and Medicaid and sorely needs the capability of doing "what if" analysis.

Against this backdrop, the Agency is now entering a period marked by a growing workload, dramatic changes in health care delivery and methods of payment, and widespread innovation in information processing technology. Rising hospital and medical costs coupled with Federal budget constraints and concern for the fiscal soundness of the Medicare trust funds have created intense pressure to stem inflationary trends in charges for medical services and to reduce unit costs for processing Part A Hospital Insurance (HI) and Part B Supplementary Medical Insurance (SMI) claims. In October 1983, HCFA initiated the Prospective Payment System (PPS) which sets limits on Medicare payments for inpatient hospital stays for 468 Diagnosis Related Groups (DRGs), and steps are being taken to establish similar ceilings on physician's fees. Concurrently, HCFA is encouraging beneficiaries to utilize lower-cost alternatives for health and medical services, including Health Maintenance Organizations (HMOs) and Group Practice Prepayment Plans (GPPPs). These new programs have greatly increased the amount and complexity of information that HCFA must collect and process in its computer systems, not only in the area of Medicare claims processing, but also in contractor management and Medicare/Medicaid statistical systems.

Faced with the need to modernize its computer and telecommunications systems to meet the needs of the next 10 years and beyond, HCFA established the "Project to Redesign Information Systems Management (PRISM)" and began working toward the definition of a long-term information systems architecture.

APPROACH TO MANAGING DATA IN HCFA

One of the first major efforts undertaken by the agency under its PRISM initiative was a mission needs analysis to identify the critical success factors for the Health Care Financing Administration. HCFA's five major missions are:

1. Formulate National Health Care Policy
2. Manage Integrated HCFA Programs
3. Operate Medicare
4. Administer the Medicare Program
5. Manage the Agency's Resources

We then subdivided each of these missions into numerous major functional areas with an eye toward eventually defining information needs relative to each. We finally selected 19 major agency functions and prioritized them in four groups. The next chart depicts these groupings (figure 2). We tried to stay with fundamental, mainline business processes that were least subject to change. The Medicare claims processing system was selected for Group 1 for the reason that it has been the major programmatic function of HCFA since the agency's inception and accounts for its largest expenditures. Because Medicare claims processing also incurs over $800 million in contractor costs, it is also a prime area for further automation and cost reduction. The Group 1 functions also include reimbursement of providers/beneficiaries, cash management, debt management, and management of the Medicare Trust Funds, as well as Medicaid cash management. The Medicaid system is also in priority Group 1 because of its size.

The Group 2 functions include management and reimbursement of the contractors, certification, integrity functions, and PRO management.

The Group 3 functions include formulation of legislative proposals, reporting on and accounting for program operations, conducting research and demonstration projects, and use of statistics to forecast health care trends.

The Group 4 functions include administrative functions and reporting and liaison activities.

Once we had identified those key functional areas of the Agency, we could then move to determining information needs associated with each, and then ultimately the data that must be gathered, processed, and formatted to produce this information -- the main idea being that we only collect and use that data necessary to carry out our business as a Federal agency. This analysis effort would also help in determining the system's architecture and necessary data structures essential to transforming data into information. The recommendations resulting from numerous in-depth studies have resulted in HCFA'S having arrived at the following set of interrelated directions.

- o A Systems Architecture
- o An Information Architecture, and
- o A Data Architecture

The overall approach is characterized by a centralized database management system for the bulk of HCFA's workload with on-line query and retrieval supported by a telecommunications network linking contractors, providers, and the government. Programmatic and administrative areas will feature subject matter databases capable of relational activity and complemented by a large complex of multipurpose work-stations linked through local area networks. Specifically,

- o The HI/SMI Claims Processing System supports HCFA's major programmatic function and accounts for most of the Agency's computer resource usage. The recommendation is to convert the Health Insurance (HI) Master file from tape to high-density magnetic disk and use a DBMS for reliable, efficient computer processing and to make on-line query/reply available to hospitals and other providers. A new HCFA telecommunications network will be required to link 102 intermediary and carrier systems directly to the HI/SMI system to support on-line and batch activity.

- The Medicare Statistical System (MSS) derives its beneficiary data from the HI Master file and utilization data from HI/SMI Processing of Part A bills, and to a lesser extent, Part B payment records. The redesign of the MSS as a disk-oriented system utilizing interactive capabilities, pre-aggregated data and sample files on disk, and mass storage devices for bulk data, is recommended. A new Medicaid statistical collection effort will result in a system comparable in size to MSS, and the planning, design, and operation of the two statistical systems will be closely coordinated.

- The Program Management Systems support contractor administration, financial operations, debt management, accreditation, and other management functions, microcomputer technology and telecommunications will dramatically change the manner in which HCFA manages its programmatic operations in the future. HCFA plans a system architecture in which microcomputers will be used in contractor offices to input financial and performance reports, in regional offices, and in central office bureaus for local computing; the microcomputers will also serve as multipurpose work stations which can access central databases. Data of purely local interest will be stored locally. HCFA anticipates that data of national or agency-wide interest will be stored in central office databases and that the update of files will be under the resource center to control the networks, assist users in the acquisition of hardware, and train users in the use of microcomputers.

- The HCFA Administrative Systems will have a similar architecture utilizing multipurpose work stations and centralized subject matter databases. However, in these systems, the main flow of data will be from Department of Health and Human Services (DHHS) budget, financial, and personnel systems to the HCFA systems. Major thrusts are to develop a consolidated debt management system, to establish interfaces between HCFA and DHHS financial systems, and to develop nationwide DHHS financial systems, and to develop a nationwide network linking long-haul telecommunications and local area networks (LANs) to support office automation, electronic mail, and preparation and dissemination of regulations and other issuances.

These major applications areas are linked in an information management architecture whose scope is described below:

Level 1- Transaction sources (end-users) include the 6600 Medicare/Medicaid providers, the 102 contractors and HCFA employees (over 2000) using the Administrative, Program Management and the Medicare/Medicaid Statistical systems. In the latter case, all information will be passed through a Local Area Network.

Level 2- This includes the data carriers or integrated nation-wide telecommunications network. The carrier can be wire, cable, microwave, fiber-optics, and/or satellite.

Level 3- These applications support system software for collecting/dispatching the transactions, network management, and support of the applications in processing the transactions.

Level 4- Includes application systems (HI/SMI, MSS, Program Management and Administrative) which process the transactions with the support of the systems software and database management system.

Level 5- The database management system is used for organizing and managing the data in a convenient way and allows for quick access of the data, and retrieval/update of the data by application programs. It also facilitates developing and maintaining application systems.

Level 6- Includes the aggregated physical data. The data is stored, maintained, and retrieved by the database management system.

Originally, we plan to implement a centralized architecture as shown in the next chart (figure 3) with the future option of extending to a distributed architecture (figure 4). In either architecture, it is envisioned that hardware/software management and operating control would be centralized for HCFA-wide control and management.

Our final area of concern is structuring the HCFA information systems environment in order to manage the Agency's data in the appropriate data architecture. What is needed is a stable structure that supports HCFA's changing data needs without the redundant maintenance of similar data in different databases and which provides the sharing of data resources among various applications areas. After considerable analysis and subsequent integration, a minimal set of unique subject matter databases was determined. These 14 unique subject matter databases are:

- o Type of Services and Charges,
- o Premium Collection,
- o Provider,
- o HI Master (beneficiary data)
- o Utilization,
- o Quality of Service and Program Integrity,
- o Contractor,
- o Medicaid Agency,
- o Health Issues and Legislation,
- o Financial,
- o Personnel,
- o Property,
- o Internal Control, and
- o Information Resource Management.

The next chart (figure 5) further delineates which application areas pertain to each of these databases, the shared use of the databases among the application areas, and finally, which of the various applications areas are their responsibility for updating and maintenance. In order for us to ensure that the right data gets to the right person at the right time, we must install a DBMS that incorporates certain necessary features. The final chart (figure 6) attempts to summarize these features and the application areas which share them.

- o Data Dictionary -- A catalog of all HCFA's data elements giving their names and structure.

- o Application Development Aids -- Software that generates source code and screen formats.

- o Ad-Hoc and Survey Report Generators -- Program Management and statistical processing requires software that allows non-ADP personnel to create and format reports.

- o Communications Support -- Data communications monitor software that allows remote users to access the database on-line.

- o Personal Computer (PC) Support -- The Program Management and Administrative areas need special data communications software that allows the database and personal computer to communicate and jointly manipulate data.

- o Security/Recovery/Backups -- Data integrity and access control software, and backup and recovery of data software.

- Distributed Data Management -- The ability to utilize multiple distinct databases at the same time.

- Multimedia Storage -- The Medicare Statistical Processing requires the ability to store a single database on different storage media.

- Remote Quick Access -- Users at a distance from the processing facility obtain quick response time after starting processing activities.

- Multi-Key Access -- Medicare Statistical retrieval requires searches of data based on the combination of several key fields.

The Program Management and Administrative areas have almost identical DBMS feature requirements. The Statistical area has the requirement of multi-media storage; except for this one requirement, the Statistical, Program Management, and Administrative areas have the same DBMS feature requirements. Finally, the HI/SMI area shares many DBMS feature requirements with other Application areas.

The HI/SMI, Statistical, and Program Management areas should utilize the same DBMS for the management of their databases because:

- There is heavy cross-utilization of data among the three application areas.

- The three Application areas need the same DBMS features.

- Database users will need to learn only one DBMS.

- A DBMS that supports HI/SMI's performance requirements and supports the Statistical area's need for multi-key retrieval and also be able to handle the other application areas performance requirements.

There is a minimal cross-utilization of data between the Administrative area and the other application areas; only the Program Management area shares data with the Administrative area. Again, there are several reasons for the Administrative area to utilize the same DBMS as the other application areas, the reason being:

- The application areas share many necessary DBMS features, and

- Database users will need to learn only one DBMS.

We are currently avidly pursuing this direction at HCFA. Our goal is to eventually arrive at an environment where management of our data has enabled us to best carry out the Agency missions with maximum efficiency. Critical to our success is a continuing assessment of our information needs and the data and technology available to meet them.

BIOGRAPHICAL SKETCH

Mr. John Parmigiani, from the Health Care Financing Administration of the Department of Health and Human Services, has had a long career in the MIS field, beginning with industrial engineering and operations research. He has spent many years in information systems, principally in the area of health care delivery and financing. He is presently in charge of the group that has policy, planning, and oversight responsibilities for information systems management and data administration in HCFA.

OVERVIEW OF HCFA DATA FLOW

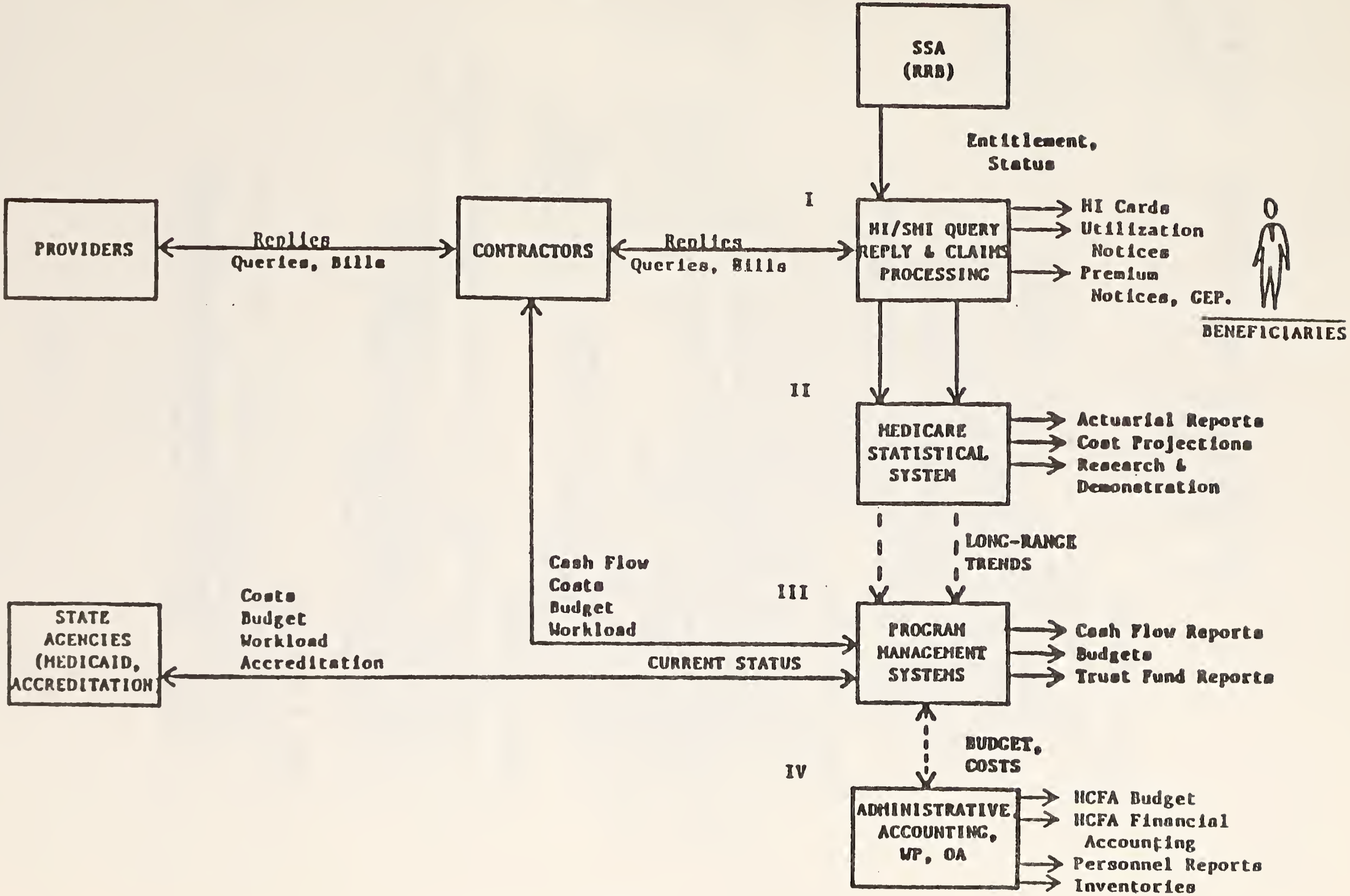
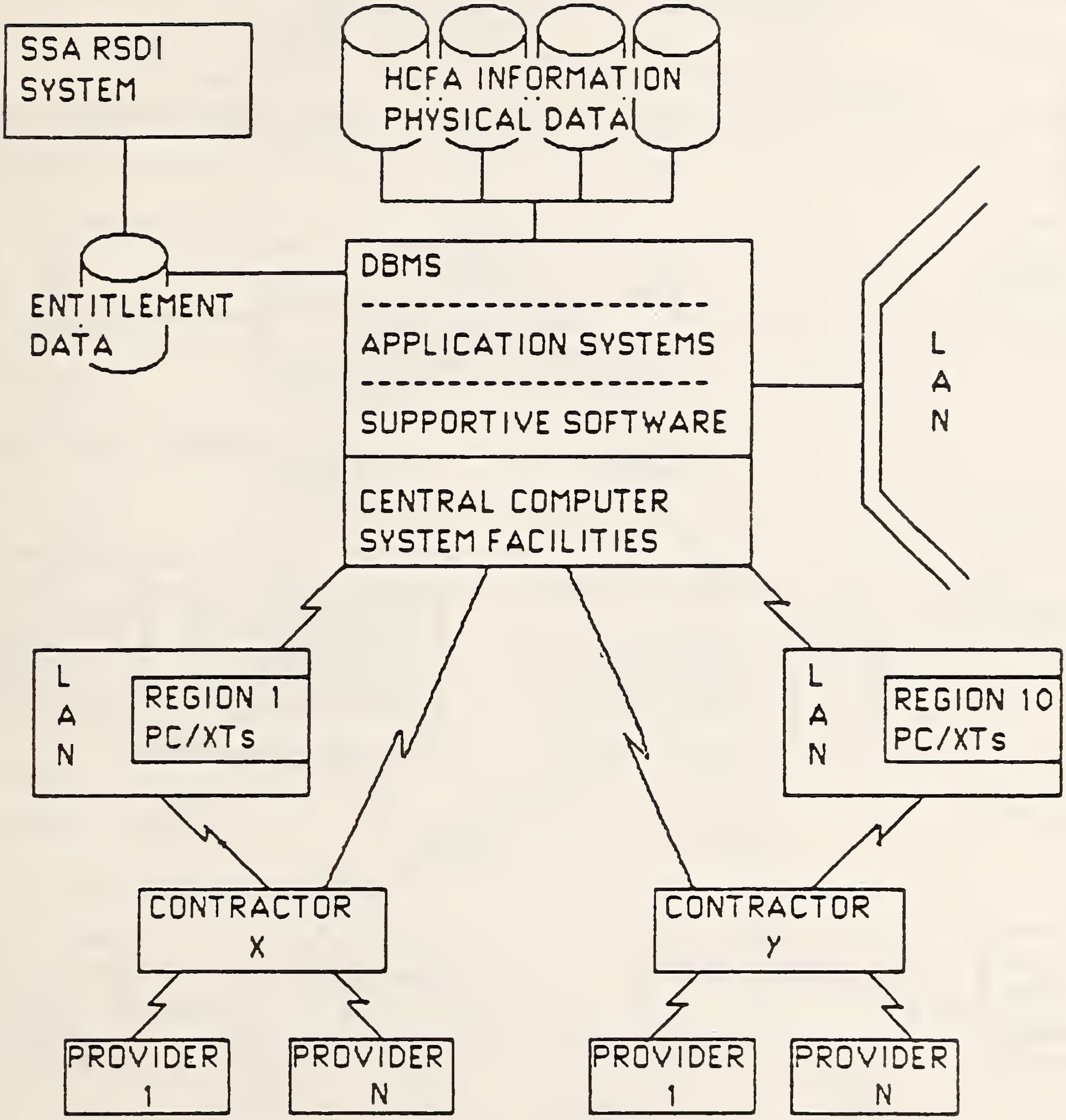


Figure 1

PRIORITIZED LIST OF HCFA FUNCTIONS

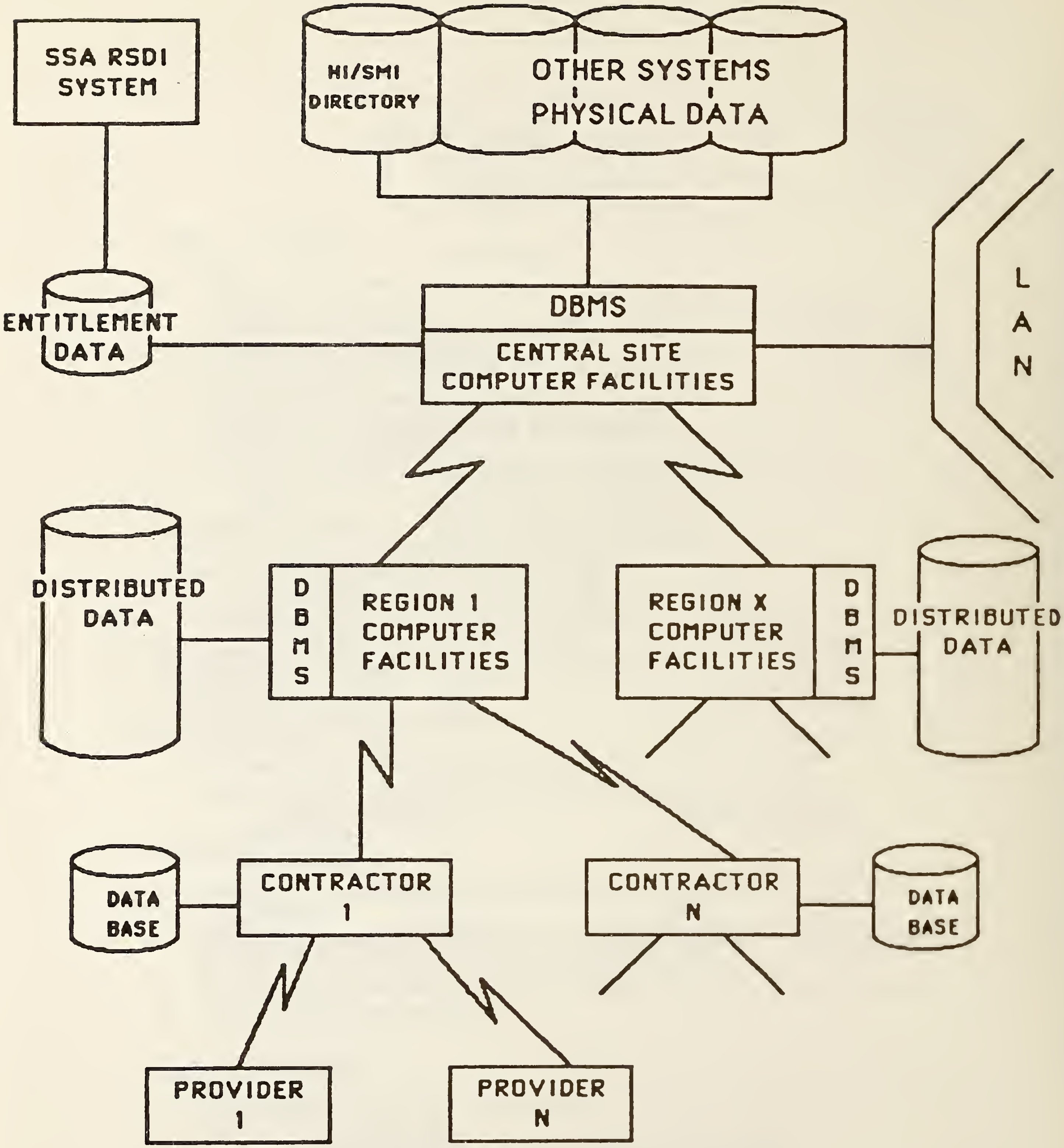
	Mission
Group 1 Functions	
• Operate Medicare	3
- Process Claims	
- Reimburse Providers and Beneficiaries	
- Manage the Trust Funds	
• Debt Management	
• Administer Medicaid at the Federal Level	4
- Reimburse State Agencies for Medicaid	
- Manage Cash Flow	
• Administer ESRD, Group Health, and HMO Programs	3
Group 2 Functions	
• Manage Intermediaries and Carriers	3
- Reimburse Intermediaries and Carriers	
- Control Administrative and ADP Costs	
- Monitor Contractor Performance	
• Formulate Medicare and Medicaid Budgets	2
• Certify Providers	2
• Process Physician Sanctions	2
• Manage PROs, PSROs	2
• Process Beneficiary/Provider Appeals	2
• Ensure the Fiscal Integrity of Medicare and Medicaid	2
Group 3 Functions	
• Formulate Legislative and Regulatory Proposals, and Health Care Policy	1
• Report Program Experience and Statistical Trends	1
• Account for Medicare and Medicaid Operations	2
• Conduct Research and Demonstrations	1
• Collect Medicare Premiums	3
Group 4 Functions	
• Manage HCFA's Monetary, Personnel, Property, and Information Resources	5
• Respond to Inquiries from Congress and the Public	2
• Maintain Liaison With the Health Care Community	1

Figure 2



CENTRALIZED ARCHITECTURE

Figure 3



DISTRIBUTED ARCHITECTURE
Figure 4

APPLICATION USERS OF HCFA DATA BASES

Subject Matter Data Bases	Application Area: HI/SMI	Statistical	Program Management	Administrative
Types Of Service And Charges	Yes	Yes		
Premuim Collection	Yes			
Provider	Yes	Yes	Yes	
HI Master	Yes	Yes		
Utilization		Yes	Yes	
Quality And Integrity			Yes	
Contractor			Yes	
Medicaid Agency			Yes	
Health Issues And Legislation			Yes	
Financial			BPO	OMB
Personnel				Yes
Property				Yes
Internal Control				Yes
Information Resource Mgmt.				Yes

Figure 5

NECESSARY DBMS FEATURES BY APPLICATION AREA

Necessary Features	HI/SMI	Statistical	Program Management	Administrative
	Application Area:			
Data Dictionary	Yes	Yes	Yes	Yes
Application Development Aids	Yes	Yes	Yes	Yes
Ad-Hoc Report Generators		Yes	Yes	Yes
Communications Support	Yes	Yes	Yes	Yes
Personal Computer Support		Yes	Yes	Yes
Security/Backup/Recovery	Yes	Yes	Yes	Yes
Distributed Data Management	Yes			
Multi-Media Storage		Yes		
Remote Quick Access	Yes	Yes	Yes	Yes
Multi-Key Access		Yes	Yes	Yes

Figure 6

DATA ADMINISTRATION POLICIES AND CONCEPTS

Moderator

Ted Albert
U.S. Geological Survey
Department of the Interior

Recorder

Joan Sullivan

DATA ADMINISTRATION: POLICIES AND CONCEPTS

Speaker

Stewart S. Morick
Price Waterhouse
Washington, D.C.

ABSTRACT

An evolutionary process is the key to success in implementing the data administration function for a given organization. The size and role of the data administration staff will be different for each Federal agency and will change with time. The framework for a variety of standards should be created initially, with the understanding that individual standards for specific areas will be filled in later. Discussion of the data administration role includes function definition, interfaces, staffing and placement within the organization, as well as a plan for implementation.

DEFINITION OF FUNCTION

The data administrator manages a resource called data. His is the human function responsible for the identification, creation, and dissemination of data usage policy within an organization. The data administrator identifies specific areas where policy is needed, locates knowledgeable individuals to write the standards, and manages their efforts in order to ensure consistency. The data administrator then disseminates the policies and standards across the organization.

The role of the Data Administrator (DA) is distinctly different from that of the Database Administrator (DBA). The DBA is responsible for the technical implementation of automated data employing the required data usage policy. The DA has a broader perspective. He is concerned with all data, whether stored in file systems, database management systems, or in manual workshops. He is concerned because eventually non-automated data becomes automated. The DA and DBA should be separate functions outside of the same organization to ensure some form of checks and balances.

PLACEMENT WITHIN ORGANIZATION

Before deciding where to place the DA function within the organization, it is important to determine first the scope of the DA role and the level of responsibility. Ideally, the DA is responsible for ALL data and reports to a senior member of the organization. This would allow for the effective use of authority across organizational lines and would yield the biggest payoff. More realistically, if the DA establishes standards just for the data processing professionals, then he should be assigned to that group. However, as the DA responsibility grows beyond that organization, he must be moved up and out to have the authority to implement the policies.

Typically, in the DA function, "business sense" is more important than "computer sense." However, it is important to identify the types of persons required, based on their interfaces within the organization. Will the DA establish data usage policy for the user community, data processing professionals, and/or operations? Will he meet with vendors or give management presentations? All of these factors should be evaluated against the organizational chart (both its formal definition and the way it is perceived from within) to place the DA function correctly. There must be a balance between responsibility and authority to implement.

FUNCTION INTERFACES

The DA role and the types of information that must be interchanged vary with the different interfaces to the organization.

The prime concerns of the end users are that data should be accurate, timely, and available in a format they can read. To that end, the DA function establishes responsibility or "custodial rights" for the data, integrity constraints, and access paths to accommodate the different views of the data. The DA selects areas for standardization and monitors service levels. The charge back environment, when used, serves as a control point to ensure satisfaction.

For management, the DA must be able to report on performance and maintain confidence.

For the applications area, the DA must make standards usable. Standards for development, maintenance, and programming support should be realistic since programmers operate in an environment where "systems are due yesterday" and "there is no time for documentation." Standards for the application areas might address edits, validity checks, test routines, and documentation.

For computer operations and systems programmers,the DA needs to identify policy that will make things run more smoothly, such as saving datasets, restarting when the system goes down, and recovering from a point of failure. Standards would be established for workbooks, log tapes, journals, etc.

The DA function may include interfacing with vendors. If so, the DA should maintain state-of-the-art knowledge in different products, such as database management systems, data dictionaries, fourth generation languages, and security software. As vendor liaison, he must coordinate purchases across organizational lines. He must be aware that packages have their own set of standards and decide how that fits in.

STANDARDS

In order to organize all these different standards for the various interfaces, the DA should initially set up a skeleton standards encyclopedia, made up of a series of volumes, each addressing a certain set of standards. For example, volume 1 might be for users and volume 2 for management. These volumes will be filled in later and will change with the environment. The following is a short list of possible standards.

- o DATA COLLECTION
- o DATA ANALYSIS
- o DATABASE DESIGN
- o PHYSICAL STORAGE
- o DATABASE DEFINITION
- o DATA CONVERSION
- o DD/DS SUPPORT
- o DOCUMENTATION
- o SECURITY
- o RECOVERY
- o VALIDATION
- o AUDIT
- o TESTING
- o DATABASE STATISTICS AND ACCOUNTING
- o CHARGE BACK
- o NAMING CONVENTIONS
- o CODING STANDARDS
- o DATA USAGE
- o EDUCATION/TRAINING
- o SYSTEMS DEVELOPMENT LIFE CYCLE
- o INTEGRITY

The primary tool on the market to implement any individual standard and assure that it is followed is the data dictionary/directory system (DD/DS). It should be required! However, it is the hardest piece of software to implement in the sense of making sure it is used properly. The DD/DS should be active. It should be the only supplier of data for any interface: "If it's not in the DD/DS, it's not in the system." Even a passive DD/DS, with manual effort, can give the effect of an active DD/DS. The policy must be established that interfaces will be generated from the DD/DS before they can be approved and put into production. A quality assurance and testing group can enforce this policy.

STAFFING

The DA function should be staffed with a small, highly experienced group. A three-person shop would be preferable to an eight-person team, which is difficult to coordinate. And, of course, the size should be limited to the role and responsibility of data administration. The best approach is to staff the function from within the organization. The team should be a group that really knows the organization, a group who has knowledge of information and where to go to get it. Data administrators need to understand the application jargon and data element names. They need to be able to identify informal data. General knowledge of computer technology or "computer literacy" is necessary, but relevant experience in application areas within the organization is more important. The DA must have his finger on the heartbeat of information and data in his organization. And, finally, the DA team must have supervisory skills. They must be able to show forcefulness in promoting standards and managing their implementation.

IMPLEMENTATION

The organization should use an information systems planning methodology. The purpose is to identify the types of data which exist, how data is used, where standards will come from, integrity constraints, rules, security, and access paths to data. The information systems planning methodology allows the organization to view information needs and requirements to determine what should then be implemented in the systems development life cycle.

The DD/DS stores the information systems plan, driving the plan down to implementation. Although the use of the DD/DS exists today, it is difficult to achieve.

Implementing the DA function is an evolutionary process. The recommended approach is to set up the function within the technical data processing shop; e.g., the DBA group. Establish baseline responsibility and guidelines. Start with a pilot project to test the forms, standards, guidelines, and conventions for the application area. Do not pick a critical production system, such as payroll, for the pilot. Instead, select a new system, preferably a "feeder system," where current manual procedures can be used in the event of system failure. Pick a system which will be a candidate for "add on" work. The data administrator may later be able to demonstrate how previous results can be reused without starting from scratch.

Finally, establish an implementation plan by project and by function. Each project will be using pieces of the DA function.

And each piece of the DA function can be applied more broadly as it is tested and refined.

BIOGRAPHICAL SKETCH

Stewart Morick is a Senior Manager for the Management Consulting Services Department of Price Waterhouse. He has an extensive background in management consulting with emphasis on information systems. His areas of expertise include database management systems, data dictionary/directory systems, information process and analysis, and fourth generation languages. Stewart Morick has taught a number of courses in both the private sector and the public sector, and he is a regular speaker at both national and international conferences.

BUILDING THE DATA ADMINISTRATION FUNCTION

Moderator

Jane Benoit
Department of Agriculture
Washington, D.C.

Recorder

Judith Newton

THE VIEW FROM BUREAU LEVEL

Speaker

Ted Albert
U.S. Geological Survey
Reston, Virginia

ABSTRACT

The United States Geological Survey became one of the first agencies to establish a data administration function in the early 1970's. A data directory was built to organize the diverse types of data collected by many independent scientific projects. Problems of support by staff, diversity and volume of data, and physical storage were addressed.

The United States Geological Survey (USGS) was one of the early organizations in the establishment of a data administration function as such. My concern when I became data administrator was to create and implement an infrastructure. At that time there were no rules. By the early seventies, the USGS began to be concerned about the quantity of their data. The major problem was the autonomy of the three major divisions of the USGS, and the lack of coordination of the data gathering mechanisms between them.

- The Water Resources Division, headed by the Chief Hydrologist of the United States, has responsibility for all water supply information, monitoring all lakes and rivers, modeling of aquifers, and doing water assessments.

- The Geologic Division has responsibility for earthquake networks, mapping the outer continental shelf, test-drilling on the North Slope, mineral resource assessment, international geology, extraterrestrial data analysis, etc.

- The National Mapping Division, which makes all the base maps of the United States, archives and processes all Landsat data, etc.

The USGS employs 8,000-10,000 people scattered worldwide, and has computers in 36 locations, including mainframes, minis, and microcomputers. All these projects may operate completely autonomously, including buying and operating their own computer, and generating large amounts of scientific data.

The main problem concerns the diffuse and varied missions of the USGS. Additionally, most data captured from satellites is digitized. This results in huge databases.

Originally, the data administrator's function was placed in the director's office. Due to budget cuts, the function was moved to the Information Systems Division.

In the beginning, as I mentioned, there were no rules. I decided to build a directory. This task took five years. The initial information collection effort resulted in an automated system called the Earth Science Information System (ESIS), a highly sophisticated product which included on-line update, remote access, key word capability, even key word generation from text. It is the only directory available for earth science data.

The next step was to add a data dictionary to list all data elements in all the databases. This was done.

Ideally, a system of standards should follow the dictionary development. But getting scientists to agree to standardize is very difficult. An agreement was instituted with the National Bureau of Standards to give USGS key-agency responsibility for all earth science data standards. Our program is now recognized internationally. Some standards have been published as FIPS and also as ANSI standards. This has led to recognition among scientists that this kind of data can indeed be standardized.

Physical storage is another problem with large amounts of data. Thousands of tapes are being stored around the country and more accumulate constantly. Much of this data is static. Laser optical disks are being considered for long-term data storage and multiple use of data.

Another problem we have been looking at is compatibility of databases. The use of artificial intelligence techniques would let diverse data sets interface with each other.

Overall, the program has been successful. We built the directory, started the standards program, raised the consciousness of the scientists with respect to data management, and started new programs which will benefit the USGS in the long term.

We did have some problems when we started to implement the program. We started from the top down, with top management support but without that of the division chiefs. They were more concerned with day-to-day problems than the big picture. Another problem is dissemination of data which is now improving.

The data administration function has moved out of the director's office into the Information Systems Division. The level of support and input, however, remains high.

The Information System Council has input into data handling. The Earth Science Information network allows scientists access to the collected data. We hope to develop the National Directory for Earth Science Data. We continue to expand through study of new technology, especially in communications.

Finally, a word of advice; start from the bottom up to gain support early in the program. Of course, upper level support is also essential.

BIOGRAPHICAL SKETCH

Ted Albert is a Data Administrator in the Information Systems Division of the USGS. He has been in his present position since 1978. He is responsible for oversight, national coordination, and overall policy relating to scientific, technical, and spacial information systems and technical databases throughout the USGS. He belongs to numerous professional and technical organizations and is the author and co-author of several papers.

THE VIEW FROM DEPARTMENTAL LEVEL

Speaker

John Coyle
Department of the Interior
Washington, D.C.

ABSTRACT

The Department of the Interior has a four-year old effort underway to improve management of information resources and the information resource throughout the Department.

The Department of the Interior (DOI) consists of many diverse bureaus with great geographical dispersion. The dispersion is abetted by structure: a small central headquarters over regional centers. Independence and decentralization characterize the organizational culture of DOI. The computer profile reflects this organization (see figure 1).

Among the data DOI must manage are such collections as (figures 2-3):

- o Land Management Information
- o Royalty Accounting Information
- o Earth Sciences Data
- o Mapping Data
- o Fire Management
- o Engineering Technical Applications
- o Wildlife Information
- o Parks Management Data
- o Indian Tribes
- o Construction Data
- o Minerals Data
- o Administrative Data

In 1980, there was a major initiative to improve management of information resources and the information resource in DOI. DOI may have been the first Federal department to establish an Office of Information Resources Management. The office is responsible for ADP management, telecommunications management, records management, library data, management analysis, office automation management, and data administration.

Three major priorities were established:

- o a long-range plan for improving management of information resources
- o program for assessing the state of information resource management

o an information resources directory

Goals of the Data Administration program included the following:

- o Assure consistent and timely data
- o Identify similar needs for data
  - reduce cost of collection
  - reduce redundancy of storage
  - reduce duplicate reporting
- o Identify conflicting and extraneous data
- o Improve Information System planning, development, documentation, and maintenance through central directory plus related directories in the Bureaus with active data dictionaries.
- o Other goals (see figure 4).

Figure 5 details the methods planned to achieve the goals of the Data Administration program. Data sharing was one of the most important, as there was almost no inter-bureau thinking about this issue.

Accomplishments have not yet measured up to goals. Figure 6 shows the progress to this point. Reasons for this include both general management and resource issues.

The need for data administration is finally being brought to the attention of top management by the realization that several projects are suffering from lack of application. An Information Resources Management Review Council has been meeting to review projects. The organization of the Department is being tightened up from a 'loose confederation' to a more consolidated profile. This will affect information management in a positive manner --specifically, through:

- o Increased cross-bureau sharing of databases
  - surface management data
  - multi bureau interest in royalty payments
  - digital mapping projects
  - geographic information systems
  - shared resources data
- o Common administrative systems with host stewardship assigned
- o Common programmatic systems in the future (perhaps)

BIOGRAPHICAL SKETCH

John Coyle is the Chief of the Program Development Division, Office of Information Resources Management, Department of the Interior. He is responsible for strategic planning, policy development, and establishing standards for information resources, including ADP, telecommunications, library services, records management, and data administration.

United States Department of the Interior

OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20240

DOI INVENTORY OF COMPUTER SYSTEMS
Source: Primarily GSA Hardware Inventory - March 1984

Bureau	Mainframes	Minicomputers		Other Computer Services Used by the Bureaus
WGS	AM -2 HON-2	PRIME - 65 DEC - 31 DG - 17 PKE - 8 HP - 11 Harris- 3 Wang - 1	SEL - 2 DPE - 3 BUR - 2 HON - 1 IBM - 1 MOD - 1 NDI - 1	
WBR	CDC-2	DEC - 8 DG - 1 HP - 5 PKE - 2 MOT - 2	MOD - 3 SYA - 1 NCR - 1 GRI - 1	
WBM	BUR-2	HP - 6 DG - 3 DEC - 1	ITD - 1 PKE - 1	WGS-GPCC
LLM	HON-2	DG - 4		
LSM				WGS-GPCC WBM-GPCC Boeing Computer Services
LMS		DEC - 5 HP - 1	PKE - 4	
FWS		HP - 1 DG - 1		
FNP		HP - 2		Boeing Computer Services
BIA		BUR - 8		WGS-GPCC WBM-GPCC Martin Marietta Computer Services
OS		HP - 2		
TOTALS	10	211		

Figure 1

United States Department of the Interior

OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20240

INFORMATION ENVIRONMENT

- o LAND MANAGEMENT INFORMATION
  - Description, Title
  - Legal Status, Lease Data
  - Resources Data: Timber, Minerals, Oid, Grazing
  - Surface Mining and Reclamation
- o ROYALTY ACCOUNTING INFORMATION
  - Production Levels
  - Royalties Collected
  - Royalties Distributed to States and Tribes
- o EARTH SCIENCES DATA
  - Water Resources
  - Geologic Structure
  - Minerals of the U.S.
  - Seismology
  - Land Sat Data
- o MAPPING DATA
  - Digitized Cartography
- o FIRE MANAGEMENT
- o ENGINEERING TECHNICAL APPLICATIONS
  - Hydro-power plant control
  - CAD
  - Economic Modeling

Figure 2

INFORMATION ENVIRONMENT (CON'T)

- o WILDLIFE INFORMATION
  - Hatchery Data
  - Bird/Duck Population
  - Other Wildlife
  - Refuge Management
- o PARKS MANAGEMENT DATA
- o INDIAN TRIBES
  - Tribe Geneology
  - Entitlements
  - Reservation Management
  - Education
- o CONSTRUCTION DATA
  - Refuges
  - Parks
  - Reservations: Schools, Housing
  - Dams
- o MINERALS DATA
  - Mine Production, Worldwide
  - Mineral Reserves, Worldwide
  - Minerals Resources, Worldwide
- o ADMINISTRATIVE DATA
  - Payroll/Personnel
  - Finance/Payments/Budget
  - Property/Space
  - Aircraft Management

Figure 3

United States Department of the Interior

OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20240

WHAT WE WERE TRYING TO ACHIEVE THROUGH ESTABLISHING A DATA ADMINISTRATION PROGRAM

- ASSURE CONSISTENT AND TIMELY DATA
- IDENTIFY SIMILAR NEEDS FOR DATA
  - Reduce Cost of Collection
  - Reduce Redundancy of Storing Data
  - Reduce Duplicative Reporting of Data
- IDENTIFY CONFLICTING AND EXTRANEOUS DATA
- IMPROVE INFORMATION SYSTEM
  - Planning
  - Development
  - Documentation
  - Maintenance
- IMPROVE KNOWLEDGE OF EXISTING INFORMATION
- IMPROVE ACCESS TO EXISTING INFORMATION
- IMPROVE RESPONSE TO MANAGEMENT REQUESTS
  - Timeliness
  - Correct
- BETTER MANAGERIAL AND EXECUTIVE DECISION-MAKING
- IMPROVED PRODUCTIVITY

Figure 4

HOW WERE WE GOING TO ACHIEVE THE GOALS

- o DEVELOP AN OVERALL STRATEGY STATEMENT
- o DEVELOP AN INFORMATION RESOURCES DIRECTORY
  - One of three highest OIRM priorities
  - Contract Dollars Available
- o DATA STANDARDS DEVELOPMENT
  - Interbureau groups to be formed
  - Coordination
- o DEVELOP AWARENESS OF NEED FOR DATA ADMINISTRATION
- o DEVELOP PEOPLE
  - Identify Bureau Cohorts
  - Training
- o DEVELOP POLICIES
  - Establishing the Program
  - Data Standards
  - Use of Data Base Management Technology
  - Data Dictionaries and Systems Development
  - Data Planning and Systems Development
  - Sharing of Data

Figure 5

United States Department of the Interior

OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20240

WHAT DID WE ACCOMPLISH

- STRATEGY DEVELOPED AND PUBLISHED, BUT . . .
- IRD FUNCTIONAL REQUIREMENTS 20% COMPLETE
  - Contract Suspended
- DATA STANDARDS ACTIVITY
  - One Standards Group Formed for Earth Sciences by USGS
  - Progress Measured in Geologic Time
- AWARENESS DEVELOPMENT
  - Some Greater Degree
  - Due to Our Efforts or General Noise in the Environment
- PEOPLE DEVELOPMENT
  - Limited Penetration Into Bureaus
    - Still Not Fully Sold
    - Lack of FTEs to Assign to Function
- POLICIES

Figure 6

STRATEGIC DATA PLANNING

Moderator

John Coyle
Department of the Interior
Washington, D.C.

Recorder

Margaret Law

STRATEGIC SYSTEMS PLANNING
CONCEPTS AND APPROACHES

Speaker

Robert H. Holland
Holland Systems Corporation
Ann Arbor, Michigan

ABSTRACT

Strategic systems planning is critical to the success of business, industry, and government organizations in managing their information resources. The MIS manager plays an important role in the implementation of strategic systems planning. The MIS manager must be able to speak in the business terms of the organization to be able to communicate information management concerns to high-level management. Strategic systems planning should address the growth needs, new services, and changes in operating philosophy for the organization. Strategic systems planning begins with a high-level management directed Business Plan, which describes the organization's missions and products. The Business Plan provides management guidance for the detailed Business Model, which defines the functions, processes, activities, information requirements, and entities within the organization. The Business Model can take approximately six to eight months to develop and can be used for a variety of purposes in information resource management (IRM), such as the development of an organization-wide Data Architecture and Information System Architecture.

Strategic data and systems planning is needed in large organizations that depend on information processing. Organizations, particularly in the private sector, are failing due to the lack of strategic data planning. Banks, insurance companies, and manufacturers trying to stay up-to-date with developments in their industries are failing because they have not done this planning adequately. The failure of these industries affects the Federal Government. In addition, strategic systems planning should have implications for the Federal Government's use of information resources.

The Federal Government should have considerable interest in understanding the directions that industry is taking in managing information resources. Organization after organization fails due to insufficient strategic data planning. For example, in the rapidly changing banking industry, the banks that cannot offer competitive services are losing their clients and going out of business. There is a tremendous pressure on the banking system to update and provide new services. On the other hand, banks that try to change too quickly find that they are unable to manage their information resources; these banks easily become over-extended and become "problem banks." This is true for many other areas of the private sector as well.

The resolution of these problems depends on the ability of the organization to recognize the need for and provide adequate strategic systems planning. It is the Management Information Systems (MIS) manager, and the support given this task within the organization, that will determine whether the strategic systems planning effort will be a success or failure.

The MIS manager has many technical responsibilities, but the success of the strategic systems planning effort will largely be determined by that manager's ability to speak in the business terms of the particular industry, such as in banking terms, rather than in technical terms such as bits and bytes.

The MIS manager's area of management includes:

- o Information processing hardware
- o Office automation
- o Networks
- o Databases
- o Software tools
- o Database management systems
- o Fourth generation languages (4GL)
- o Information systems

In addition to these areas of responsibility, it is important that the MIS manager understand the business or industry of the organization, and be able to communicate technical needs through business planning issues.

The business planning issues that the MIS manager should address are:

- o Ability to meet the information needs associated with the growth in government, business, or industry such as the service levels which include the following: (1) number of transactions; (2) volume of transactions; (3) the ripple effect that will be experienced if information availability is cut back, for example, by 25 percent; (4) present data distribution; (5) requirements for future data distribution; etc.

- o Diversification of services as new information services are being provided in business, industry, and government, such as: (1) quality control sampling, when each item cannot be checked due to higher volume of products produced; (2) preventive maintenance, where emphasis is placed on avoiding breakdown of services rather than on repairing breakdowns once they have occurred; etc.

- o Reorganization within government, industry, or business, emphasizing any changes to the organization's existing charter or mission, and changes in the information reporting structure.

- o Broad changes in the organization's operating philosophy (e.g., emphasis on preventive maintenance rather than fixing failed systems after breakdowns) so that the synthesis and integration of data can be provided where needed across the organization.

The Business Plan is an informal description of missions and sub-missions within the organization, which would include descriptions of: (1) the products and services, (2) the operating policy, (3) the expected rate of growth, and (4) any foreseen reorganization or shifting of responsibilities. The Business Plan would demonstrate management's general direction for the organization that would be formalized and expanded in the Business Model.

The Business Model is a structured, high-level functional description of the organization's many management functions and their level of integration to provide the products or services that the organization produces. The Business Model includes: (1) all functions from the executive to operational levels, (2) processes performed within each function, (3) activities performed within each process, (4) the information requirement of each activity, and (5) the entities for each information classification. The resulting total of functions, processes, activities, information requirements, and entities can be a very large number. The data gained through the Business Model can then be used for many purposes, such as to define the Data Architecture of the organization's subject-oriented databases and to define the organization's Information Systems needs.

Problems experienced in the development of a Business Model are typically: (1) not receiving adequate support or cooperation needed in the different management areas and (2) limited information accessibility, where the Business Model has not been automated.

The Business Model provides: (1) documentation of the way every business function works, (2) resource information for future organization planning without the need of redeveloping the information, and (3) lower project development time and costs by allowing the reconciliation of the information needs of multiple projects.

The present use of separate application-oriented databases within organizations will be described in terms of a "fruit salad" analogy. In many existing information resource centers, the information from all the application areas is mixed together in a proliferation of user transactions, reports, and duplicate data. When this same information is clustered into subject area databases, instead of being indiscriminately mixed, a data architecture can be created. With the subject databases, the data architecture provides the foundation for an information system. The information system can, in turn, be organized into project modules that have planned user transaction formats and reports. The project modules, in turn, represent changing collections of business function activities. The information maintained about each business function activity may be used by one or more project modules.

The primary benefits of developing a detailed Business Model was in ascertaining the "time and precedence sequence" in which data is needed to support each activity. Through the development of a good Business Model, it is possible to know the scope of each activity and area of work and the interactions of data used and produced in the multiple activities. The Business Model should be extended through at least three levels of detail to supply sufficient information for the model to be implemented.

Six to eight months should be allowed for the Business Model project depending on the size of the organization:

1. Define business functions; preparation time one month; deliverable is a functional Business Model diagram.

2. Define business processes, activities, information requirements, and time/precedence sequence; preparation time two months; deliverable is a detailed Business Model.

3. Identify and cluster business entities; preparation time one to two months; deliverable is a Data Architecture.

4. Determine milestones, events, and project modules; time of study one to two months; deliverable is in Information System Architecture.

5. Prepare implementation plans and final report; preparation time one month; deliverables are the Implementation Plans and the Final Report.

A successful Strategic Systems Planning project should be performed by staff within the organization that know the organization, not solely by an outside contractor. Contractors can be useful in supplying guidance and limited help. The involvement of staff from the business production end of the organization, with their expertise about the organization, can be a valuable asset in developing a complete and workable product. It is important to establish commitment and support within the organization for the implementation of the Strategic Systems

Plan. Cooperation from each functional area will be needed to provide and validate information in the plan.

The benefits of developing and using a Strategic Systems Plan are:

- o Information resource plans consistent with the organization's goals.

- o Organizational commitment to information resource management (IRM) goals.

- o Reduced project development time and maintenance costs.

- o An integrated data architecture design, which is derived from the functions of the organization, for all information used and produced by the organization.

- o Manageable implementation projects in the development of a comprehensive information system.

- o A comprehensive approach to planning data migration.

- o A planning tool for managing many organization-wide activities.

The potential benefits of developing a good Business Systems Plan could make a substantial difference in the success of the organization and its enterprise.

STRATEGIC DATA PLANNING/U.S. POSTAL SERVICE

Speaker

William Leftwich
U.S. Postal Service
Washington, D.C.

ABSTRACT

A major undertaking for any organization is to bridge the gap between strategic data planning and logical database design. Some of the approaches taken by the U.S. Postal Service (USPS) to do this include: "bottom-up" process defining existing data elements, data architecture or logically grouping of data elements, defining information architecture through data flow diagrams, and business process view of data needs and their relationships from a "top-down" view. The highlights of the "lessons learned" are addressed.

The USPS has completed the strategic data planning effort and is moving toward logical database design. Once a strategic data planning effort such as one that utilizes the Business Systems Planning (BSP) methodology is nearing completion, it is important to look toward its uses and consequences. The following problems were identified as requiring consideration after BSP nears completion.

- o Bridging the gap to logical database design.
- o Relating the corporate database model to individual application projects.
- o Creating a shared data resource in a multi-project environment.

There were several phases of the USPS's work in strategic data planning. The first phase, which was titled, "First Attempts," occurred from April, 1980 through September, 1983 when USPS was attempting to find the correct direction in which to proceed.

USPS began with a "bottom-up" approach of data element definition, identifying and defining the elements in the ongoing application projects. While many data elements were identified, USPS later felt that this effort tended to focus on the existing systems and application projects rather than on the corporate data resource as a whole.

In the same time frame, a data architecture was undertaken by attempting to develop a BSP data plan/data group structure through a logical grouping of the data elements that had previously been defined. USPS personnel later felt that the structure developed from these groupings tended to be more intuitive (i.e., elements were grouped because they looked like they belonged together) rather than analytical. Again the view was generally "bottom-up" which resulted in gaps and overlaps. The resulting data architecture did not provide a clear view of the data entities, relationships, and the business use necessary to support logical database design.

Also, as part of this initial effort, USPS tried to define an information architecture in an attempt to relate the BSP systems/subsystem structure to the data resource. This effort was not considered successful. It did not contribute to the logical database design process, and it did not clearly relate the corporate view of the data resource to the individual application projects.

Following this initial stage of work, a period of reassessment and redirection began, which started in September 1983 and extends into the present. This revised strategic data planning has taken four major forms: (1) a business process view, (2) a revised information architecture, (3) a revised data element definition standard, and (4) a revised data architecture.

The business process view was undertaken to gain a better understanding of the data needs and relationships within USPS. The result was a corporate view of the acquisition, movement, storage, and use of data throughout the USPS enterprise. A "top-down" view of data was provided in which the data structure had been decomposed down to each individual project area. A map of the data was constructed to show where the data flowed.

The revised information architecture resulted in a corporate business process model, or data flow diagram, which showed both a broad/shallow view of the architecture that evolved to broad/deep views. The broad/deep views showed data transactions for individual applications. The new information architecture provides a framework for the integration of individual application projects and for planning data migration.

The revised data element definition standard provides a structure for categorizing the various data relationships. The data relationships clarified are:

- "User view" or "root/role" element relationships.
- Business entity and entity relationships with their groupings of data.
- Data element/data store relationships.

The revised data architecture, which is now underway, provides data groupings based on business entities and entity relationships. A high-level logical database design has begun, along with designs for the supporting subject databases.

The USPS learned several lessons through this entire process. The "lessons learned" were the following:

- The importance of stewardship of the data or taking responsibility for maintaining and sharing the information resource that belongs to all of USPS. The Postmaster General assumed the role of prime steward of USPS data in order to resolve any conflicts about use of data within USPS.

- How to avoid the methodology trap. While many methodologies are appealing and have something to offer the user, it is important to devise a methodology that meets the needs of the organization rather than relying on a methodology that may not fully meet those needs.

- The need to avoid delays to application projects during the strategic data planning effort. While the planning effort is necessary, application projects should not be held up while the planning and data collection are going on.

- How to deal with the "hot" project, where the project is moving ahead so fast that no one has any time to talk to the strategic data planners. The best approach was to leave the "hot" project alone, not offering any interference or immediate help, and let the manager of the project come to the planners for help with information about other aspects of USPS. An attempt should be made to anticipate the needs of the "hot" project, which would run into trouble eventually because it was out of communication with other data resources.

- The need to avoid buying computer hardware without a full understanding of the problem and real expertise in the area. USPS found that many novice users rapidly considered themselves data processing experts once they had talked to a few salesmen and been "sold" on the virtues of various hardware devices. Later these users found that the hardware did not solve their problems and could cause more problems later.

These pitfalls should be avoided for an effective strategic data planning effort. USPS learned these lessons the hard way but now they feel they are on course and have a valuable information tool as a result of their efforts.

BIOGRAPHICAL SKETCH

Mr. Leftwich has worked in several areas involving information management and data processing in his 11 years with the USPS. He is currently acting as General Manager, Data Administration Division, Office of Data Management. In this capacity, he is directly responsible for development and maintenance of the corporate Information Systems Architecture, Data Architecture and Directory of Information Resources and, in addition, provides data definition and integration planning support to individual application projects and supports the logical database design process.

He has a BA in Political Science from Memphis State University and a MA, also in Political Science, from Miami (Ohio) University.

ISSUES IN MANAGING DATA

Moderator

Jane Benoit
Department of Agriculture
Washington, D.C.

Recorder

Martha Gray

DATA ADMINISTRATION/PROJECT LIFE CYCLE

Speaker

Susan Arnold Church
Federal Bureau of Investigation
Washington, D.C.

ABSTRACT

Some experiences as data administrator for the Federal Bureau of Investigation are described, in particular, how the data administration activities are integrated into the project life cycle. The discussion described how the review process insures that the database architecture gets built the way that it was planned. The composition of the review board that is responsible for the reviews is also described.

In the FBI, the data administrator reports to the top management of the Technical Services Division. Systems development is centralized, with few exceptions, in the Technical Services Division. I have been data administrator for one and one-half years and have a staff of two, including myself. A previous data administrator had participated in a study which selected ADABAS as the primary DBMS. Before this we had used the Generalized Information Management System (GIMS) DBMS which was developed by TRW for the CIA.

The responsibilities of the data administrator are primarily strategic database planning, what Martin and Holland call "top-down database planning." The basic purpose of this is to develop the subject databases which separate the data from the software that accesses the data. The goals are to reduce redundant data collection, data storage, and data inconsistencies.

Because databases were already being designed using ADABAS, another technique was chosen to get the data administrator involved as quickly as possible. This meant getting the data administrator involved in the subject database design or what Martin and Holland call "bottom-up database design." The applications programmers proceed with their database design, synthesizing the data requirements from the user views, and forming a conceptual data model. This conceptual data model is based on a relational data model because we believe that this allows us to get the most reliable and flexible databases. Next we look at the DBMS data model. Because ADABAS uses a tabular model, we do not get a bad fit with the relational model. From this fit we develop a logical data model. Next, the physical data model is developed taking into consideration performance constraints, transaction volumes, and data usage requirements.

I will next review the normal project life cycle to highlight where database development fits in (see figures 1 and 2). In the FBI, the data administration staff is not large enough to perform each subject database design. The data administration staff does not review the process until the conceptual data model has been developed by the project. The database review of the conceptual data model is the first data administration review point. At this point, the following areas are considered: are the scope and interfaces of the database appropriate; does the model correctly fit the relational model; and have the designers insured the auditability of the data and considered all of the data integrity, security, and privacy issues.

The next review point is when the logical data model has been developed. The data administrator looks at how the designers have handled the constraints of the DBMS and how they have implemented the data integrity, auditability, security, and privacy requirements given in the conceptual data model. When the physical data model has been developed, the data administrator looks at how the implementation addresses the performance requirements and how the test plan covers the data integrity, auditability, security, and privacy requirements. The database administrator develops a testing environment for the designers who program a test database. Results of testing are reviewed to see if the test plan was carried out. After this review is complete, the database goes into a production mode and some time later is evaluated for how well it has technically implemented all of the requirements.

The Database Review Board that handles this review process is composed of the following members:

Data Administrator - Chair of the STUDY and EVALUATION phase reviews
Database Administrator - Chair of the DESIGN phase reviews
Information Systems Auditor - Chair of the TEST phase reviews
ADP/Telecommunications Security Officer
Systems Analysts - representing interested/affected applications

The critical factors for data administration's success seem to be having three kinds of support:

1. management support - both organizational authority and functional authority. This authority needs to be clearly defined in a charter.
2. automated tools - data modeling tools, a Database Management System (DBMS) and a Data Dictionary System (DDS). Of these, the Data Dictionary System is probably

the most important followed by the Database Management System.
3. database administration support - necessary for DDS/DBMS integration and configuration management. In our organization, the database administrator does not report to the data administrator, but we do work closely together.

To give a little more detail on the data dictionary system, let me review what we do with the system. First, we have built an inventory of many environments, including ADABAS systems, other database systems, and manual systems. Second, we have developed an on-line directory for systems analysts and end-users. We are now adding the natural language query system, INTELLECT, so that end-users can use the data dictionary system without even using the screen menus. Third, we use the DDS to record our strategic data planning entities and our relational data modeling entities as conceptual entities in the DDS. From these we can develop the physical database design directly, calling upon the data standards that are also recorded in the DDS. Fourth, we use the DDS to interface with the auditors for handling internal audits.

There are three organizational interfaces that I have found extremely useful. The first is being placed organizationally on the Planning and Administration Staff. This is where long-range automation planning, budget planning, and budget execution is accomplished and allows the data administrator to present strategic data plans and see that they get funded and implemented at the project level.

The second useful organizational interface is the Systems Review Board. The Systems Review Board is a good way for the data administrator to review overall project plans and progress. The third is the Data Access Policy Committee which serves as a forum for promoting data sharing. The composition of these boards is given in figure 3.

As a final note, there are two areas that will provide challenges to the future of data administration. The first is rapid prototyping. We need to determine how rapid prototyping will fit into the system life cycle without sacrificing database quality. The second is end-user computing. We need to interface the data dictionary system with end-user databases.

BIOGRAPHICAL SKETCH

Ms. Susan Arnold Church is Data Administrator for the Technical Services Division at the Federal Bureau of Investigation. She holds a B.A. degree in mathematics from Washington College in Chestertown, Md., and an M.S. degree in computer science from the

Johns Hopkins University. She was awarded the ICCP's Certificate in Computer Programming in 1979 and the Certificate in Data Processing in 1980. She began her career with the Federal Government as a mathematician at the National Security Agency in 1971, and joined the FBI in 1976.

LIFE CYCLE REVIEW POINTS

SYSTEM LIFE CYCLE PHASES*	DATA BASE REVIEW POINTS
STUDY: SERVICE REQUEST; SYSTEM REQUIREMENTS DEFINITION; SYSTEM DESIGN ALTERNATIVES	CONCEPTUAL DATA MODEL
↓	
DESIGN: SYSTEM EXTERNAL SPECIFICATIONS; SYSTEM INTERNAL SPECIFICATIONS	LOGICAL DATA MODEL PHYSICAL DATA MODEL
↓	
PROGRAM: PROGRAMMING; TESTING	TEST DATA BASE
↓	
IMPLEMENT: CONVERSION; IMPLEMENTATION	
↓	
EVALUATE: POST IMPLEMENTATION REVIEW	PRODUCTION DATA BASE

---

*SYSTEM DEVELOPMENT METHODOLOGY(SDM)/70, AGS Management Systems, Inc.

Figure 1

DATA BASE REVIEW BOARD

REVIEW CRITERIA	
•Conceptual Data Model	STUDY PHASE -appropriate data base scope and interfaces -relational model -completeness of security, privacy, integrity and audit requirements definition
•Logical Data Model	FUNCTIONAL DESIGN PHASE -consideration of DBMS data model and other constraints in logical data model -implementation of security, privacy, integrity and audit requirements with the DBMS
•Physical Data Model	DETAILED DESIGN PHASE -consideration of data usage and other performance requirements in physical data model -adequacy of Test Plan in covering security, privacy, integrity and audit controls
•Test Data Base	TEST PHASE -adequacy of Test Plan execution
•Production Data Base	EVALUATION PHASE -data base performance in terms of technical, operational and economic measures of effectiveness

Figure 2

OTHER ORGANIZATIONAL INTERFACES

DATA ACCESS POLICY COMMITTEE

MEMBERS:

DATA ADMINISTRATOR, Chair of standing committee

USER DIVISION REPRESENTATIVES, representing data base sponsors

GENERAL POLICY:

Data Base "Sponsors" are responsible for authorizing data access

COMMITTEE CHARTER:

Encourage data sharing within security, privacy, and data integrity constraints;
Review data base access requirements and access controls;
Recommend general policy and extensions to specific data bases to management.

SYSTEMS REVIEW BOARD

MEMBERS:

TECHNICAL SERVICES DIVISION DIRECTOR, Chair of Systems Reviews

CHIEFS OF SYSTEMS DEVELOPMENT and OPERATIONS

CHIEF OF PLANNING AND ADMINISTRATION (ADP/T budget)

ADP/T SECURITY OFFICER

INFORMATION SYSTEMS AUDITOR

DATA ADMINISTRATOR

BOARD CHARTER:

Provide for system control throughout the management cycle

Figure 3

STRATEGIC DATA PLANNING/IMPLEMENTING SYSTEMS

Speaker

Marianne Russek
Federal Reserve Board
Washington, D.C.

ABSTRACT

A description will be given of how the Federal Reserve Board has used two strategic data planning products, in particular, the business model and a data architecture, in various activities, especially implementing systems. The Federal Reserve Board has been working vary hard to separate data from the work that is done to promote data sharing (see figure 1).

The Federal Reserve System is comprised of 12 reserve banks and districts across the U.S., the Federal Reserve Board of Governors in Washington and an Automation Program Office in Dallas, Texas. The Federal Reserve Banks' functions include loans to banking institutions, operating a nationwide network for clearing checks and electronic payments, supplying as much coin and currency as the public needs to carry on its business, selling Treasury bills, bonds, and notes, and regulation of other banking institutions. The Federal Reserve Board of Governors handles complex regulation activities, does economic research that is the backbone of setting monetary policy, and handles truth in lending responsibilities. The Automation Program Office coordinates automation activities in the 13 sites.

In 1980, a long-range plan was developed which specified that the Federal Reserve would standardize on hardware, software, a database management system, and a data dictionary. A business model of the Federal Reserve System was completed. The business model committee was co-chaired by a member of the Board of Governors and a member of the Automation Program Office. The model, finished in 1981, had 21 functions, 100 processes, 500 activities, and 90 information requirements or entities. From this, eight projects were specified. They would be resource shared, that is, they would be developed by one or two banks but when completed would be used by all 12 banks and in some cases by the Board of Governors. Six out of eight are now in production in at least one bank, and many are at more than one bank.

The Board of Governors needed a business model which was different from that of the banks because of the Board's different responsibilities. However, the co-chair had trouble convincing top management that a different model was needed so a different approach was used.

Since 1979, there have been standards in place which specified that for every application project there would be a data administration representative for data analysis and a database administration representative for the logical database design. They also specified that there would be eight database quality reviews. The first of these was for data element definitions. At the Federal Reserve Board, users have to sign off on all data element definitions.

For two projects that were in the design phase, the data administration office offered to do a business model during the design phase. These projects covered four large functions. Figure 2 shows a detailed listing from the business model. From this, they ran the entity analysis program (see figure 3 for the results) and developed the subject database architecture. Note that every entity has a four character abbreviation and each subject database has a two character abbreviation. Every database must start with the two character subject database abbreviation (see figure 4). These abbreviations are enforced through the data dictionary. All COBOL records are generated from the data dictionary as well as all database management systems structures.

The Federal Reserve Board also uses the business model to help manage the scope of application projects. What Dr. Holland said about projects being divided into project modules that can be accomplished in six months to one year should be stressed. One of the projects suffered from the mythical labor months syndrome and did not get completed when planned. This is a good way to give data administration a bad name. The data administration office also used the business model to work with applications programmers and users and found the model a good way to show users what the project would do. The business model was also a good tool for training new applications programmers to bring them up to speed more quickly.

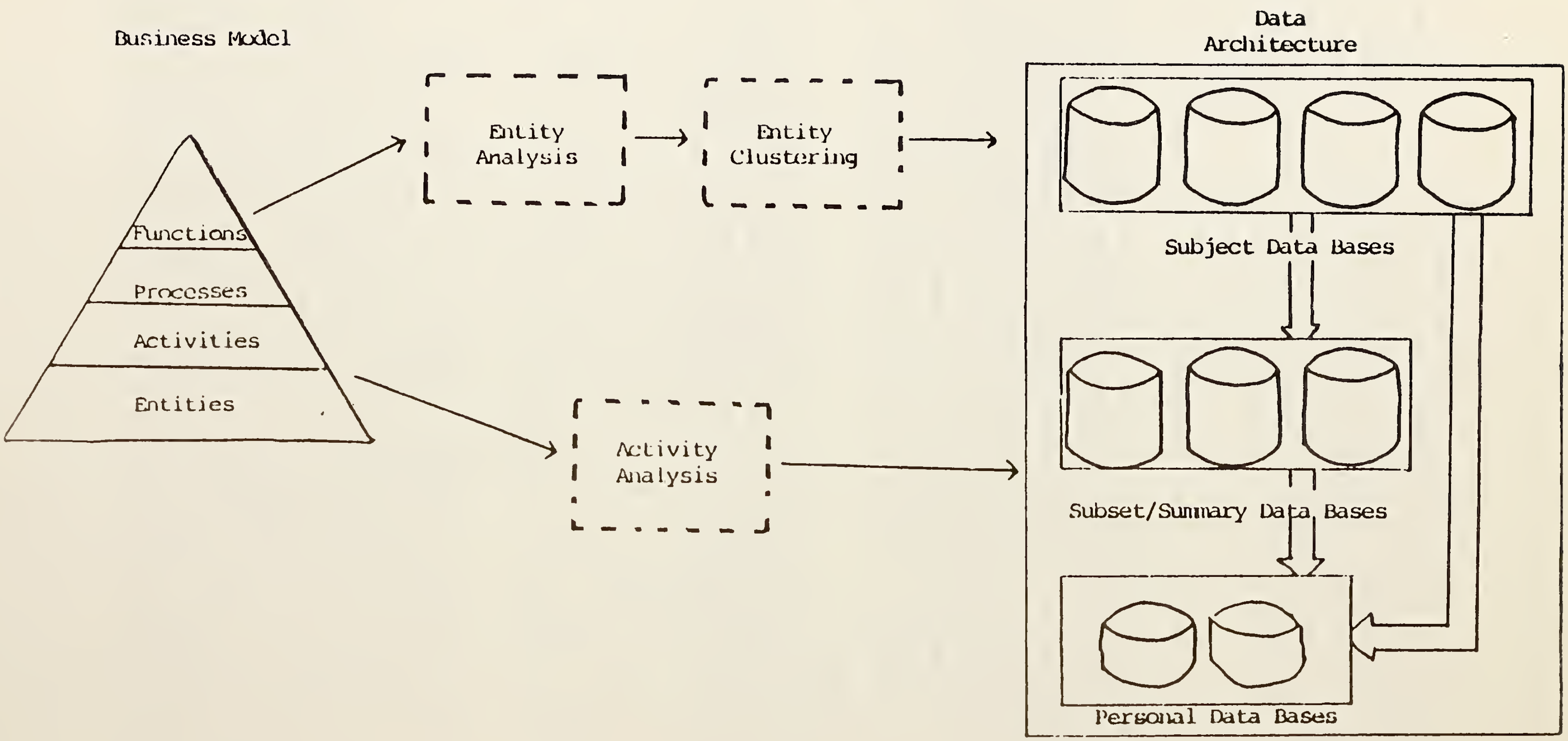
This project did not, however, help to get any more functions modeled. In 1984, a new senior director came to the Board and issued a directive to finish the business model. The data administration office was given three staff years to finish the model, but in fact they only used one and a half staff years. The reason that it took that much time and effort was that they interviewed every line manager. After they consolidated their model, they submitted it to top management for approval. There are still four small offices left to be modeled, all belonging to the senior management function. These should be completed in spring of 1985. Data administration's staff believe that the data architecture will remain intact after these offices are done, and that the subject databases and entities listed in figure 5 are accurate.

At this time, the Board is not interested in new application projects which will generate subject databases. Rather, the priority is getting data from the mainframe to office systems and personal computers. The data administration section is using the data architecture to explain the costs of doing this for the users and to help manage the users' expectations. Figures 6 and 7 show the views of data administration on the differences between subject databases and summary databases. They use the data architecture to control data redundancy and to help develop pictures for the users to help explain various applications. This is very important as the Board moves to automate offices.

By now, the business model is secure and in place. New systems are built on the foundation of the business model and the data architecture. Other architectures such as office automation and communications are related to this data architecture. The business model is used for informing new users, training new applications people, and for planning contingency processing. The data architecture is used not only to ensure that new databases are stable but also that new systems will be well integrated.

BIOGRAPHICAL SKETCH

Ms. Marianne Russek has been the Data Administrator for the Federal Reserve Board since August 1982. Prior to that time, she has worked for the Federal Government and private firms as a systems analyst and systems and applications programmer.



STRATEGIC DATA PLANNING PRODUCTS

Used to Implement Systems

Figure 1

```
                         FEDERAL RESERVE BOARD
                     STRATEGIC DATA BASE PLANNING
          REPORT NUMBER 18                  DATE: 10/15/84
                                            TIME: 19:30:04        PAGE:    8
OPTION: FUNCTION THRU ENTITY                                      SEQUENCE: FUNCTION

   PROCEDURE   400  SURVEY BENEFIT TRENDS TO ASSURE BENEFIT PROGRAMS ARE UP-TO-DATE AND
                    EFFECTIVE

 PROCESS   600  PREPARE AND DISTRIBUTE PAYROLL

  ACTIVITY  100  PRODUCE EMPLOYEE PAYROLL

   INFO.RQMT    14  FR-400, PERSONNEL ACTION REQUEST
   INFO.RQMT    27  SALARY SCHEDULES
   INFO.RQMT    40  BENEFIT ENROLLMENT FORMS
   INFO.RQMT    52  LWOP INFORMATION
   INFO.RQMT    53  TIME & ATTENDANCE REPORTS (FOR WAE'S)
   INFO.RQMT    54  OATH OF OFFICE FORMS
   INFO.RQMT    55  EXTRA WORK FORMS (OVERTIME CLAIMS; NIGHT, SUNDAY, HOLIDAY DIFFERENTIAL)
   INFO.RQMT    56  TAX FORMS
   INFO.RQMT    57  SAVINGS BOND FORMS
   INFO.RQMT    58  DIRECT DEPOSIT FORMS
   INFO.RQMT    59  SAVINGS DEPOSIT FORMS
   INFO.RQMT    60  GARNISHMENTS
   INFO.RQMT    61  TAX LEVIES
   INFO.RQMT    62  IRS DIRECTIVES
   INFO.RQMT    63  RETIREE COLA CHANGES
   INFO.RQMT   202  LUMP SUM PAYMENT MEMO

   ENTITY   0004  EMPLOYEE
   ENTITY   0010  BOARD ACCOUNT
   ENTITY   0011  BENEFIT
   ENTITY   0018  PERSONNEL TRANSACTION

   PROCEDURE   100  ASSEMBLE PAYROLL INFORMATION
   PROCEDURE   200  INCORPORATE ALL LEGISLATIVE AND MANAGEMENT ACTIONS THAT AFFECT PAYROLL
   PROCEDURE   300  PROCESS PAYROLL TRANSACTIONS

  ACTIVITY  200  DISTRIBUTE PAYROLL TO EMPLOYEES AND INSTITUTIONS

   INFO.RQMT    64  CREDIT UNION TAPE
   INFO.RQMT    65  EFTS TRANSMISSION
   INFO.RQMT    66  BANK INFORMATION
   INFO.RQMT    67  EARNINGS STATEMENTS
   INFO.RQMT    68  CHECKS
   INFO.RQMT    69  EMPLOYEE AND BOARD BENEFITS PAYMENTS
   INFO.RQMT    70  AUTHORIZATIONS TO RICHMOND FRD

   ENTITY   0004  EMPLOYEE
   ENTITY   0007  BOARD ACCOUNT TRANSACTION
   ENTITY   0010  BOARD ACCOUNT
   ENTITY   0011  BENEFIT
   ENTITY   0013  CUSTOMER
```

Figure 2

The Board uses PAF programs developed by the Federal Reserve Bank of Dallas to produce business model reports.

FEDERAL RESERVE BOARD
STRATEGIC DATA BASE PLANNING
REPORT NUMBER 19 DATE: 02/19/85
TIME: 18:33:34

OPTION: ENTITY AFFINITIES
MINIMUM AFFINITY: 0.01

PAGE: 2
SEQUENCE: ENTITY/AFFINITY

FIRST ENTITY	SECOND ENTITY	PR CT	E1 CT	AFF.	REV. AFF.	AVG. AFF.
E0003 BOARD ORGANIZATION	E0022 SUBSCRIPTION	4	289	0.01	0.80	0.41
E0003 BOARD ORGANIZATION	E0031 ACCOUNT GENERAL LEDGER	4	289	0.01	0.31	0.16
E0003 BOARD ORGANIZATION	E0035 ACH SERVICES	2	289	0.01	0.13	0.07
E0003 BOARD ORGANIZATION	E0038 BUILDING	5	289	0.02	0.17	0.10
E0003 BOARD ORGANIZATION	E0040 CASH SERVICES	2	289	0.01	0.13	0.07
E0003 BOARD ORGANIZATION	E0042 CHECK SERVICES	2	289	0.01	0.13	0.07
E0003 BOARD ORGANIZATION	E0047 COMPUTER JOB	34	289	0.12	0.83	0.48
E0003 BOARD ORGANIZATION	E0050 DATA PROCESSING STANDARDS	100	289	0.35	0.82	0.59
E0003 BOARD ORGANIZATION	E0051 EXAMINATION	8	289	0.03	0.24	0.14
E0003 BOARD ORGANIZATION	E0053 FUNDS SERVICES	2	289	0.01	0.13	0.07
E0003 BOARD ORGANIZATION	E0054 ORGANIZATION	77	289	0.27	0.56	0.42
E0003 BOARD ORGANIZATION	E0055 NETWORK	36	289	0.12	0.92	0.52
E0003 BOARD ORGANIZATION	E0056 NONCASH SERVICES	2	289	0.01	0.22	0.12
E0003 BOARD ORGANIZATION	E0057 ORGANIZATION STANDARDS	12	289	0.04	0.35	0.20
E0003 BOARD ORGANIZATION	E0062 REGULATION LAW	75	289	0.26	0.54	0.40
E0003 BOARD ORGANIZATION	E0065 SECURITIES SERVICES	3	289	0.01	0.20	0.11
E0003 BOARD ORGANIZATION	E0066 TT&L SERVICES	2	289	0.01	0.22	0.12
E0003 BOARD ORGANIZATION	E0077 SUPERVISION SERVICES	7	289	0.02	0.17	0.10
E0003 BOARD ORGANIZATION	E0079 LEGAL STANDARDS	3	289	0.01	0.27	0.14
E0003 BOARD ORGANIZATION	E0082 CORRESPONDENCE	10	289	0.03	0.19	0.11
E0003 BOARD ORGANIZATION	E0084 LEGAL LIBRARY	15	289	0.05	0.27	0.16
E0003 BOARD ORGANIZATION	E0086 RECORDS	90	289	0.31	0.66	0.49
E0003 BOARD ORGANIZATION	E0087 REGULATION TRANSACTION	8	289	0.03	0.35	0.19
E0003 BOARD ORGANIZATION	E0089 SERIES PANEL	5	289	0.02	0.83	0.43
E0003 BOARD ORGANIZATION	E0092 SAVINGS BOND	42	289	0.15	1.00	0.58
E0003 BOARD ORGANIZATION	E0098 PROGRAM/VISIT	2	289	0.01	0.18	0.10
E0003 BOARD ORGANIZATION	E0100 SUPPLIES	3	289	0.01	0.38	0.20
E0004 EMPLOYEE	E0001 BOARD ACCOUNTING STANDARDS	5	114	0.04	0.33	0.19
E0004 EMPLOYEE	E0003 BOARD ORGANIZATION	56	114	0.49	0.19	0.34
E0004 EMPLOYEE	E0005 RESEARCH SERVICES	2	114	0.02	0.03	0.03
E0004 EMPLOYEE	E0006 VENDOR	20	114	0.18	0.16	0.17
E0004 EMPLOYEE	E0007 BOARD ACCOUNT TRANSACTION	12	114	0.11	0.92	0.52
E0004 EMPLOYEE	E0008 PERSONNEL STANDARDS	53	114	0.46	0.69	0.58
E0004 EMPLOYEE	E0009 COURSE	17	114	0.15	0.39	0.27
E0004 EMPLOYEE	E0010 BOARD ACCOUNT	34	114	0.30	0.41	0.36
E0004 EMPLOYEE	E0011 BENEFIT	10	114	0.09	0.91	0.50
E0004 EMPLOYEE	E0012 CONTRACT	10	114	0.09	0.28	0.19
E0004 EMPLOYEE	E0013 CUSTOMER	10	114	0.09	0.20	0.15
E0004 EMPLOYEE	E0014 MONETARY POLICY	3	114	0.03	0.18	0.11
E0004 EMPLOYEE	E0015 EQUIPMENT	11	114	0.10	0.19	0.15
E0004 EMPLOYEE	E0016 CUSTOMER SPECIAL	14	114	0.12	0.13	0.13
E0004 EMPLOYEE	E0017 PUBLICATION	4	114	0.04	0.09	0.07
E0004 EMPLOYEE	E0018 PERSONNEL TRANSACTION	21	114	0.18	0.81	0.50
E0004 EMPLOYEE	E0019 SERIES CUSTOMER	2	114	0.02	0.02	0.02
E0004 EMPLOYEE	E0021 DATA PROCESSING PROJECT	1	114	0.01	0.02	0.02
E0004 EMPLOYEE	E0034 ACCOUNTING TRANSACTION	3	114	0.03	0.12	0.08

Figure 3

ADMINISTRATIVE INFORMATION RETRIEVAL SYSTEM

System Code A1

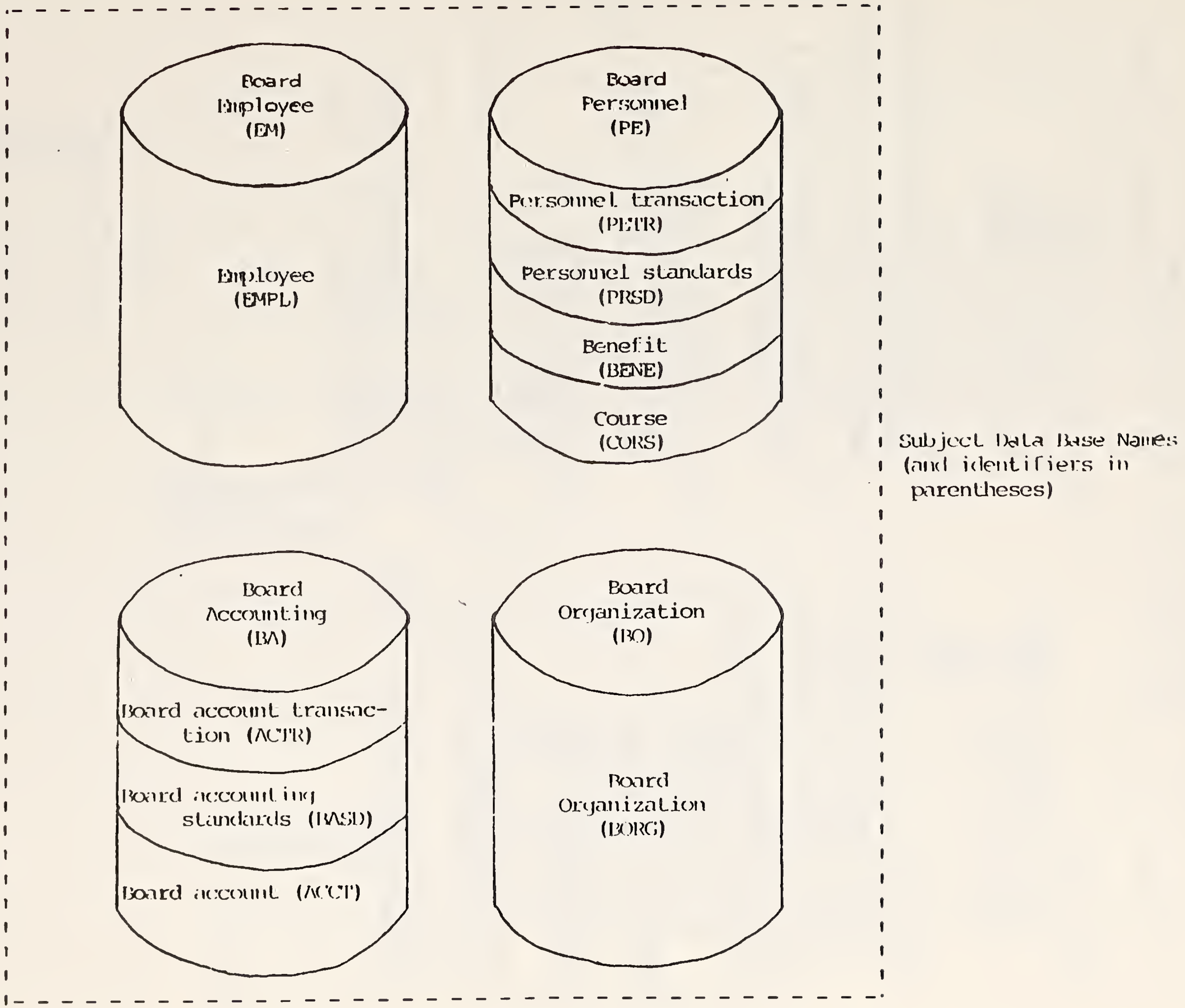


Figure 4

FEDERAL RESERVE BOARD SUBJECT DATA BASES AND
ENTITIES FOR INTERNAL SUPPORT
(Subject data base and entity prefixes shown in parentheses)

LEGAL (LG)
- Legal Standards (LGSD)
- Legal Library (LEGL)
- Contract (CONT)
- Regulation Law (RGLW)

PACS (PA)
- PACS Statistics (PAST)
- PACS Expense (PAEX)
- PACS Standards (PASD)
- Organization (OPGN)

PURCHASE ORDER/VENDOR (PV)
- Purchase Order (PODR)
- Vendor (VEND)

SUPPLIES (SU)
- Supplies (SUPL)

EMPLOYEE (EM)
- Employee (EMPL)

CORRESPONDENCE (CR)
- Correspondence (CORR)

ORGANIZATION (BO)
- Board Organization (BORG)

DATA PROCESSING (DP)
- DP Project (DPPJ)
- DP Standards (DPSD)
- Computer Job (CMJB)

COMMUNICATIONS (CO)
- Communication Transaction (COTR)
- Network (NETW)
- Communications Services (COSV)

PUBLICATION/SUBSCRIPTION (PS)
- Publication (PUBN)
- Subscription (SUPN)

BOARD ACCOUNTING (BA)
- Board Account (ACCT)
- Account Transaction (ACTR)
- Board Accounting Standards (BASD)

EQUIPMENT/BUILDING (EB)
- Equipment (EQUP)
- Building (BLDG)
- Equipment/Building Transaction (EBTR)

PERSONNEL (PE)
- Course (CORS)
- Personnel Standards (PRSD)
- Benefit (BENE)
- Payroll Transaction (PRTR)

Figure 5

Data Architecture

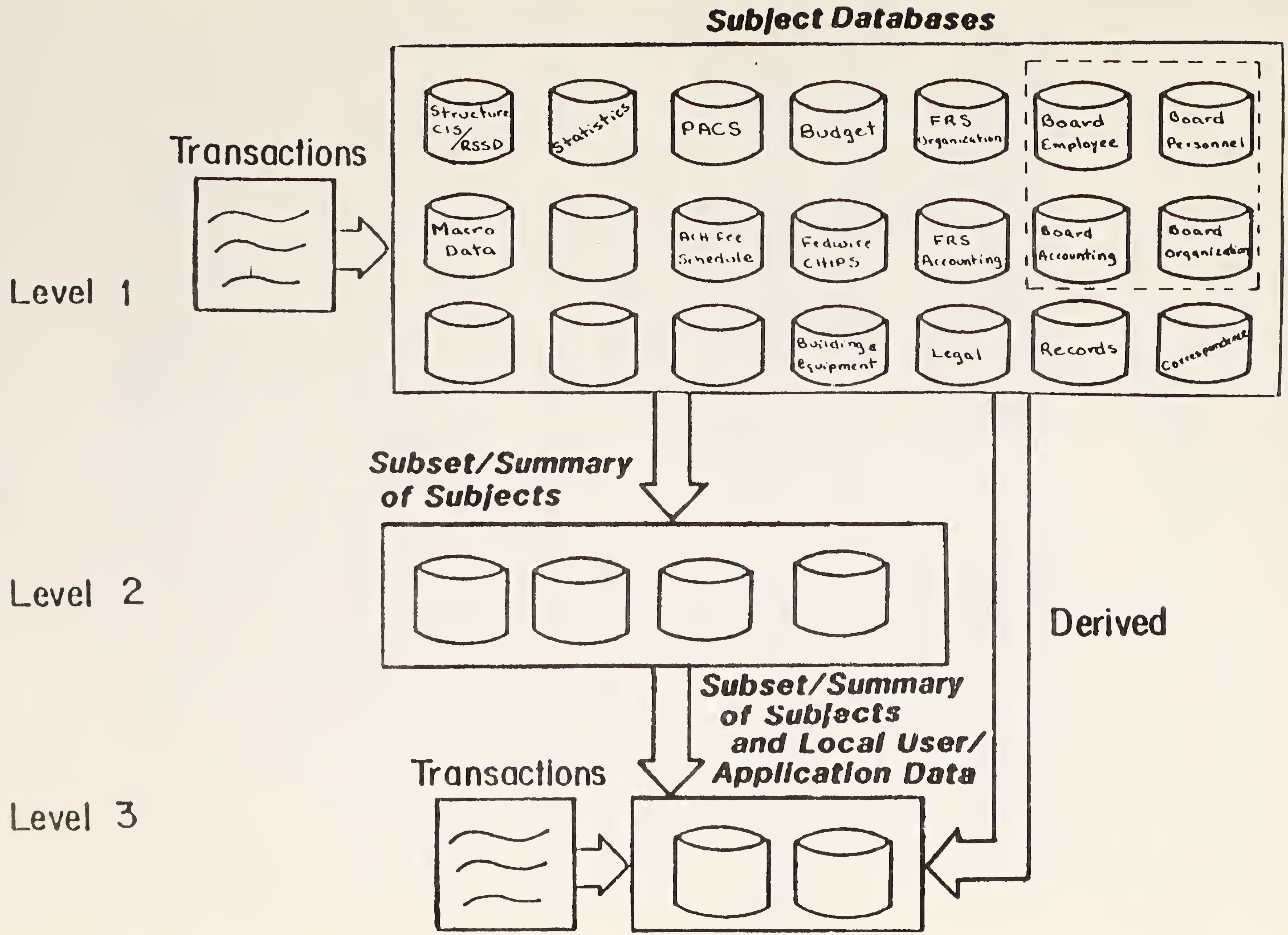


Figure 6

Data Architecture Perspective

Subject Data Bases	Summary Data Bases
- Production	- Management Information
- Transaction-driven	- Decision-support
- Operational	- Analytical
-- Require logical DB design	-- Require logical DB design
-- Complete data continously updated	-- All summary/selected data periodically updated
-- Hierarchical structures	-- Relational structures
-- Inquiries keyed by primary keys	-- Can be searched using multiple keys via powerful user languages
--Simple updates in real time	-- Complex updates (due to indices) offline in periodic runs
-- Large and mainframe	-- Large-mainframe to small-micro
-- Shared	-- Shared or private
- Transaction volume -- high	- Transaction volume -- low
- Design objective -- efficient processing	- Design objective -- ease of use

Figure 7

THE INFORMATION CENTER AND DATA ADMINISTRATION

Speaker

Ron Shelby
Department of the Interior
Washington, D.C.

ABSTRACT

Information centers and data administration can be allies in serving users. To demonstrate this, the interactions between an information center and data administration are described and examined. Data administration support of the information center is discussed. Alternatives for the organizational placement of the information center are also given.

Before everyone becomes discouraged listening to all of the successes of the previous speaker and noticing that their own work environment is not as far along as the Federal Reserve Board, let me emphasize that there are things that data administration can accomplish before strategic data plans and data architectures are completed. The data administrator can set data policy, decide what data standards are needed, and support the information center and its users.

What follows is a description of a scenario wherein data administration supported an information center in meeting end-users data access needs.

In 1980, I was appointed the data administration manager of an insurance company. This data administration unit started supporting applications development in 1981 using a data dictionary/directory system. We developed systems through the data dictionary/directory rather than trying to document systems after they were already developed. In 1983, a data inventory was created to support end-user data access. This inventory was entered into the data dictionary. Also in 1983, the information center and data administration began to report to the same manager. As that manager, I decided to ensure that these two functions worked in harmony to support the corporation.

The objectives of this presentation are straightforward:

1) Outline the interactions between the information center and data administration.
2) Describe how data administration can support the information center's users.
3) Outline the management choices for information center placement in an organization.

Figure 1 is a broad view of data administration and the information center showing the missions of both organizations, their functions, and the tools they use. The function 'manage existing data' means that data administration should help the corporation to manage the data resource effectively by managing metadata describing existing data. If you look over this figure, you will see that there is not much overlap between these two organizations. Note that neither a fourth generation language nor a database system were listed as tools of the information center. This will be discussed later.

Figure 2 highlights the match-up between the functions and tools of data administration and the information center. One of the main functions of the information center is end-user support. The information center can become totally consumed with applications support, or problem solving, and will totally bypass data administration. Data administration can serve a very useful role in helping the information center users locate information and data, especially if a data dictionary contains a data inventory and information center personnel know how to use the data dictionary.

While data planning software and data design software are very specific database-oriented tools, the data dictionary/directory is a multi-use tool. The data inventory contained in a data dictionary is listed as a tool of the information center, since information center users are the primary end-users of this inventory. Originally, the information center staff felt that data administration was a constraint on them. It was very gratifying to see that attitude change over an 18-month period as the information center staff started to use the data inventory in the data dictionary and to view it as one of their tools.

Figure 3 illustrates the forces that tend to draw data administration and the information center closer together. First and most important is that users need to access and manipulate data that has been captured elsewhere. In the insurance industry, users need operational data that has been summarized and placed in a summary database. The users will not know about this data unless they find out about it in the data dictionary. The data dictionary has to be tailored to allow users to access it and locate information about data easily. Figure 4 shows an overview of this process. You should also note that users need to improve data design skills when developing applications. Users such as scientists, accountants, and actuaries, for example, have been writing programs for a long time, whether or not data processing professionals knew about it. These people need help to improve their data design skills. Improving the quality of their data design will help them have more stable applications.

If the user has an application, such as a unit's budget, that needs to be viewed using different scenarios (appropriate for spread-sheet software), then they should go to the information center. If a user wants to do end-user computing but needs to access data, the information center can help show where the data is available by using the data dictionary. The separation between what is appropriate for systems development work and what is appropriate for the information center needs to be made clear by data administration policies. The one constraint that data administration placed on the information center is that no one develops a large, important application in a Fourth Generation Language (4GL) as an end-user application. The data administration office can guide the users to the appropriate data processing support area.

Forces that pull the information center and data administration apart are given in figure 5. Data administration and the information center do have natural tendencies that will drive the functions apart if management doesn't intervene. If data administration tries to control all of the data all of the time, they will have a war with the information center and will likely fail. Without management control, the information center pursues its natural tendency to solve each problem as it arises. If the image of data administration is that they want everything to be externally controlled, documented, and approved, they will be avoided by the information center and their users. Data administration has to be viewed as part of the solution to the information center's problems. This can be done by working with the information center and serving the information center users with the data dictionary.

An organization where debates over whether an application belongs on a mainframe or a micro computer are frequent, probably lacks a clear information planning direction. Management planning and data planning can overcome these problems.

Figure 6 summarizes the kinds of support that data administration should give the information center. The first is a data inventory using the data dictionary/directory. This should not just be an inventory of the databases but should be to the data element level. Such an inventory takes time, money, and commitment; building the inventory requires a thorough knowledge of the data dictionary and of the organization's systems. Second, the data dictionary interface must be designed for use by end-users and the information center. There should be an interface by subject area, by organizational unit, and by application system. If the data dictionary is not convenient to use and if the users cannot easily locate the data, it will not be used. Finally, data administration must be able to provide consulting help for the users and the information center. Data administration must be able to provide training and help in the use of the data dictionary, and the ability to create files for the users to access the data itself. Using a 4GL can be very beneficial, but data administration must constrain the users from putting a database in the 4GL until it is thoroughly documented in the data dictionary. This constraint will only work if data administration provides a good level of support for users who are documenting a database. If documentation is easily done, users spend less time trying to avoid it.

Figures 7 and 8 summarize criteria for use in deciding where the information center should report. The information center should report to the data administration function if the organization is data dependent. If access and reuse of data is critical to the organization as with a health insurance organization, then the information center should report to the data administrator. If the organization is scientific or has a lot of stand alone computing, the information center should report to the information systems director. In the second case, you are moving data more than sharing data.

It is rarely a good fit to have the information center reporting to the systems development manager. Still, this seems to be done frequently. The only justification for this structure that comes to mind is in an organization that is building expert systems for users. Large expert systems take years to build. Once the expert systems are done, an information center would provide support for the users of these expert systems. In this case, it is a good fit to put the information center under systems development since system use and enhancement require effective communications between systems development and end-users.

Placing the information center under an area outside of the information systems area altogether is rarely advisable. Only when an organization does not have very many important information-driven needs would it be advisable to place the information center outside the information systems area altogether. Most of the time, placing the information center outside the information systems area is only a short-term solution. Users might be content with a spread-sheet program for a while, but when they want to know where their budget is and why they can't get a copy of the data, end-user computing is no longer stand alone. Once end-user computing is not stand alone, the information center needs to be brought into the information systems department to provide access to the information and services users need. Placing the information center in data administration is likely to benefit both functions.

SUMMARY

While it is important to define your organization's data architecture, data administrators don't have to wait until an architecture is done to serve their organizations. Start today to build a data dictionary/directory system, support it, and get a good user front-end on it. Then you can sell the tool to the information center user community. This will build credibility for data administration and help the information center fulfill its mandate.

Finally, keep in mind that this metadata is not just documentation. It is the gateway to the data and information that the organization has already collected and paid for. Managers like the idea of reusing data and information since it is less costly collecting the information again. You have to sell the metadata for what it is worth. In large organizations, it is worth quite a lot.

BIOGRAPHICAL SKETCH

Mr. Ron Shelby is Data Administrator at Department of the Interior in Washington, D.C. Prior to joining Interior late in 1984, he was with Travelers Insurance (Canada). At Travelers, he established the Data Administration function, and managed the use of a data dictionary/directory for support of systems development and maintenance, end-user data location, data element standardization, and business functional modeling. Prior to leaving Travelers, Mr. Shelby had management responsibility for Data Administration, Data Base Administration, the Information Center and the Program Source Librarian functions.

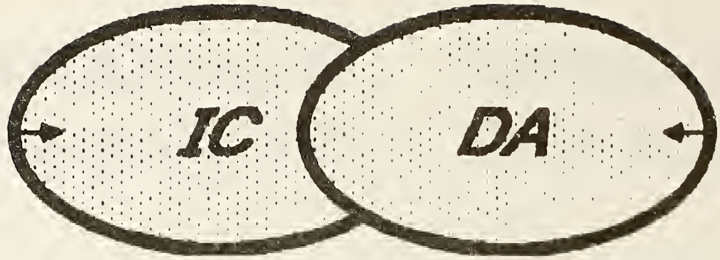
Mr. Shelby has been active in user groups dealing with the subject of data administration and database, and has addressed these groups frequently on data-related topics.

INFORMATION CENTER		DATA ADMINISTRATION
End-User Computing Support	***MISSION***	Data Resource Management
Computing Tools Support		Data Policy
Computing Tools Selection	***FUNCTIONS***	Data Planning
User Application Consulting		Data Design
Information Location Support		Manage Existing Data
Training Media		Data Planning Software
Hardware & Software Inventory	***TOOLS***	Data Design Software
Demo Hardware & Software		Data Dictionary/Directory
Data Inventory		Data Base Systems

Figure 1

INFORMATION CENTER		DATA ADMINISTRATION
End-User Computing Support	MISSION	Data Resource Management
Computing Tools Support		Data Policy
Computing Tools Selection		Data Planning
	FUNCTIONS	
User Application Consulting		Data Design
Information Location Support		Manage Existing Data
Training Media		Data Planning Software
Hardware & Software Inventory		Data Design Software
	TOOLS	
Demo Hardware & Software		Data Dictionary/Directory
Data Inventory		Data Base Systems

Figure 2



CENTRIPETAL FORCES

1. Users need to access and manipulate data captured elsewhere.

2. Need to understand what data is available in certain subject areas.

3. Need to have documentation facilities and consulting available when they decide to share their applications with other users.

4. Users need to improve data design skills in developing applications.

Figure 3

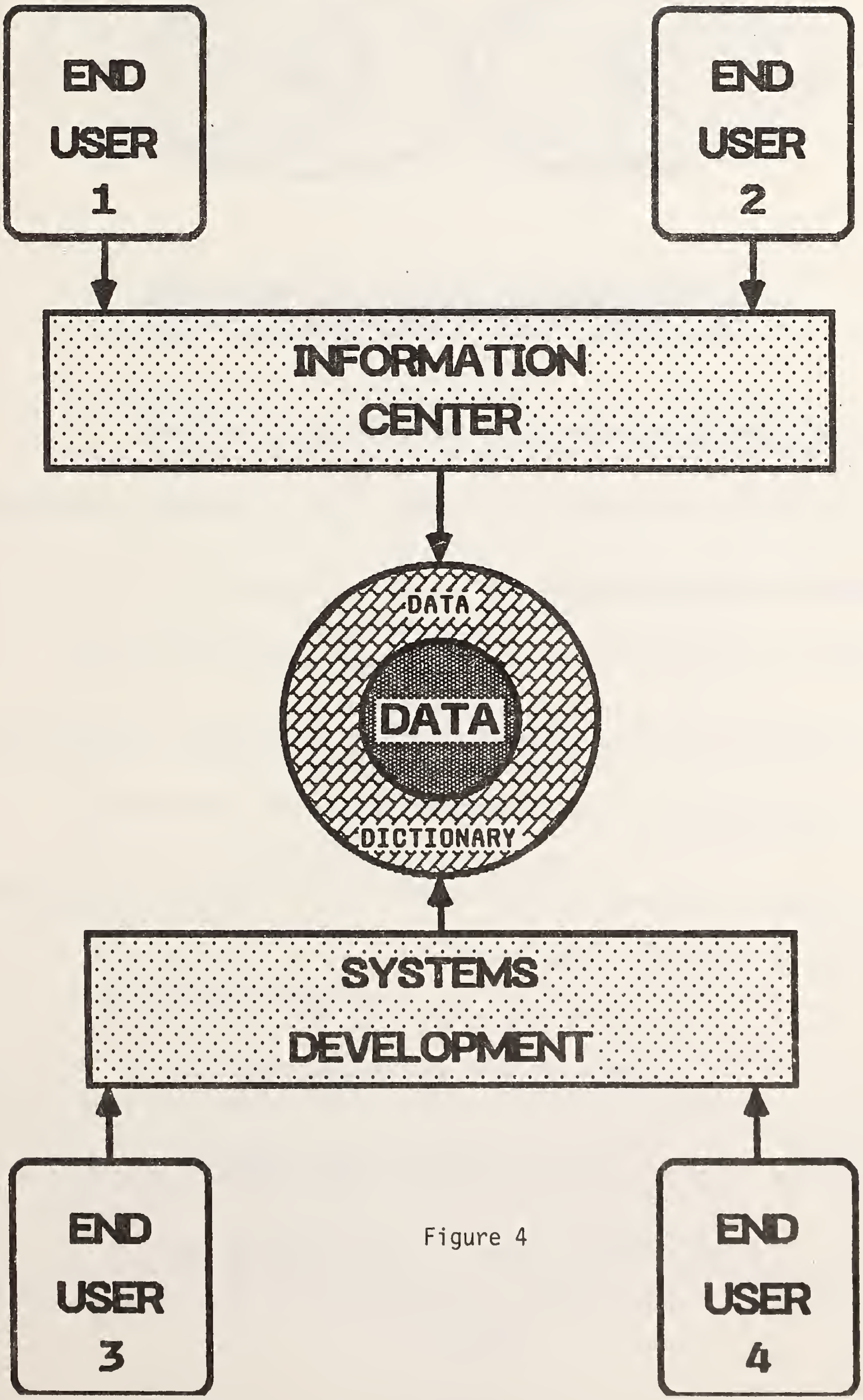
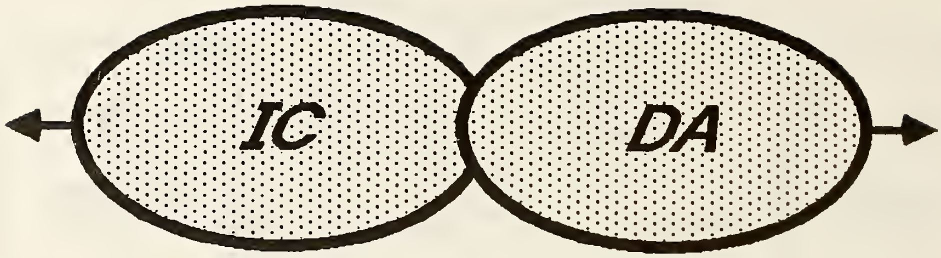

END
USER
1
END
USER
2
INFORMATION
CENTER
DATA
DATA
DICTIONARY
SYSTEMS
DEVELOPMENT
END
USER
3
END
USER
4

Figure 4

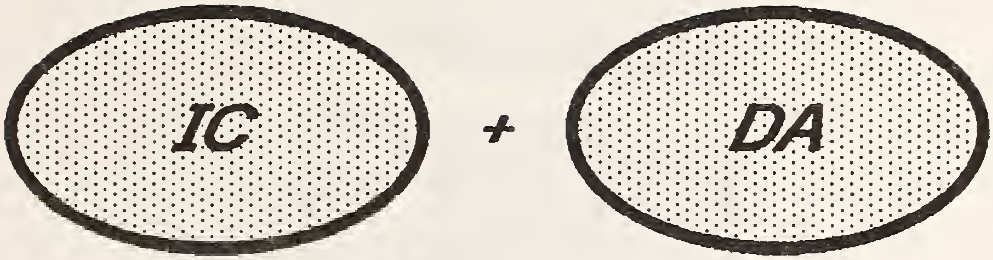


CENTRIFUGAL FORCES

1. IC's tendency to solve each problem as it arises.
2. DA's tendency to want everything to be externally controlled, documented and approved.
3. A lack of user needs analysis and planning in the IC.
4. A tendency for both areas to debate mainframes versus micros, and centralized versus decentralized data processing.
5. The lack of an overall data plan that includes policies for the end-user computing area.

Cause: Lack of clear direction and planning of the information needs of the organization.

Figure 5



DA SUPPORT OF THE IC

1. Data inventory in a data dictionary/directory.

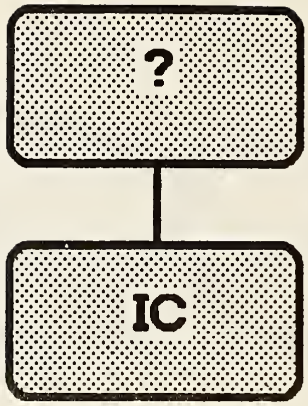
2. Dictionary interface for users and IC.

   - by subject
   - by organizational unit
   - by application system

3. Consulting help.

   - Dictionary use training/help
   - Data copy access

Figure 6



WHERE WILL THE IC REPORT?

Reports to Data Administrator, if

- locating, accessing and manipulating operating results data is crucial to product design, pricing, marketing or other key mission areas

- IC users frequently need access to data captured by others to perform their personal computing work

Reports to Informations Systems Director, if

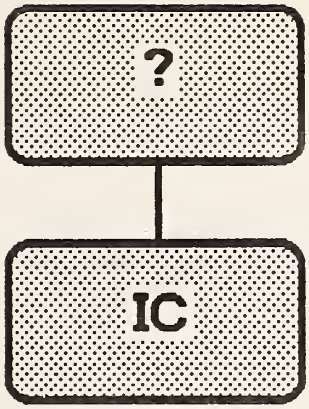
- end-user computing doesn't require the IC to report to the DA

- most end-users are writing their own applications and capturing their own data

- microcomputers are being networked for communications reasons

- host-system based personal computing predominates

Figure 7



WHERE WILL THE IC REPORT?

Reports to Systems Development Manager, if

- end-user computing is confined to pre-defined data retrieval to support queries and reports determined and built during application systems development
- specialized, complex expert systems need to be designed and constructed
- end-user built applications standalone, and the end-users need substantial help in programming

Reports outside the information systems area, if

- end-users are self-sufficient in their own applications
- microcomputer use is intended to be standalone
- end-users are to have only computing tools selection and training services

Figure 8

CENTRALIZED VERSUS DECENTRALIZED DATA ENVIRONMENT

Speaker

Dr. Ingeborg Kuhn
Veterans Administration
San Francisco, California

ABSTRACT

Dr. Kuhn described the changing environment of data processing in the Veterans Administration. She described the past centralized system and the new decentralized system that is currently being installed. Included is a description of the management issues for software development and the functions of the database administrator in this new environment.

The Veterans Administration (VA) has been doing something very exciting in the last few years. The Department of Medicine and Surgery supports a large network of 170 hospitals and medical facilities throughout the country. Until two years ago, all management information and administrative data was reported to centrally controlled, centrally located information systems. All data was produced manually in each hospital, keypunched and shipped to Austin, Texas, where the computers were located. The systems were all batch processed which meant that there was a time delay getting the reports back to the hospitals or even to the central office. The time delays for the reports and corrections to the reports meant that the information these systems generated was of little value to the local hospitals because the information was always out-dated. Because of the system design, it was very difficult to implement any kind of changes. Any new system took a long time to develop and implement, and changes to an existing system could take months or even years to implement.

Another serious problem with these batch systems was the accuracy or the validity of the data. Monthly reports on staffing levels, for example, were quite often made on the basis of "well it looked good last month so we'll use it again." Also, the people who designed the system had a centralized view of the data definitions and data element standards, but there was no assurance that those people in the field who were entering the data had the same definitions in mind. Quite often they did not.

A few years ago there was an underground effort to improve the existing system. It was called underground because at that time the only computers allowed at the hospitals were word processors. A small group of programmers were recruited and placed around the country. These programmers worked on their 'word processors' to develop admission transfer discharge, clinical scheduling, pharmacy, and laboratory systems. This underground effort surfaced in February 1982 when the Decentralized Hospital Computing Program (DHCP) was formalized. The DHCP was based on the concept that the clinical data produced at the local hospital will be for the local hospital. Clinical service data is the first step, but in the future there will be a totally integrated hospital information system including clinical data, clinical service data, administrative systems, and even clinical decision making systems.

The management of the computer and information systems and the data has been shifted to each local hospital. Computers are being placed in each hospital and the data entered on-line at each hospital. The data for centralized reporting is still developed locally. The time frame for this change was very short; within two years, the procurement was completed (RFP written and contract let). Computing is now in place and software is running at all 170 sites. The first applications that are being implemented are clinical service applications, pharmacy, laboratory, admission transfer discharge, and clinical scheduling.

There are several factors involved to make this work. First of all, the programs were written in the ANSI standard MUMPS language. Second, standards were developed for programming conventions to assure exportability to all hospitals and assure the integration of modules developed at different times and different places. The basic tool that has been used is a VA developed file manager. Within the file manager is a data dictionary structure. As an application is developed, the data dictionary is developed at the same time. The data dictionary has a technical orientation, but it also provides for user-oriented descriptions of each data element that is in the information system.

While software is being developed on a national basis, any site can install the software and add their own tailoring as long as certain conventions, rules, and regulations are followed. The local site can add data elements and create new applications following these rules and using the File Manager as their basic tool.

At each hospital, the database is integrated. The patient database is used by all clinical modules of the system. This eliminates having one database in one place and a different one in another. As the administrative system is developed, there will be one personnel file, one file for inventory use, and one file for the fiscal systems. We are all following the same structures, i.e., all data dictionaries look alike. All the software can be integrated because of the programming conventions and naming conventions used.

The VA approach to management of a decentralized environment is somewhat evolving. This is truly a decentralized effort to the point where the database administrator is even in the field. In the central organization is the Medical Information Resource Management Office (MIRMO). I represent the Information System Centers of which there are six throughout the VA corresponding to the six medical regions in the VA. These centers are where the software development is done, software management handled, interaction with the users takes place, data elements are defined, and also where the database administration function is housed.

MIRMO has many organizations. The two that the Information System Centers deal with most are the Information Reports Management Office and the Field Systems Support Group. The Information Reports Management Office works with the database administrator to establish the data administration policy for the program. Additionally, there is also a package coordinator in the field, currently in Albany. He is a software developer who works with all of the software developers to solve problems of package integration, file content, and data element definitions.

All of the DHCP software is public domain software. It is saving the taxpayers' money. The software is developed in the VA and has saved the VA from buying 170 licenses for a commercial information systems. The initial hardware was estimated to cost $60 million but because of competitive bidding and centralized procurement, it was acquired for $40 million. The $40 million has put fairly sophisticated computer systems in all 170 sites in the country.

The File Manager is the basic tool. It provides automatically a basic data dictionary for each application. Coordination and control for these data dictionaries is handled by the database administrator and the package coordinator. Still to be completed is a method to insure that the established policies are being followed and that developed packages follow the basic design principles. All software is developed with input from the end user, not with a centralized directive on how it should be done. These end users are not only at the six Information System Centers but at all 170 sites. This means that coordination is a major task. There is an elaborate electronic mail and conferencing system which helps, especially when travel dollars are cut so the coordinators cannot meet with the users and developers face-to-face.

The responsibilities of the database administrator are outlined as follows, although this is still evolving. The first is developing policies regarding database management. Also we coordinate the development and maintenance of DHCP data dictionaries. Automatically there is a technical data dictionary that is produced for each application, but we are in the process of automating a supplement for additional user documentation. The user documentation will be available to the users on-line, and a hard copy version will be distributed. We also are concerned with the proper placement of data elements, the definitions of the data elements, and eliminating unnecessary redundancy in the data elements. We are responsible for the more technical aspects of database administration such as name-spacing and file numbering assignments. Name-spacing is the method of assigning routine names used by the software. If the routine names are the same for two packages, it is impossible to bring up both at the same installation. We also develop tools for data element documentation and report modeling, and develop criteria for data element documentation and standards.

We are near to issuing our first policy circular. We selected the issues that are most pressing, those dealing with software development. The first area addressed in the policy circular is the classification of software. In the VA, there are three classifications of software. There is nationally distributed software, Class 1, that is often mandatory for all sites to install. The admission transfer discharge package is an example of this classification. This package is required because it generates reports that are used by the central system in Austin. Class 1 software has had extensive testing and verification and is supported by the Information Systems Center. Class 2 does not have the support of Information Systems Centers. The software may have been verified to see that it conforms to standard programming conventions but it does not have continuous support. Class 3 software is everything else. If an installation develops software, they may release it for use but the software is distributed with a 'buyer beware' label. It is not tested and is not supported.

The policy circular also addresses file numbering conventions, the management and assignment of name-spaces, modification to data dictionaries, data element and routines, and DHCP programming conventions. In the future, there will probably be policy circulars on software release management and standards for data documentation.

There have been procedures established for modifications to data dictionaries and data elements. For the data dictionaries, any local facility can add data elements for internal use as long as certain prescribed rules are followed. If a data element is to be added for external use, the local site must assume responsibility for the validity and accuracy of the data element. The issue of how to monitor this has not been worked out yet. Local development of software is encouraged as long as the procedures and conventions are followed.

SUMMARY

The benefits of a decentralized environment seem to be as follows:

- o Data available for immediate local use.
- o Greater incentive for accurate data capture (because the data are used on the local level).
- o Capability to add unique local data needs.
- o User defined data leads to increased validity.

The disadvantages of a decentralized environment:

- o Lack of central control over data element definition.
- o Need for reconciliation between 'agency standards' and 'user standards'.

BIOGRAPHICAL SKETCH

Dr. Kuhn is the Acting Director of The Information Services Center, Department of Medicine and Surgery, Western Region, in the Veterans Administration. She has extensive professional history as a consultant, a co-founder of a computer software and management consulting firm, a research associate at one university and a lecturer at another, and a Federal employee at several different agencies. She is also the author of a number of publications.

COST JUSTIFYING DATA ADMINISTRATION

Moderator

Frankie E. Spielman
National Bureau of Standards
Washington, D.C.

Cost Justifying the Data Administration Function

by Dr. Ronald F. Voell
Vice President, Corporate Data Resources Management
Bank of America
San Francisco, California
March 24, 1985

Introduction

As data administration evolves in an organization, management will usually ask questions about this function's contribution to cost effectiveness. All too frequently these questions are answered by data administrators in a way that does not instill very high confidence in senior management. Data administration, management is told, has benefits. However, these benefits are all intangible. Or, if management is given quantified benefits, the numbers are taken from the literature and represent the experience of other companies. Actual hard savings within the business are rarely presented.

While it is true that some data administration benefits will be intangible, many can be quantified. It may be difficult to quantify a value of improved data accuracy, but if storage space is reduced because data redundancy is decreased, that benefit should be quantifiable and be reflected by reduced costs. The key is to examine each benefit in detail to identify if it is really intangible, and then to see if the organization doesn't collect some data which can be used to calculate savings. The data administrator needs to be innovative.

When benefits being realized by other companies are used, those benefits need to be very carefully examined. When this is done, the data administrator will find that quantified benefits may be based on soft mathematical or research rigor. Some will be derived from hunches or assumptions which may not be valid in all settings. In addition, the organizational structure of the company from which the benefits were derived may affect the magnitude of the benefits. Data administration is implemented with varying degrees of responsibility and authority in each company. The benefits being realized by a company which has rigorously implemented data administration for several years will not be a valid predictor of the savings that will be realized in a recently-implemented, less rigorous setting.

The bottom line is this -- there is no substitute for investigating benefits that data administration is achieving in your organization and quantifying them for your management. The purpose of this paper is to present techniques and ideas for answering management's questions about data administration's contributions. The numbers that are presented represent actual research performed by the author. Although results compare closely with the results of other researchers, the emphasis of the paper is not those numbers, but on techniques that can be adapted and used in other organizations to quantify the actual benefits being achieved by data administration.

Data Administration's Evolution Affects Savings

Data administration evolves over time. The methodology that can be applied for measuring and the magnitude of the benefits that are to be achieved will depend on the evolutionary phase of the function. A phase theory similar to G. Gibson and R. Nolan's stage theory[1] was proposed by the author in 1979[2,3] to identify data administration growth patterns. It is important to understand this evolutionary process in order to apply the proper benefit measurement technique. The following is presented as a digest of the more detailed discussions in references 2 and 3.

Data administration growth can be measured in terms of its manpower. At first the department will be small, having perhaps only one or two people assigned to it. It will remain at this size for a year or so. After this phase, data administration will undergo a year or so of slow growth. This will be followed by a period of rapid growth where the department may exceed four or five times its original size. The growth will then tend to flatten out, perhaps even decreasing. The whole process may take about five years. Management decisions could extend or shorten the time.

If one were to plot this labor growth, the result would be a logistic curve with the various points of labor changes roughly relating to the evolutionary phases that data administration moves through: small at initiation phase, slow growth at expansion phase, faster growth at formalization phase, and topped-out growth at maturity phase. The labor changes are being driven by changing responsibilities in data administration.

The initiation phase begins the day top management establishes a data administration department. Major attention will need to be given to establishing a charter, selecting and installing a data dictionary/directory, developing standards, and training personnel. Data administration will also have a weak policing role. It will be responsible for overseeing system development to assure that programmers comply with the numerous design standards that the department will be issuing. Data administration's major areas of attention generally place the department in a staff position in the organization. That is, data administration does not contribute directly to development of data processing's product, the application system.

The assignment of line responsibilities, which may include an active role in file and database design, direct control over the data dictionary, security, defining data, etc., marks data administration's evolution to the expansion phase. The technical needs of data administration begin to change during the expansion phase. Where data administration was previously involved with conceptualizing, teaching and policing, it is now doing part of the job.

Data administration's work load is not usually too heavy at this point. Not many systems are documented in the data dictionary; therefore only a few systems need support. A low level of conversion may be underway. Data administration slowly becomes a department to which the organization looks to solve database technology issues. And, as significant percentages of application systems become documented, the control of new development and the maintenance of documented applications becomes more critical and difficult. At this point data administration moves into the formalization phase.

The formalization phase is identified by the centralization of data definition and database design expertise in data administration. The strong dependence of the organization on this expertise and data administration's relatively small size cause it to be become a roadblock which slows application development. Rapid growth becomes a necessity. The informality of the "mean and lean" stance no longer works as rapid expansion occurs. More and more, the data administrator's role becomes one of managing the department in contrast to managing data. Forecasting personnel requirements, personnel acquisition, project estimating, scheduling, tracking the status of projects, and solving problems replace his or her concerns about how data will be controlled. Data control is delegated to the personnel in the data administration department.

As technical expertise grows in data administration, information must be disseminated if interfaces are to be maintained with the programming departments and data administration. In the expansion phase, the department's training role diminished. In the formalization phase, the role must be expanded again in order to maintain interfaces. A strong return to training becomes important. Data administration line responsibilities also continue to grow. Concerns over how new applications and changes to existing documentation affect the integrity of databases force data administration to become heavily involved in overall application planning design reviews. Security issues expand auditing responsibilities. Data administration, therefore, begins to reassume a policing role. All of this causes it to become a function which must be closely teamed with the programmer and project to jointly develop system design approaches.

At this point, large percentages of application systems and databases will be under data dictionary management control. The nature of uncontrolled past system development begins to evolve. Redundant information can be identified. Problems with nonstandard data definitions in existing systems create major interface issues as new systems are developed. Questions over what data goes in what databases repeatedly surface. The problem that these concerns identify is the lack of a long-range system architecture. Without this architecture, every new application system builds more databases geared to the unique requirements of that individual system. Data administration will then create long-range data architecture plans. These plans, together with an adequately staffed and managed department to implement them, mark the point where data administration moves from the formalization phase to the maturity phase.

In the maturity phase, labor growth slows and may even decrease. Many of the master files and databases used by the organization are documented in the data dictionary. So are most data elements. Some will have been established as standard definitions. Data administration will find that some existing databases fit the plan. Other databases do not. But slowly, as new designs are developed in compliance with the architecture plan, and old designs become obsolete and are cancelled from the data dictionary, the documentation in the dictionary becomes an important tool. Redundancy disappears, and so does confusion.

The movement of the data administration department away from a preoccupation with a database for application systems and toward a database as a system and data as a resource may move the department away from a pure line function and more into a hybrid staff-line function. This hybrid will probably be necessary because data administration will have divergent objectives -- managing to a strategic plan and designing new systems. Implementing strategic and tactical objectives by one department is usually difficult. A pure line relationship might cause designs to evolve that compromise the plan. The department will need to be broken into two functions -- one responsible for managing the architecture, which reports into the organization at a relatively high-level staff position, and one that designs and manages data in support of applications at a middle management line level.

The "Best" Methodology Changes As Data Administration Evolves

Eight methods of cost justifying data administration will be discussed in this paper. These methods are:

1) More and less method
2) Other user experience method

3) Dogs of the year analysis method
4) Project comparison method
5) Project estimating method
6) Development cost analysis method
7) Failure reporting analysis method
8) Data modeling method

There are several points that should be made about these methods. First, the mathematical rigor differs with each method. The above list has been organized in rigor order. Method 1 (more and less) is the least rigorous, while the data modeling method is the most rigorous. The second point has already been touched on. Methods cannot be used at just any stage in the evolution of the data administration function. During the initiation phase, data administration is in a planning and implementation mode. Real benefits are not being achieved. The benefits cannot be measured for the organization because benefits don't exist. But they can be predicted. Methods 1 and 2 are the only appropriate techniques for this phase.

During the expansion phase the management system of the department is very informal. Quantified benefits are therefore difficult to derive and where they do exist, they are soft. Methods 3 and 4 are useful at the expansion phase.

During the formalization phase, a management system evolves and provides the first accurate measurement data that may be used to calculate data administration's benefits. The systems that provide useful data are the labor, failure reporting, computer usage, estimating, and dictionary system. Methods 5, 6 and 7 interrelate data from these systems and could be used at this phase.

Finally, during the maturity phase, strategic plans which identify a data architecture are developed. The data models that result make Method 8 appropriate at this phase.

The third point concerning the methodologies is that benefits are not fixed, but change as the data administration function evolves. Benefits that are calculated in one phase will not always be valid in another phase. Benefit assertions need to be stated in the context of a data administration's evolutionary phase.

The last point to consider is that data administration is not introduced into a data processing environment while all other things are held constant. Productivity tools are continually being introduced, e.g. structured design, training, on-line compilers, etc. It is quite logical to conclude that part of the productivity improvement findings discussed here was caused by these other factors. The important point is that productivity increases caused by data administration methods are difficult to isolate.

Methodologies and Findings[4]

This section defines the logical basis for the various methodologies used by the writer to identify the benefits of data administration. It then reports on the observed results and where possible, attempts to explain why the results were observed.

Method 1: More and Less

When data administration is being proposed for an organization and even after it is in its early stage of implementation, lists of what the organization is going to be able to do more of and what it will experience less of are going to be drawn up and presented to management. A typical list might be:

1) A consistent corporate definition of data
2) Less dictionary problems
3) Improved control of data definitions
4) More efficient labor utilization
5) Better systems quality
6) Lower maintenance cost
7) Faster implementation of systems
8) Faster response to change
9) Better use of hardware resources
10) Reduced data redundancy
11) Reduced impact of personnel attrition

While these are not quantified benefits, they usually satisfy the organization's need for justifying the department, at least for a while. During the expansion phase, however, the data administrator needs to examine the "more and less" benefits and determine which are tangible. Measurable benefits need to be thought about. How can they be quantified? What data already exist in the organization and what data need to be collected to quantify it? How can data be interrelated to derive new insights?

Method 2: Other User Experience

These is a dearth of other user experience concerning quantified data administration benefits to the organization. One hears numbers being exchanged in conferences, but few of these numbers appear in the literature. This author has found quantified benefits published but they do not contain an explanation of the approach used to develop these data. Reference 5 is the only known publication. This paper, of course, is another source. The reader might watch the literature for the results of research which, it is understood, is being performed by Marilyn Parker[6]. She has published a paper on a methodology but no results have been released.

Method 3: Dogs of the Year

This method gives only a rough indication of the benefits of data administration. It identifies the ten worst programs (dogs) in the data processing shop by determining the yearly frequency of recompiles to correct bugs and then assigns the program to either the data administration-controlled or uncontrolled class.

In one organization where this methodology was used, none of the ten worst programs were developed under data administration control procedures. The top ten were reported to management for possible rewrite. The first data administration controlled programs were at the bottom of this list. The program at the top of the list had an average time between compiles of 3.7 working days. Over a period of one year, it was fixed and compiled in production over 70 times.[7] Although this statistic does not quantify a direct saving, it indicates to management that data administration makes sense.

Method 4: Project Comparison[8]

This method is based on a comparison between two project estimates, one of which assumes data administration involvement and the other which assumes no data administration involvement with a project. (This is in contrast to Method 5, which compares estimates and actuals using a rather detailed estimating model.) This method requires a team effort between data administration and development management. The team nature of the effort lends credibility to the results. The development manager needs to have had experience working both with and without data administration support. In addition, the development manager must be a "believer."

A development manager will notice that from a programmer standpoint, several areas begin to stand out as being easier to accomplish with data administration. The automated generation of parts of the program by the data dictionary produced observed savings. Testing of these programs will be more likely to be successful and less costly because of reduced errors. Some interface files may already be documented. Over a period of time, the development manager will form opinions about the magnitude of these savings. This methodology uses these observations to estimate the costs of a project both with and without data administration, and then compares results.

This kind of study often occurs naturally in organizations. Frequently projects are proposed when the organization isn't sure it wants to involve data administration. These projects frequently result in dual estimates. When the two estimates were made on an actual medium-size project, the benefits were calculated as 7%.

Method 5: Project Estimating

This method is based on a detailed estimating model which was used to predict project labor costs. The model was used for each project. Actual labor data were captured for each project and compared with the model . The model was modified during the first four months of its use so that it accurately agreed with the results observed in the labor data collection system.

Understanding the model itself is not as important as understanding how such a model would be constructed. Although the process is straightforward, it does require a detailed analysis of the process used by data administration to support a project. Once these processes are understood, they need to be related to the work variable which sizes the process and a labor standard to accomplish the work variable. By way of illustration, one work process might be "define the data elements". One could conclude that this process should be related to the estimated number of data elements and some observed work standard, let's say 1.1 hours per data element. The labor needed to define data elements would therefore be 1.1 times the number of data elements.

This analysis is repeated for each process. Some processes are a constant, that is, they are independent of project size, e.g. "submit a work authorization form". Other processes are difficult to relate to a work variable, e.g. "prepare a specification". These must be related to some indirect work variable. I found that the number of files or databases used in the project was related to labor spent developing the specification.

After all processes are defined and estimated, the individual factors are algebraically combined and a useful estimating tool exists. [9, 10] This model was originally developed to provide rapid project and cost-to-complete estimates. It was not intended to be a benefits measuring tool. But if one reflects on what one has available with such a model, it is clear that it can also measure benefits. The model accurately estimated data administration project costs for the first year. If data administration improves productivity, then as data administration begins to expand and mature, the model will over-predict labor costs. And that is exactly what was observed. At the end of the second year, total labor expended on projects was 9% below the predicted value and was 18% below the third year.

Method 6: Development Cost Analysis

Another source of information which demonstrated the cost benefits of operating under data administration procedures is the system development labor collection system. Development labor costs can be used to compare systems developed using data administration controls with those that were not. But direct comparisons between these costs are not valid because project size varies. Costs need to be normalized.

One way of normalizing project costs so they can be compared is to divide the labor cost by the number of "standard programs" that were written for each project. Standard programs are computed by dividing the total number of lines of code written for the system by some constant, in this case 1000. The value of the constant is not important, but after it is selected it is always used. The division of the development costs by the number of standard programs yields a factor which allows the direct comparison of project dollars per standard program.

These data were collected for about a year. It was found that there was a relationship between the normalized project costs and the size of the project. Small projects had higher normalized project costs than did medium or large projects. This was because the project fixed costs had to be spread over a smaller amount of total costs. This observation on small projects was found regardless of whether data administration did or did not participate in the project. Cost data were therefore segmented by project size as well as by the application of data control procedures. Projects with labor costs below $4,000 were defined as small, between $4,000 and $40,000 as medium, and above $40,000 as large.

The findings are that system development using data administration procedures average about 7% less labor costs than system development not using these procedures. This saving was attributed to the fact that data were better understood by the project personnel when data control procedures were followed. Less redundant work was required, and the data dictionary automatically generated accurate parts of the project code. In addition, data control procedures allowed the project to more rapidly and accurately react to changes identified during development.

Small projects had a slightly higher savings than medium or large projects. This was attributed to the fact that smaller projects were more likely to use data or files which already existed and were understood.

Method 7: Failure Reporting Analysis

Another source of positive data administration benefits data is the data processing failure reporting system. One company had a system which listed each program that had failed, and the labor and computer time associated with correcting the bug. The system was also used to capture data about the time required to perform minor modifications to programs. The data were in machine-readable form and contained information on all program bugs over a one-year period. That included more than 3,000 problems. Data administration had been part of this organization for only three years. Many of the programs (about 75%) had been written before data administration, a data dictionary, and standards had been in place. This provided an excellent opportunity to compare problem data on two classes of programs: programs which were subject to data administration controls and those which were not subject to those controls.

The failure data suggested the following comparison questions between the two classes of programs:

1) How do the failure rates compare?
2) When a failure occurs, are there any differences between the labor/computer time required to fix the failure?
3) Are there any differences between the amount of labor required to make minor modifications?

Programs were written which took each failure in the failure reporting system and checked the failed program against the data dictionary. If the program was in the dictionary, the statistics were assigned to the data administration-controlled class; if it was not in the dictionary, it was assigned to the uncontrolled class (or, more precisely, to the programmer-controlled class).

The findings concluded that no differences could be seen in the programs' reliability. Programs developed with data controls and standards failed as often as those without the controls. This was unexpected. One would think that controls would produce a product that was less failure-prone. But if one reflects on these similar failure rates and remembers that a comparison was being made between old programs (the uncontrolled ones) and young programs (the controlled ones), an important factor should be noted: new programs have a greater likelihood of failure in their early life as time sweeps out bugs. On the other hand, older programs have lived through this early phase of life where use has identified many programming errors. The fact that a young program failed only as frequently as a mature one is therefore encouraging. This would seem to indicate that as new programs approach the age of older ones, the newer ones will be more reliable. The similar failure rates found in this research, then, might be viewed positively.

Another finding was that significant differences were observed in the amount of labor required to fix bugs. Five times more labor was required to find and fix failures if data administration had not been involved in the development. A similar unfavorable, but less dramatic, comparison was found in computer time to validate fixes. It was determined that 10% more computer time was required to repair a failure in the data uncontrolled class system. These differences in labor and computer time were attributed to the more standardized and readable code in the data-controlled class. It would seem reasonable to conclude that standard names throughout the code made it easier for programmers to understand the code and locate problems. This understandability made it more likely that they fixed the problem the first time; therefore, there would be less computer time needed to test the fix.

A third finding was that minor modifications were also less costly on systems developed with data controls. About 14% less labor was expended in modifying systems developed in the data administration class. This was attributed to more readable codes. But it was also the result of files and data definitions being defined and automatically available to the programmer. In contrast, the uncontrolled systems required more research into data in order to make modifications.

Method 8: Data Modeling [11]

A basic premise of data modeling is that it organizes data and designs databases using relational rules into structures which tend to be independent of application design. Those who believe this premise conclude that the resulting data structures and application codes will require fewer changes as new applications are added which interface with these existing applications. Now in fact, the organization will have two kinds of applications: those based on the model and those developed before a model was used. This methodology measures how many lines of code in existing applications need to be rewritten when a new application interfaces with files or databases used by that existing application. The measurements were made from two perspectives: existing applications based on the model and those not based on the model.

The findings were that for interfaces with non-model developed systems there was about a 50% probability that 20% or more of the code would require rewriting. When interfacing with model-based systems, there was a 15% probability that 10% or more of the code would require rewriting. Using the minimum percent change and computing expected value, this translates into an 85% savings for a part of the new development effort.

Contrasting The Benefits Results

The last section presented methodologies for quantifying data administration benefits. One way of addressing the creditability of the methodologies is to compare the results that were computed at the different phases of data administration's evolution. The commonly accepted assumption is that saving will not exist during the initiation phase but will occur at latter phases and continue to grow. Do the measured results follow that pattern? Another way of addressing credibility is to compare the results with those of other researchers.

Table 1 provides the comparison of benefits results that were measured using the various methodologies.

Table 1

Data Administration Benefits at Different Phases of Group Evolution

	Phase		
Saving Areas	Expansion	Formalization	Maturity
New Development/ Modifications	9%(4)*	9-18%(5) 7%(6) 14%(7)	85%(8)**
Maintenance		80%(7)	

* (4) Signifies the methodology being used
** Methodology 8 only addressed the part of the saving associated with changing the existing system

No savings are presented in this table for the initiation phase because at startup, there are none. At formalization, three methodologies were used which all demonstrated fairly close results. The savings for new development and modification averaged about 12%. The methodology applicable at maturity only examined those savings that resulted from a part of the development effort and were not directly comparable with the results derived from the other methodologies. Maintenance savings were computed only at one phase and therefore cannot be contrasted.

Not much can be derived from an analysis of the trends in Table 1. There is an increase as was expected in savings observed in new development as the function moved from the expansion to the formalization stage (9 to 12%). But the savings that occur in maturity are not available for comparison. The fact that the benefits trend seems to be congruent with what intuition predicts lends at least some credibility to the methods.

A number of other researchers have documented or, through personal contacts with this researcher, reported cost benefits of data administration. Table 2 contrasts their findings with this research. The methods used to compute their findings are unknown, but the various findings also are in fair agreement. This researcher was unable to find any other quantifiable reports of savings in computer test time or reliability.

Table 2

Comparison of Benefit Results with Other Researchers

	Savings			
	1*	2	3	4
New Development	7-14%	3-15%	17%	30%
Modification	14%	10-14%	-	-
Maintenance	80%	-	-	-
Changes to Existing Systems When New Systems Interface	85%	-	-	-

*Source:
1. This researcher
2. Ronald Ross, Data Dictionaries and Data Administration
3. A large multinational petroleum company
4. One of the Big Eight auditing firms

Summary

Top management understands quantifiable results. Subjective "mores", "lesses" and "betters" don't impress them, at least not for very long. Sooner or later, management wants to know, "How much better are we performing and how much money or time are we saving?" Data administrators shouldn't throw their arms up and view this as an impossible task. They have powerful tools for answering that question. The information in the data dictionary can be related to the tremendous amount of performance information that is created throughout the data processing organization. In fact, that may be a part of the problem. Data processing organizations are so information-rich that they don't know what data they have. The innovative data administrator looks to these data and discovers new interrelations which have hitherto escaped the organization.

This paper discusses eight possible methods for measuring the cost benefits of data administration. It then reports on the results that were measured in actual business settings. The author is not recommending that these exact results be used to predict benefits in other business settings. Each company is unique. The purpose is, however, to demonstrate methods that a data administrator might use in order to quantify cost benefits. Finally, this paper contrasts the findings of the author with those of others.

Footnotes

1. G. Gibson and R. Nolan, "Managing the Four Stages of EDP Growth," Harvard Business Review, Jan-Feb 1974, pp 76-88.

2. R. Voell, "The Data Administrator's Role in Long Range Planning," Proceedings of the Sixth Annual Information Management Symposium and Conference, Session 3.12, Red Dot Verbatim Reporters, 1979.

3. The paper in reference 2 has been retitled and used by several organizations. IBM uses it in several of its classes retitled, "Data Administration Evolution". POSP, Inc. has published it under the title, "Planning Data Administration," 1980.

4. The formula used to compute savings that was used throughout this paper is:

$$\frac{\text{Cost Without D.A.} - \text{Cost With D.A.}}{\text{Cost Without D.A.}}$$

   Cost with Data Administration participation in a project always includes the direct cost of that involvement.

5. R. Ross, Data Dictionaries and Data Administration, Amacom, New York, 1981.

6. Marilyn Parker of IBM, Los Angeles, has been reported working on collecting data based on a similar methodology.

7. This observation appears at variance with the reliability finding under the Failure Report Analysis method reported below. It is pointed out that the former method analyzed all failures whereas this method only analyzed the worst programs.

8. An unpublished paper was presented by the author on this methodology in 1978 to the LEXICON Users Group entitled, "Modeling Data Administration."

9. The model that resulted was:

$$C = 1.1 \left[1.17E + 5.25 (I + O) + 12.5 (F + 2.12D) + 1.5T + 4.25P (1 + 1.71y) + 8y + 16 \right]$$

where

C = Cost in labor hours
E = Number of data elements
I = Number of unique inputs (screens, cards, etc.)
O = Number of unique outputs (screens, reports, etc.)
T = Number of transactions (IMS)
F = Number of files
D = Number of databases
P = Number of programs
y = 1 if data base project, otherwise 0

This is an example of an estimating model. The author does not suggest that this model is valid in any organization except the one for which it was developed.

10. A similar technique, known as Function Point Analysis, was developed by IBM and will be discussed in a GUIDE Publication due to be released in 1985.

11. The approach and results reported in this methodology were taken from notes the author took at the Third Data Administrator's Users Conference, 1983, in San Francisco. During the conference, a conference attendee from the floor representing the EG&G Corporation related the above methodology findings.

USING CONVENTIONS AND STANDARDS

Moderator

Roy G. Saltman
National Bureau of Standards
Gaithersburg, Maryland

Recorder

Leonard Gallagher

NATIONAL AND FEDERAL STANDARDS EFFORT

Speaker

Helen M. Wood
Chief, Information Systems Engineering Division
Institute for Computer Sciences and Technology
National Bureau of Standards

ABSTRACT

Because of the increased use of packaged software for data management and applications development, there is a renewed interest in Federal, national, and international standards for database languages, data dictionary systems, computer graphics, data interchange, and interfaces to programming languages. Along with such specific standards, there is a need for guidance documents on data administration, logical database design, use of standard codes and representations, selection of DBMS and graphics systems, and applications development. This presentation discusses the Information Systems Engineering program within the Institute for Computer Sciences and Technology at the National Bureau of Standards and identifies standards, research activities, and guidance projects in these areas.

Rapid increases in the costs associated with software development and maintenance are driving organizations to alternative methods of data management and applications development, including packaged software, DBMS's, and application generators. Consequently, the advantages of standards for software facilities and interfaces are beginning to be recognized in much the same manner as in the computer communications field.

Users of sophisticated data management software want to be able to export their data to powerful graphics software systems. Organizations who employ independent data dictionary systems want to control data used by a DBMS, COBOL programs, and so-called "fourth generation languages." Those who buy into data management technology expect these expensive software facilities to support constantly changing requirements. Just as for hardware systems, the days of user dependence on one vendor for all of their software needs are over.

The Institute for Computer Sciences and Technology (ICST) is a center of technical expertise in information technology. ICST provides scientific and technical guidance on the effective use of computers and the application of information technology. It develops guidelines, standards, technology forecasts, research reports, and other documents to help managers and users of computers and networks. ICST conducts research in computer sciences technology as required to fulfill its role of technical advisor to the Federal Government in effective management of information technology. ICST also sponsors conferences, workshops, seminars, and user groups to exchange information on current issues in information technology. Activities addressing data administration are carried out within the Information Systems Engineering Division in ICST. This report describes the activities of this Division, all of which impact the data administration function.

The goals of NBS's program in Information Systems Engineering are to help Federal agencies improve their data management and applications development and to support U.S. industry in the international standards arena. Program activities fall into four major areas: (1) data administration, (2) data management systems, (3) computer graphics, and (4) applications development.

DATA ADMINISTRATION

The data administration activity develops guidance on strategic data planning, data naming conventions, data modeling, data interchange, and data administration tools. In addition, it produces standards used for facilitating the interchange of data both within government and across industry. Until recently, the major emphasis has been concentrated in the area of data element representations and data interchange. However, this workshop is part of the effort to expand to include the other aspects of Data Administration management.

Sixteen standards have been produced identifying widely used data elements and representations, many in the area of geographic location data. For example, the FIPS PUB 55 contains over 155,000 entries providing unique codes for populated places and other location entities throughout the United States, and it specifies ZIP code, Congressional District, and Metropolitan Statistical Area for many of these places.

Other FIPS PUBS include: (1) FIPS PUB 95 which provides a list of codes for identifying Federal and Federally-assisted organizations; (2) FIPS 104, which implements the American standard codes for countries; and (3) FIPS PUB 19-1 which provides a catalog listing and brief description of many sets of codes that are in wide use in the U.S. and that might be used in Federal data systems. ICST has also been active in the development of ANSI Business Data Interchange Standards. These are uniform standards for inter-industry electronic interchange of business transactions.

Recently a report, NBS SPECIAL PUB 500-122, was published detailing an iterative methodology for Logical Database Design. This report is being described in a later Tools and Techniques panel. Last year (May 1984), a paper on Naming Conventions was presented at the Trends and Applications - 84 Conference.

As to the future, there are numerous products that ICST plans to produce which directly relate to the Data Administration function. These products, to be developed over the next five years, are listed and described as follows:

1. Workshop Proceedings: This will be a publication documenting the results of this workshop.

2. Strategic Data Planning: Describes the different approaches and methodologies commonly taken to develop strategic data plans.

3. Integrating Conventions and Standards into the Organizations: Describes the major issues and activities that an organization must consider and evaluate when establishing and setting up the Data Administration conventions and standards.

4. Using Data Dictionaries and other Automated Support Tools: Describes the requirements for the automated tools needed to support the Data Administration function and to facilitate the sharing of an organization's data.

5. Issues in Managing the Data: Describes the major issues associated with Data Administration and provides alternatives for managing an organization's data. Some of the issues center around Information Centers, Data Directories, Distributed data, and Impact Analysis.

6. Cost/Benefits Analysis and Data Life Cycle Management: Describes the approaches to managing data throughout its life time from the initial inception during requirements definition to the final disposal after the data is no longer needed. It also evaluates the process of analyzing the cost associated with the acquisition and use of data versus the benefits to be derived from the data.

7. Planning, Organizing, and Implementing the Function: Describes the approaches that can be taken during the building of the Data Administration function. It describes the different alternatives and steps for evaluating the alternatives.

8. Prototype Environment: This will be a prototype project set up in the ICST laboratories. It will illustrate and demonstrate typical Data Administration activities that can be performed using automated tools.

DATA MANAGEMENT SYSTEMS

The data management systems activity develops standards and guidelines to support the effective selection and use of sophisticated database management software and hardware. Emphasis is placed on developing urgently needed national and international standards, including:

Information Resource Dictionary System (IRDS) which specifies the most commonly needed facilities of a data dictionary system.

Database Language NDL, for network structured databases, and Database Language SQL, for relational databases, specify essential structures and operations for conforming DBMS products.

Data Descriptive File (DDF) which provides a media independent format for the interchange of structured data.

Preliminary cost-benefit studies have identified over $250 million in expected cost savings government-wide from these standards through lower database conversion and training costs.

NBS is supporting the development of testing and measurement techniques needed to verify conformance to these emerging standards. Recent Publications in this arena include: Guideline for choosing a Data Management Approach (FIPS PUB 110) and a Guide to Performance Evaluation of Database Systems (NBS Special Pub 500-118). The requirements for distributed database management systems are also being addressed. A Fourth Database Directions Workshop is scheduled for October 1985.

COMPUTER GRAPHICS

The emergence of computer graphics as an invaluable tool for conveying technical information, technical training, and general communication of information is well known. Now, as graphics technology becomes ubiquitous on mainframe and microcomputers, the demand has grown for graphics-based systems that are transparent to programmer and end-user. Emerging standards in this field promise benefits including host computer portability, display-device independence, ease of application program design, and portability of graphics databases.

NBS is actively participating in the development of graphics standards and conformance testing and measurement techniques including:

Graphical Kernel System (GKS) - an ISO standard addressing 2-D graphics functions for computer programmers;

GKS 3-D Extensions - which extend GKS to cover three-dimensional graphics;

Programmer's Hierarchical Interface to Graphics (PHIGS) - supports high-level programming capabilities not addressed by GKS for applications requiring very high performance and interaction;

Computer Graphics Metafile (CGM) - for transporting graphics pictures among different devices; and

Computer Graphics Interface (CGI) - which defines the interface between device-independent graphics software and device-dependent drivers.

Additional projects in this area include: development of standard bindings of graphics standards to major programming languages, benchmarking techniques for evaluating the performance of graphics systems, and an assessment of microcomputer-based graphics systems.

APPLICATIONS DEVELOPMENT

Until recently, this program activity has focused on the development of traditional programming language standards and validation tests. In response to the upsurge of interest in higher-level programming languages and techniques such as applications prototyping, this area has been broadened to develop guidelines on the use of such emerging technology. Emphasis is on identifying criteria, including performance considerations, for selecting the most appropriate tools for the job. The tools include applications generators, query languages, and report generators. Life cycle requirements, such as machine and programmer portability, are also addressed.

STANDARDS AND ICST INFORMATION EXCHANGE PROGRAM

There are several service activities ICST has set up or participates in to help in the effort to provide assistance and information interchange to the Federal Government. These activities center around information exchange, standards or guidelines, and research.

FEDERAL DATA MANAGEMENT USERS' GROUP (FEDMUG)

The Federal Data Management Users' Group (FEDMUG), sponsored by ICST, meets three to four times a year to provide a government-wide forum for the sharing of technical information among Federal data managers. FEDMUG also provides a basis for presentation by ICST of forthcoming products such as standards and guidelines in the area of data management and to receive feedback from agencies on their plans and needs.

PUBLICATIONS

To aid in the dissemination of information and implementation of standards activities, ICST produces numerous publications. The primary documents published by ICST usually are produced as either FIPS Publications or NBS Special Publications. These publications are generally classified as Federal standards, guidelines, technology forecasts, or research reports.

NATIONAL/INTERNATIONAL STANDARDS REPORT

Members of ICST participate in National and International Standards Committees. Each of these two groups of committees develop voluntary standards that may affect the Federal Program. Each of the groups is described below.

The AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI) is afederation of approximately 180 organizations representingtrade, professional, commercial,organized labor, and consumer interests. It serves as the national coordinating institute for the development of national standards and provides an independent mechanism for approving, coordinating, and managing programs of national standards. In the arena of Data Administration, there are basically four ANSI committees currently working towards standards. They are: Information Resources Dictionary System (IRDS), Electronic Business Data Interchange, Databases, and Data Representations.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) is an organization responsible for writing international standards. It's made up of representatives of standards bodies from the participating countries. Ideally, this should be a standard that can be adopted in every country of the world.

INFORMATION SYSTEMS ENGINEERING LABORATORIES

Information Systems Engineering laboratories provide environments for experimentation and research leading to the solution of Federal information systems problems. The labs provide expertise and facilities to assist agencies in exploring new technology and methods, as well as a test-bed for cooperative efforts with industry. Currently, labs in this area address Database, Knowledge-Based Systems, and Computer Graphics technology.

ELECTRONIC BULLETIN BOARDS

ICST operates dial-up electronic bulletin boards for information exchange. To reach, simply dial the numbers listed below; the systems provide instructions for operation and use.

Microprocessors: (301) 948-5718
Data Management: (301) 948-2048

Terminal requirements: ASCII; 300 or 1200 baud; 7 or 8 data bits; no parity; 1 stop bit.

DATA ELEMENT STANDARDS IN DEFENSE INTELLIGENCE

Speaker

Carl Fritzges
Defense Intelligence Agency
Washington, D.C.

ABSTRACT

The intelligence community first developed data element standards in the late 1970's. Each standard data element is kept in a catalogue called IDEAS, Intelligence Data Element Authorized Standards, together with its name and description. IDEAS is maintained by a committee composed of various members of the intelligence community who meet quarterly to consider modifications and additions. This presentation describes IDEAS and future plans for data administration.

This morning I'll be telling you about the data standardization efforts that have been ongoing at the Defense Intelligence Agency (DIA), but first let me make one point of clarification. Very often there is confusion about what data standards really deals with. Sometimes data standards is confused with programming standards. Things like naming conventions for records or fields within programs and restrictions on the use of "go-to" statements. That is not what I will be talking about today. I will be discussing data element standards. Data element standards deal with the definition of basic elements of information and how those units of information are stored in databases, including the specification of codes to represent specific values. It also deals with the application of data element standards, known in DoD and at DIA as Data User Identifiers (DUI's).

Data element standards at DIA are maintained in a hard copy document known as the Intelligence Data Elements Authorized Standards (IDEAS). This document, developed over ten years ago, contains approximately 1000 data elements and data chains (combinations of data elements). Part of this document was extracted from the DoD standard 5000.12M, and the rest were developed through usage in existing intelligence databases.

New standards and changes to existing standards are developed, reviewed, and approved through the Defense Intelligence Data Element Standards Committee (DIDESC). This committee, chaired by the DIA, consists of voting representatives from the Unified and Specified commands, the military services, and other intelligence organizations as appropriate. DIA has one voting representative. The DIDESC meets at least quarterly to review and approve (or disapprove) proposed new standards and changes to existing standards which are submitted by working groups or individuals within participating organizations. At DIA, we have a Data Standards Working Group (DSWG) which determines DIA's position on data standards issues. The chairman of that group is the voting representative for DIA on the DIDESC. Standards in IDEAS are either DoD approved, DIDESC approved, or DIDESC working standards.

Data elements in IDEAS are generic in nature. For example, date may be defined as six numeric characters in the sequence YYMMDD. By itself, "Date" has little or no meaning. But when used in a specific application such as "date of birth of an employee," it becomes more meaningful. This is what we call a data use identifier or DUI.

The IDEAS document is maintained by the Data Standards office at DIA with a collection of COBOL batch programs. We are currently in the process of revalidating and updating IDEAS so that the document will be even more useful.

In the near future, DIA plans to restructure and maintain IDEAS under a database management system. Longer range plans are to make it available on-line throughout the DoD intelligence community.

Now, I would like to discuss the effect, as we see it, of today's rapidly changing technology on data element standards. Database management systems, improved and increasing usage of networking, and automated message handling systems are just a few examples of changing technology that are causing people to look at data standardization with new interest. These and other technological developments allow for direct access of databases by end-users. This requires the data be stored in a standard format, and where possible, in a code that is more easily understood by the end-user. Lower cost of data storage allows for the use of longer, more human-readable codes.

SUMMARY

In summary, data standards are becoming more important to those who wish to gain direct access to other databases. It is becoming more and more feasible to store longer, more human-readable codes; but until these codes can be incorporated into existing databases, dual or optional standards will probably have to be maintained.

BIOGRAPHICAL SKETCH

Mr. Fritzges has been with the Defense Intelligence Agency for 19 years as a computer programmer and systems analyst. He has participated in the development of several computer systems designed to support the collection, processing, and dissemination of intelligence information with the intelligence agency.

Currently he is assigned to the Data and Message Standards Branch at DIA where he is responsible for the maintenance of the Intelligence Data Elements Approved Standards (IDEAS). He also serves as a chairman of the Data Standards Working Group within DIA and is the DIA voting representative to the Defense Intelligence Data Element Standards Committee (DIDESC).

DATA ADMINISTRATION TOOLS AND TECHNIQUES

Moderator

Margaret Skovira
Department of the Treasury
Washington, D.C.

Recorder

Gary Fisher

MANAGING A DATA DICTIONARY/DIRECTORY SYSTEM

Speaker

Harold Boylan
Department of the Navy
Washington, D.C.

ABSTRACT

The role of a data dictionary/directory system is important in the achievement of strategic goals where manpower, personnel, and training systems in the Navy are concerned. An overview of the Navy's ongoing project to produce automated systems in these areas emphasizes the need for a data dictionary/directory system and how such a system can be managed and administered.

The Navy's manpower, personnel, and training business is information dependent.

- o Manpower concerns what type of jobs are required to perform the Navy's mission, and how many personnel are required to staff the Navy at the levels authorized by Congress (figure 1).

- o Personnel deals with how many persons are in the Navy, where they are located, how much they get paid, etc.

- o Training plans for the instilling of technical knowledge in the personnel needed to staff the positions required to perform the Navy's mission.

The Navy has been transformed from a purely personnel-oriented organization to a highly technical one. Consequently, higher management levels require more information on personnel to maintain the level of readiness required by the mission. Whereas, before the shift to highly technical jobs, a head count of able-bodied seamen was sufficient to determine strength, now the personnel reports must contain information to allow commanders to determine if their nuclear-powered ships can get out of port. The Navy relies on highly technical people to run the fleet.

The systems that provide this information (the Navy's Personnel/Payroll information systems) have over 200 automated interfaces and 30 systems. Information deficiencies in these systems have been identified (figure 2). Some other real problems faced with in these systems are--

- o they are extremely complex;

- o the value of the information contained therein is not recognized;
- o there is no way to separate overhead attributed to students, patients, prisoners, etc., from the 650,000 active duty personnel;
- o data is not scarce and is not allocatable to operating costs under the current rules of accounting;
- o there is no commonality of data (many sources are disregarded in the development of new software, and data is collected through redundant methods); and
- o there is no current technology in the older systems.

This office has developed a definition of Information Resources Management (IRM) for the Navy to provide a basis for organizing resources in attacking the problems currently faced in information needs (figure 3). It consists of the following elements:

- o a strategic information plan
- o objectives and goals
- o central control of data
- o implementation of a quality control plan

The role of a data dictionary/directory system in this situation is to centrally administer and support data management goals, prototypes, standards, and security. In the Systems Interface Project and Data Registration Project, the Navy is using a data dictionary to capture information about all data elements, processes, relationships, functions, etc., see figure 4. To date, it has run into several severe obstacles such as difficulty in individual organizations being able to find all of their data and in the naming of data elements among different organizations.

Another project, the Data Flow Analysis and Systems Interface Inventory Project (figure 5), developed the organizational and data models upon which information needs were projected. The first task in this project was to develop a model of the information flow architecture, i.e., determining what organizations perform specific functions, figure 6. This was essentially a model of the physical architecture. Secondly, a data architecture was modeled from the flows identified in the information flow architecture and stored in the data dictionary, figure 7. A separate automated tool was used in the design of the data architecture to facilitate transferring the information to the data dictionary.

With the data architecture as a guide, the software development life cycle could be checked at specific milestones against the data dictionary, and determine how to improve control over the development and standards used in producing applications. Instead of writing data element standards, the Systems Interface Project is trying to influence the procedures that lead to "de facto" standards. However, other standards are produced for using the data dictionary and defining organizational functions.

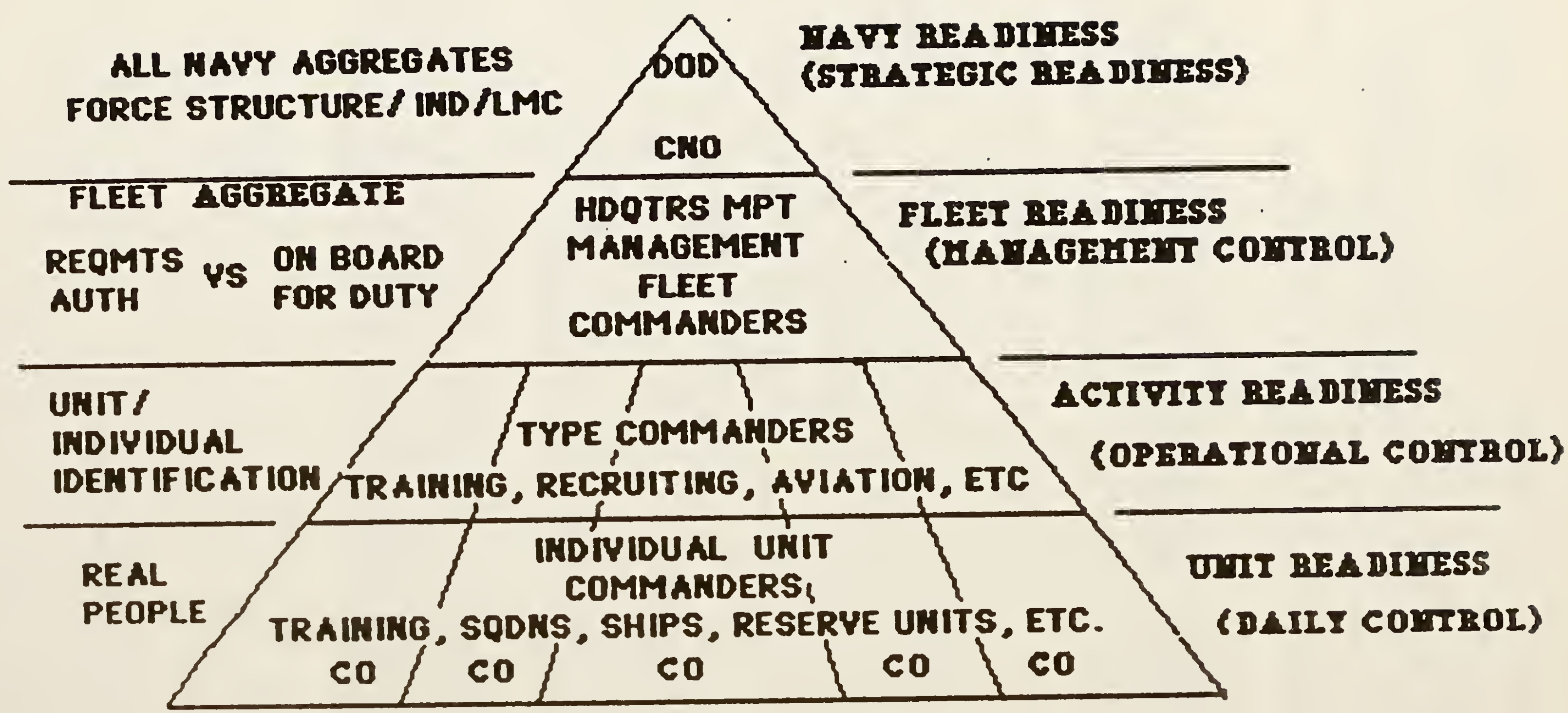
Finally, the data dictionary will become the core controller for the configuration management function (figure 8) and will be used extensively in the auditing and quality control of software. The goal is to make the corporate data conform to the requirements of management and users while maintaining central control over the data resource.

BIOGRAPHICAL SKETCH

Commander Boylan is the Director of the Data Resource Management (DRM) Program for the Deputy Chief of Naval Operations, Manpower, Personnel, and Training. He established the DRM Program to encompass strategic plans, information flow and data architectures, enterprise models, implementation plans, and management policy. He has been instrumental in promoting acceptance of IRM and DRM concepts among many of the large organizations which currently manage Navy's active duty, reserve, retired, and civilian personnel. Commander Boylan is also responsible for the creation of centralized data administration offices within these organizations to implement DRM policies, develop data standards for over 8000 data elements, maintain automated dictionaries, and resolve data management issues.

CDR Boylan graduated from the U.S. Naval Academy in 1968 with a B.S. in Systems Engineering. He received an MS in Computer Systems Management from the Naval Post Graduate School in 1976. Since then he has completed over two years of post graduate work at George Washington University in Financial Management.

MODEL OF INVENTORY MANAGEMENT



INVENTORY MANAGEMENT

ASSETS=PEOPLE

RECRUITS
STUDENTS
OTHER INDIVIDUALS
FORCE STRUCTURE
INACTIVE

Figure 1

THERE ARE INFORMATION DEFICIENCIES IN THE MPT BUSINESS

- SEPARATE REGULAR AND RESERVE PERSONNEL DATA BASES DO NOT ADEQUATELY SUPPORT MOBILIZATION
- FRAGMENTED AND REDUNDANT TRAINING SYSTEMS RESULT IN POOR QUALITY INFORMATION
- LACK OF ACCURATE AND TIMELY DATA CAUSED DECERTIFICATION OF THE NAVY PAY SYSTEM BY GAO
- FLEET COMMANDERS CAN NOT MANAGE MANPOWER RESOURCES WITHOUT ACCESS TO MPT DATA BASES
- PERSONNEL EXPENDITURES AND END-STRENGTH LEVELS CAN NOT BE RECONCILED DUE TO INCONSISTENT DATA

Figure 2

IRM - INFORMATION RESOURCE MANAGEMENT

THE STUDY, MANAGEMENT AND CONTROL OF INFORMATION AS A CORPORATE ASSET.

DRM - DATA RESOURCE MANAGEMENT

THE PROMULGATION OF PROCEDURES, STANDARDS, AND POLICIES RELATING TO DATA MANAGEMENT.

DA - DATA ADMINISTRATION

THE GROUP RESPONSIBLE FOR IMPLEMENTING AND MAINTAINING DRM STANDARDS, PROCEDURES, AND POLICIES

Figure 3

DATA & INFORMATION RESOURCE DIRECTORY DATA-REGISTRATION PROJECT (DIRD)

* COLLECT AVAILABLE DOCUMENTATION ON DATA ELEMENTS FOR EACH AIS (DATA DICTIONARY/RD/RECORD LAYOUTS/ETC)

* ENTER DEFINITION IN DIRD WITH KNOWN ATTRIBUTES

o NAME	o USE-OF-ELEMENT
o CONSTRUCTION	o REFERENCES
o ALIAS	o METHOD-OF-UPDATE
o CATALOGUE	o SOURCE
o DESCRIPTION	o EDIT-CRITERIA
o SPONSOR	o PERMISSIBLE-VALUES

* VERIFY DIRD DEFINITION WITH SOURCE DOCUMENT

* VALIDATE DIRD DEFINITION WITH ADP MANAGERS & USERS

Figure 4

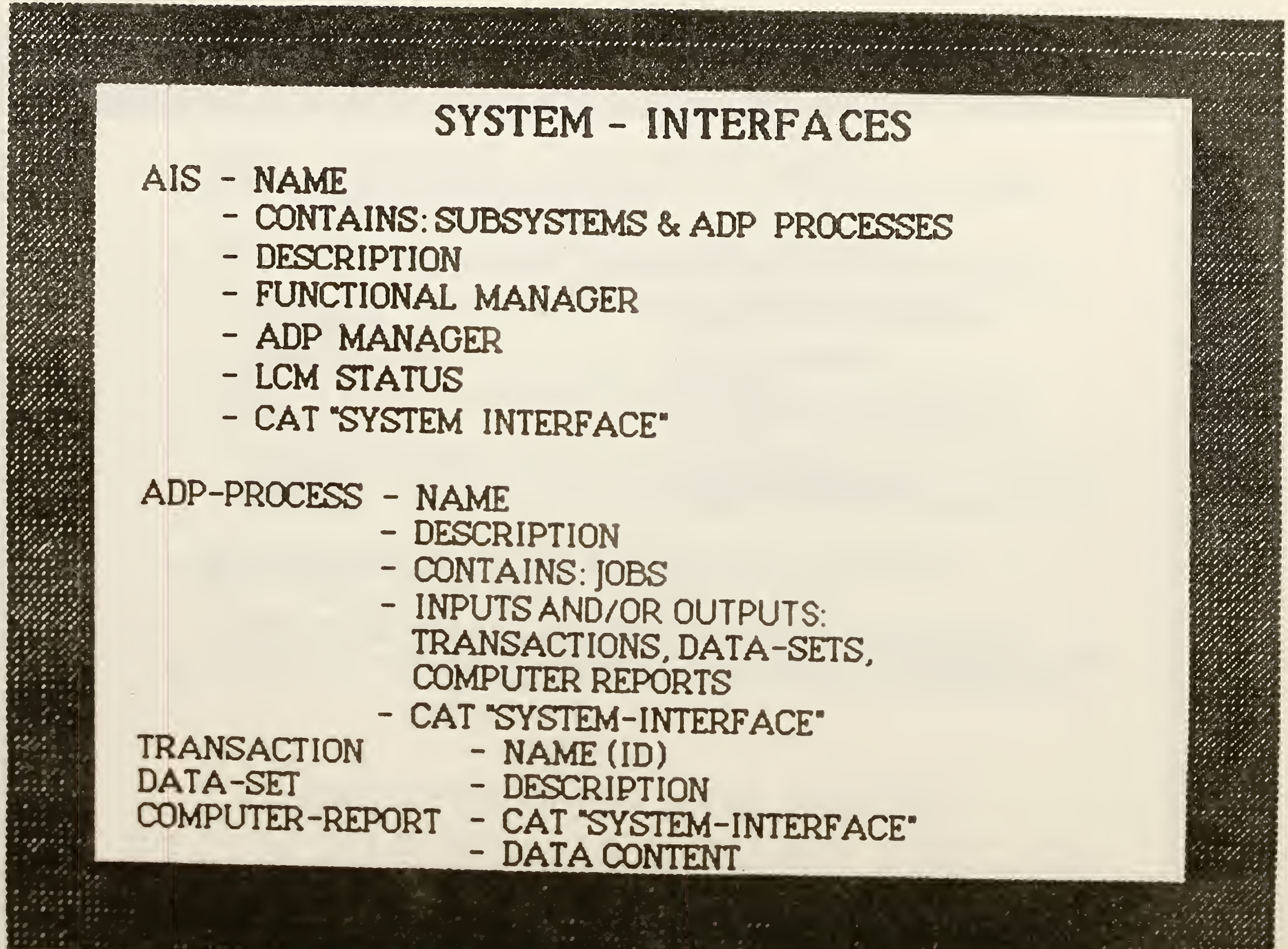
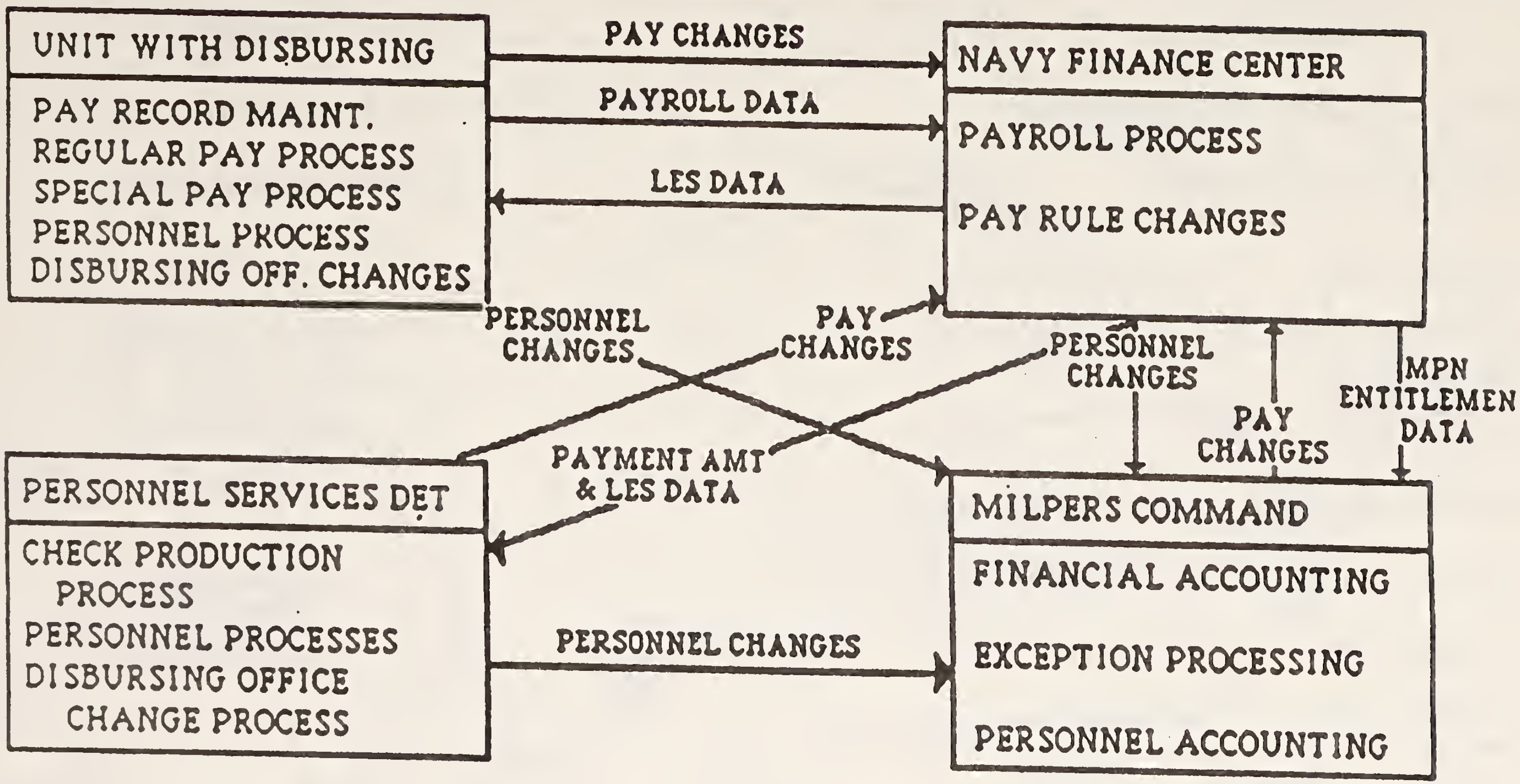

SYSTEM - INTERFACES
AIS - NAME
- CONTAINS: SUBSYSTEMS & ADP PROCESSES
- DESCRIPTION
- FUNCTIONAL MANAGER
- ADP MANAGER
- LCM STATUS
- CAT "SYSTEM INTERFACE"
ADP-PROCESS - NAME
- DESCRIPTION
- CONTAINS: JOBS
- INPUTS AND/OR OUTPUTS: TRANSACTIONS, DATA-SETS, COMPUTER REPORTS
- CAT "SYSTEM-INTERFACE"
TRANSACTION
DATA-SET
COMPUTER-REPORT
- NAME (ID)
- DESCRIPTION
- CAT "SYSTEM-INTERFACE"
- DATA CONTENT

Figure 5

INFORMATION FLOW ARCHITECTURE

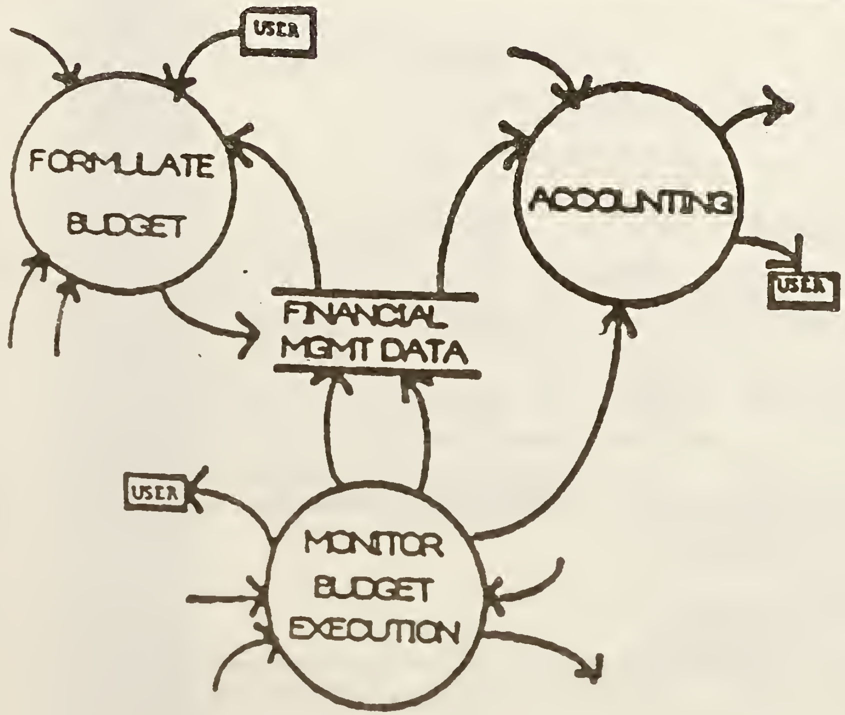


IS COMPOSED OF:

1. ORGANIZATIONS 2. FUNCTIONS 3. DATA FLOWS

Figure 6

DATA ARCHITECTURE

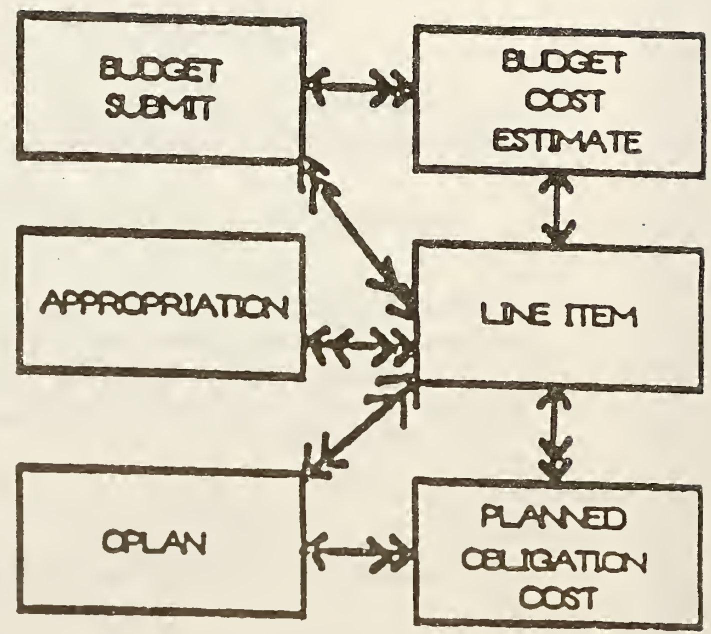


FUNCTIONAL MODEL

(A SET OF LEVELED DATA FLOW DIAGRAMS)

COMPOSED OF:

1. MPT FUNCTIONS
2. DATA FLOWS
3. DATA STORES



LOGICAL DATA MODEL

(ENTITY RELATIONSHIP DIAGRAM)

COMPOSED OF:

1. ENTITIES
2. RELATIONSHIPS

Figure 7

CONFIGURATION MANAGEMENT

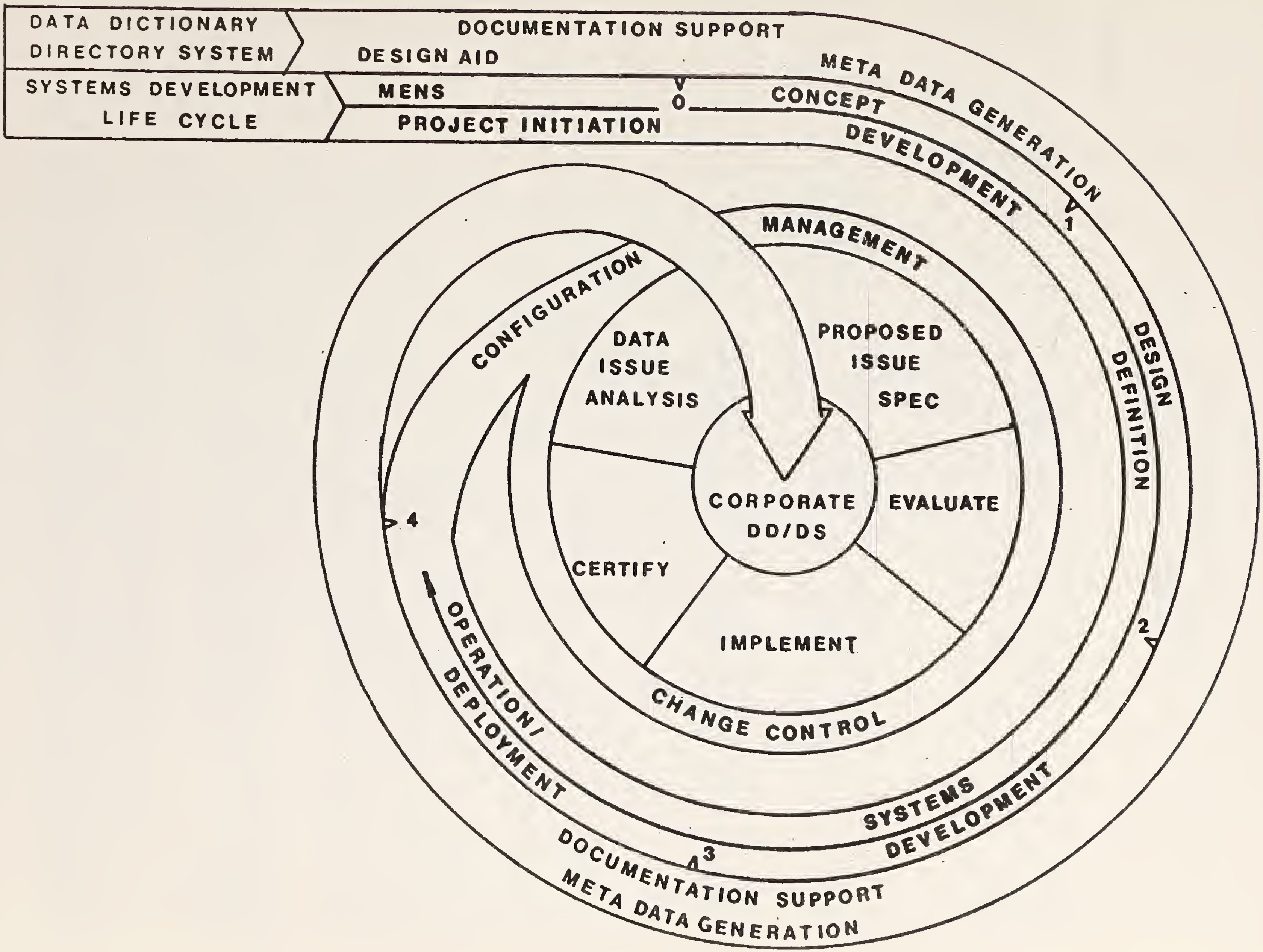


Figure 8

LOGICAL DATABASE DESIGN

Speaker

Joan Sullivan
National Bureau of Standards
Gaithersburg, Maryland

ABSTRACT

A guided tour of logical database design is presented with side trips concerning advice and examples. For those organizations that already have a logical database design methodology, this presentation will provide food for thought. For those that do not presently have a methodology, this should whet the appetite for investigating tools and techniques.

This is a guided tour of logical database design. It is based on NBS Special Publication 500-122, "Guide on Logical Database Design," and is intended to assist analysts in designing large and complex database systems.

Database design is not generally understood by the public. With the advent of personal computers and personal databases available through dBASE II and similar products, a novice user can generate a database application in a matter of hours or minutes. This familiarity and ease of use for simple data models leads to high expectations for designing large databases, as if the design were simply a matter of scaling things up. It is important to note that sharing data throughout an organization adds a whole new dimension of difficulty to the problem of designing a database. Some of the problems associated with large, shared, databases include the complexity of the logical model (thousands of data elements, hundreds of records and relationships), concern for performance, support for multiple applications, overlapping needs and views of the data, conflicting naming conventions, shared responsibility for data integrity, simultaneous access and update, a variety of security needs, requirement for high reliability in a real-time environment, micro/mainframe links and distributed processing, links to other information systems, and an overwhelming volume of data and metadata.

Generally, a large database design cannot be accomplished by one person. It requires a team of users and specialists in database techniques working over a period of months or years in a highly organized manner.

The scenario for this guided tour is described as follows:

- o The application specialist or systems analyst for a Federal agency has just been handed a newly-completed

Business Systems Plan (BSP). The BSP describes 24 databases and 37 processes. Still, not all requirements are known.

- o The database design team is funded and staffed. Priorities have been established among the candidate databases and processes identified by the BSP.
- o The team is responsible for 3 of the 24 subject oriented databases which provide data for 11 of the processes.
- o During design, the team must allow for interfaces to the other databases and processes to be designed at a later date.

The design team will build two distinctly different models of the data and information needs in the organization (see figure 1). The first model (the two boxes on the top) documents the current information system, including plans for future needs. This model is process-oriented and consists of the local information-flow models and the composite global information-flow model.

The second model (the two boxes on the lower half of the diagram) is a <u>data-structure</u> model. It consists of the conceptual schema and the external schema which document the organizational view and the users' views of the data's structure.

During this modeling process, the team will collect information such as the types of data needed, the volume of records and transactions, relationships between collections of data, frequency and priority of access paths, security, privacy, and integrity constraints. This information will come from interviews, reports, forms, existing computer applications, etc.

The local information-flow model or LIM (figure 2) depicts the information needs of a single organizational unit, person, function, or event (the center box). Other boxes represent organizational units which exchange data with that unit. This model does not worry about the exchange of information beyond the focal point, i.e., among the other boxes. The concern here is to limit complexity and to avoid speculation as to how others might use information. The objective of the LIM is to <u>focus</u> on the data needs of a single unit. Some units (such as management) may deal with summary data and information packages. Other units (such as clerks and technicians) may concentrate more on individual data elements. In any event, the LIM documents the information needs of a unit on the level at which the information is used and understood.

The global information-flow model or GIM (figure 3) is an interconnected collection of all the LIM's. This model tracks data as it crosses organizational boundaries or flows through functions and events. The GIM consolidates LIM's, resolving definition and naming conflicts. The GIM will refine the boundary of automation. This may reduce the scope of the logical database design and therefore reduce the effort expended in subsequent phases. After all, it is not feasible to automate <u>all</u> functions. Additionally, the GIM will define the interfaces of the database with other databases and systems (both automated and nonautomated).

As the current information model is being documented, another model, the composite data structure, is being built. This model is called the conceptual schema or CS (figure 4) and describes the logical structure of the data required by an organization. The CS is not concerned with how data is collected (such as input forms) or how data is distributed (such as periodic reports), but rather, what data should exist in the database and how it should be grouped and interrelated. Normalization as well as other types of analysis are employed to refine the CS to satisfy certain technical goals.

One method for building the conceptual schema is to use entity-relationship diagrams to represent real-world objects (entities), their identifying characteristics (key attributes), and their interactions (relationships) with other objects (figure 5). As the model is developed in greater detail, additional attributes (data elements) will be assigned to the entities.

The external schema or ES (figure 6) extracts from the CS those entities, relationships, and attributes needed by a given LIM. Local names (synonyms) may be used. The primary function of an ES is to help users and programmers interact with the database by presenting a simplified view of the database in terms which are familiar to them. The building of the ES also serves as a completeness check, verifying that the data needs of each function to be automated are addressed in the composite structure. Of course, some data needs will, by design, be excluded from the database and will be provided by other means, perhaps even by manual procedures.

Basically, the procedure involved in logical database design is a top-down hierarchical analysis of goals and functions of the organization (figure 7). Although detailed information about data elements is needed for the final logical database design, preliminary analysis will focus on data groupings and identifiers as well as the broader mission-oriented functions performed. Initial interviews will be held with administrators and planners to gain an organizational perspective of the data. Later interviews with managers and specialists will focus on increased detail for functions and data elements. Eventually, interviews will be used to collect information on (or to verify) data element definitions, functional dependencies among data elements, and use of the data by various functions. All these activities are supported by the use of automated tools (see figure 8).

A useful analogy for explaining the logistics involved in logical database design (figure 1) is contained in the following situation:

- o Instead of dining at the NBS cafeteria during the course of the workshop, suppose all attendees decided to celebrate fast food week by sending out to McDonald's for lunch. What is involved?

- o A team of individuals would circulate among the attendees to get each person's order.

- o The team leader would then consolidate the orders into one group order with composite requirements for hamburgers, fries, etc.

- o The team would go to McDonald's, place the order, and watch the confusion as the friendly, courteous staff converts numbers into packages.

- o The team leader would check orders as they were finished and mark the composite list to verify that (allowing for substitutions) what was ordered is what was received.

- o The team would then return to NBS, extract individual orders, and verify from each individual that they received their proper order.

As a database designer, the process is even more difficult. Each individual order for information is taken, requirements are combined into a single organizational view, and then the model is switched. Entities, or data groups, are abstracted and checked against the organizational view to make sure that information is still intact. Finally, entities are extracted to conform to a particular user's view of the data. Many of these processes may be performed concurrently. Interviews may take place as soon as the mission of the team is understood. Forms may be gathered and analyzed. One word of advice is to employ an automated tool such as a data dictionary to record the information gathered. This will keep the detail of this analysis to a manageable level. A data dictionary can be used to generate a variety of reports and cross references for the design team and for users.

As the information flow model develops, feedback may be obtained from users to make sure that the model incorporates their views of the data, and that understanding between the design team and the users is facilitated.

Some questions and answers that usually pop up on a tour like this are--

- o Where does the information come from? The information used to analyze information requirements is gathered from interviews, and analysis of forms and documents.

- o Where do the interviews start? The Business Systems Plan identified major processes that can be traced to the responsible organizations. In addition, organization charts, statements of mission, etc., point to organizational entities that can give guidance.

- o How do you make the intuitive leap from forms and reports to entities and attributes? Look for natural divisions and groupings among types of real-world data. Do not try to force groupings that are difficult to comprehend or do not seem right. Normalization techniques should eventually be used to refine these groupings.

- o What are the deliverables from logical database design? A data dictionary populated with the information used to derive the models, and the drawings used to express the model to users (i.e., local information flow model, global information flow model, conceptual schema, and external schema).

This wraps up the tour. In parting, you may find it helpful to look at a map of where we have been. Figure 9 shows logical database design in the context of the information systems life cycle. Under data activities you see the two types of models we have discussed. One final point to remember is that you cannot design a shared database unless you understand the shared data.

BIOGRAPHICAL SKETCH

Joan Sullivan is a computer specialist in the Database Architecture Group at the National Bureau of Standards. She has assisted Dr. David K. Jefferson in producing the NBS Special Publication Guide on Logical Database Design. Current activities include the study of fourth generation languages/prototyping tools and knowledge-based systems, and standards meetings for SQL (Structured Query Language).

Prior to joining the Database Architecture Group at NBS, Joan supervised a group of programmers supporting administrative systems such as personnel, payroll and financial management. These systems were largely COBOL file systems migrating towards an integrated database implementation with ties to microcomputer subsystems.

Joan received a Bachelor of Arts in mathematics from Trinity College in Washington, D.C., and completed a year of graduate studies in theoretical mathematics.

DIAGRAM OF THE FOUR LDD PHASES

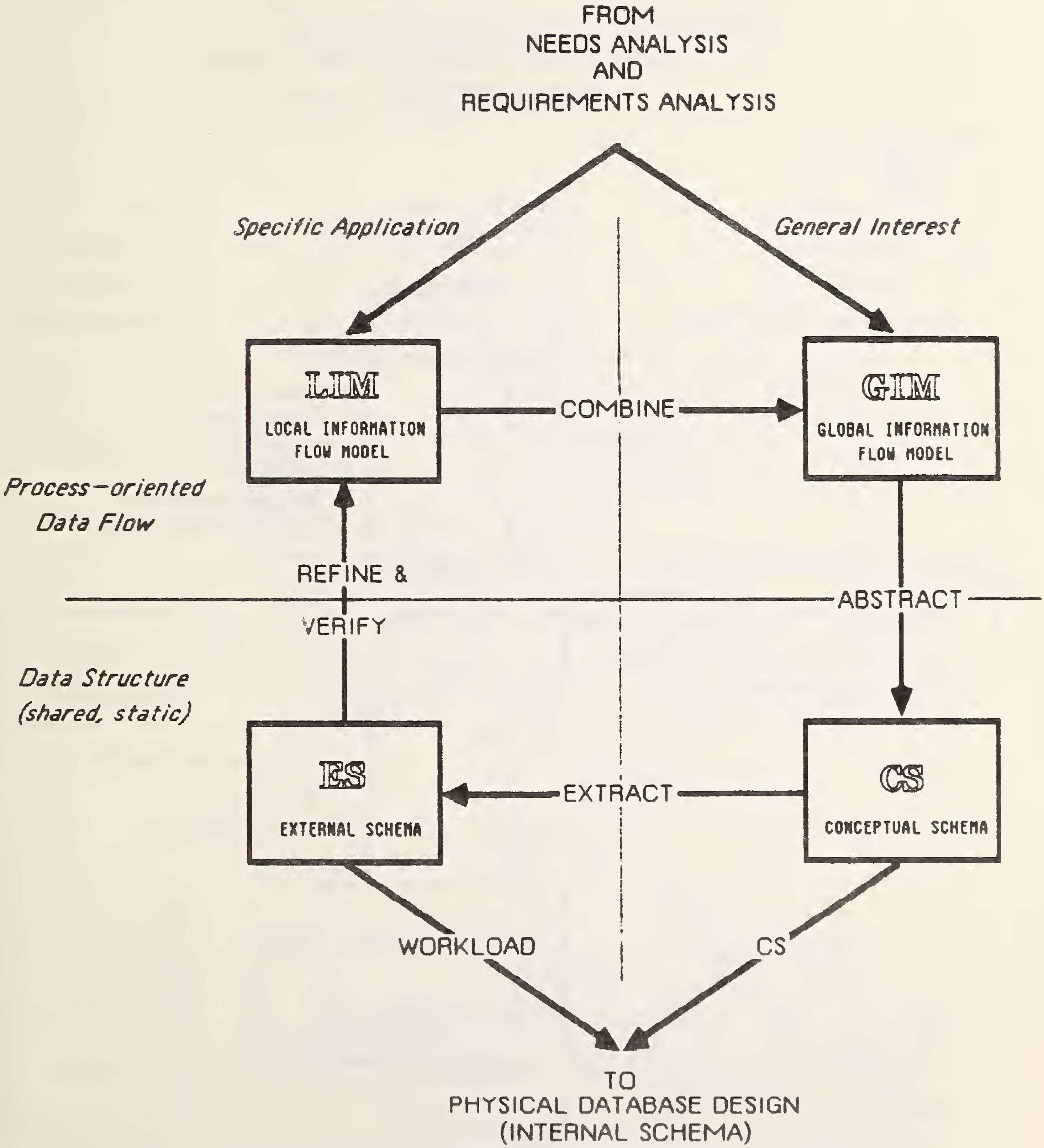
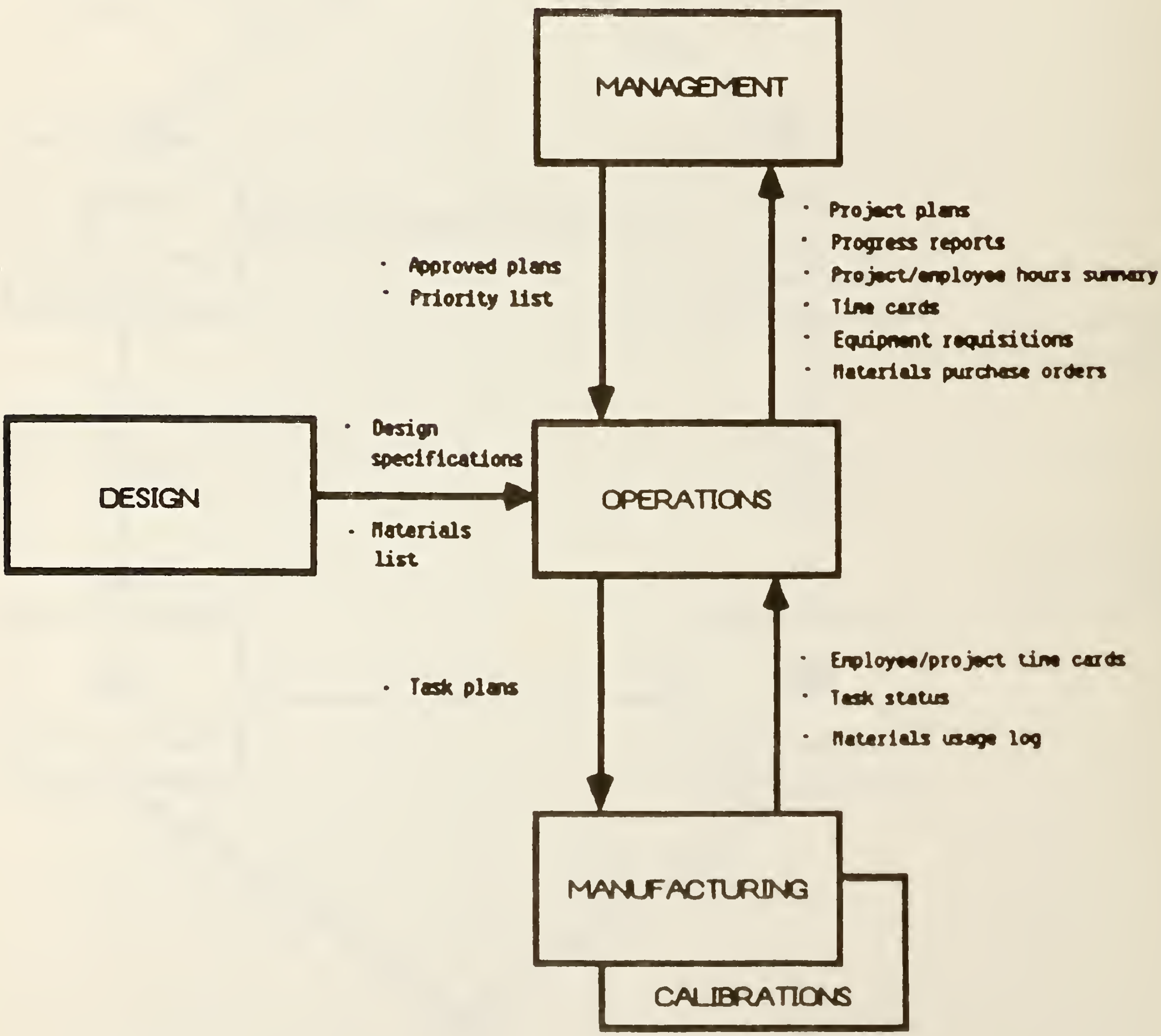


FIGURE 1

INSTRUMENT FABRICATION DIVISION

Local Information-flow Model
OPERATIONS Unit



<u>NOTES</u>

OPERATIONS is responsible for coordinating the efforts of MANUFACTURING and CALIBRATIONS, scheduling tasks, ordering materials and equipment, reporting material and labor spent on each project.

FIGURE 2

AGENCY FINANCIAL MANAGEMENT SYSTEM

Global Information-flow Model

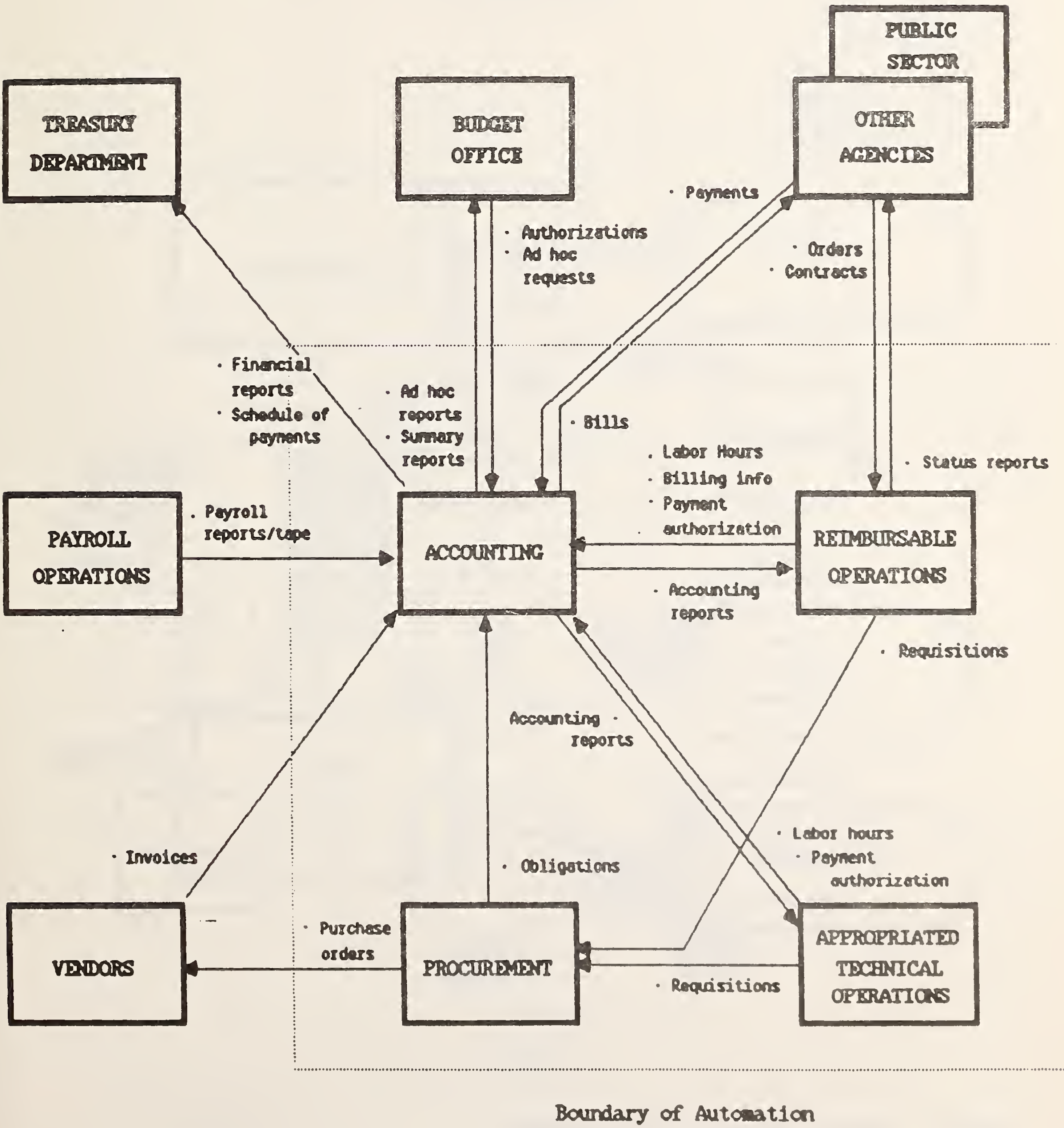


FIGURE 3

AGENCY FINANCIAL MANAGEMENT SYSTEM

ENTITY-RELATIONSHIP DIAGRAM OF CONCEPTUAL SCHEMA

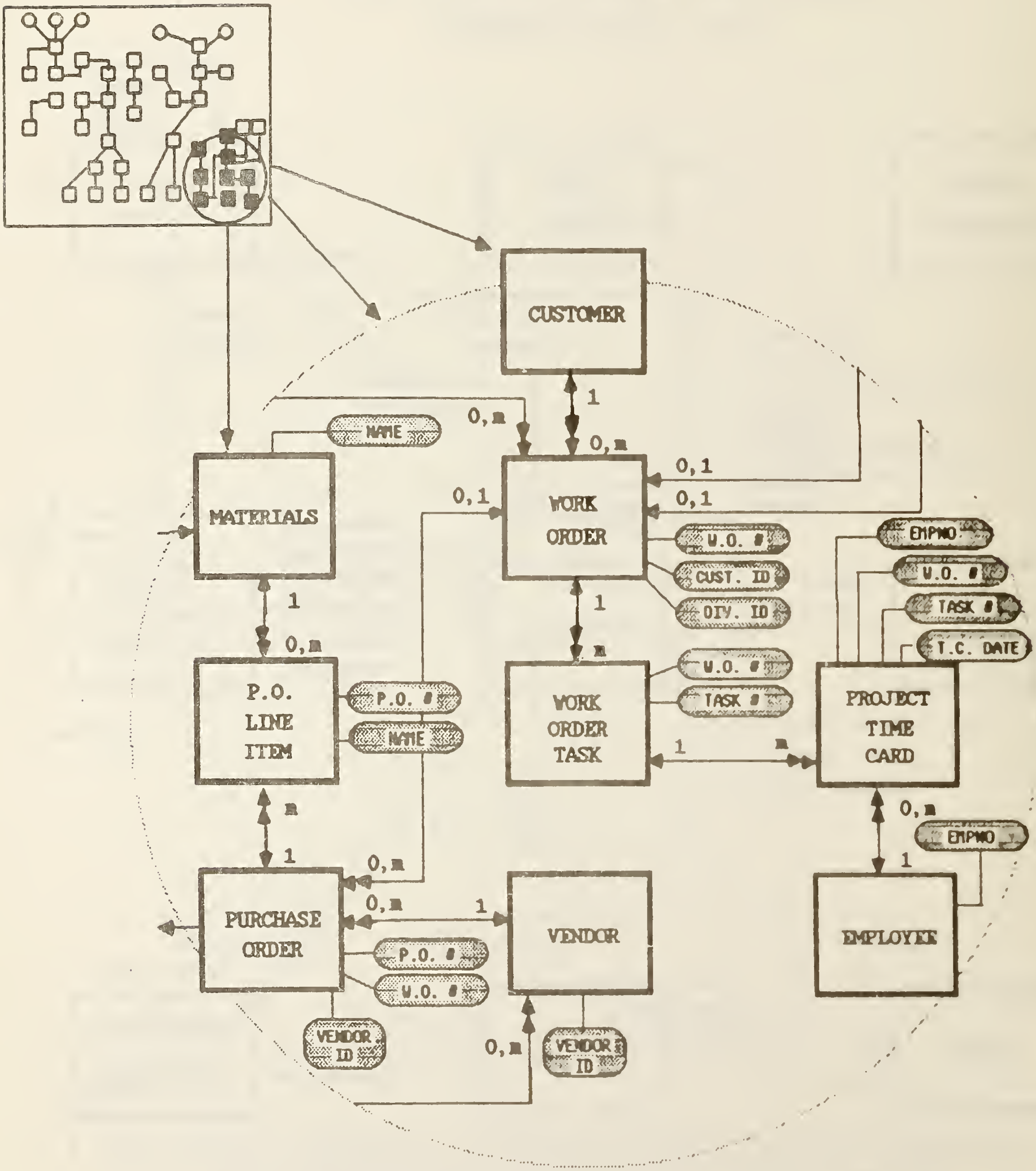


FIGURE 4

NOTES: Non-key attributes are not shown.
Data dictionary reports list all attributes.

ENTITY-RELATIONSHIP-ATTRIBUTE DIAGRAM

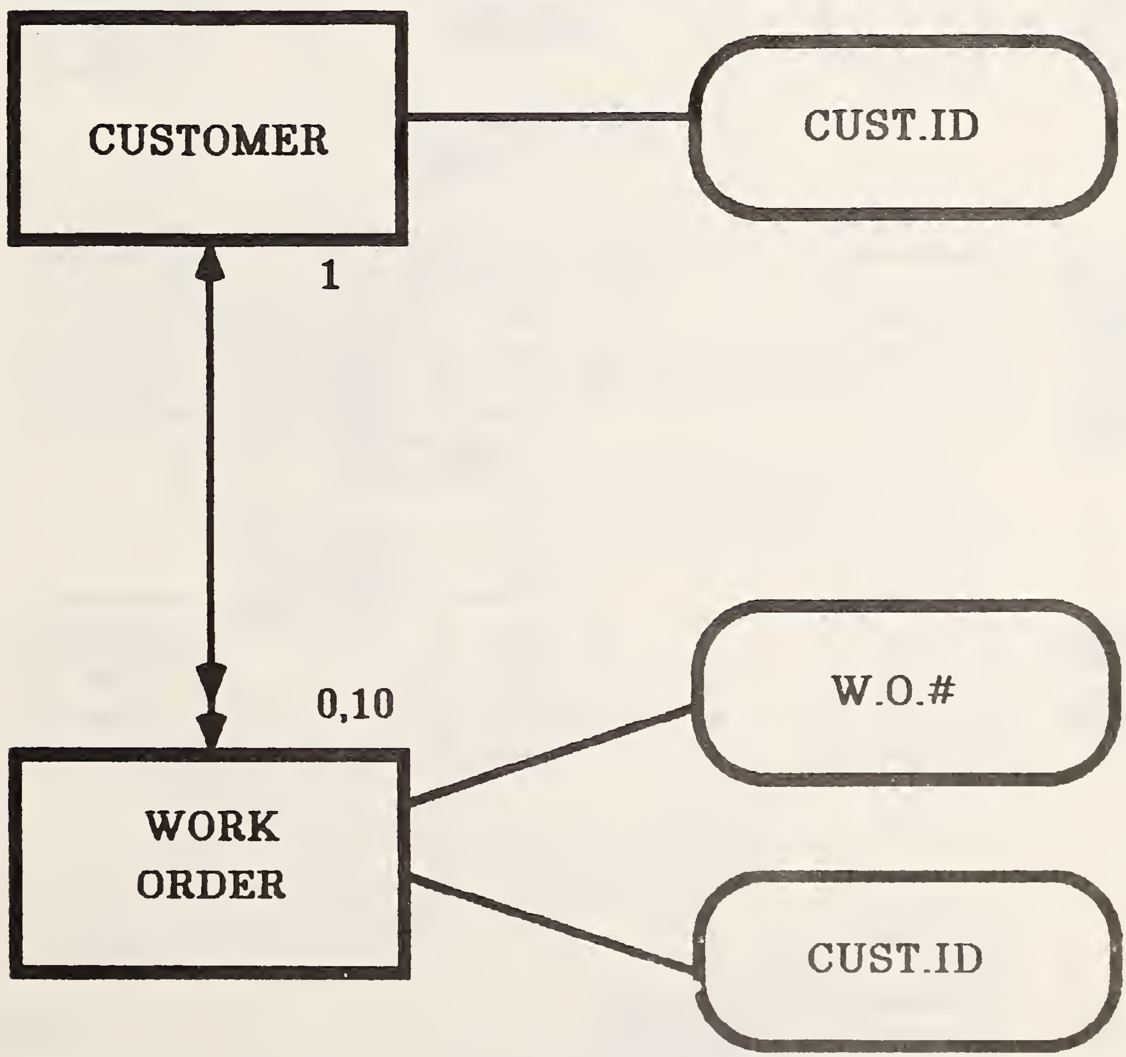


FIGURE 5

AGENCY FINANCIAL MANAGEMENT SYSTEM
EXTERNAL SCHEMA

Function : Close Out Work Order

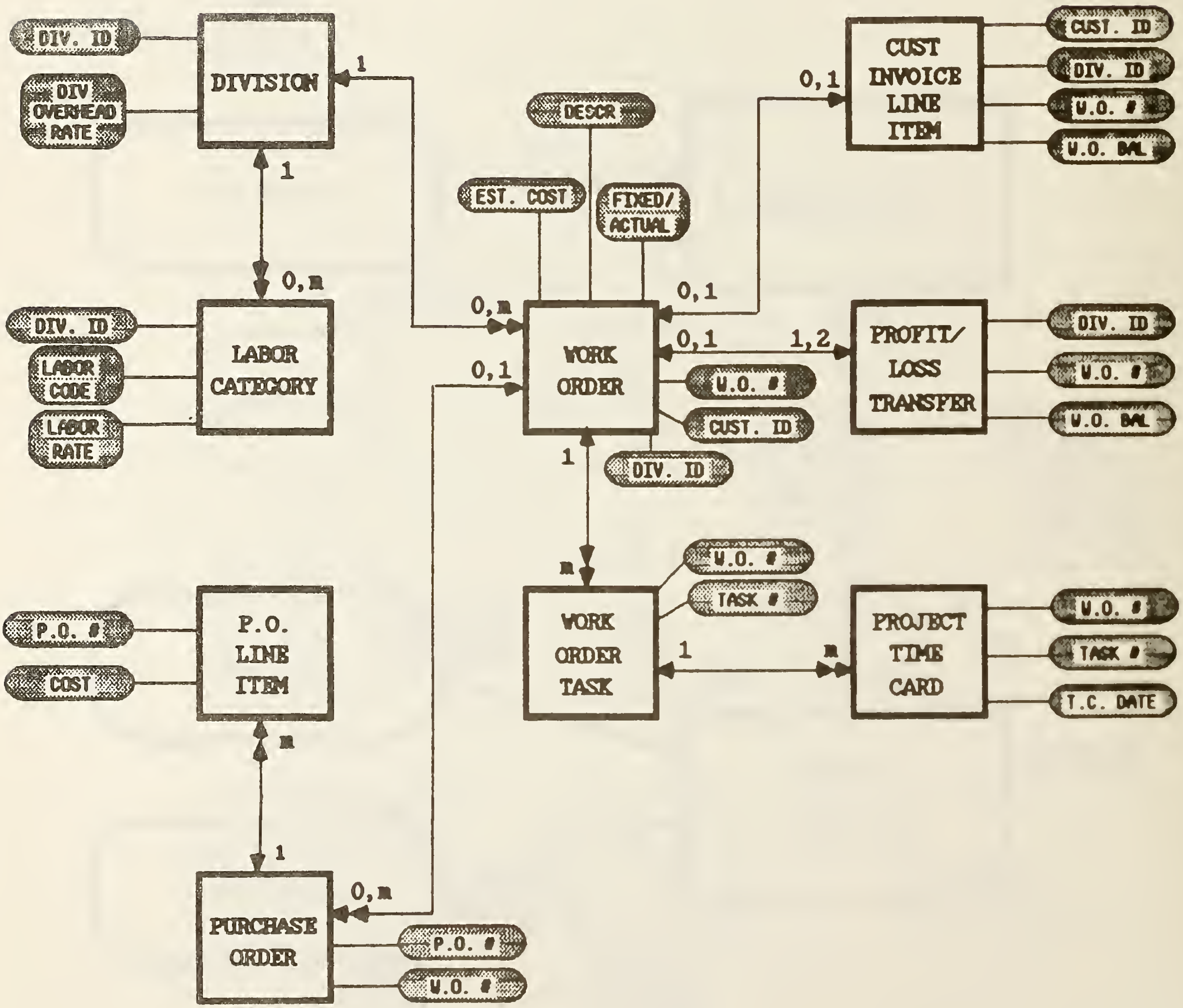


FIGURE 6

NOTE : Entities, relationships and attributes not used by this function are not shown. Complete details are available from the data dictionary.

LOGICAL DATABASE DESIGN METHODOLOGY

LDD PRACTICES

DESIGN STRATEGY

- o MANAGEMENT DIRECTION (DERIVED FROM BSP)
- o HIERARCHICAL, TOP-DOWN APPROACH
- o ITERATIVE REFINEMENT
- o CLEARLY DEFINED STEPS FOR ANALYSTS AND DESIGNERS
- o SERIES OF CHECKPOINTS
    - o PROGRESS REVIEW FOR DESIGNERS AND MANAGERS OF LDD
    - o SYNCHRONIZATION WITH OTHER PARALLEL LIFE-CYCLE PHASES

ANALYTICAL METHODS

- o DIFFERENTIATION OF VARIOUS POINTS OF VIEW
    - o ORGANIZATIONAL COMPONENTS
    - o FUNCTIONAL COMPONENTS
    - o EVENT, CONTROL AND DECISION STRUCTURES
- o DETECTION OF REDUNDANCIES, INCOMPLETENESS
- o NORMALIZATION PROCEDURES

STANDARDS

- o A MODE OF NOTATION (GRAPHIC OR SYMBOLIC)
- o A SPECIFICATION LANGUAGE
- o NAMING CONVENTIONS

FIGURE 7

LOGICAL DATABASE DESIGN METHODOLOGY

LDD TOOLS

DATA DICTIONARY

- TO RECORD, STORE AND PROTECT DESCRIPTIONS OF INFORMATION RESOURCE
- TO PROVIDE A VARIETY OF CROSS REFERENCE REPORTS FOR ANALYSIS
- A FRAMEWORK FOR ENFORCING STANDARDS
- A CONTROL POINT FOR COORDINATING OTHER LIFE-CYCLE PHASES

DESIGN AIDS

- CONSISTENCY CHECKERS
- GRAPHICS PREPARATION
- NORMALIZATION ROUTINES

FIGURE 8

INFORMATION SYSTEMS LIFE CYCLE

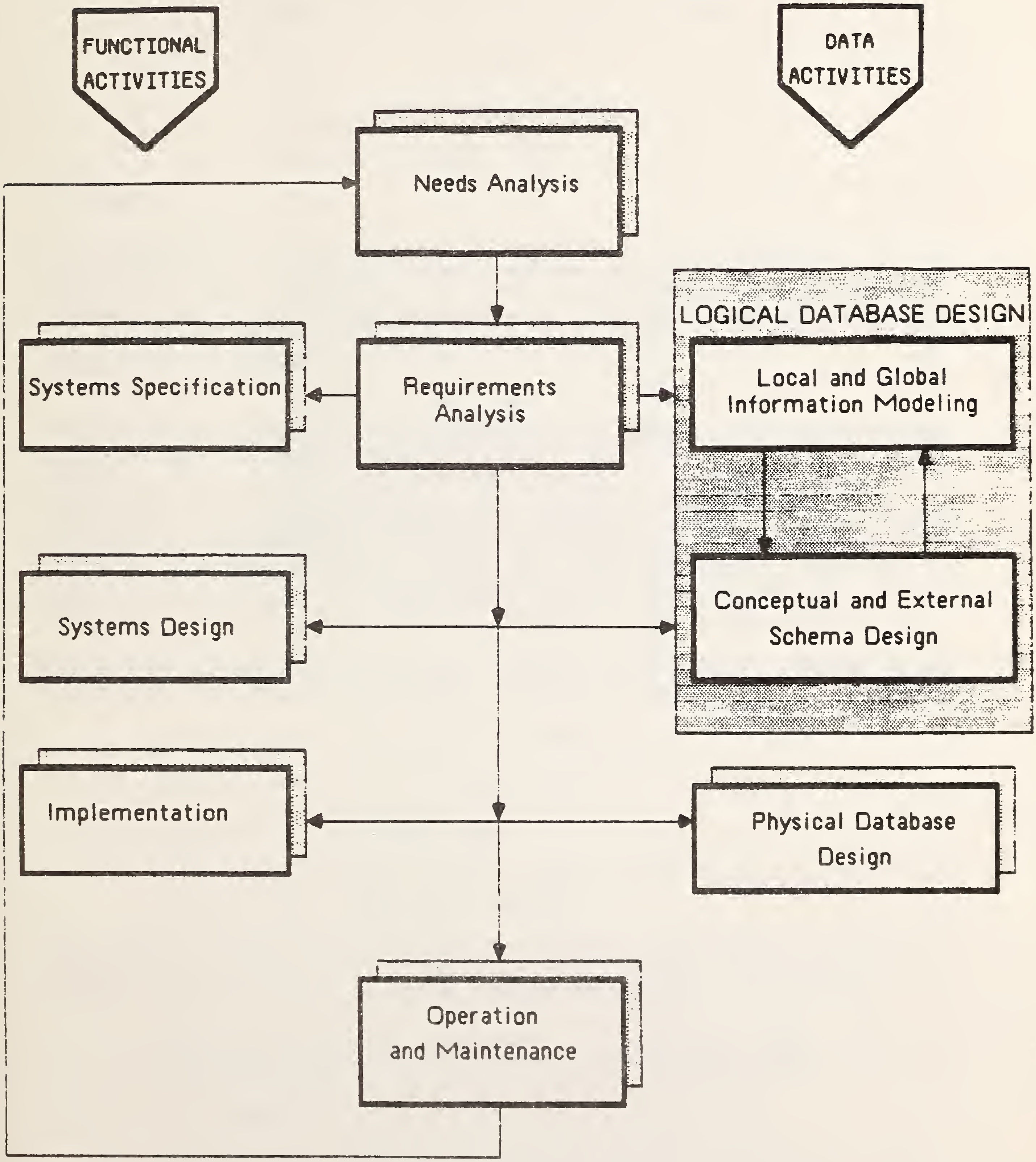


FIGURE 9

MARINE CORPS STANDARD SUPPLY SYSTEM

Speaker

Capt. David Hering
Marine Corps Logistics Base
Albany, Georgia

ABSTRACT

The application of data dictionary/directory systems in the development of automated supply systems for the Marine Corps is central to the four areas of human performance monitoring, data engineering, application development, and systems engineering. The results of building a Corps-wide data model for supply systems is discussed.

Marine Corps logistics management is based on information provided by the Marine Corps Unified Materiel Management System (MUMMS), the Direct Support Stock Control System (DSSC), the Supported Activities Supply System (SASSY), and the Marine Corps Integrated Maintenance Management Systems (MIMMS). All of these systems are antiquated (written in assembly/COBOL language for batch processing) and were produced with mid 1960's and early 1970's technology. In the late 1970's, an on-line data entry mechanism, of sorts, was added in front of SASSY and DSSC.

The Marine Corps Standard Supply System (M3S) project is designed to replace the majority of these systems. Objectives of M3S are--

- o provide real-time inquiry capabilities;
- o reduce paperwork by 40 percent;
- o reduce training costs; and
- o reduce maintenance costs, both in the system upkeep and Marine Corps logistics changes.

The Systems Engineering branch of the M3S Development Office is responsible for the implementation of M3S. The branch is concentrating in four areas of system development. They are--

- o human performance monitoring through project control, estimating, and work breakdown structuring tools;
- o data engineering using data dictionary tools;
- o application development through program generation; and
- o systems engineering.

A major step in project management was initiated by bringing an Integration Support Contractor on board. The presence of an independent contractor was designed to give continuity to the project since many changes in tours of duty take place over the course of a long-term project. In addition, the contractor was to lend an outside non-parochial view of the world to put the M3S problems into perspective. The contractor was to provide an analysis of data requirements, standardized system engineering, and analytic support in project planning, management, interface definition, and control.

The eventual goal in Phase 2 is to build a world-wide Marine Corps data model for field site systems.

M3S architecture consists of a central policy for hardware procurement and seven field sites utilizing distributed data, databases, CPU's, and software. These items are controlled centrally so the same configuration exists at all seven sites.

The first database model of M3S was designed using DATAMANAGER as the data dictionary for cataloging the information uncovered in analysis. Five thousand data elements were defined, of which 1600 are currently used. The other 3400 elements have not been used as yet. Originally, 20,000 data elements were identified and catalogued before synonyms and homonyms were resolved. The model consists of three levels (figure 1): the normalized data structures at the raw data level; the decision support structures extracted from the normalized structures; and the personal database structures for microcomputers extracted from subsets of the decision support structures and normalized structures.

The first model has been used to date to implement these projects:

- o support an Air Force database for a 500 bed hospital (The database elements were set up in two days, and programmed logic was in place in three weeks.)
- o create a Navy medical logistics system in six months
- o establish a system to assist in prepositioning ships geographically.

Eventually, the information stored in the data dictionary will be fed to automated tools to generate source code from requirements. The system can currently generate reports from user views recorded in the data dictionary.

Some of the problems encountered included the following:

- pressure from management to corrupt the database structure to facilitate retrieval of information (This led from the shock of having so much information available from M3S which was not available before and which is now more accurate and timely.)

- converting an application before it was ready to be converted (It is difficult to say when enough is known about a system to make the decision.)

- a general lack of technical expertise in management, analysts, and programmers

- conflicts among Marine Corps requirements, DoD requirements, and the data model.

The positive result of utilizing a data dictionary is that many of these problems could be overcome by referencing information stored in the dictionary and presenting it as evidence to support the arguments in favor of or against each side.

In summary, the M3S project has used/is using the data dictionary to assist in normalizing the data, document outside forces that interact with the system, and supply information to keep the project in the same general direction of development.

BIOGRAPHICAL SKETCH

Captain David Hering is a Data Systems Software Officer in a database management office where he is involved in the development of the Marine Corps Standard Supply System (M3S). In his 18 years with the Marine Corps, Captain Hering has served as a program systems officer, as head of a programming branch, and as a programming instructor before taking his present position. Captain Hering holds a BS degree in data processing from the University of North Carolina.

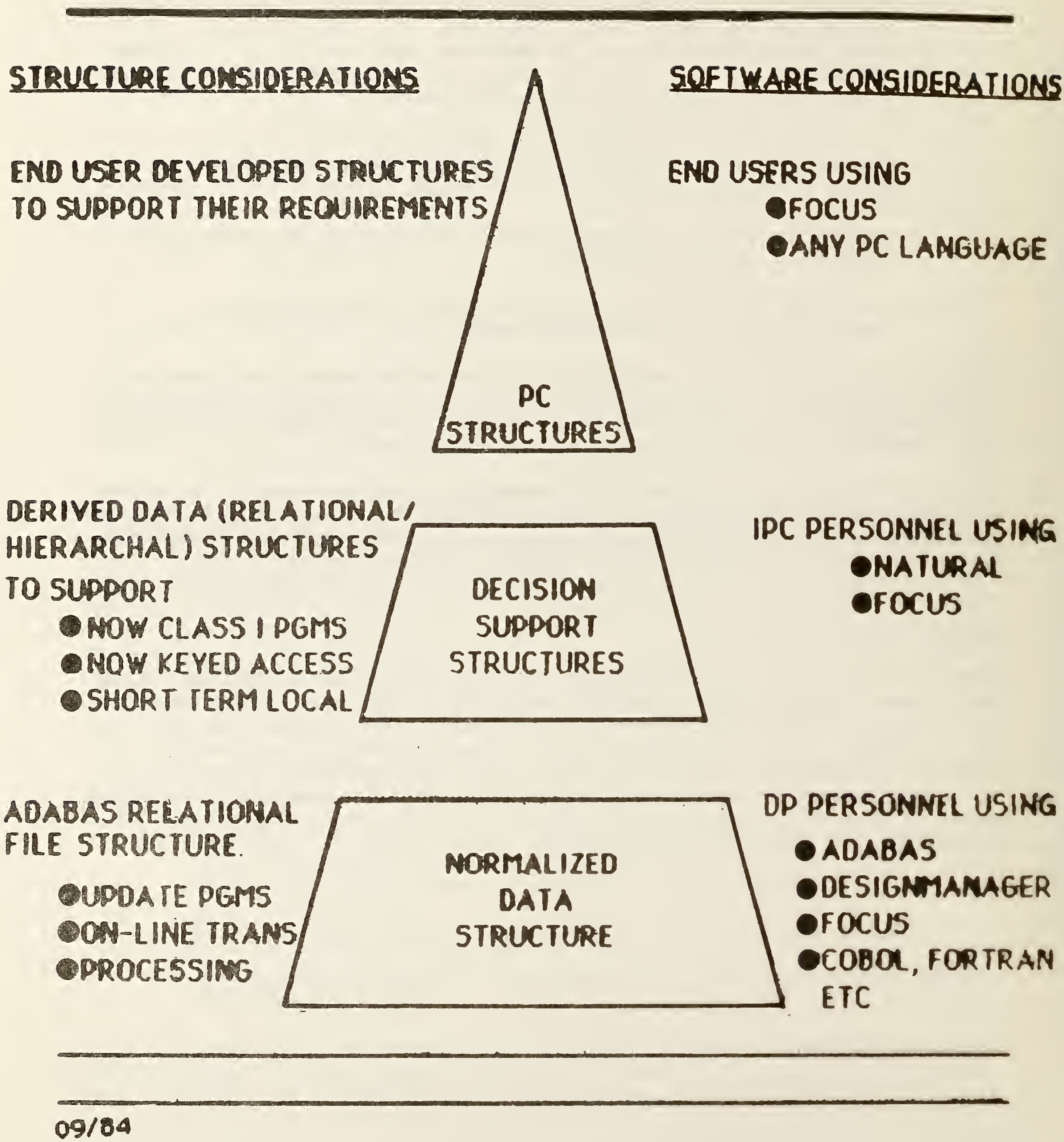

STRUCTURE CONSIDERATIONS
SOFTWARE CONSIDERATIONS
END USER DEVELOPED STRUCTURES TO SUPPORT THEIR REQUIREMENTS
PC STRUCTURES
END USERS USING
●FOCUS
●ANY PC LANGUAGE
DERIVED DATA (RELATIONAL/HIERARCHAL) STRUCTURES TO SUPPORT
●NOW CLASS I PGMS
●NOW KEYED ACCESS
●SHORT TERM LOCAL
DECISION SUPPORT STRUCTURES
IPC PERSONNEL USING
●NATURAL
●FOCUS
ADABAS RELATIONAL FILE STRUCTURE.
●UPDATE PGMS
●ON-LINE TRANS
●PROCESSING
NORMALIZED DATA STRUCTURE
DP PERSONNEL USING
●ADABAS
●DESIGNMANAGER
●FOCUS
●COBOL, FORTRAN ETC
09/84

Figure 1

DEMONSTRATION OF A DATA MODELING TOOL

Speaker

William Kurator
U. S. Postal Service
Washington, D.C.

ABSTRACT

The selection of a data modeling methodology and automated tools is a difficult task as the experience of the U. S. Postal Service illustrates in this presentation. However, there are tools and methodologies to fit numerous styles of organizations. One such methodology is the Curtice-Jones associative data model. The demonstration of a tool which automates this model is part of the presentation.

The U. S. Postal Service developed a Business Systems Plan (BSP) in 1980 to cover financial data, data administration, application administration, and several other areas of importance to the organization. A methodology was needed to assist in performing the analysis of the BSP requirements and design a database model for each functional area of the BSP. Several methodologies were examined. Among them were Peter Chen's entity relationship diagram model, Yourdon's data object model, the Holland-Ross models, and finally the Curtice-Jones model from Arthur D. Little Company.

A severe criticism of all models was that the terminology used in each caused a great deal of confusion. The terms defined for different pieces of each model were sometimes the same and sometimes not. There turned out to be no standard terminology for database modeling. Over simplifying, it appears that a methodology is based in large part on the uniqueness of its terminology.

The Curtice-Jones' associative data model was chosen for its ease of understanding and usage as compared to the other models. One of the basic principles of the Curtice-Jones' model is that data does not mean anything by itself. It has meaning only in relationships or association with other pieces of data. The fundamental concepts of the Curtice-Jones' model involve entities, identifiers, domains, assertions, and data elements (figures 1-5). Normalization is not part of the model. However, using an established procedure, the Curtice-Jones' model can be normalized for use in relational database management systems. Domains are standardized for data elements. Assertions contain keys, data items, and associators tied together in relationships to give meaning to the data structure.

An automated tool, ADL/IRMA, is associated with the Curtice-Jones' method. The tool provides an easy to use interface to the diagramming methods used, as well as facilitating understanding between the end user and analyst.

ADL/IRMA is menu driven with high reliance on function keys. Capabilities include the following:

- o enter and display logical database structures
- o print reports (data structure, element definitions, domain definitions)
- o process data flow descriptions
- o design screens and displays

ADL/IRMA supports logical database design by storing data element descriptions, domain definitions, assertion templates, and produces various reports. Data flow diagrams as described by Gane and Sarsen are also supported.

The U.S. Postal Service is currently using the tool on several finance, logistics, and production systems.

BIOGRAPHICAL SKETCH

Mr. Kurator is employed with the U.S. Postal Service and currently is working on a logical database design for the entire Postal Service Finance System. Prior to the U.S. Postal Service, he worked in various areas of Computer Science and Operations Research for both the government and private industry. At the Department of Energy he worked in energy modeling. The large national energy models involved the use of linear programming and state-of-the-art network theory.

Mr. Kurator has a B.S. degree in mathematics from the Purdue University and a M.S. in Numerical Science from Johns Hopkins.

ENTITY

An Entity (occurrence) is an object (real or abstract) to which the data base refers.

Distinguish between entity occurrence and entity class.

Part refers to an entity class.

The Empire State Building refers to an entity.

Figure 1

ENTITY IDENTIFIER

An entity identifier is a symbol string which has been assigned to an entity and is used to refer to that entity within the date base.

Note: The assignment can be

one-to-one (good)
one-to-many (not so good)
many-to-many (ugh!)

Figure 2

DOMAINS

The set of correspondences between members of an entity class and their identifiers is called a Domain.

Data Standardization means Domain Standardization (Application development projects can define new data elements; only data administration can define new domains.)

Distinguish between domains and data elements in your data dictionary.

Figure 3

ASSERTION

A data base assertion is the representation of a relationship or mapping between entities in two domains (or between entities in the same domain).

Figure 4

DATA ELEMENT

A Data Element is a Key or Target, suitably defined over a domain, in a data base assertion template.

Figure 5

APPENDICES

APPENDIX A: SAMPLE JOB DESCRIPTIONS

DATA ADMINISTRATOR

Immediate Supervisor

V.P. of Information Resource Management

Job Summary

The Data Administrator manages the staff assigned to do data planning, analysis, modeling, documentation, and the mapping of database designs against the strategic plan. Provides coordination between users, Project Managers, Analysts, and management.

The Data Administrator maintains the Data Dictionary and establishes standards for its use.

The Data Administrator is responsible for the education of management, systems analysts, and users on data planning, data analysis, modeling, documentation, and logical design. Provides data modeling support to all project team system development efforts.

Provides logical database designs and performance specifications to Database Administration and verifies any required database design changes with project and user management.

Maintains the strategic plan.

Duties and Responsibilities

- Manages the development of standards, methods, and guidelines for data planning, analysis, data modeling, documentation, and logical database design.

- Manages the coordination between users, project management, analysts, and management.

- Manages the logical database designs and the use of logical design software.

- Manages the establishment of the Data Dictionary and develops standards for its use.

- Plans and manages the education of the staff on data planning, analysis, modeling, documentation and logical design.

- Manages the staff in providing data modeling support to all project team system development efforts.

- Provides logical database designs and performance specifications to database administration and verifies any required database design chagnes for the project and user management.

- Provides an awareness of contemporary methods of data modeling and evaluates their application in the current organizational setting.

- Manages the security and privacy of the data in all logical design.

- Manages the maintenance of the strategic plan.

- Provides the resolution of all data definition and usage issues.

Background Attributes

- A college degree of its equivalent in systems analysis, programming, or business administration.

- Five to seven years experience, preferably in data processing, including at least three years as a Senior Data Analyst.

- Excellent written and verbal communication skills and can express ideas concisely and clearly.

- Analytical ability - grasps concepts, quantifies and reassembles ideas, processes, tasks, etc., into improved systems.

- Knowledge and understanding of the DP standards regarding phased system development.

- Has the demonstrated skill to develop a DP strategic plan.

- Has the demonstrated skill to maintain a project control system.

- Knowledge and understanding of company personnel policies and practices.

- Knowledge and understanding of company business policies and procedures.

SENIOR DATA ANALYST

Immediate Supervisor

Data Administrator

Job Summary

Under the direction of the Data Administrator, the Senior Data Analyst investigates the stated problem and recommends solutions for review. (This is accomplished within the guidelines of Data Administration standards.) The Senior Data Analyst, working with project management and systems analysts, provides data modeling support in the analysis and design phases of systems development. The Senior Data Analyst will interface with users and Database Administration to provide problem resolution.

The Senior Data Analyst supervises the maintenance of the Data Dictionary and participates in the establishment of standards for its use.

The Senior Data Analyst participates in the development of training for management, analysts, and users on data planning, data analysis, modeling, documentation, and logical design.

The Senior Data Analyst participates in the maintenance of the strategic plan and in the communication of its status.

Duties and Responsibilities

- Prepares the analysis and implementation of logical database designs.
- Supervises the use of logical design software.
- Prepares requirements analysis on Data Dictionary support projects.
- Participates in the planning for and development of training for staff on data planning, data analysis, data modeling, data documentation, and logical design.
- Provides logical database designs and performance specifications to Database Administration and verifies any required database design changes for the project and user management.
- Provides data modeling support to project team systems development efforts.
- Analyzes and supervises the implementation of security and privacy of the data in all logical designs.

- Supervises the maintenance of all logical data models.

- Maintains the strategic plan.

- Provides direct interface between users, project management, and analysts.

- Participates in the development of standards, methods, and guidelines for data planning, data documentation, and logical database design.

- Develops project plans - identifies, estimates, prioritizes project tasks - for data analysis projects.

Background Requirements

- A college degree or the equivalent business experience in systems analysis, programming, or business administration.

- Three to six years of business experience and/or training.

- Demonstrated excellent written and verbal communication skills and can express ideas concisely and clearly.

- Analytical ability - grasps concepts, decomposes, quantifies and reassembles ideas, processes, tasks, etc., into improved systems.

- Knowledge and understanding of DP standards regarding phased systems development.

- Has the skill to maintain a DP strategic plan.

- Has the skill to maintain a project control system.

- Knowledge and understanding of company personnel policies and practices.

- Knowledge and understanding of company business policies and procedures.

DATA ANALYST

Immediate Supervisor

Data Administrator

Job Summary

Under the direction of the Data Administrator, the Data Analyst investigates a stated problem and prepares solutions for review. Data Analysts participate in the analysis and design phases of systems development through direct assignment to a project team. Data Analysts, in cooperation with Systems Analysts, may provide direct user interface, analysis, problem solving, and troubleshooting.

The Data Analyst develops user views and inputs them into database design tools, in order to develop logical databases. The Data Analyst communicates the databases to Database Administration and verifies any required database design changes to the project and user management.

The Data Analyst participates in the maintenance of the strategic plan.

Duties and Responsibilities

- Prepares the analysis and implementation of logical database designs and the use of logical design software.
- Analyzes and implements the Data Dictionary for developing applications.
- Participates in the training of the staff on data planning, data analysis, data modeling, data documentation, and logical design.
- Provides logical database designs and performance specifications to Data Administration and verifies any required database design changes for the project and user management.
- Provides data modeling support to all project team systems development efforts to which they are assigned.
- Analyzes and implements the security and privacy of the data in all logical designs.
- Develops plans of action for the tasks in his project, identifying priorities and estimating the completion dates.

Background Attributes

- A college degree of its equivalent business experience in systems analysis, programming, or business administration.

- One to three years of training and/or business experience.

- Excellent written and verbal communication skills and can express ideas concisely and clearly.

- Analytical ability - grasps concepts, quantifies and reassembles ideas, processes, tasks, etc., into improved systems.

- Has ability to maintain a DP strategic plan and to maintain a project control system.

- Knowledge and understanding of company personnel policies and practices, and business policies and procedures.

DATABASE ADMINISTRATOR

Immediate Supervisor

V.P. of Information Resource Management

Job Summary

The Database Administrator manages the staff assigned to do physical database design.

The Database Administrator is responsible for the education of management, analysts, data center operations, and users in physical design and performance tuning. Provides support to all project team system development efforts.

Provides physical database designs from the logical database designs developed by Data Administration.

Provides the database description and program specifications block control blocks to the project team for each database developed.

Duties and Responsibilities

- Manages the development of standards, methods, and guidelines for implementation of the database environment.
- Supervises the activities related to the design of the physical databases.
- Consults with Data Administration, users, project managers, analysts, and management on the applicability and use of the database environment.
- Assists in the performance, monitoring, and tuning of the database environment.
- Plans the education and training of the DBA staff.
- Supervises the staff in generating, maintaining, and controlling database description and program specification block control blocks for the project team.
- Arranges for the allocation of disk space required for each database.
- Supervises the implementation of security and privacy, strategies for data in the database environment.
- Participates in the evaluation and selection and support of appropriate software products.
- Develops plans of action for the tasks in their project,

identifying priorities, and estimating the completion dates.

Background Attributes

- A college degree or its business experience equivalency in systems analysis, programming, or business administration.
- Five to seven years experience in data processing, at least three years of which were as a Database Specialist.
- Excellent written and verbal communication skills and can express ideas concisely and clearly.
- Analytical ability - grasps concepts, quantifies and reassembles ideas, processes, tasks, etc., into improved systems.
- Understands company database software.
- Has the demonstrated skill to develop a DP strategic plan.
- Has the demonstrated skill to maintain a project control system.

SENIOR DATABASE SPECIALIST

Immediate Supervisor

Database Administrator

Job Summary

With minimum direction from the Database Administrator, the Senior Database Specialist investigates the stated problem and prepares solutions for review. The Senior Database Specialist accomplishes this within the guidelines of Database Administration standards. The Senior Database Specialist also participates in the development of Database Administration standards and guidelines.

The Senior Database Specialist is responsible for the development of physical database designs from the logical database design developed by Data Administration. The Senior Database Specialist coordinates the creation of database description and program specification block control blocks needed by the project team for each database that is developed and, in addition, will define the backup recovery approach for each application being developed. The Senior Database Specialist will consult with the project team in analyzing the impact an on-line system will have on database performance.

The Senior Database Specialist participates in the education of management, analysts, data center operations, and users in physical design and performance tuning.

Duties and Responsibilities

- Prepares the analysis and implementation of physical database designs and the use of physical design software with minimum direction from the database administrator.

- Is responsible for the performance, monitoring, and tuning of each database.

- Participates in the education of the staff on physical design and performance tuning.

- Coordinates the generation and maintenance of the database description and program specification block control blocks for the project team.

- Provides physical database designs and performance specifications to the project team.

- Assists in the analysis and implementation of security and privacy of the data in all physical designs.

- Arranges for the allocation of disk space required for each database.

- Develops plans of action for the tasks in his project, identifying priorities, and estimating the completion dates.

- Participates in the selection of access methods for each database.

- Participates in the evaluation, selection, and implementation planning of appropriate software products.

- Participates in the development of database administration standards, guidelines, and procedures.

- Participates in the analysis, design, and review of on-line systems, programs, and transactions.

Background Attributes

- Four to seven years of training and/or experience in Database Administration.

- Excellent written and verbal communication skills.

- Has the demonstrated skill to maintain a DP strategic plan.

- Has the demonstrate skill to maintain a project control system.

- Understands company database software.

- Has an in-depth knowledge of the systems development process.

- Experience in performing a staff or consulting role.

- Experience in on-line systems development.

DATABASE SPECIALIST

Immediate Supervisor

Database Administrator

Job Summary

Under the direction of the Database Administrator, the Database Specialist investigates the stated problem and prepares solutions for review.

The Database Specialist participates in the development of physical database designs from the logical database design developed by Data Administration. Under the direction of the Database Administrator, provides the database description and program specification block control blocks to the project team for each database that is developed.

The Database Specialist participates in the education of management, Analysts, data center operations, and users in physical design and performance tuning.

Duties and Responsibilities

- Prepares the analysis and implementation of physical database designs and the use of physical design software under the direction of the Database Administrator.

- Assists in the performance monitoring and tuning of each database.

- Participates in the education of the staff on physical design and performance tuning.

- Participates in the generation and maintenance of the database description and program specification block control blocks for the project team.

- Provides physical database designs and performance specifications to the project team.

- Assists in the analysis and implementation of security and privacy of the data in all physical designs.

- Arranges for allocation of disk space required for each database.

- Develops plans of action for the tasks in the project, identifying priorities and estimating the completion dates.

- Participates in the evaluation and selection of appropriate software products.

- Participates in the selection of access methods for each database.

Background Attributes

- One to three years training and/or experience.
- Average written and verbal communication skills.
- Has the skill to maintain a DP strategic plan as well as a project control system.
- Understands company database software.

This article presents some personal ideas on approaches for establishing and measuring the value and cost of information and how this analysis can be used as a management tool in Information Resources Management (IRM). It also addresses some information problems and discusses how they reduce information value and/or increase its cost. Most of the examples are drawn from the U.S. federal government, although the logic should apply to most commercial environments as well.

Information value and cost measures for use as management tools

by M. J. Chick

Both corporate and governmental organizations spend literally hundreds of billions of dollars a year for resources to process raw data into information. It is now generally recognized that information is not a free good. Additionally, the dollar cost impacts of managerial, operational, and administrative decisions and actions taken on the basis of processed information are probably even higher.

Poor management and operational decisionmaking about the data to be processed into information and the resources to be used for that processing can result in a number of negative consequences:

- Ineffective support of organizational missions, goals, and objectives
- Significant excess costs
- Significant worker and organizational productivity and efficiency losses
- Loss and/or non-attainment of information value

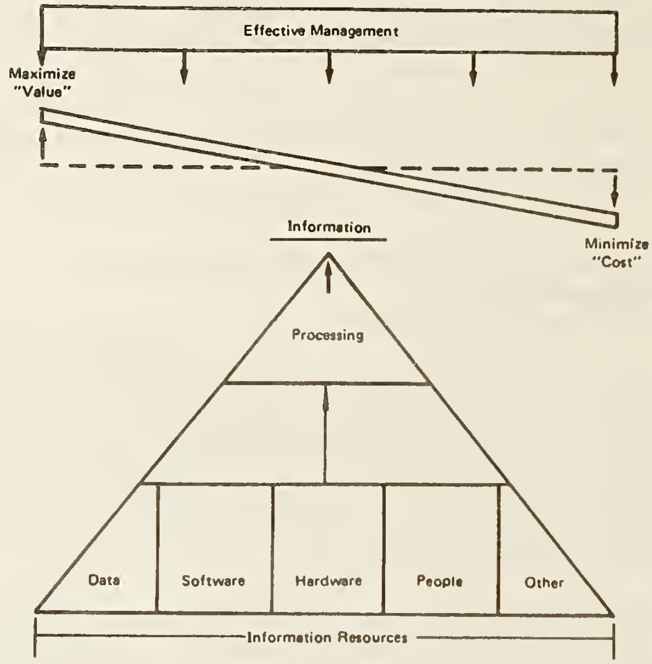
Recognition of the fact that information must be managed as a valuable and costly resource has been slow and piecemeal. The Paperwork Reduction Act of 1980 (Public Law 96-511) has formally set the framework for what we now call Information Resources Management (IRM) in the federal government.

Objectives of Information Resources Management (IRM)

Presently, IRM means different things to different people and organizations. No universally accepted definition supported by common terminology exists.[1] I define IRM as *an approach to applying appropriate and effective management philosophy, methodology, and techniques to decisions about data and information and other information resources (equipment, software, personnel, etc.).* The objectives are to assure that information produced from information resources has *maximum "value"* to the organization and at the same time, is produced at *mimimum "cost"* through effective management (Figure 1).

The author is an employee of the federal government, and the information presented in this article is in the public domain. The views expressed by the author are personal and not intended to reflect GAO policy.

Figure 1 Objectives of IRM



Definitions of value and cost

The terms "information value" and "information cost" are an extremely important part of IRM. Since these terms are often confused (and sometimes considered synonomous by information managers who do not have an accounting, management, or economics background), careful definition is essential. The following terms are defined with the help of *Webster's* and *Random House* dictionaries:

Costs

Cost represents an outlay, expenditure, or price paid to acquire, construct, or manufacture capital assets and commodities as well as other expenses incurred for operating a business, running an organization, and accomplishing institutional missions, goals, and objectives. Costs include expenditures for raw materials, direct labor, and other related expenses, as well as depreciation and amortization of capital assets.

Information costs

The costs incurred in acquiring, and/or producing information. This includes the cost of the resources used to produce information and other related expenses incurred in its production, storage, and dissemination. This production of information, from an accounting standpoint, is similar to the production (manufacture) of a commodity. Both involve converting something "raw" (unfinished) to a finished product, by applying resources such as direct labor (people), equipment, and overhead (see Figure 2).

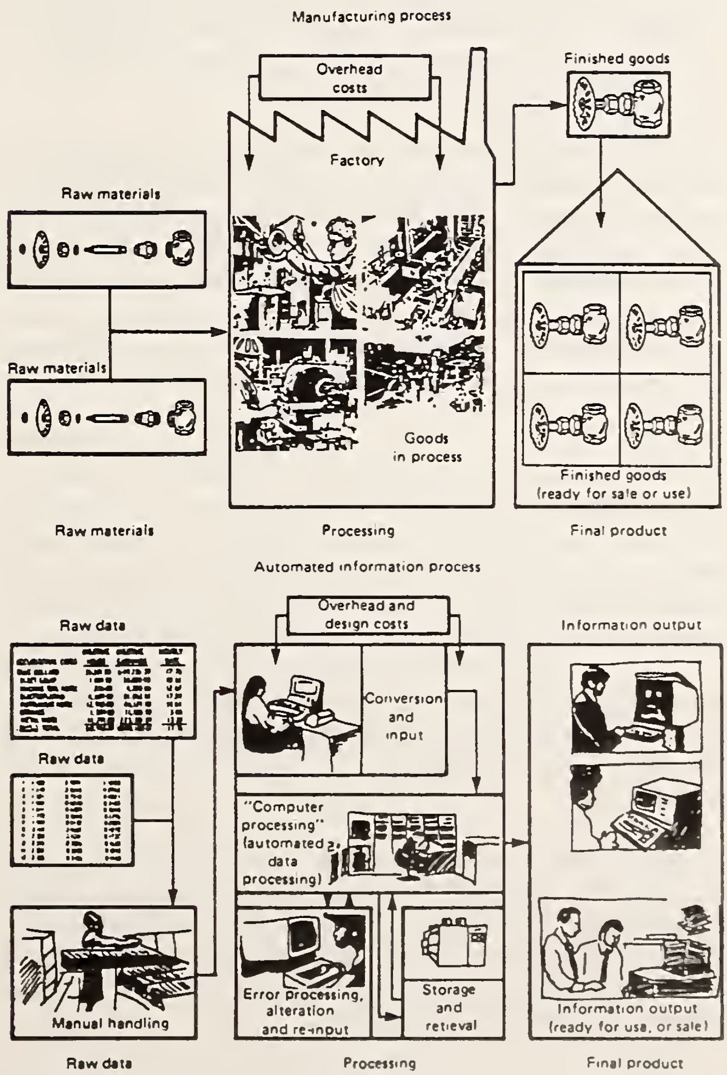
Value

Value represents monetary, attributed, intrinsic, and/or relative worth, merit, usefulness, importance, and/or utility of a good, service, product, principle, item, or entity. The value of something can be evidenced by a willingness or need to pay for, barter in exchange for, or otherwise need to use or have it available for use or for other purposes.

Information value

The value attributed to information produced or acquired by organizations, entities, and person.

Figure 2 Comparison of the production of a commodity and information



Underdeveloped arts

In studying and participating in the evolution of IRM in the U.S. federal government, it has become clear to me that the concepts of information value and information costs are still underdeveloped. Although research has been performed in both of these areas, few criteria involving the use of these concepts as management tools and considerations have been provided to those responsible for implementing IRM. This problem seems to be compounded by several factors including:

1. *Terminology problems*–As mentioned above, there is confusion between the terms *value* and *cost* as well as with other related but non-synonomous terms such as expense, asset, commodity, resource, and many others. Terminology problems in the information resources management arena seem to be magnified when the concepts of information as a commodity, resource, or asset are introduced.

2. *Limitations of traditional cost accounting systems*–Today's accepted accounting methodologies do not accumulate and present financial information about information *costs.* A possible exception are accounting systems designed for organizations that are in the business of producing information (often for sale).

3. *The often intangible nature of information*-This can cause a *cost* allocation problem. The concept of managing information as a commodity has many valid points. However, information has some unique characteristics, such as (a) electronic representation which can not be seen by the naked eye, (b) potential simultaneous uses of the same information by many, even while it still resides in storage (inventory), and (c) unknowns involved in the number of times information may be used and by how many users.

4. *A lack of consensus of the notion of information value*–Besides attaining consensus, there is an apparent need for more research in developing *approaches for assigning monetary measures to represent information value for use as a management tool.* Many people perceive information value from their own personal perspectives. Few, if any, have attempted to synthesize these perspectives for use as management tools in making various decisions about information.

Despite these problems, development of methodologies, approaches, and techniques for measuring both information value and information costs are needed for effective management and decisionmaking at all organizational levels.

Before I discuss some methods of measurement and their potential applications in management, I believe it is worthwhile to first review the concepts of information costs and values[2] as I am using them here. In the area of "costs," I suggest the "information executive" obtain the services of a knowledgeable accountant in order to integrate the two arts, i.e., information theory and accounting theory.

Information cost theory

In general, accounting principles and approaches applicable to commodity manufacturing, including asset capitalization, can be applied to establish a mechanism to measure the costs of information production. Such a mechanism should identify the total costs of resources applied to producing information in sufficient detail to disclose:

- The costs incurred at each information processing step (e.g., collection, input, processing, retrieval, etc.)

- The elements of costs incurred at those steps and in total (e.g., labor, hardware and software depreciation and amortization, supplies, training, travel, etc.).

Figure 3 shows one perspective of the "typical information cycle." It depicts the costs involved at each processing step, and also shows examples of some of them. The information produced from this cycle bears these costs in total. Accepted accounting rules of asset capitalization and expense should also be applied. In actual application, management may decide to make estimates of information costs for use as a management tool in lieu of establishing an accounting mechanism.

It is necessary to decide the level of detail required to determine actual information costs. Generally, the greater the level of detail and the closer to the actual transaction the costs are recorded, the more accurate the accounting system. However, the more detail (e.g., accounting for actual transactions) the more the effort will cost. The level of detail is a management decision that should be based on a cost-benefit analysis.

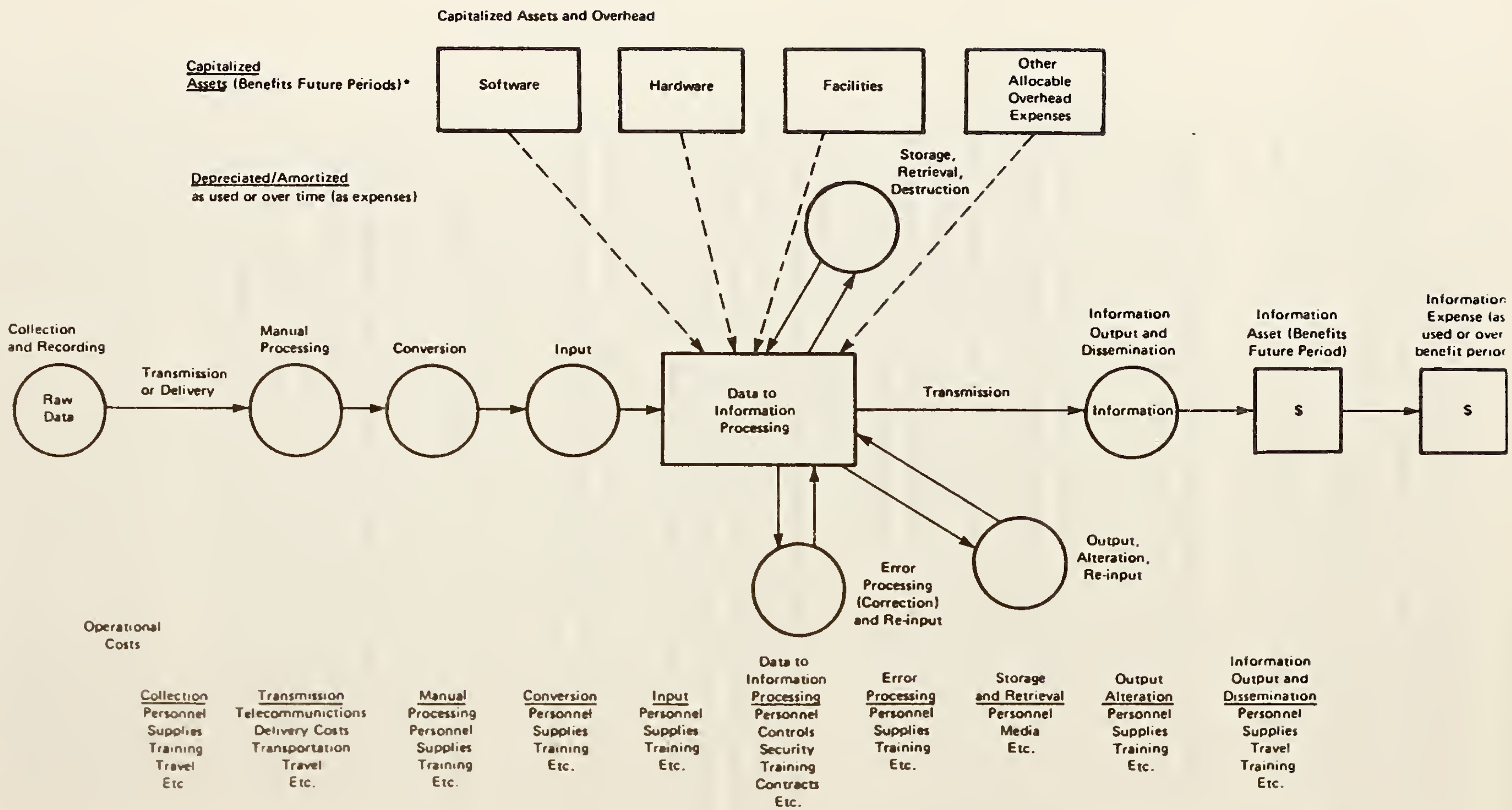
The information cost *estimating* approach has both advantages and disadvantages. On the positive side, it should be less costly to estimate costs than to account for them. On the other hand, extreme care is needed to assure that:

- All elements of significant cost are included in the estimates

- The accounting concepts of a going concern be observed. These include the necessity of reporting expenses (and revenues) in the period of benefit, and related fixed asset capitalization and depreciation theories be applied when preparing estimates, and

- The metholodology for cost estimating is fully disclosed.

Figure 3 Accumulation of costs incurred during a "typical information cycle"



Unnecessary and excess costs

Any discussion of management and costs should include a definition and brief discussion of unnecessary and excess costs. In the information business:

- *Unnecessary costs* can apply to costs that should not be incurred for information production. These costs of operation can be reduced with appropriate management attention and action because they are variable or semi-variable in nature. Fixed costs, most often attributed to depreciation of capitalized assets acquired or contructed for use over a relatively short period of time (e.g., 3-5 years), can also be reduced in the short-term and could also fit this category. These unnecessary costs could be eliminated at the end of the useful life of the capitalized asset, or sooner. The trend toward shorter useful lives of software and hardware, caused in part by rapid advances in information technology, may result in a more ''variable'' nature in some ''fixed cost'' categories because decisions to continue to incur them will have to be made more frequently.

- *Excess costs* apply to long-term fixed costs of operation (often called ''sunk costs''). They are allocable to the production of information that, in general, has little or no value. These excess costs, often referred to in terms of efficiency or productivity losses, cannot be immediately reduced, but could often be applied to more productive activities, including the production of more valuable information. They can usually be eliminated or reduced eventually.

- *Effectiveness* problems caused by poor-quality information, lack of needed information, and other reasons (discussed later). They involve such things as overpayments, failure to collect revenue, and poor decisionmaking in meeting organizational missions, goals, or objectives. These are another form of *unnecessary* or *excess costs*, depending on whether they can be eliminated or reduced in the near term. Examples of these types of undesirable situations are presented in Figure 7.

Information value theory

I have found the National Science Foundation and others to be extremely valuable sources of thinking on the value of information (as well as cost). There are many theories about what information value means. There is a lack of consensus on the notion of information value.

It is difficult, or in some cases impossible, to assign dollar values to many of the ''indicators of information value'' contained in the literature. However, they deserve some coverage in order to (1) set the tone for my discussion of value measures for management purposes, (2) provide the information executive with an appreciation of the valid thoughts on this subject, and (3) illustrate the varying philosophical perspectives that exist in this very ''soft'' area. Figure 4 lists some basic indicators of information value contained in the literature and my comments on the apparent feasibility of assigning a dollar measure to the value indicator for IRM purposes.

Figure 4 Some possible indicators of information value

Indicator	Comments
1. Positive impact on income factors (return on investments, revenues, and/or net profit) resulting from information	Feasible
2. Willingness to pay (or exchange something else of value) for information	Feasible
3. Motivation for information production and use (added value)	Sometimes feasible
4. Reduction in costs resulting from information use (does not include information production costs)	Feasible
5. Productivity and efficiency improvements from information use	Sometimes feasible
6. Impact of information withdrawal or problem (negative value)	Sometimes feasible
7. Use of information	Difficult
8. Extensive citation (or use) of information	Difficult
9. Usefulness and impact of information use as related to well defined organizational goals (effectiveness)	Sometimes feasible
10. Multiple and different uses of the same information	Difficult
11. Continued expenditures (costs) for producing information over a period of time	Difficult
12. User perceptions of value of information produced	Difficult

Value in a commercial environment

One of the information value indicators shown in Figure 4, "positive impact on revenues, net profits and/or return on investment (income factors)," is appropriate for use by most organizations operating in a commercial environment. Almost all are in business for producing revenues. Most want to make a profit (revenues exceeding expenses). In such an environment, it is often possible to assess and estimate the impact of having (or not having) certain information by relating it to the revenue and profit-producing capabilities (income factors) of the organization.[3]

For example, an automobile manufacturer could affect its revenue/profit-producing capability by effectively producing and using information about:

- Customer habits (marketing information)–the preferences on the size of automobile being purchased would be an example of this type of information

- The impact of price changes on sales (marketing and economic information)

- Production technology (manufacturing, technological, and economic information)–the debate about robotics versus direct labor production modes would apply here

- Quality characteristics of component parts and their manufacturers (engineering and acquisition information)

- Unsold automobiles (inventory and marketing information)

It is clear that the appropriate generation and use of specific classes of information can make a contribution to producing revenues, maximizing net profits, and containing operating costs. Therefore, it has a measurable value. Justification for any new information requirement of most commercial organizations should include an estimate of the impact such information will have on the income factors mentioned–a determination of information value.

There are situations where the main product generated for sale is information itself. Examples are newspapers, books and magazines, mailing lists, etc. An appropriate measure of value would, of course, be the willingness of the consumer to pay (or exchange something else of value) for the product. Decisions to produce or eliminate information should be based, in part, on the effect those decisions have on the consumer's willingness to pay for the product (information value). It is obvious that a decision made by a technical publication to eliminate state-of-the-art material and to retain just the advertisements would severely reduce its value (measured by willingness to pay), perhaps even eliminate it.

Problem of using income factors in the federal environment

Simply state, the federal government has only a few organizations which have a primary mission of collecting revenue (some then selling that information). Almost all federal activity is expenditure-oriented. However, it would be appropriate to apply the notion of measuring the impact of information on income factors to determine its value to the Internal Revenue Service (IRS) and the Bureau of Customs, both of the Treasury Department. The Government Printing Office, National Library of Medicine, the National Technical Information Service, and the U.S. Geological Survey, among others, do sell information. My previous comment regarding "willingness to pay" as a measure of information value applies, at least in part, to their activities.

However, an examination of the federal budget would show that most major departments and agencies are service and expenditure-oriented and collect little or nothing in the way of significant revenues. For instance:

- The Department of Health and Human Services provides assistance to the needy and elderly

- The Department of Defense provides the means to protect this country's security

- The Department of Agriculture works to improve farm income, maintain our production capability, and ensure food quality

- The Department of Labor assists Americans who want to work and is concerned with working environment, discrimination, unemployment, and the like

- The Environmental Protection Agency works to control and eliminate pollution to our air, water, etc.

The fact is that, in most cases, measuring information value based on contribution to income and willingness to pay is so difficult in government that it can not be related to revenue-producing missions, goals, and objectives.

Other approaches to measuring information value

As Figure 4 shows, many other indicators of information value exist besides the impact on income factors and willingness to pay indicators. In situations where income factors are not a consideration, such as in a major portion of federal activity, other approaches could be used. Care must be taken to attach the appropriate perspective to these measurements. They should be used to demonstrate continuing or new information needs or correction of information problems, but they are *not necessarily* comparable to the cost of producing the information *(information cost).* A brief discussion of them follows, while a more detailed presentation of one indicator is discussed.

Motivation for information production and use

Sometimes, the motivating factors behind creating new or changing existing information needs involves improved organizational performance in meeting its missions, goals, and objectives. In an income-producing environment, such motivating factors can involve items previously discussed. However, even in an expenditure-oriented environment, new information needs and uses can result in measurable reductions in expenditures.

In December 1982, representatives of the General Accounting Office (GAO) testified at hearings held by a Senate Governmental Affairs Subcommittee[4] on the agency's view of computer matching to detect error, waste, and fraud in government benefit programs. Computer matching is really jargon for the comparison (processing) of existing data contained in separate files to create new information. This new information, used properly and legally, can have great and measurable value, since it can disclose potential payments that exceed the appropriate amount or should not have been made at all.

In the Social Security Administration (SSA) alone, overpayments identified by routine computer matching currently exceeds $100 million a year in only one of its many benefit programs. Overpayments in needs-based benefit programs[5] are probably in the billions of dollars a year. Creating new information to identify and reduce them certainly has a measurable value.

Information that improves user productivity and/or reduces user cost

In defining and distinguishing between value and cost, I stated that information *cost* relates to the costs incurred in acquiring and/or producing information. Once produced, however, anything that can be done to make information more easily and appropriately usable, requiring less time of the user or manager, increases information *value*.

For instance, a manager has to determine expenditure trends. Information can be presented in many ways and trends can be determined each way. However, as Figure 5 demonstrates, the way information is presented can add to or detract from user productivity. Each exhibit shows the same information. Which one decreases the analysis and is easiest to use for determining trends?

Exhibit I really just presents "data." In order to obtain "information," the user must make additional computations manually. This creates the need for additional user resources in order to attain the needed information. Exhibit II does present "information" (already computed). However, this presentation requires additional manual analysis in order to determine the required trend information, since it merely lists figures by month. It is not difficult to see from Exhibit III that the highest expenditures are in the first half of the calendar year. This information obviously has the most value for trend analysis, since it increases user productivity and reduces the cost of use. Measuring information value in this context would involve the use of productivity and efficiency measures as well as an analysis of the cost of information use.

Figure 5 Information presentation for increased value

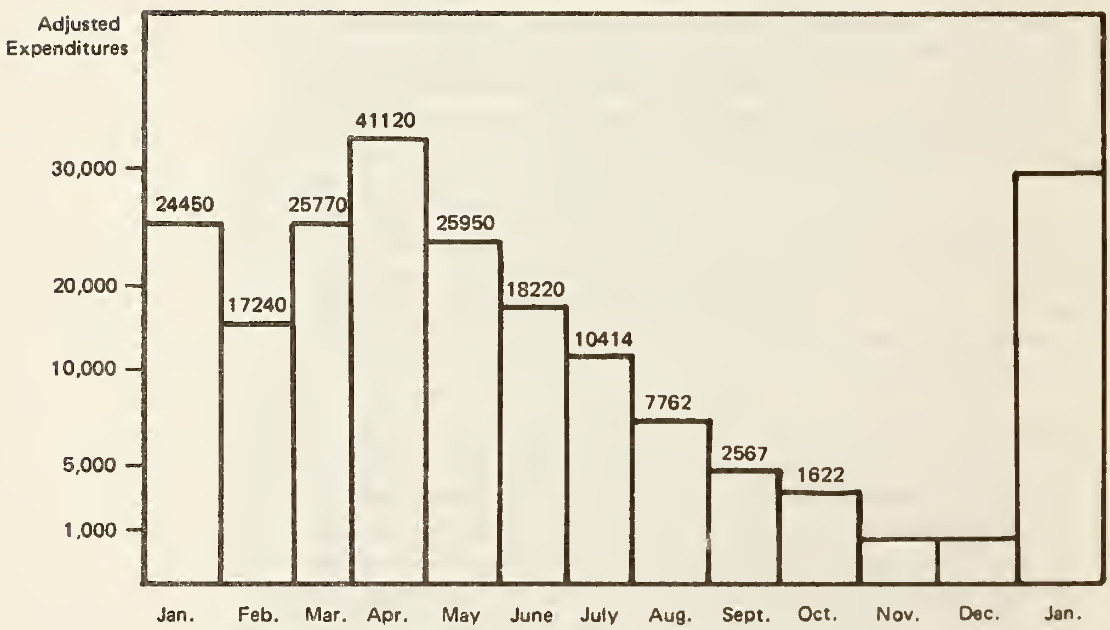
EXHIBIT I

Month	Data	Information
January	8212 ÷ 20 × 14 =	?
February	4331 ÷ 9 × 5 =	?
March	3113 ÷ 11 × 2 =	?
April	4224 ÷ 14 × 4 =	
May	6210 ÷ 12 × 9 =	
June	7212 ÷ 11 × 11 =	
July	6967 ÷ 10 × 8 =	*Compute*
August	8211 ÷ 15 × 7 =	*It Yourself*
September	7991 ÷ 2 × 8 =	
October	7413 ÷ 17 × 11 =	

EXHIBIT II

Month	Adjusted Expenditur
January	$24,450
February	$17,240
March	$25,770
April	$41,120
May	$25,950
June	$18,220
July	$10,414
August	$ 7,762
September	$ 2,567
October	$ 1,622

EXHIBIT III



Impact of information withdrawal

Potentially useful measures of information value involve identifying and measuring the effects of (1) withdrawing information availability from an organization, or not having or using information which is needed to effectively meet an organization's missions, goals and objectives, and (2) information problems that reduce or eliminate information value, increase costs, and need to be corrected.

The concept of negative value involves measuring information value by its converse; that is, by asking the question "what is the dollar impact which may occur because of the failure to attain information value (either by not having it, or by problem conditions which inherently reduce or eliminate its value)''?

Impact of information withdrawal or non-availability

Imagine a profit-oriented, retail outlet (like Sears, for instance) or a military inventory supply manager (such as the Navy Aviation Supply Office), both responsible for positioning the right amount of inventory (in total and at each location) to meet the demands of their customers.[6] Basic information needed by both organizations include (1) the quantity of inventory stored at each location and in total, and (2) the requirements or demands of customers for meeting future needs for each commodity. Figure 6 depicts this basic and oversimplified equation for a given commodity. It is obvious that, if an inventory manager did not have ''requirements'' information, any decision to purchase more stock, or dispose of existing quantities, would be merely a guess and most often wrong.

In this scenario, the costs associated with purchasing unneeded inventory, or disposing of needed inventory, are both measurable in dollars. Dollar measurements for negative value are to be used as an indicator of the value of good ''requirements'' information, and could include:

- The unnecessary cost of purchasing inventory that is not needed
- The excess cost of storing unneeded quantities of inventory
- The unnecessary cost of disposing of needed inventory that must be repurchased
- The impact of not having the inventory at the right location at the right time

A similar analysis of information value can be made in the scenario of having good but modularized ''requirements'' and ''inventory'' information for each location, but no information produced or made available in the aggregate.

Figure 6 Hypothetical inventory supply information (one commodity)

	LOCATION			TOTAL
	A	B	C	
ANTICIPATED FUTURE REQUIREMENTS IN UNITS (DEMAND)	2,000	2,500	1,000	5,500
INVENTORY AVAILABLE (IN UNITS)	100	17,000	1,000	18,100
NET REQUIREMENTS	1,900	(14,500)	0	(12,600)

Impact of problems that reduce information value

Taking the concept of negative value measurements one step further, it would be useful to measure the loss of information value (and also increases in cost) when information problems are detected. In the information arena, there can be problems with any or all of the major resources (i.e., data and information, software, hardware, people, etc.). Some problems may impact the value of information; others impact just information costs; still others may affect both.

Figure 7 depicts several major types of information problems. It shows categories of potential problems, a definition of each, commentary, and brief examples of each. Some of the examples are taken from an older GAO report on automated decisionmaking problems.[7] The report on automated decisionmaking discussed detected problems in information and software which, at one agency, resulted in lost information value and excess and unnecessary costs in the tens of millions of dollars. These problems were allowed to go uncorrected for a minimum of five years, and maybe longer.

Automated decisionmaking is defined as computer applications that initiate action without manual review and evaluation (through output) on the basis of programmable decisionmaking criteria. These are established by management and incorporated in computer instructions (see Figure 8). I elected to use automated decisionmaking here because:

- Unreviewed actions initiated by computer are significant. For instance, in 1976, GAO estimated such unreviewed actions cost over $40 billion a year; in 1981, they reported to be a minimum of $126 billion

- The effect of information problems are accentuated in such applications due to the absence of manual review

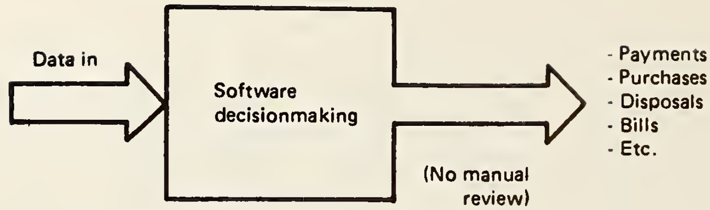
However, research of more recent GAO reports show repeated examples of these and other information problems, emphasizing the need for new and aggressive approaches to IRM.

When problems such as those depicted in Figure 7 are detected, measurements like value (and cost) should be used to assure correction of the most significant problems detected. Such measurements would provide a basis for establishing priorities and the assignment of personnel and other resources to problem solution in order to assure that the most significant problems get appropriate and timely attention.

Figure 7 Impacts of major categories of information problems

Major category of problems	Definition	Information value	Information costs	Other impacts	Examples
Data quality problems	Incomplete, inaccurate, obsolete, untimely or inconsistent data and/or information	Reduced value	Excess costs	Mission, safety, etc.	Quality problems caused millions in overpayments at one agency and unauthorized issue of radioactive material by another (FGMSD 76-5, April 26, 1976). Costs of processing poor quality data are excess.
Software problems	Application software containing incorrect, obsolete, or otherwise inappropriate processing criteria, and/or inadequate technical controls over data in software	Reduced or eliminated value	Excess costs	Mission, safety, etc.	Obsolete software criteria caused millions of dollars in unnecessary transportation of military equipment. Overpayments were made to veterans because of inadequate software controls. (FGMSD 76-5, April 26, 1976)
Unnecessary data processing	The processing and output of data and information that is not needed, and unnecessary data processing steps	No value exists in the former; unnecessary steps may or may not reduce value	Excess and unnecessary costs	None (unless unneccessary steps delay output)	An agency corrected individual inputs several times rather than correcting all data element errors simultaneously. (GGD 83-8, October 14, 1982)
Data and information use difficulties	Information and data access problems and poorly designed or non-standard output	Eliminated or reduced value	Excess and unnecessary costs	Mission, safety, etc.	An agency experienced a success rate of only 8 percent in attempts to access the system and data. (CED 81-116, June 19, 1981)
Data storage problems	Unnecessary storage of information beyond its useful life or premature data destruction	Eliminated or no value	Excess costs of storage beyond useful life	Mission, safety, etc., when data destroyed prematurely	Data and information was decades beyond its useful life at an excess cost of millions. (ID 81-39, August 6, 1981)
Duplications in data processing	The collection, processing, output, and/or dissemination of data and information that duplicates data in the same or other systems, or systems that duplicate the support of a program or function	Duplicate information may have no additional value	Excess and unnecessary costs	None (unless duplication slows use)	An agency developed at least six separate systems to support functions already being supported by another system. (CED 82-28, June 9, 1981)
Non-use of needed information	Failure to use available and needed information or failure to acquire or produce it	No value achieved	Excess and unnecessary costs of producing information not used	Mission, safety, etc.	An agency failed to use information acquired about vendors supplying poor quality equipment. Tens of millions were unnecessarily spent in repairing defective equipment so it could be used. (PLRD 82-115, September 2, 1982)

Figure 8 Automated decisionmaking



Effective management tools[8]

More often than not, management theory is broken down into discrete functional activities. Examples of such categories include planning, organizing, budgeting, directing, staffing, and controlling. IRM can be viewed as applying effective and integrated management concepts and tools to the resources used to produce information (as well as to the information itself). A GAO team responsible for developing criteria for performing information resources management studies developed a matrix to depict this (''adapted'' version is shown in Figure 9).

Presenting concepts, methodologies, approaches, and ideas about measuring information costs and values to be used as tools in managing information and information resources would not only help to better define the application of IRM but also stimulate further thinking in this area. The tools could be very valuable in performing these management activities, as well as in reaching specific decisions about information needs and uses, timing and quality considerations, technology and obsolescence determinations, and much more.

Representing information value

Assigning absolute, consistent, and uniform dollars to represent the value of information being produced and (hopefully) used by an organization will be very difficult because of some of the problems mentioned previously. However, some determination of information value in dollar terms is needed for management purposes, for such things as:

- Periodically confirming the continued need for information currently being produced
- Establishing priorities and allocating resources for providing new information
- Establishing a basis for taking management actions to assure that perceived information value is being attained
- Identifying problems that result in information value losses or reductions (in addition to excess or unnecessary costs and poor effectiveness)

Figure 9 IRM matrix (adapted)

MANAGEMENT ELEMENTS

	Plan	Organize	Direct	Budget	Control	Evaluation
Information Resources						
Data and Information						
Software and Procedures						
Hardware and Operations						
People						
Other–Media, Forms, R&D, etc.						

- Establishing priorities, allocating resources, and establishing targets for correcting information problems that reduce or eliminate information value

- Providing a basis for applying sound management principles, as part of IRM, to information (planning, directing, controlling, etc.)

- Establishing a basis for protecting the information being produced (involving effective internal controls and security measures)

Further, information value (and costs) measured in dollars would be very useful in applying traditional functions of management to the management of information resources and to information. These functions include planning, organizing, directing, controlling, and evaluating.

Information research should include value and cost

Section 3504 (b), (6) of the Paperwork Reduction Act (PL 96-511) requires "planning for, and conduct of research with respect to Federal collection, processing, storage, transmission and use of information." The keys to effective implementation of the Act's major objectives of improving the management of information and information resources include (1) defining the elements of management and resources to be managed, and (2) making maximum use of management tools and techniques for implementing effective and integrated management approaches.

The National Science Foundation and others fund or perform research in information areas. From a management perspective, one of the areas in which such research should be focused is the application of information value and cost approaches to IRM. Improved information management probably could save billions of dollars, and make the government much more productive and effective. Breakthroughs made in this area could also be applied in corporate environments.

A final word

Although dollar measurements of information value are desirable and useful, there will be some times when such measurements will not be possible. Value indicators such as the extent of information use, different uses made, user perspectives, and mandated and legislated information will not be measurable. (Exactly how does one place a dollar value on information that will be helpful in preventing full-scale nuclear war?) The major goal of management is to make sure that information being produced has value, and where feasible and appropriate, the value be measured to demonstrate its significance.

Acknowledgments

The author wishes to thank individual members of GAO, the National Science Foundation, the Congressional Research Service, the Association of Federal Computer Users, the Association for Federal Information Resources Management, and specific management consultants for their encouragement and support in this evolving, but elusive environment (IRM).

Cited references and notes

1. Some U.S. federal government organizations are currently attempting to arrive at a definition of IRM. They include, among others, the Congress, certain legislative agencies, the Office of Management and Budget, and the Association for Federal Information Resources Management.

2. I am drawing a clear distinction between cost and value and hope to maintain it. So, for example, reducing the *cost* of information production (a worthy objective) does not necessarily increase the *value* of the information produced (an equally worthy objective).

3. Even most ''non-profit'' organizations in the private sector collect revenues and incur costs. Comments on information value measures based on income factors would therefore also apply to them.

4. Subcommittee on Oversight of Government Management.

5. E.g., Aid to Dependent Children, Medicaid, food stamps, Supplement Security Income and Section 8 Housing.

6. Sears' customers are, of course, the individual consumers. The customers of ASO are Navy facilities which overhaul and repair Navy aircraft and major assemblies (such as aircraft engines).

7. "Improvements Needed in Managing Automated Decisionmaking by Computers Throughout the Federal Government" FGMSD 76-5 (April 23, 1976).

8. There are several categories of management tools. They include (1) quantitative measures (e.g., rates of error), (2) time-related measures (e.g., response time), (3) quantitative measures (e.g., numbers of transactions processed per labor hour), and (4) financial measures. The concept of information value and costs here are financial measures.

General references

1. L. G. Becker, "Information Resources Management (IRM): An Approach to Implementation of the Paperwork Reduction Act (PL 96-511)," *Issue Brief IB81166*, Congressional Research Service, The Library of Congress (1983).

2. S. H. Black and D. A. Marchand, "Assessing the Value of Information in Organizations: A Challenge for the 1980's," *The Information Society Journal*, 1, No. 3 (1982).

3. Y. Braunstein, "Information as a Commodity: Public Policy Issues and Recent Research," *Information Services: Economics Management and Technology*, Westview Press (1981).

4. D. Carroll, "Information Resource Management — A Costing Perspective," *IRM Magazine* (1983).

5. M. J. Chick, "The Cost of Information," *The GAO Review* (Summer 1973); "Audit and Management Responsibilities for the Paperwork Reduction Act of 1980," presentation to the National Institute for Management Research (1981); and "Accounting for Information Costs Essential to Effective Management," *IRM Magazine* (December 1982).

6. J. Griffiths, "The Value of Information and Related Systems, Products, and Services," *Annual Review of Information Science and Technology (ASIS)* (1982).

7. F. W. Horton Jr., *The Information Management Workbook*, Information Management Press (1981).

8. King Research, Inc., "Value of the Energy Data Base," Department of Energy DOE/OR/11232-1 (1982).

9. F. Machlup, "Users, Value and Benefits of Knowledge," *Knowledge: Creation, Diffusion, Utilization,* Sage Publications, Inc. (1979).

10. National Science Foundation, "Summary of Awards FY 1981 and FY 1982," Division of Information Science and Technology, **NSF-83-10.**

11. D. S. Price "The Cost of Information: A Prerequisite for Other Analysis" (1973).

12. Cited reports of the U.S. General Accounting Office.

13. Statement (GAO) ***, Accounting and Financial Management Division before the Subcommittee of Government Management, Senate Committee on Governmental Affairs on Computer Matching to Detect Error, Waste, and Fraud in Govenment Programs (December 16, 1982).

14. W. R. Synnot and W. H. Gruber, *Information Resource Management,* John Wiley & Sons, New York (1981).

15. S. Taylor, "Government By Computer," *Dollars and Sense* (March 1981).

16. D. R. Vincent, "Information Economics," *Information Executive,* 1 (1983).

The author is with the U.S. General Accounting Office, 441 G St. N. W., Washington D.C. 20548.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** *(See instructions)*	1. PUBLICATION OR REPORT NO. NBSIR-86/3324	2. Performing Organ. Report No.	3. Publication Date FEBRUARY 1986

4. TITLE AND SUBTITLE

Data Administration Workshop Proceedings

5. AUTHOR(S)

Frankie E. Spielman

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

This Special Publication constitutes the proceedings of a two-day workshop on Data Administration, held at the National Bureau of Standards, Gaithersburg, Maryland, on March 27-28, 1985. The workshop was sponsored by the National Bureau of Standards under the auspices of the Federal Data Management Users' Group (FEDMUG).

The purpose of the workshop was to provide a forum for Federal, State, and local government Program Managers, Information Resource Managers, Data Processing Managers, and Data Administrators to hear nationally prominent speakers and to discuss and share data administration ideas and experiences.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

administration; concepts; conventions; costs; data; issues; planning; tools; standards

13. AVAILABILITY

[X] Unlimited
[ ] For Official Distribution. Do Not Release to NTIS
[ ] Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
[X] Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

240

15. Price

$22.95