1983

Information System Integrati on: A Metadata Management Approach

Larry Kerschberg
University of South Carolina

Donald Marchand
University of South Carolina

Arun Sen
University of South Carolina

Information System Integration: A Metadata Management Approach

Larry Kerschberg, Donald Marchand, and Arun Sen
Department of Management Science
and
Institute of Information Management, Technology, and Policy
University of South Carolina

ABSTRACT

This paper deals with the integration and control of organizational information systems. The framework for this integration and control is that of Information Resource Management (IRM) - the management not only of information but also of the processes that specify, generate, distribute, and consume information.

Within the context of IRM, the concept of metadata -- data about data -- is introduced. A tool to manage metadata, the Information Resource Dictionary System, is defined and a data model and data architecture are presented for the system. Also, support tools to aid in information integration are discussed. Lastly, the Enterprise Administration function is proposed to ensure not only the proper use of the Information Resource Dictionary System but also the implementation of corporate information system integration and control policies.

INTRODUCTION

The evolution of corporate information systems is being influenced by several potentially divisive trends that are application specific. These trends are reshaping the nature and scope of corporate information systems. In particular, we are referring to Electronic Data Processing Systems (EDP), Management Information Systems (MIS), Decision Support Systems (DSS), Office Information Systems (OIS), and Personal Information Systems (PIS).

EDP Systems and Management Information Systems address the record-oriented, transaction-processing, view of the world (Thierauf and Reynolds, 1982). A typical corporation may have several large-scale applications, each having access to a large, shared database that is managed by a Database Management System (DBMS). In addition, it might have several heterogeneous applications developed using distinct DBMS on distinct hardware. Decision Support Systems aid strategic planners in solving unstructured problems (Sprague, 1980).

Another important trend is that of Office Information Systems, also called Office Automation Systems. These systems are designed to improve office productivity by providing a set of tools to support office activities. These tools include word processing for document preparation, electronic messaging for mail and document distribution, electronic "spreadsheets" for decision support, and tools for activity management such as calendars, PERT, and reminder files. Typically, these OIS's are stand-alone systems, or they may be loosely-coupled to the corporate MIS. However, because OIS is so new, its true relationship to MIS is not clear.

An even newer phenomenon within the corporation is the Personal Information System (PIS). It is not uncommon to have executives bring personal computers to the workplace. These computers also have word processors, personal filing systems, and the capability of interacting with corporate databases and those of the "information utilites." The relationship of the PIS to the other information systems is unclear, and effective management policy guildelines will be needed to determine the evolution and interrelationships of organizational information systems.

The goal of this paper is to study the interrelationships among EDP, MIS, OIS and PIS in a modern corporation. Of particular importance is the sharing of information. Information sharing is important for several reasons: 1) to plan for future information systems, 2) to support information systems communication, 3) to aid organizational restructuring due to mergers, expansions, etc., and 4) to institute effective control over information resources.

Our approach to information sharing is by means of information integration, that is, the management of knowledge (metadata) of the information resources available to the organization. This paper discusses the concept of metadata management within the framework of information resource management. A tool to support metadata management, and the management structure required to implement the concept are presented.

In the remainder of this paper, the second section provides an historical perspective on the evolution of information systems in organizations. The third section discusses information resource management as a framework for information integration. The last section presents our conclusions and suggests areas of future research.

HISTORICAL PERSPECTIVE

In order to understand the evolution of information systems, it is important to review the major concepts, tools, and management policies that have influenced the development of modern information systems. In fact, we will use the (concept, tool, policy) paradigm to discuss major advances.

The (concept, tool, policy) triple is used to stress that an organization perceives the need for a tool to support some organizational concept. Once the tool is developed it must be managed so as to be used effectively to meet organizational goals. Oftentimes, the (concept, tool, policy) triple does not develop simultaneously. Frequently it is led by a new technology in search of meaningful applications. Only later is the appropriate management policy instituted.

Stage 1 - Physical Control of Information

The first stage in the development of the information management function, occurred from 1900-1960 and concerned the physical control of information. Standard operating procedures were instituted at the supervisory and middle levels of management to manage: 1) records and reports, 2) mail and correspondence, 3) vital records, and 4) the corporate library.

The triple which denotes this period is (physical control, manual tools, supervisory level management).

Stage 2 - Management of Automated Technology

The advent of computers and their application in organizations led to a separate development of the data processing, telecommunications, and office administration functions. Each function had its own management structure.

Initially, data processing applications focused on automating routine, repetitive applications. These systems are referred to as EDP systems and Management Information Systems (MIS), and are designed to provide managerial support at the operational and tactical levels of management. Henceforth, we will not distinguish between EDP and MIS, and will simply refer to both as MIS. The triple to denote these concepts is (data files, file systems, systems manager).

As more and more applications were developed, each supported by its own set of files, it became apparent that much of the data in these files was redundant, and the need arose for data sharing. The notion of data sharing allows several applications to access and manipulate a nonredundant pool of shared data - the database. The tool to support this concept is the Database Management System (DBMS) with its facilities for data definition and manipulation, its view (subschema) mechanism to tailor a data representation to an application's needs, and a concurrency control mechanism to allow concurrent access to data while maintaining data integrity. The management function is called Database Administration (DBA) (Gillinson, 1982). Our triple to denote this concept is (database, DBMS, DBA).

As organizations increasingly view data as a resource to be managed, the database administration function will evolve to that of data administration. Because this area is so new, the tools to support this concept are not yet available. Data dictionary systems (DDS) - originally a subsystem of the DBMS - should play a major role in the development of tools for the triple (data resource, DDS, Data Administration). Data dictionary systems will be discussed in later sections.

Another important concept arising from MIS research, and evolving in parallel to data management, is the Decision Support System (DSS) (Sprague, 1980). Such systems are designed to aid decision makers in unstructured problem solving. The tools employed by DSS include mathematical programming packages, statistical packages, financial analysis programs, interactive query facilities, and graphics packages. Some of these tools require a high degree of training in their application and effective use. In response to the need for trained specialists, management has instituted the information center - a group of experts, trained in the use of specialized tools, who help users to obtain and manipulate information for decision-support. Thus, we have the triple (decision-support, DSS, Information Center) to represent the influence of Decision Support Systems.

In the area of Office Information Systems (OIS), the key concept is the integration of diverse media such as text, voice, image, and message to support office activities and to automate the flow of information throughout the organization. Research and development in Office Information Systems (Morgan, 1980) has been directed toward 1) new hardware architectures to support media integration 2) integrated software to support word processing, electronic messaging, computer conferencing, and personal computing, and 3) office models (Ellis and Nutt, 1980) that support office concepts such as forms, form flows, office procedures, etc. The triple we use to denote office automation is (media integration, OIS, Office Administration).

Lastly, we have the growing influence of personal workstations and personal computers (PC) which office workers can use to tap into the corporate information system. The PC user may obtain data from the corporate database, and merge that data with data obtained from external information utilities such as Dun and Bradstreet, Standard and Poor, Dow/Jones, CompuServe, and the Source, to create a personal information system. Although the PC's have various tools such as word processors, electronic spreadsheets, database or personal file systems, and graphics, these act as "stand-alone" tools which should be integrated.

Another phenomenon involving personal computers is the availability of hardware boards which plug into the PC to make it look like another computer and operating system for which a particular software package has been written. Thus, we see the emergence of software driving the purchase of hardware. Further, the cost of hardware is becoming so inexpensive that one can posit the universal personal computer. However, the drawback is the multitude of operating systems, each having distinct file formats and organizations, and thereby reinforcing the incompatibility of software packages.

In dealing with the personal computer phenomenon, organizations have set up consultants to aid users in selecting PC's and User Groups to share information, programs, and to develop future plans. Our triple to characterize the personal computer in organizations is (personal information support, PIS, User Groups).

Clearly, the relationship between the PIS and the organization is not well-defined. The organizational information system could act as an information utility for the PIS, or it might have an expanded role as software supplier and software integrator. These roles will be discussed shortly.

In this section we have focused on three major trends affecting the corporate information system: management information and decision support systems, office information systems, and personal information systems. We now turn to our proposed framework to integrate these various information systems.

INFORMATION RESOURCE MANAGEMENT

In order to properly address the relationships among the various organizational information systems, one needs a framework in which to study the issues. We propose that the proper framework is that of Information Resource Management (IRM).

The Database Directions III Workshop (Goldfine, 1982), in its report "Information Resource Management - Strategies and Tools," defines IRM as:

> "Information Resources Management (IRM) is whatever policy, action, or procedure concerning information (both automated or non-automated) which management establishes to serve the current and future needs of the enterprise. Such policies, etc., would include considerations of availability, timeliness, accuracy, integrity, privacy, security, auditability, ownership, use, and cost effectiveness."

Within the framework of IRM one is interested not only in the information resource but also in the processes which define, create, manipulate, distribute, consume, archive, and eventually destroy information - the information life cycle. Thus, IRM focuses on all the processes which govern the organizational information resource. These processes may be quite diverse and, among others include:

1) Database Design (Lum, et.al., 1979),
2) Application Program Development (Mylopoulos, Bernstein, and Wong, 1978),
3) Software Configuration Management (Sibley, Scallan, and Clemons, 1981),
4) Information System Documentation (Teichroew and Hershey, 1977),
5) Information Systems Planning (Sakamoto and Ball, 1982),
6) Heterogeneous Database Communication (Ozkarahan and Kerschberg, 1982),
7) Office System Specifications (Shu, et.al., 1982), and
8) Information Requirements Specification (Davis, 1982; Sen and De, 1981).

In order to manage IRM processes effectively, we must understand the underlying events, activities, actions, and data requirements of these processes.

Since the IRM concept is relatively new, the tools and management policies to support it are in their infancy (Horton and Marchand, 1982). The scope of IRM is quite broad, so we wish to focus on metadata, and the tools and management functions to support metadata management. In particular, we propose a triple (metadata, Information Resource Dictionary System (IRDS), Enterprise Administration). The following sections explore the notion of metadata and how it can be managed by an Information Resource Dictionary System (IRDS). Of particular interest is the data model for the IRDS; its features are presented and a candidate model is proposed.

Lastly, we define and develop the Enterprise Administration function required to manage the metadata concept and the associated IRDS tools.

IRM and Metadata

The "self-conscious" examination of IRM processes leads to the specification of metadata - data about data. We define metadata as follows:

Definition 1: Metadata is a formal, implementation-independent description of the data and information objects used by the enterprise. These objects may be database descriptions, data files, forms, directories, programs, standard operating procedures, etc.

Thus, metadata, in its most general context, is a complete description of the enterprise - its data, procedures, policies and goals. In most enterprises, metadata is distributed throughout the organization and may be in both manual and machine-readable form, as for example, organizational charts, standard operating procedure manuals, job classification and responsibility documents, telephone directories, paper forms, database descriptions, program libraries and the accumulated knowledge and expertise of key organizational figures.

We stress that enterprise experts play a key role in an organization because they distill knowledge of how the organization really works, as compared to how it was designed to function. This discrepancy between the real and ideal functioning of the enterprise indicates the very real need for metadata management tools to increase organizational effectiveness (Black and Marchand, 1982). The next section introduces a key tool for metadata management.

The Information Resource Dictionary System (IRDS)

Modern data dictionaries were originally used in conjunction with database management systems to serve as repositories of information regarding 1) data item definitions, 2) database descriptions (schemes and subschemes), 3) database constraints, 4) program libraries and modules, etc. A survey of data dictionary systems is provided by Allen, et. al. (1982). As organizations focus on the IRM function, the data dictionary is finding new applications in the form of the IRDS. Quoting once again from Goldfine (1982) we have:

> "An Information Resource Dictionary System (IRDS) is an information system with automated support which documents the information environment of an enterprise, supports the operational aspects of that information environment, illustrates the interrelationships of the information environment components and documents the locations of all components of the information environment. The Information Resource Dictionary (IRD) is the actual database manipulated by the IRDS software."

Clearly, the notion, scope, and proposed capabilities of the IRDS make it an extremely powerful tool. At present, the IRDS is a concept that must be developed into an effective tool. In the following sections we address the issues related to the features of such a system and explore several of them in-depth. In particular we wish to address the following:

a) The IRDS Data Model
b) The IRDS Architecture
c) IRDS Support Tools.

The IRDS Data Model

One of the drawbacks of data models of existing data dictionary systems is that they either 1) support rather primitive modeling constructs, or 2) have a sizable collection of constructs that address a specific information environment - the database environment with objects such as data items, records, files, databases, processes, modules, and appropriate predefined relationships among these objects. Therefore, it may be necessary to define a new class of data models to support the IRDS. Below we explore the requirements for the IRDS data model:

1) The Temporal Dimension - the "real world" is event-driven, and information systems must respond to both external and internal events. The notion of time is intimately connected with events because it is important to determine when events occur and their (possible) duration. Thus, it is important to be able to specify events, a temporal ordering among events, and possibly logical relationships among events. Several researchers have proposed such mechanisms (Bubenko, 1980; Codd, 1979; De and Sen, 1982; Sen and Kerschberg, 1983).

2) Enterprise Semantics - the IRDS data model must support a number of information environments, each having specialized concepts. Thus, the model should provide facilities to define such concepts. In addition to the data abstraction mechanisms of aggregation and generalization (Smith and Smith, 1977), the model should provide for the specification of complex objects (or conglomerates) that represents enterprise objects such as forms and form operations (Tsichritzis, 1980; 1982) as well as generalized transactions triggered by events and involving conglomerates. Another important feature is the ability to model the status of a complex object with respect to the many transactions which may act upon it. Finally, the notion of messages (Hewitt, 1977; Langefors, 1977) and message passing among conglomerates is a useful modeling construct for some environments that model the flow of information within the organization.

An IRDS data model, endowed with the above modeling mechanisms, would support a variety of information environments such as: management information systems, office information systems, personal information systems, the information system development process, an information locator application, etc.

A Candidate IRDS Data Model

In recent research in enterprise modeling (Sen and Kerschberg, 1983) we consider the enterprise as consisting of a collection of business functions. Each of these functions, in turn, consists of a collection of activities with components at the three levels of management - operational, tactical, and strategic (Anthony, 1965). Activities are specified as collections of views which incorporate three aspects: 1) organizational dynamics specified in terms of events, states and their temporal and logical interrelationships, 2) forms, form operations, and form routings, and 3) data requirements expressed in terms of Functional Data Model data abstractions (Hecht and Kerschberg, 1981; Shipman, 1981; Sibley and Kerschberg, 1977). Each view is performed at a particular workstation.

The view mechanism is a conglomerate of three distinct yet interrelated components: events, forms, and data. Events trigger actions which in turn operate on forms. Forms, in turn, may be routed from station to station (message passing) as a means of coordinating activities associated with a business function. In addition, forms have attributes which take on data values as they flow through the organization; the forms are data abstractions which are a source of data requirements for database support.

Figure 1 depicts a view called "special-terms-approval" which is performed at the sales-supervisor-desk as part of the sales-order-processing activity. The "es-section" specifies events, possible component events, synonyms (as-clause), and both temporal and logical relationships among events. The uses-clause indicates which forms are used by events.

The "form-section" specifies operations and routings involving forms. It is assumed that the forms have been defined previously by means of a form management package. Notice that both when and where clauses are used for event-dependent and data-dependent conditions, respectively.

The "data-section" is intended to capture the minimal number of data abstractions relevant to the view. In this example the properties "bonus-amount" and "date-approved" are aggregated to the entity sets "item" and "sales-order", respectively. The overall database specification is obtained by integrating the collections of views for the various business functions.

This modeling approach leads to new ways of thinking about their entities, their properties, and how they relate to other entities. For example, entities "pick up" property values as they are operated upon by the "real-world" - in fact we have a temporal history of the entity. The view mechanism allows the modeler to encapsulate a collection of events, states, forms and operations. The flow of forms allows views to communicate and request service from other views.

The IRDS Architecture

In the context of the ANSI/SPARC three-schema database architecture (Tsichritzis and Klug, 1977), each information environment would be modeled as a collection of external schemas, while the schema at the conceptual level would represent the integration of the concepts from the external

```
view     special-terms-approval
          workstation  sales-supervisor-desk
          activity  sales-order-processing
          activity level 1
          hierarchy  sales

es-section
         event  sales-order-form-arrives-in-intray
                as    e1
         event  sales-order-form-picked-from-intray
                as    e2
         event  order-approving as    e3
                composite components
                item-approving as     e31
                bonus-accounting as   e32
                uses
                (sales-order-form, bonus-form)

         relationships
                start  e3 cause (e31 where
                       SOF.item-info.bonus = 'T')
                e31 and e32 cause end e3

form-section
         form sales-order-form asSOF
           route SOF to sales-order-desk
                 when end e3

         form bonus-form as BF
         forms-op
           update BF.bonus-total by
             let BF.bonus-total =

             (BF.bonus-total +
           SOF.item-info.qty * SOF.item-info.bonus)
         when start e32 and
         where
         for SOF.salesperson-id = BF.salesperson-id
         and for each SOF.item-info,
                SOF.item-info.bonus - 'T'

data-section

         entityset item;
           key (item) = (item-id);
           bonus-amount:  item->  dollars

         association sales-order;
           data-approved:  sales-order->  date x time

end-view  special-terms-approval
```

Figure 1. An Enterprise View

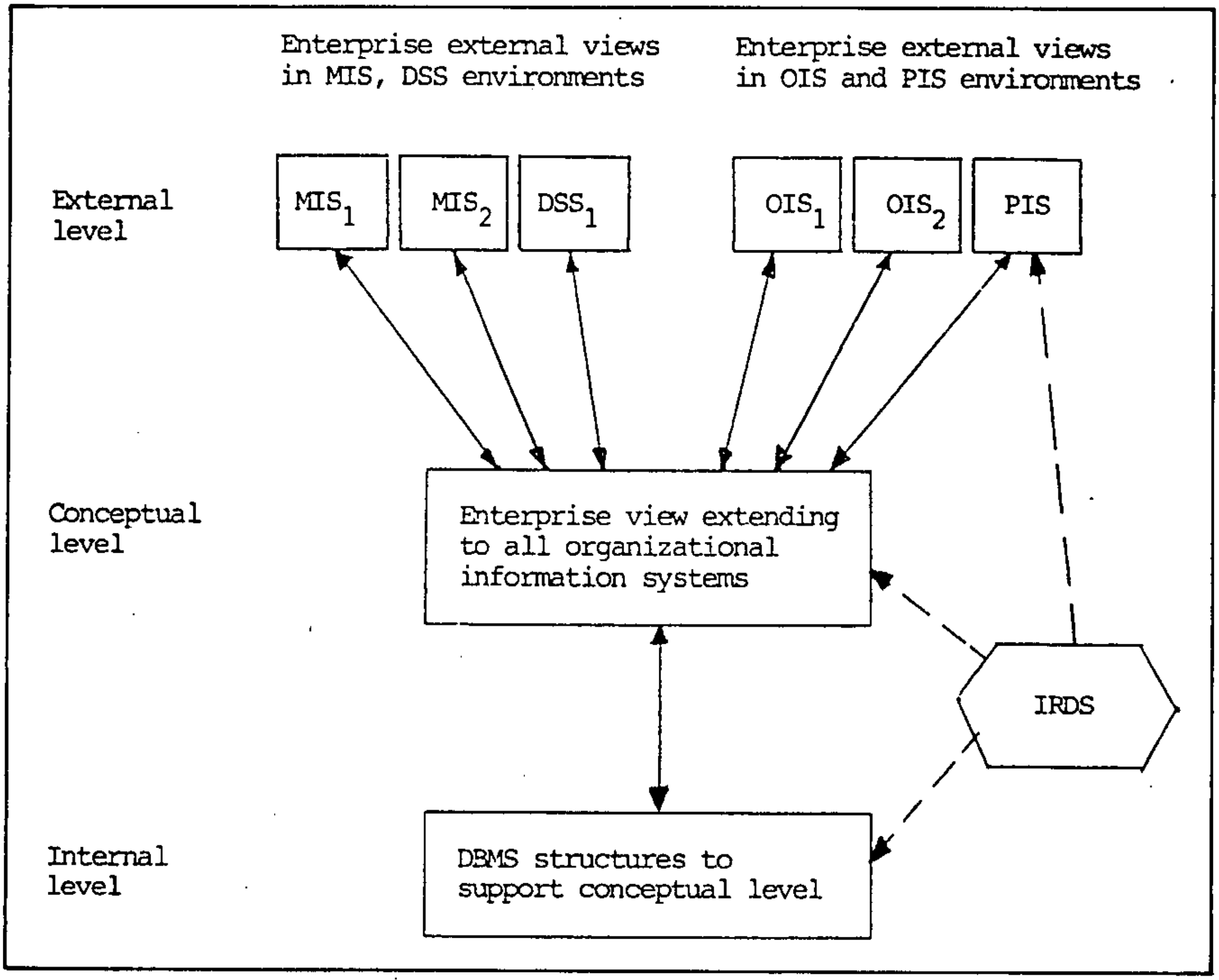


Figure 2. IRDS Architecture

---

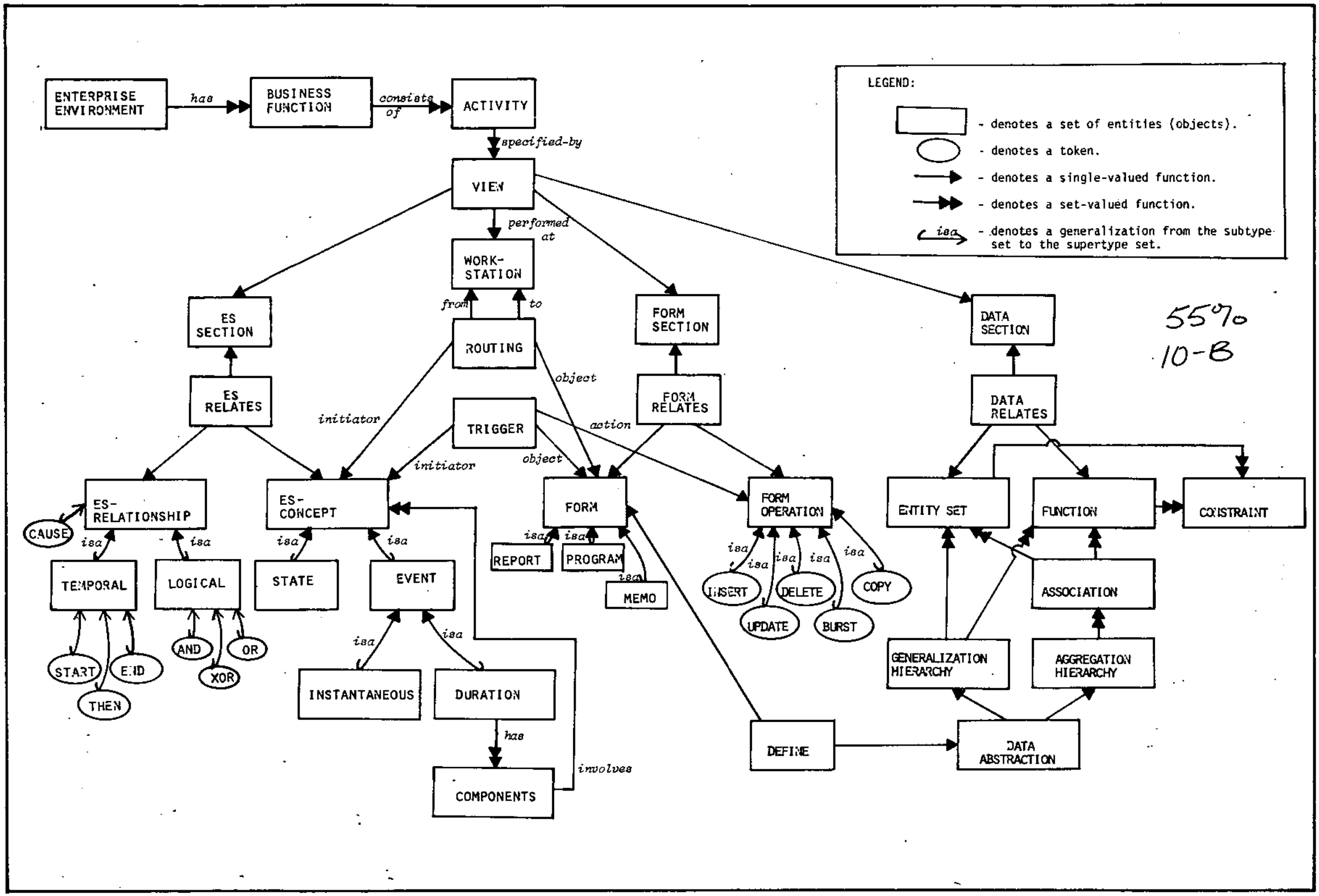
schemas (Navathe and Schkolnick, 1978). Figure 2 depicts a three-schema IRDS architecture. Thus, the interrelationships among information environments can be represented at the conceptual level of the IRDS.

The enterprise model can be used to support the various information environments, for example, the MIS, DSS, OIS and PIS. Each environment has its own events, activities, views, forms and form routings. In this context forms are a very general notion that can represent computer reports, graphics, a memo (letter), or a spreadsheet.

Once an enterprise external schema has been specified for an information environment, it can be mapped to a conceptual level representation in terms of a data model supported by a DBMS. Figure 1 provides an external view for (possibly) an MIS or OIS environment.

In order to show how the enterprise modeling concepts are mapped to the conceptual level of our architecture, we provide a Functional Data Model (FDM) schema for the enterprise model, as depicted in Figure 3.

An application in an information environment is represented by an enter-


ENTERPRISE ENVIRONMENT
has
BUSINESS FUNCTION
consists of
ACTIVITY
specified-by
VIEW
performed at
WORK-STATION
from
to
ROUTING
initiator
object
TRIGGER
initiator
object
action
ES SECTION
ES RELATES
ES-RELATIONSHIP
CAUSE
isa
TEMPORAL
START
THEN
END
LOGICAL
AND
XOR
OR
ES-CONCEPT
STATE
EVENT
INSTANTANEOUS
DURATION
has
COMPONENTS
involves
FORM SECTION
FORM RELATES
FORM
REPORT
PROGRAM
MEMO
FORM OPERATION
INSERT
UPDATE
DELETE
BURST
COPY
DEFINE
DATA SECTION
DATA RELATES
ENTITY SET
FUNCTION
CONSTRAINT
ASSOCIATION
GENERALIZATION HIERARCHY
AGGREGATION HIERARCHY
DATA ABSTRACTION
LEGEND:
- denotes a set of entities (objects).
- denotes a token.
- denotes a single-valued function.
- denotes a set-valued function.
isa - denotes a generalization from the subtype set to the supertype set.
55%
10-B

prise schema, i.e., the specification of business functions, activities and views. In the mapping from the external to conceptual level, a view specification would be transformed to "instances" or "occurrences" in the entity sets in Figure 3, with the functions providing the relationships among the entity sets.

In the FDM schema of Figure 3, each "box" denotes a set of entities, and is termed an entity set. There are three types of arrows: single-headed, double-headed, and those labeled "isa". Single-headed arrows denote single-valued functions from the domain entity set to the range entity set. For example, every view in VIEW is "performed-at" exactly one workstation in WORKSTATION. Double-headed arrows denote set-valued mappings, as for example, an activity in ACTIVITY is specified by one (or more) views in VIEW. The "isa" arrow denotes a generalization from a subtype entity set to a supertype, as, for example, REPORT "isa" FORM.

Certain entity sets, such as ROUTING and TRIGGER, actually represent associations (relationships) among other entity sets. For example, a routing entity in ROUTING signifies that an "es-concept initiates a routing of a form from one workstation to another." Similarly, a trigger entity in TRIGGER means that an "es-concept initiates a trigger which performs a form operation on a form."

Lastly, the circled names denote tokens (singleton sets). For example, START, END, and THEN are specific instances of temporal es-relationships used to relate es-concepts. Note also that, for the sake of clarity, many entity set properties (functions) have been omitted.

The conceptual schema in Figure 3 is a representation of enterprise metadata that can now be managed by a DBMS. Support tools would access the metadatabase to manipulate this information to support IRM functions. These tools, in effect, are metadatabase applications.

The proposed IRDS architecture is similar to that of the System Encyclopedia Manager (SEM) (Teichrow, Macasovic, Hershey, Yamamoto, 1980) of the ISDOS project. In addition, the Information System Language Definition System (ISLDS) allows the user to define a language to specify an information environment. The underlying data model that supports the system is the Entity-Relationship Model (Chen, 1976).

Research in the use of the ISLDS to specify "rich" semantic data models comparable to the Enterprise and Functional Models has been inconclusive (Teichroew, Germano, Luca, 1981). Although one can express the syntax of a language in terms of a data model (cf. Figure 3), it is more difficult to capture the operational (or behavioral) semantics. These would have to be expressed as transactions on the metadatabase.

Our IRDS architecture differs from the SEM in that, in addition to documenting information requirements, we expect the IRDS to play an active role in enterprise administration, as will be discussed in the section entitled Enterprise Administration.

IRDS Support Tools

The IRDS support tools are applications that access and manipulate the (meta)database schema of Figure 3 supported at the conceptual level of the IRDS architecture depicted in Figure 2.

Although the range of possible support tools is vast, we will focus on tools to support information integration in MIS, DSS, OIS and PIS. In particular we examine the areas of:

- Heterogeneous Information System Communication,
- OIS Scheme Analysis,
- A PIS Information Center.

Heterogeneous Information System Communication

There is a growing need for diverse, heterogeneous information systems to communicate. A large corporation may have several large-scale applications, developed independently on distinct hardware and running under distinct DBMS's. These applications, however, may support a single business function, say service-order-processing. In addition, the corporation may wish to integrate these applications by having the databases communicate. The IRDS's role is one of coordinator; the enterprise model captures both temporal and behavioral semantics of the business function. The forms to invoke transactions at such application would be stored in the IRDS, and appropriate data translations among the databases would be handled by the IRDS.

In another context, we may consider the organizational MIS and OIS as examples of heterogeneous information systems. Both may contain related information about the same enterprise entity. For example the MIS may contain employee records which include name, salary, medical history, skills, etc. On the other hand, the OIS may contain reports and memos written by these employees. The IRDS metadatabase schema can support a query of the form:

"Find the names and skills of employees who have written memos or reports on the subject of information utilities."

The IRDS would first decompose the query into two subqueries, one to the OIS to obtain employee identifiers for those having written on the subject of information utilities, and the second to the appropriate MIS to obtain the names, salaries and skills of those employees. Second, the IRDS would determine the appropriate query language to be used for each system, and format the subqueries for processing. Thus, the IRDS must have detailed information regarding the location of the desired information, and the local database system characteristics, e.g., data model, database schema, query language, etc.

Organizations will rely increasingly on external information provided by information utilities. This information will arrive in a pre-specified format (e.g., a VisiCalc ledger sheet) and will have to be convered by the IRDS into formats required by the various information environments. The converse translation will occur when an internal query requests external information from one or more utilities.

Admittedly, the task of specifying diverse information environments is difficult. The authors are involved in a study to implement an Information Locator System for the state of South Carolina. We intend to use the pardigm of the Enterprise Model and the IRDS architecture to guide us in obtaining information requirements and structuring the system. The experience obtained through this project will help us to formulate a methodology for the Enterprise Model and to aid in developing a prototype implementation.

OIS Schema Analysis

The enterprise model of the IRDS is well-suited to represent business functions in terms of activities represented by views. The metadatabase schema can support various applications for schema analysis. Several representations of OIS schemas are useful. They are:

1) A diagram (Sen and Kerschberg, 1983) showing the flow of forms from workstation to workstation for one or several (superimposed) business functions. This diagram can be generated from the ROUTING entity set and associated entity sets such as BUSINESS FUNCTION, ACTIVITY, VIEW, WORKSTATION, ES-CONCEPT, AND FORM. Recall that ES-RELATIONSHIP specifies the logical and temporal interrelationships among entities in ES-CONCEPT.

2) An event-state diagram (De and Sen, 1981; Sen and Kerschberg, 1983) to show the temporal and logical relationships among the ES-CONCEPT entities for a view.

3) An event-driven simulator (Nutt and Ricci, 1981) to estimate the workload on each of the workstations.

4) A schema restructuring tool to allow the merging and splitting of view for office streamlining (Cook, 1980).

The above tools present different aspects of an OIS environment, each obtained from transformations of the metadatabase.

A PIS Information Center

The IRDS can play an active role in linking a personal information system to the organizational information system. Below we outline some of the services provided by the IRDS:

1) The PIS Starter Kit consists of a set of programs, downloaded from the corporate information system or supplied via floppy-disk, to make the new workstation a part of the overall organizational information system. Included in the starter kit would be file transfer programs and a menu-driven information-service dictionary/directory to organizational information resources.

2) IRDS subscription forms for information utilities would allow the user to create a profile of interests so that the IRDS could route copies of relevant externally obtained information to the workstation.

3) IRDS messaging and data translation facilities would allow documents to be sent from and received at the PIS workstation. Thus, the IRDS must translate among diverse document formatting languages on different types of hardware.

4) The PIS would have some database management capabilities for local database support, and, in addition, would have access to corporate databases through the IRDS. Data would be requested by means of forms and the queries would be handled by the IRDS. The resulting information would be provided according to a user-specified format. The IRDS, in turn, could monitor database usage for reporting purposes.

Enterprise Administration

The Information Resource Dictionary System (IRDS) is the tool component of the triple (metadata, IRDS, Enterprise Administration). Since the IRDS is an integrated repository of enterprise knowledge, it has the potential of serving as an integration and control mechanism for organizational information systems. The Enterprise Administration function is the management function that oversees the use of the IRDS in the organization.

Recent articles suggest the role of Enterprise Administrator (EA) (Durchholz and Will, 1982; Goldfine, 1982). This role will evolve in conjunction with the expanding role of IRDS in the organization. We study the enterprise administration function for two cases: a passive IRDS and an active IRDS (Kent, 1980).

Enterprise Administration With Passive IRDS

In the early stages of IRM, the IRDS will play a passive role, that is, it will be consulted to support information management decisions. During this phase the EA will be a consultant to the Chief Information Officer. The EA would be a liaison not only with the existing functions of Database Administration, Data Administration, and Network Controller, but also with the evolving functions of Office Administration (in the context of OIS) and PIS Administration.

The EA would 1) provide impact studies of proposed applications, 2) propose interfaces among existing systems to promote information sharing, 3) assist in enterprise schema analysis for reorganization and streamlining, and 4) assist auditors in understanding existing applications (Durchholz and Will, 1982).

Enterprise Administration With Active IRDS

An active IRDS would control the development, maintenance, and integration of information systems. Therefore, Enterprise Administration will be a line function. The organizational chart for the various administration functions is shown in Figure 4. The role of the EA would be those specified in earlier, and, in addition, the following: 1) control the evolution of enterprise applications including the IRDS, 2) provide standards and protocols for interfaces to existing internal and external information systems, 3) provide enterprise information services to users through the IRDS, and 4) control procurement of hardware and software according to corporate IRM policies.

CONCLUSIONS

This paper has focused on the need for corporate information system integration. The triple (metadata, Information Resource Dictionary System, Enterprise Administration) denotes the concept, tool and management function, respectively, needed to implement such integration.

Through a better understanding of metadata, that is, data regarding information and the processes governing information usage, one can design information integration tools. The primary tool is the IRDS, an information system to support information resource management. The IRDS data model, with its view mechanism to capture events, states, form operations and routings as well as data abstractions, permits the information modelor to describe information environments.

The IRDS three-schema architecture allows several information environments to be integrated at the conceptual level by means of the functional database schema. This schema can also be used to build support tools (applications) to effect the integration of management, office, and personal information systems.

The role of Enterprise Administration will evolve from a passive, consultative role to an active, vital role as the IRDS matures from its role of serving as a simple repository of metadata to actually controlling information resources.

Metadata management and Enterprise Administration are seen as a first step in the development of knowledge management in an organization, where the primary focus is no longer exclusively on the physical control of data resources and technology, but on the actual content and use of information within management and operations. Increasingly, managers in business or-

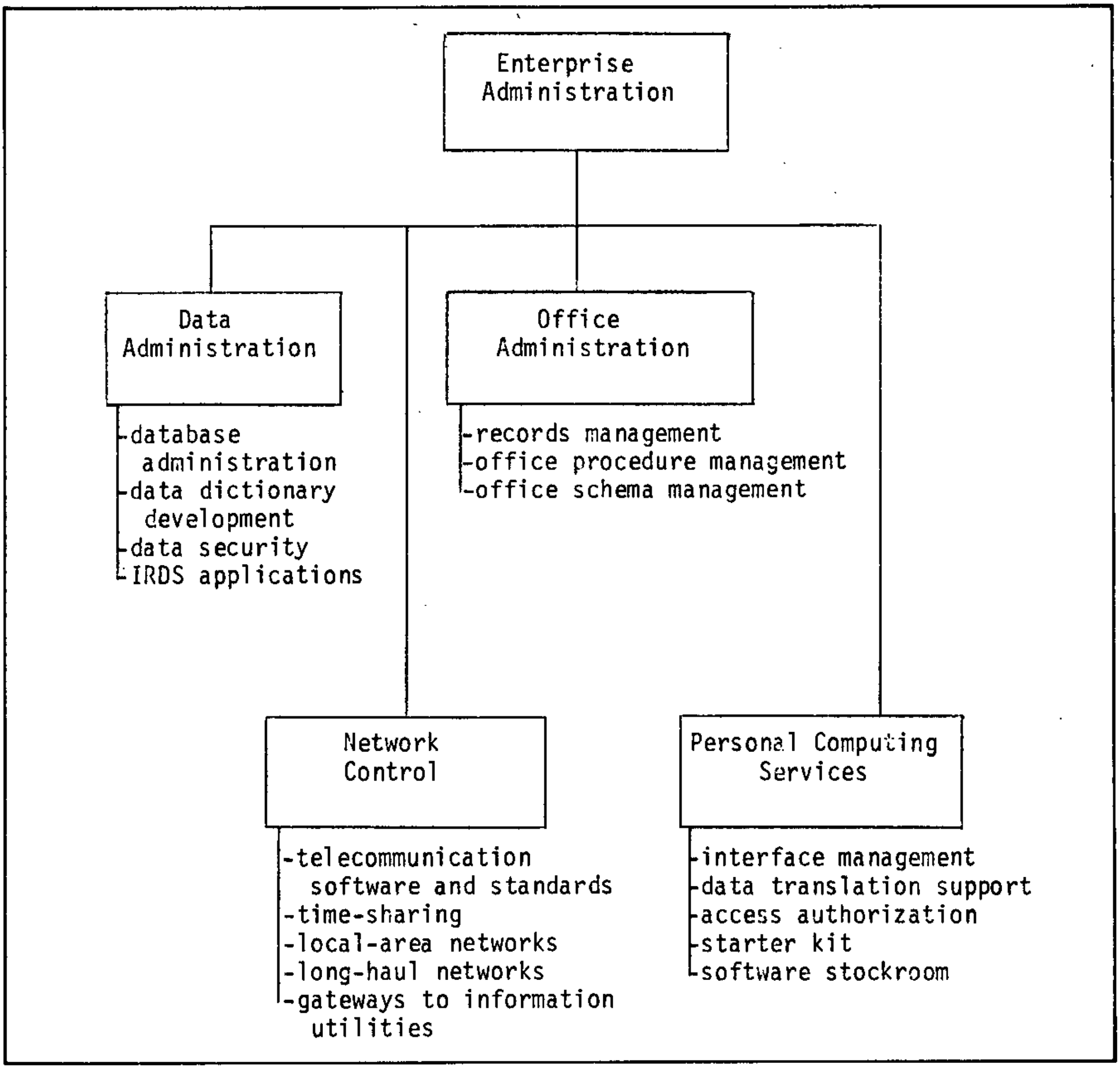


Figure 4. Enterprise Administration

ganizations will define organizational productivity on the basis of how well they guide their business in the acquisition and use of knowledge. How knowledge is applied to products and markets will determine the long-term expansion or contraction of the firm's production workforce and, therefore, will define the basis for economic growth and profitability. Recognition of the firm's dependence on effective knowledge management will become an essential management attribute and influence the basic operating philosophy of senior managers in the firm.

REFERENCES

Allen F.W., Loomis, M.E.S., and Mannino, M.V. "The Integrated Dictionary/Directory System," ACM Computing Surveys, Volume 14, Number 2, June 1982, pp. 245-286.

Anthony, R.N. "Planning and Control Systems: A Framework for Analysis," Graduate School of Business Administration, Harvard University, Boston, Massachusetts, 1965.

Black, S.H. and Marchand, D.A. "Assessing the Value of Information in Organizations: A Challenge for the 1980's" The Information Society Journal, Volume 1, Number 3, 1982, pp. 191-225.

Bubenko, J.A. "Information Modeling in the Context of System Development," Proc. of the IFIP Congress 80, October 1980, pp. 395-411.

Chen, P.P.S. "The Entity-Relationship Model: Toward a Unified View of Data," ACM Trans. Database Syst. Volume 1, Number 1, 1976, pp. 9-36.

Codd, E.F. "Extending the Database Relational Model to Capture More Meaning," ACM Trans. Database Syst. Volume 4, Number 4, December 1979, pp. 397-434.

Cook, C.L. "Streamlining Office Procedures - An Analysis Using the Information Control Net Model," Proc. National Computer Conference, AFIPS Press, 1980, pp. 555-565.

Davis, G.B. "Strategies for Information Requirements Determination," IBM System J, Volume 21, Number 1, 1982, pp. 4-30.

De, P. and Sen, A. "Knowledge Representation of a Data Base by a Semantic Network," International Journal of Policy Analysis and Information Systems, Volume 6, Number 1, 1982, pp. 25-45.

Durchholz, R. and Will, H.J. "The Impact of Database Management Systems Standardization on Auditing," Computers and Standards, Volume 1, 1982, pp. 49-59.

Ellis, C.A. and Nutt, G.J. "Office Information Systems and Computer Science," ACM Computing Surveys, Volume 12, Number 1, March 1980.

Gillinson, M.L. "The State of Practice of Data Administration--1981," Comm. ACM, Volume 25, Number 10, 1982, pp. 699-706.

Goldfine, A.H. editor. "Data Base Directions, Information Resource Management-Strategies and Tools," NBS Special Publication 500-92, September 1982.

Hecht, M.S. and Kerschberg, L. "Update Semantics for the Functional Data Model," Database Research Report No. 4, Bell Laboratories, Holmdel, New Jersey, 1981.

Hewitt, C. "Control Structure as Patterns of Passing Messages," Artificial Intelligence, Volume 8, 1977, pp. 323-363.

Horton, F.W. and Marchand, D.A., editors. Information Management in Public Administration, Information Resources Press, Arlington, Virginia, 1982.

Kent, W. "Splitting the Conceptual Schema," Proc. 6th VLDB Conference, Montreal, Canada, October 1980, pp. 10-14.

Langefors, B. "Information Systems Theory," Information Systems, Volume 5, 1977, pp. 207-219.

Lum, V., Ghosh, S., Schkolnick, M., Jefferson, D., Su., S., Fry, J., Teorey, T., and Yao, S.B. "1978 New Orleans Database Design Workshop Report," Proc. 5th VLDB Conference, Rio de Janeiro, Brazil, 1979.

Morgan, H.L. "Research and Practice in Office Automation," Proc. IFIP Congress 80, October 1980.

Mylopoulos, J., Bernstein, P.A., and Wong, H.T.K. "A Language Facility for Designing Interactive Database Intensive Applications," Proc. ACM SIGMOD International Conference on Management of Data, Austin, Texas, 1978.

Navathe, S.B. and Schkolnick, M. "View Representation in Logical Database Design," Proc. ACM SIGMOD International Conference on Management of Data, Austin, Texas, 1978, pp. 144-156.

Nutt, G.J. and Ricci, P.A. "An Office Modeling System," IEEE Computer, May 1981.

Ozkarahan, E.A. and Kerschberg, L. "A

Heterogeneous Distributed Database System Architecture Incorporating Data Semantics and a Rational Database Machine Interface. Technical Report #6, Department of Computer Science, Arizona State University, November 1982.

Sakamoto, J.G. and Ball F.W. "Supporting Business Systems Planning Studies with the DB/DC Data Dictionary," IBM System J, Volume 21, Number 1, 1982, pp. 54-80.

Sen, A. and De, P. "A Formal Procedure for Requirement Analysis in Data Base Design," Proc. of the Fourteenth Hawaii International Conference on Systems Sciences, Honolulu, Hawaii, 1981.

Sen, A. and Kerschberg, L. "Enterprise Modeling for Information System Design, Technical Report: TR-83-1, Institute of Information Management, Technology and Policy, University of South Carolina, 1983.

Shipman, D.W. "The Functional Data Model and the Data Language DAPLEX," ACM Trans. Database Syst., Volume 6, Number 1, March 1981, pp. 140-173.

Sibley, E.H. and Kerschberg, L. "Data Model and Data Architecture Considerations," Proc. National Computer Conference, AFIPS Press, June 1977, pp. 85-96.

Sibley, E.H., Scallan, P.G., and Clemons, E.K. "The Software Configuration Management Database," Proc. National Computer Conference, AFIPS Press, 1981, pp. 249-255.

Shu, N.C., Lum V.Y., Tung, F.C., and Chang, C.L. "Specification of Forms Processing and Business Procedures for Office Automation," IEEE Trans. on Software Engineering, SE-8, Volume 5, September 1982, pp. 499-512.

Smith, J.M. and Smith, D.C.P. "Database Abstractions: Aggregation and Generalization," ACM Trans. Database Syst., Volume 2, Number 2, June 1977, pp. 105-133.

Sprague, R.H., Jr. "A Framework for the Development of Decision Support Systems," MIS Quarterly, Volume 4, Number 4, December 1980, pp. 1-26.

Teichroew, D. and Hershey III., E.A. "PSL/PSA: A Computer-Aided Technique for Structured Documentation and Analysis of Information Processing Systems," IEEE Trans. on Software Engineering, SE-3, No. 1, January 1977, pp. 42-48.

Teichroew, D., Germano, F., and Luca, S. "Applications of the Entity-Relationship Approach," in Entity-Relationship Approach to Information Modeling and Analysis, P. Chen, ed., ER Institute, 1981, pp. 1-17.

Teichroew, D., Macasovic, P., Hershey, E.A., and Yamamoto, Y. "Applications of the Entity-Relationship Approach to Information Processing System Modeling," in ERA Approach to Systems Analysis and Design, P. Chen, ed., North Holland, 1980, pp. 15-38.

Thierauf, R.J. and Reynolds, G.W. Effective Information Systems Management, Charles E. Merrill Publishing Company, Columbus, Ohio, 1982.

Tsichritzis, D.C., and Klug, A., editors, "The ANSI/X3/APARC DBMS Framework-The Report of the Study Group on Data Base Management Systems, AFIPS Press, 1977.

Tsichritzis, D.C. "Form Management," Comm. ACM, Volume 25, Number 7, July 1982, pp. 453-478.